T.C
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

REGRESYON ÇÖZÜMLEMESİNDE KELEBEK DAĞILAN ARTIKLAR İÇİN DEĞİŞEN VARYANSLILIK SORUNU VE DÜZELTİLMESİ

DOKTORA TEZİ

Reşit ÇELİK

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı: Prof. Dr. M. Aydın ERAR

OCAK 2012

Reşit ÇELİK tarafından hazırlanan REGRESYON ÇÖZÜMLEMESİNDE KELEBEK DAĞILAN ARTIKLAR İÇİN DEĞİŞEN VARYANSLILIK SORUNU VE DÜZELTİLMESİ adlı bu tezin ...DOKTORA..... tezi olarak uygun olduğunu onaylarım.

Prof. Dr. M. Aydın ERAR

Tez Yöneticisi

Bu çalışma, jürimiz tarafından .....İstatistik..... Anabilim Dalında

..Doktora..... tezi olarak kabul edilmiştir.

Başkan: : Prof. Dr. M. Aydın ERAR

Üye : Prof. Dr. Nalan CİNEMRE

Üye : Prof. Dr. Ali H. Büyüklü

Üye : Prof. Dr. Sahamet Bülbül

Üye : Prof. Dr. Gülay Kuroğlu

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

ÖZET

Bu çalışmanın amacı, doğrusal regresyon çözümlemesinde ortaya çıkan kelebek dağılan artıklar sorunu için değişen varyanslılığı dengelemede kullanılan yöntemlerin karşılaştırılması ve alternatif yöntemlerin araştırılmasıdır. Bu amaç doğrultusunda, dönüşümler, ağırlıklandırma ve her ikisini bir arada kullanan çeşitli yöntemler uygulama verisiyle karşılaştırılmıştır.

Giriş bölümünde, klasik doğrusal regresyon analizi ve varsayımları üzerinde durulmuştur.

İkinci bölümde, değişen varyanslılığın tanımı, nedenleri, belirlenmesi ve sonuçları, özel olarak da değişen varyanslılık kalıplarından biri olan kelebek dağılan artıklar anlatılmıştır.

Üçüncü bölümde, varyans fonksiyonları ve değişen varyanslılığın dengelenmesi konusunda literatürde var olan önemli ağırlıklandırma ve dönüşüm yöntemleri üzerinde durulduktan sonra bu çalışmanın önemli bir parçası olan alternatif yöntemler verilmiştir.

Dördüncü bölümde, önceki bölümlerde verilen yöntemler, kelebek dağılan artıkları olan bir veri kümesine uygulanarak sonuçlar değerlendirilmiştir. Bu bölümde, ayrıca, önerilen yöntemlerin geçerliliği farklı örneklem büyüklüğündeki yapay veriler ile de incelenmiştir.

Son bölümde, genel bir değerlendirme yapılarak sonuçlar özetlenmiştir.

STATING AND CORRECTING THE HETEROSCEDASTICITY FOR BUTTERFLY DISTRUBUTED ERRORS IN REGRESSION ANALYSIS

SUMMARY

The aim of this study is to compare general statistical methods for stabilizing variances of butterfly distributed errors, emerge in linear regression analysis and investigate the alternative methods to fulfill the homoscedatic errors. For this purpose, weighting, transformations and several methods which use both transformations and weighting were compared on the same data.
The the first part of this thesis focuses on the classical linear regression and its assumptions.

In the second part, definition, causes, specification and results of heteroscedasticity are emphasized and in particular, the butterfly distributed errors which are one of the heteroscedasticity patterns are explained.

In the third part, after emphasizing on variance functions and the important weighting and transformation methods in the literature about the stablizing heterogeneous errors variances, some alternative methods, which are also the important part of this study, are given.

In the fourth part, the methods mentioned in previous sections are applied with a specific data set which has butterfly distributed errors and the results are evaluated. Also, the validity of proposed methods are evaluated with artifical data sets with different sample sizes.

In the final part, the results were summarized with genaral evaluation.

ÖNSÖZ

Uzun soluklu bir çaba sonucu ortaya çıkan bu eserin oluşmasında, değerli fikirlerini ve zamanını esirgemeyen danışman hocam sayın Prof. Dr. Aydın ERAR'a, üzerimde çok emeği olan değerli hocam sayın Prof. Dr. Mehmet Genceli'ye, desteğini her zaman hissettiğim bölüm başkanım, sayın hocam Prof. Dr. Ali Hakan BÜYÜKLÜ'ye, anlayışı ve desteği için sayın hocam Prof.Dr. Nalan CİNEMRE'ye, çalışma arkadaşım sayın Yard. Doç. Dr. İbrahim DEMİR'e, çalışma arkadaşım sayın Araş. Gör. Ömer BİLEN'e, çalışma arkadaşım sayın Araş. Gör. Fulya GÖKALP'e, ve son olarak zor zamanlarda hep yanımda olan Aileme
Teşekkürü borç bilirim.

Ocak 2012 **Reşit Çelik**

İÇİNDEKİLER

ÖZET ii

SUMMARY iii

ÖNSÖZ iv

ÇİZELGE LİSTESİ vii

ŞEKİL LİSTESİ ix

KISALTMALAR LİSTESİ x

1. GİRİŞ 1
1.1. Klasik Doğrusal Regresyon Modeli ve Varsayımları 1
1.2. Klasik Doğrusal Regresyonda Parametre Tahmin Edicileri ve Özellikleri 3

2. DEĞİŞEN VARYANSLILIK 5
2.1. Değişen Varyanslılığın Nedenleri 5
2.2. Değişen Varyanslılığın Saptanması 6
2.2.1. Grafiksel Yöntemler 6
2.2.1.1 Kelebek Dağılan Artıklar için Değişen Varyanslılık 9
2.2.2. Değişen Varyanslılığın Belirlenmesinde Kullanılan Sayısal Yöntemler 11
2.2.2.1. Spearman Sıra Korelasyonu 11
2.2.2.2. Değiştirilmiş Levene Testi 11
2.2.3.3. White Testi 12
2.3. Değişen Varyanslılığın Sonuçları 13

3. VARYANS FONKSİYONLARI ve DEĞİŞEN VARYANSLILIĞIN DENGELENMESİ 15
3.1 Varyans Fonksiyonu Tahmini 15
3.2. Karesel Artıklarla Ağırlıklı En Küçük Kareler 16
3.3. Mutlak Artıklara Dayanan Yöntemler 17
3.4. Ardışık Artıklara Dayanan Yöntemler 19
3.5. Değiştirilmiş (Modified) Ardışık Artıklar Yöntemi 20
3.6. Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Yöntemi 21
3.7. Chatterjee ve Machler Tarafından Önerilen Iteratif Yöntem 22
3.8. Tekrarlı Gözlemler için Ortalama ve Varyans Fonksiyonları 23
3.9. Bootstrap Yöntemi 23
3.10. Jackknife Yöntemi 25
3.11. Parametrik Olmayan Düzleme 26
3.12. Dönüşümler 29
3.13. Ağırlıklandırma ve Dönüşümlerin Birleştirilmesi 32
3.14. Kelebek Dağılan Artıklar için Değişen Varyanslılığı Dengelemede Alternatif Yöntemler 32

3.14.1. Dışsal Değişken ve Merkezileştirme 32
3.14.2. Merkezi Dışsal Değişken ile Doğrudan Ağırlık Tahmini 36

4. UYGULAMA 38
4.1. Hava Kalitesi İndeksi 38
4.2. Model Seçimi ve Artıklar 40
4.3. Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler 43
4.4. Chatterjee ve Machler Tarafından Önerilen İteratif Yöntemin Sonuçları 44
4.5. Değiştirilmiş Ardışık Artıklar Yöntemi Sonuçları 45
4.6. Parametrik Olmayan Düzleme 46
4.6.1. Mutlak Artıklarla Parametrik Olmayan Düzleme 46
4.6.2. Karesel Artıklarla Parametrik Olmayan Düzleme 48
4.7. Tekrarlı Gözlemler için Ortalama ve Varyans Fonksiyonları 50
4.7.1. Alt Gruplara Karşılık Gelen Bağımlı Değişken Değerlerinin Standart Sapmaları 50
4.7.2. Karesel Artıklar 51
4.8. Jackknife Yöntemi Sonuçları 52
4.9. Bootstrap Yöntemi Sonuçları 54
4.10. Box-Cox Dönüşümleri 55
4.11. Değişen Varyanslılığın İki Parçalı Ağırlıklandırma ile Dengelenmesi 57
4.12. Dışsal Değişken ve Merkezileştirme Sonuçları 58
4.13. Yapay Veriler ile Önerilen Yöntemlerin Sınanması 60
4.13.1. Yaklaşık Simetrik Veri n = 270 62
4.13.2. Yaklaşık Simetrik Veri n = 1339 65
4.13.3. Yaklaşık Simetrik Veri n = 8057 69
4.13.4. Sağa Asimetrik Veri n = 203 72
4.13.5. Sola Asimetrik Veri n = 223 75
4.13.6. Merkezi Dışsal Değişkenle Doğrudan Ağırlık Tahmini ve Ağırlıklı En Küçük Kareler 78

5. SONUÇ ve ÖNERİLER 81

KAYNAKLAR 83

ÖZGEÇMİŞ 86

ÇİZELGE LİSTESİ

Sayfa No

Çizelge 3.1. Çekirdek Fonksiyonları 28
Çizelge 4.1. Tam Küme Parametre Tahmin Edicileri ve Model Seçimi 40
Çizelge 4.2. En Küçük Kareler Analiz Sonuçları 42
Çizelge 4.3. Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Sonuçları 43
Çizelge 4.4. Chatterjee ve Machler İteratif Yöntemi Sonuçları 44
Çizelge 4.5. Değiştirilmiş Ardışık Artıklar ile Ağırlıklı En Küçük Kareler Sonuçları 45
Çizelge 4.6. Mutlak Artıklar İçin Parametrik Olmayan Regresyon ile Ağırlık Tahminleri ve Ağırlıklı En Küçük Kareler Sonuçları 47
Çizelge 4.7. Karesel Artıklar için Parametrik Olmayan Regresyon ile Ağırlık Tahminleri ve Ağırlıklı En Küçük Kareler Sonuçları 49
Çizelge 4.8. Bağımlı Değişken Standart Sapmaları için AEKK Sonuçları 50
Çizelge 4.9. Karesel Artık Standart Sapmaları için AEKK Sonuçları 52
Çizelge 4.10. Jackknife Parametre Tahmin Edicileri ve Artık Çizimleri 53
Çizelge 4.11. Bootstrap Yöntemi İçin Parametre Tahmin Edicileri ve Artık Çizimleri 54
Çizelge 4.12. Box-Cox $\lambda = 0 \rightarrow$ Logaritma Dönüşümü için Analiz Sonuçları 55
Çizelge 4.13. Box-Cox $\lambda = 0.5$ için Dönüşüm Sonuçları 56
Çizelge 4.14. İkiye Parçalama ile Elde Edilen Ağırlıklar ve AEKK Sonuçları 57
Çizelge 4.15. Merkezi Dışsal Değişken ile Tahmin Edilen Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri 59
Çizelge 4.16. Merkezi Dışsal Değişken ile Ağırlık Tahmini ve AEKK Sonuçları 59
Çizelge 4.17. n = 270 için En Küçük Kareler Sonuçları 63
Çizelge 4.18. Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri 64
Çizelge 4.19. n = 270 Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu Tahmini ve AEKK Sonuçları 64
Çizelge 4.20. n = 1339 için En Küçük Kareler Sonuçları 66
Çizelge 4.21. Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri 67
Çizelge 4.22. n = 1339 için Standart Sapma Fonksiyonu ile Ağırlık Tahmini ve AEKK Sonuçları 68
Çizelge 4.23. n = 8057 İçin En Küçük Kareler Sonuçları 69
Çizelge 4.24. Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri 71
Çizelge 4.25. n = 8057 için Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu ile Ağırlık Tahminleri ve AEKK Sonuçları 71
Çizelge 4.26. Sağa Asimetrik Veri için En Küçük Kareler Sonuçları 73
Çizelge 4.27. Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri 74

Çizelge 4.28. Sağa Asimetrik Veride Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu Tahmini ve AEKK Sonuçları 74
Çizelge 4.29. Sola Asimetrik Veri için EKK Sonuçları 76
Çizelge 4.30. Standart Sapma Fonksiyonu Tahmininin Özet İstatistikleri 77
Çizelge 4.31. Sola Asimetrik Veri için Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu Tahmini ve AEKK Sonuçları 78
Çizelge 4.32. Mutlak Merkezi Dışsal Değişken ile AEKK 79

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1. Sağa Megafon 9
Şekil 2.2. Sola Megafon 10
Şekil 2.3. Galaksi Dağılan Artıklar 10
Şekil 2.4. Kelebek Dağılan Artıklar 11
Şekil 3.1. Ham Artıklar 33
Şekil 3.2. Karesel Artıklar 33
Şekil 3.3. Mutlak Artıklar 33
Şekil 3.4. Mutlak Artıklardan Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu 36
Şekil 4.1. Partiküller Maddenin Kükürt ve Sıcaklıkla Değişimi 41
Şekil 4.2. Sıcaklık- Partiküller Madde Değişimi 41
Şekil 4.3. Kükürt-Partiküller Madde Değişimi 42
Şekil 4.4. Mutlak Artıkların Bağımlının Tahmin Değerleriyle Çekirdek Düzlemesi 47
Şekil 4.5. Karesel Artıkların Bağımlı Tahminleri ile Çekirdek Düzlemesi 48
Şekil 4.6. Merkezi Dışsal Değişken ile Mutlak Artıkların Dağılımı 59
Şekil 4.7. Artık Dönüşümlerinin Merkezi Dışsal Değişken ile Fonksiyonları 61
Şekil 4.8. n = 270 için Yapay Veri Grafiği 62
Şekil 4.9. Standart Sapma Fonksiyonu 64
Şekil 4.10. n = 1339 için Yapay Veri Grafiği 66
Şekil 4.11. n = 1339 için Standart Sapma Fonksiyonu 67
Şekil 4.12. n = 8057 için Yapay Veri Grafiği 69
Şekil 4.13. n = 8057 için Standart Sapma Fonksiyonu 70
Şekil 4.14. n = 203 için Yapay Veri Grafiği 72
Şekil 4.15. Sağa Asimetrik Veri için Standart Sapma Fonksiyonu 74
Şekil 4.16. n = 223 Sola Asimetrik Veri Grafiği 76
Şekil 4.17. Sola Asimetrik Veri için Standart Sapma Fonksiyonu 77

KISALTMALAR LİSTESİ

AEKK	: Ağırlıklı En Küçük Kareler
AIC	: Akaike Bilgi Kriteri
AKT	: Artık Kareler Toplamı
AKO	: Artık Kareler Ortalaması
BIC	: Bayes Bilgi Kriteri
EKK	: En Küçük Kareler
HKİ	: Hava Kalitesi İndeksi
PM	: Partiküller Madde
ortc	: Ortanca
S.E	: Standart Hata
SBC	: Schwarz'ın Bayes Kriteri
VIF	: Varyans Şişme Değeri
YAIEKK	: Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler

1. GİRİŞ

Klasik doğrusal regresyon analizinde varsayımlar bozulduğunda, ağırlıklandırma ya da dönüşümler kullanılarak bu varsayım bozulmaları giderilmeye çalışılır. Değişen varyanslılık bu varsayım bozulmalarından biri olup, kelebek dağılan artıklar da değişen varyanslılığın özel bir halidir.

Bu çalışmanın amacı, kelebek dağılan artıklar için değişen varyanslılığın dengelenmesinde kullanılan yöntemlerin sınanması ve alternatif yöntemlerin araştırılmasıdır.

Çalışma kapsamında, giriş bölümünde, klasik doğrusal regresyon analizi ve varsayımları üzerinde durulmuştur. Çalışmanın ikinci bölümünde değişen varyanslılık, kaynakları, saptanması, özel olarak da kelebek dağılan artıklara değinildikten sonra değişen varyanslılığın sonuçları üzerinde durulmuştur. Üçüncü bölümde, varyans fonksiyonları ve değişen varyanslılığın dengelenmesi konusunda literatürde var olan önemli ağırlıklandırma ve dönüşüm yöntemleri üzerinde durulduktan sonra bu çalışmanın önemli bir parçası olan alternatif yöntemler verilmiştir. Dördüncü bölümde, önceki bölümlerde verilen yöntemler, kelebek dağılan artıkları olan bir veri kümesine uygulanarak sonuçları değerlendirilmiştir. Bu bölümde, ayrıca, önerilen yöntemlerin geçerliliği hem uygulama verisinde hem de farklı örneklem büyüklüğündeki yapay veriler ile incelenmiştir. Son bölümde, genel bir değerlendirme yapılarak sonuçlar özetlenmiştir.

1.1. KLASİK DOĞRUSAL REGRESYON MODELİ ve VARSAYIMLARI

p sayıda bağımsız değişken için çoklu doğrusal regresyon modeli

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \qquad \textbf{i = 1,.., n} \qquad \textbf{(1.1)}$$

şeklinde ya da matris biçiminde $Y = X\beta + \varepsilon$ olarak yazılır. Burada Y, nx1 boyutlu bağımlı değişken gözlem değerleri vektörü; X, nxp boyutlu tasarım matrisi; β,

(p+1)x1 boyutlu parametre vektörü ve ε_i, nx1 boyutlu gözlenemeyen hata terimi vektörüdür.

(1.1) modelinin matris formunda, parametre tahmin edicileri,

$$\hat{\beta} = (X'X)^{-1} X'Y \quad \textbf{(1.2)}$$

ile ve uyum tahminleri de,

$$\hat{Y} = X\hat{\beta} \rightarrow \overbrace{X(X'X)^{-1}X'}^{H} Y \quad \textbf{(1.3)}$$

ile elde edilir. Artık değerleri ise $e = Y - \hat{Y}$ ile bulunur. Burada eşitlik (1.2) ve

$$H = X(X'X)^{-1}X' \quad \textbf{(1.4)}$$

matrisi kullanılarak

$e = Y - HY$

ya da

$$e = (I - H)Y \quad \textbf{(1.5)}$$

yazılabilir. Burada H, nxn boyutlu eşgüçlülük (idempotent) özelliği olan "şapka" matrisidir.

Parametre tahmin edicilerinin varyansı,

$$V(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} \quad \textbf{(1.6)}$$

ile ve $\hat{\sigma}^2$'nın tahmin edicisi, artık kareleri ortalaması (AKO),

$$\hat{\sigma}^2 = AKO = \frac{1}{n-p-1}\left(Y'Y - \hat{\beta}'X'Y\right) \quad \textbf{(1.7)}$$

ile verilir.

(1.1) modelindeki gözlenemeyen hata terimleri ε_i için,

- $E(\varepsilon_i) = 0$ olduğu,
- $\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$ için gözlemlerin bağımsız olduğu,
- $V(\varepsilon_i) = \sigma^2$ ile varyansların eşit olduğu,
- $\varepsilon \sim N(0, \sigma^2)$ sıfır ortalama ve σ^2 varyans ile normal dağıldığı,
- Bağımsız değişkenlerin rastlantı değişkeni olması durumunda $E(x_{ij}, \varepsilon_i) = 0$

olduğu varsayılır (Drapper ve Smith, 1998).

Herhangi bir regresyon problemine başlarken tam karma model tahmini yapılmalı ve artıklar incelenmelidir. Artık incelenmesi regresyondaki varsayımlardan sapmaları belirler (Bagnall ve diğ., 2006).

Varsayım bozulması tipine göre en küçük kareler ve en çok olabilirlik yöntemlerine alternatif bazı yöntemler aşağıdadır:

- Veride değişen varyanslılık söz konusu ise *ağırlıklı en küçük karelere,*
- Hatalar ilişkili ise *genelleştirilmiş en küçük karelere,*
- Parametrelere bazı kısıtlamalar uygulanırsa *koşullu en küçük karelere*,
- Veri kümesinde çoklubağlantı varsa *ridge regresyonu* gibi yanlı tahmin yöntemlerine,
- Tüm değişkenler rastgele bir hataya bağlı ise *genişletilmiş en küçük karelere (Expanded Least Squares)*,
- Veri normalden başka bir dağılım gösteriyorsa ya da değişen varyanslılık, aykırı değerler ya da yüksek kaldıraç noktaları söz konusu ise o zaman bazı sağlam (robust) yöntemlere başvurulur (Meloun ve diğ., 2002).

1.2. KLASİK DOĞRUSAL REGRESYONDA PARAMETRE TAHMİN EDİCİLERİ ve ÖZELLİKLERİ

En küçük kareler en eski ve en yaygın biçimde kullanılan tahmin sürecidir. En küçük kareler, çoğunlukla hata kaynaklarına ilişkin bilgi olmadığında deterministtik modellerde kullanılır. En küçük kareler tahmin edicileri, gözlenen değerler ile tahmin değerleri arasındaki farkların kareleri toplamını minimize eder. En küçük kareler tahmin edicileri parametrelerin doğrusal fonksiyonlarının herhangi bir kümesi için minimum varyanslı, toplamsal ve yansız tahmin ediciler verir. Ortalama, parametrelerin doğrusal olmayan bir fonksiyonu olsa bile en küçük kareler tahmin edicileri tutarlı ve asimptotik olarak yansız kalır. Normallik varsayımının gerçeklenmesi durumunda ise en küçük kareler tahmin edicileri en çok olabilirlik tahmin edicileridir.

En çok olabilirlik tahmin edicileri, tam bir olasılık yoğunluk fonksiyonu belirlenmesine dayanır. Bu tahmin ediciler küçük örneklem büyüklüğü için

optimal özelliklere sahip değildir; ama artan örneklem büyüklüğüne koşut olarak asimptotik olarak tutarlı, yansız ve etkindirler (Silverman ve Kot, 2000).

2. DEĞİŞEN VARYANSLILIK

Basit ve çoklu regresyon yöntemlerinde, bağımlı değişken Y'nin $X = x_i$'ye göre koşullu dağılımlarının aynı varyansa sahip olduğu varsayılır.

Hata terimlerine ilişkin varyansların, $\hat{y}_i$ ya da bağımsız değişkenlerden birinin çeşitli düzeyleri için istatistiksel olarak farklı olması değişen varyanslılık olarak tanımlanır. Değişen varyanslılık, sağlık, ekonomi ve biyoistatistik gibi pek çok alanda karşımıza çıkar (Basu, 2005).

Harvey (1976), hata terimlerine ilişkin varyansın aynı olmadığı regresyon modellerinde, hata terimlerine ilişkin varyansın bağımsız değişkenlerin bir ya da bir kısmının fonksiyonu olduğunu belirtmiştir.

Carroll ve Ruppert (1982), varyansların bağımsız değişkenlerin parametrik bir fonksiyonu olduğu değişen varyanslı doğrusal bir modelde, regresyon parametreleri tahmininin asimptotik varyanslarla ağırlıklandırılmış en küçük kareler ile yapılabileceğini belirtmiştir.

2.1. DEĞİŞEN VARYANSLILIĞIN NEDENLERİ

Değişen varyanslılık aşağıdaki nedenlerle ortaya çıkabilir:

1. Bağımsız değişken değerleri değiştikçe, $V(y_i)$'nin de değişmesi sonucu değişen varyanslılık ortaya çıkabilir. Örneğin, gelir arttıkça kişilerin istedikleri gibi harcayabilecekleri miktar da artar ve gelirlerini harcayabilecekleri seçenekler genişler; dolayısıyla, harcamanın varyansı artan gelire bağlı olarak büyür.
2. Bağımlı değişkenin kesikli olması durumunda ortalama ile varyansın ilişkili olması değişen varyanslılığa neden olabilir.
3. Değişen varyanslılığın bir diğer nedeni modelleme hatasıdır. Farklı bir model tahmini ya da bazı önemli değişkenlerin model dışı bırakılması değişen varyanslılığın kaynağı olabilir (Gujaratti, 2001).

4. Değişen varyanslılık aykırı değerlerin örneklemdeki 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.
5. Veri derleme teknikleri ya da ölçme araçları geliştikçe varyans küçülebilir.
6. Oberg ve Davidian'a (2000) göre, normal dağılım ve eş varyanslılık varsayımlarının bozulması daha çok tekrarlı ölçümlü verilerde görülmektedir.
7. Kişilerin herhangi bir konudaki deneyimleri arttıkça davranış hataları da zamanla azalmakta dolayısıyla varyans da küçülmektedir.

2.2. DEĞİŞEN VARYANSLILIĞIN SAPTANMASI

Artık grafikleri, değişen varyanslılığın saptanmasında en çok kullanılan araçlardır. Artık grafiklerine ek olarak, özellikle ekonometride kullanılan, pek çok sayısal test bulunmaktadır.

2.2.1. Grafiksel Yöntemler

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 artık çizimleridir.

1. (1.3) Eşitliğinden elde edilen $\hat{y}_i$ tahmin değerleri yatay eksende ve bunlara karşılık gelen $(y_i - \hat{y}_i)$'den elde edilen artıklar dikey eksende gösterildiği grafikte megafon biçimindeki bir görüntü, artık varyanslarının bağımlı değişkenin tahmin değerleri ile değiştiğinin kanıtıdır. Değişen varyanslılığın saptanmasında çoğu zaman artıkların, bağımsız değişkenlerle olan grafiklerinden de yararlanılır.

2. Değişen varyanslılığın belirlenmesinde ham artıklarla yapılan çizimler çoğu zaman yeterli olmaz ve değişen varyanslılığın modellenmesi konusunda bilgi vermez. Cook ve Weisberg (1982) kare artıkların $\hat{y}_i$ değerlerine karşı çizimlerini tartışmış; ilk maddede yer alan genel çizimlerin yorumunun kolay olmadığını, pozitif ve negatif artıklarla belli bir kalıbın ortaya çıkmasının her zaman olası olmadığını belirtmişlerdir. Cook ve Weisberg'e (1982) göre sorun karesel artıklarla çalışıldığında ortadan kalkar ve örneklem büyüklüğü göreceli olarak ikiye katlanır. "Kama" şeklinde bir karesel artıklar görüntüsü varyans değişkenliğinin ortalamayla

arttığının göstergesidir. Bununla birlikte bu tür durumlarda büyük artık değerleri ölçek sorununa yol açar.

3. Carroll ve Ruppert'e (1988) göre, ölçek sorununu gidermek için mutlak artıkları, mutlak artıkların logaritmalarını ve artıkların karelerinin küp köklerini kullanmak daha yararlıdır. Artıkların kareleri varyans tahmini gibi düşünülürse mutlak artıklar da standart sapmanın tahmini gibi düşünülebilir. Buradaki amaç değişkenliğin yapısının mutlak artıklar ve onların dönüşümleriyle analiz edilerek anlaşılması ve modellenmesidir.

4. Değişen varyanslılığın belirlenmesi için, çoğu zaman, mutlak artıkların logaritmasına karşı $\hat{y}_i$ tahmin değerlerinin logaritmasının çizimi yararlı olur (Carroll ve Ruppert, 1988). Mutlak artıkların logaritması standart sapmanın logaritmasının yerine; tahmin değerlerinin logaritması ise ortalamanın logaritması yerine geçer. Eğer, standart sapma için

$$\sigma_i = \sigma \mu_i(\beta)^{\theta} \qquad \textbf{(2.1)}$$

gibi ortalama üslü bir fonksiyon söz konusu ise çizim nadiren θ parametresi ile doğrusal bir ilişki gösterir. Burada σ ortak standart sapma, σ_i i'inci kitle bağımlı değişken değerlerinin standart sapması; $\mu_i(\beta)$, i'inci kitle için $\hat{y}_i$ tahmin değerlerinin ortalaması θ, üs parametresidir.

5. Veriler eş varyanslı olsa bile, H matrisi eşgüçlü olduğundan artıklar $(1-h_{ii})$ ile orantılıdır. Kaldıraç (Leverage) özelliğine sahip büyük h_{ii} değerleri ortalama ile yakından ilişkiliyse, artık grafiklerinde tamamen değişen varyanslı bir görünüm ortaya çıkar. Bu nedenle ham artıkların yerine, kendi varyansları ile ölçeklenen,

$$r_i = e_i \Big/ \left[\hat{\sigma}(1-h_{ii})^{1/2} \right] \qquad \textbf{(2.2)}$$

Student-türü artıkları kullanmak daha duyarlı ve doğru olur. Buradaki amaç, yatay eksen boyunca çizilen bileşendeki hafif değişmelerle değişen varyanslılığı ortaya çıkarmaktır. Carroll ve Ruppert (1988), varyansların θ skaleri ile,

$$\sigma_i^2 = \sigma^2 g^2(\mu_i(\beta), z_i, \theta)$$

modeline uygun olduğunu düşünmüşlerdir. Burada σ_i^2, i'inci kitlenin artık varyansı; g, varyans fonksiyonu; $\mu_i(\beta)$, i'inci kitle ortalaması; z_i, varyansın fonksiyonu

olduğu bağımsız değişken ve θ, eşitlik (2.1)'deki standart sapma fonksiyonundaki üs parametresidir. θ'nın sıfır olması değişen varyanslılığın söz konusu olmadığı anlamına gelmektedir. Öte yandan,

$$s_i = \frac{\partial}{\partial \theta}\left\{ g^2\left[\mu_i(\beta), z_i, \theta\right]\right\}_{\theta=0} \qquad \textbf{(2.3)}$$

tanımlanırsa, küçük θ için,

$$\begin{aligned} \text{varyans}\left(\hat{\sigma} r_i\right) &= (1-h_{ii})\sigma_i^2 + \sum_{i\neq k}\sigma_k^2 h_{ik}^2/(1-h_{ii}) \\ &\simeq \sigma^2\left[1+\theta\left((1-h_{ii})+\sum_{i\neq k} s_k h_{ik}^2/(1-h_{ii})\right)\right] \end{aligned} \qquad \textbf{(2.4)}$$

olur. Burada H matrisinin köşegen öğeleri dışındaki h_{ik} değerleri küçük ise, $\text{varyans}(\text{r}_\text{i}) \simeq 1+\theta(1-h_{ii})s_i$ yazılabilir. Burada önerilen; ham, mutlak ya da Student-türü artıkların $(1-h_{ii})s_i$'lere karşı grafiklerinin çizimidir. Örneğin (2.1)'deki standart sapma fonksiyonunun karesi alındığında, $\sigma_i^2 = \sigma^2 \exp\left[2\theta f(x_i, \beta)\right]$ üstel varyans modeline ulaşılır. Bu nedenle $\hat{y}_i$ değerlerine karşılık çizilecek Student-türü artıkların grafiği üstel varyans modelinin algılanmasında; $\hat{y}_i$ değerlerinin logaritmalarına karşı Student-türü artıklarla çizimi ise (2.1) formülüyle verilen ortalama-üslü varyans modelinin belirlenmesinde kullanılabilir. Çizim, artıkların bağımsız değişkenlerden birine ya da bağımsız değişkenlerin fonksiyonlarından her herhangi birine karşı çizimi için de geçerlidir.

6. Örneklem büyüklüğü çok büyük olduğunda standart sapmanın yapısını anlamak için tahmin değerlerinin logaritmasının yatay eksende yer aldığı artık çizimleri ile çalışılabilir. Standart sapmaların yapısının belirlenmesindeki yollardan biri de veri kümesini yatay eksende sıraladıktan sonra eşit gruplara bölmek, her bir gruba karşılık gelen mutlak artık ortalamalarını bulmak ve bu ortalamalara karşılık çizim yapmaktır (Carroll ve Ruppert, 1988).

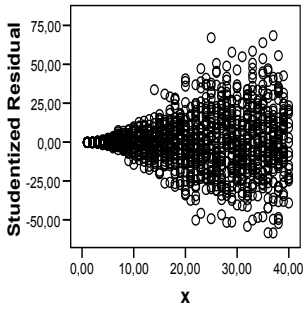
7. Büyük veri kümelerinde yoğunluk etkisinin giderilmesi konusunda ikinci yol, aynı zamanda değişkenlik ile ilgili olarak parametrik olmayan regresyon teknikleriyle model hakkında bilgi edinilmesidir. Parametrik olmayan yöntemler kullanıldığında değişkenlik hakkında bilgi edinme yollarından biri saçılım grafiği "scatter plot"tur. Artıkların mutlak, log-mutlak ve karelerinin küp köklerinin çizimi, varyans yapısının anlaşılmasında ve ortaya konmasında serpilme diyagramı gibi işlev görür. Mutlak

artıkların çeşitli dönüşümlerine uydurulacak fonksiyonlar (eğriler) standart sapmanın davranışının anlaşılmasına yardımcı olur. Mutlak artıkların logaritması kullanıldığında göz önünde tutulması gereken husus, çok küçük mutlak artıkların logaritmalarının büyük negatif değerlere neden olması dolayısıyla yanlış ve aşırı yorumlara yol açmasıdır. Bundan kaçınmanın yollarından biri en küçük mutlak artıkların az bir kısmının atılmasıdır (Carroll ve Ruppert, 1988).

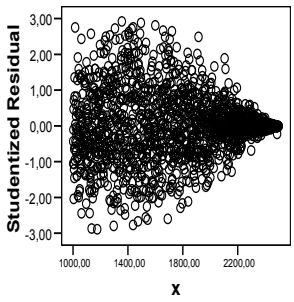
8. Artık varyanslarının davranışını ortaya koyabilmenin bir başka yolu, hareketli ortalama, medyanlar ve budanmış ortalamalar kullanılarak yapılabilir. Düzleyici olarak "spline" ya da "çekirdek" (kernel) düzlemesi kullanılabilir. Çekirdek düzlemesi yapılırken fonksiyon olarak uniform çekirdek kullanılabilir; bant genişliği ise deneme ya da hataya bağlı olarak seçilebilir. Çizimlere kabaca göz atılması, genelde şiddetli değişen varyans yargısına neden olur. Pratikte, birkaç büyük artık yokmuş gibi çizime bakıldığında, çizime ilişkin izlenim aynı kalıyorsa, ilk teşhise güvenilebilir (Carroll ve Ruppert, 1988).

2.2.1.1 Kelebek Dağılan Artıklar için Değişen Varyanslılık

Değişen varyanslılıkta en çok karşılaşılan durum, artık grafiklerindeki megafon biçiminde görüntüdür. Megafon biçimindeki şekiller (Şekil 2.1 ve 2.2); "galaksi" biçimindeki (Şekil 2.3) ya da "kelebek" biçimindeki (Şekil 2.4) diğer iki kalıp kadar istatistiksel çıkarsamayı bozmaz. Kelebek şeklindeki bir saçılım grafiğinin iki ucu geniş, ortası ise dardır. Galaksi ise bunun tersine orta kısım geniş ve uçlar dardır (Darlington,1997).



Şekil 2.1. Sağa Megafon

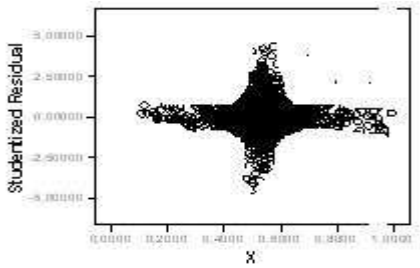


Şekil 2.2. Sola Megafon

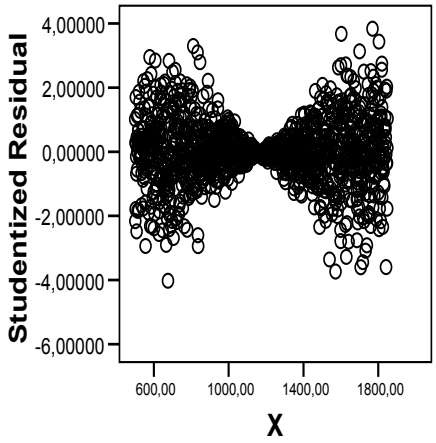
A değişkeni, B ve C değişkenlerinin bir fonksiyonu olsun. A, B'nin doğrusal artan ve C'nin de doğrusal azalan bir fonksiyonu olduğunda modelin artıkları kelebek şeklinde bir dağılıma gösterir.

Örneğin, bilgisayar tecrübesinin bir fonksiyonu olarak standart olmayan bilgisayar klavyeleriyle çalışma durumu göz önüne alınsın. Tecrübenin sağ el ya da sol el kullanıyor olmakla etkileşimli olduğunu varsayılsın. Böyle bir durumda solaklar standart olmayan klavyeler kullanımı konusunda deneyimleri artıkça daha olumlu; sağ ellerini kullananlar ise daha az olumlu gelişme kaydederler. İki regresyon doğrusunun ortalama tecrübe civarında birleştiği varsayılsın. O zaman durum-tecrübe saçılım grafiğinin doğruların birleştiği orta kısmı daha dar, sağ ve sol uçlarda ise çok daha geniş olacaktır. Araştırmacı el kullanma durumunu gözden kaçırıp yer vermez ise regresyon modelinde kelebek dağılan artıklar nedeniyle değişen varyanslılık ortaya çıkar. Bundan dolayı tecrübe için $V(\hat{\beta})$ olduğundan çok daha küçük tahmin edilecektir.

Regresyon modelinde yer alan bir değişkenle ilişkili olan başka değişkenlerin dışlanması da kelebek dağılan artıklara neden olabilir (Darlington, 1997). Kelebek dağılan artıklar için değişen varyanslılığı ortaya koyan yapı örneklerinden biri de atmosferik verilerin aylara göre sıralanarak modellenmesidir.



Şekil 2.3. Galaksi Dağılan Artıklar



Şekil 2.4. Kelebek Dağılan Artıklar

2.2.2. Değişen Varyanslılığın Belirlenmesinde Kullanılan Sayısal Yöntemler

Değişen varyanslılığın belirlenmesi konusunda özellikle ekonometri ve zaman serileri literatüründe geçen pek çok skor testi vardır. Burada, bu testlerin üç tanesine değinilecektir.

2.2.2.1. Spearman Sıra Korelasyonu

Carroll ve Ruppert (1988), sayısal tanı için Student-türü artıklarla yatay eksende yer alacak değişkenlerin Spearman sıra korelasyonu katsayılarının çok kullanılışlı olduğunu belirtmişlerdir. Bu korelasyon, değişken değerlerinin büyükten küçüğe sıralanması ve küçükten başlanarak sıra numarası verilmesine dayanır. Böylece Spearman korelasyonu iki değişkenin sıra numaraları korelasyonudur. Spearman korelasyonu monoton dönüşümlerden etkilenmez; dolayısıyla artıkların kareleri ya da logaritmalarıyla çalışıldığı gibi, $\hat{y}$ değerleri ya da bağımsız değişkenlerin logaritmaları ile de çalışılabilir. Pozitif katsayılar artan varyansı işaret eder. Sıra korelasyonuna benzer olarak artık çizimlerinden görsel yorumlar edinilebilir.

2.2.2.2. Değiştirilmiş Levene Testi

Değişen varyanslılığın saptanmasında kullanılan bu test, hata terimlerinin normal dağılıma uymasını gerektirmemektedir. Hata terimleri normal dağılmasa bile hataların büyüklüğüne dayanan çok sağlam bir testtir. Karesel artıkların (e_i^2) büyük olması büyük sapmalara neden olacağından s_y^2' de büyük olacaktır. Veri kümesi, n_1 ve n_2 şeklinde en az 2 gruba bölünür. Varyanslar, bağımsız değişkenin artan değerlerine karşılık büyürse, o zaman birinci grup artık kareleri toplamının ikinci

grup artık kareleri toplamından daha küçük ya da büyük olması beklenir. Test, birden çok değişken için de yapılabilmesine rağmen tek bağımsız değişken için uygulanması daha kolaydır. Testte, pozitif ya da mutlak değerli artıklar kullanıldığından sağlamdır. Testin işleyişi iki örneklem t testine dayanır. Mutlak sapmalar genellikle normal dağılıma uymaz; ancak tek tek her bir gruptaki birim sayısının 10'dan fazla olması durumunda yaklaşık olarak t dağılımına uyarlar.

H_0 : İki grubun varyansı arasında istatistiksel olarak anlamlı bir fark yoktur hipotezinin testi için,

n_1 : küçük x_i değerlerine karşılık gelen gruptaki birim sayısı,

n_2 : büyük x_i değerlerine karşılık gelen gruptaki birim sayısı,

$\tilde{e}_1$: 1. gruptaki artıkların ortancası,

$\tilde{e}_2$: 2. gruptaki artıkların ortancası,

$d_{i1} = | e_{i1} - \tilde{e}_1 |$ 1. grubun mutlak sapması,

$d_{i2} = | e_{i2} - \tilde{e}_2 |$ 2. grubun mutlak sapması,

olmak üzere,

$$s^2 = \frac{\sum_{i=1}^{n_1} (d_{i1} - \bar{d}_1)^2 + \sum_{i=1}^{n_2} (d_{i2} - \bar{d}_2)^2}{n_1 + n_2 - 2} \text{ iken}$$

$$t = \frac{\bar{d}_1 - \bar{d}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \textbf{(2.5)}$$

test istatistiği kullanılır. Burada, $t > t_{\alpha, n_1+n_2-2}$ ise H_0 hipotezi ret edilir.

Veri kümesinde aykırı değerlerin (outliers) olması, çoğu zaman maskeleme (gizleme) yüzünden değişen varyanslılık hakkında görsel olarak yanlış kararlara neden olabilir. Aşırı aykırı değerler atıldıktan ya da düzeltildikten sonra değişen varyanslılık olup olmadığı konusunda yargıya varılabilir. Değiştirilmiş Levene testinin aykırı değerler ve pürüzler olması durumunda bile doğru sonuç verdiği Paulson (2006), tarafından gösterilmiştir.

2.2.3.3. White Testi

Test, eş varyanslılık varsayımının geçerli olup olmadığının incelenmesinde en çok kullanılan testlerden biridir. Normallik varsayımına dayanmayan yardımcı bir

regresyon modelinin tahminini gerektir. 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:

$$e_i^2 = \alpha_0 + \alpha_1 x_{i1} + \ldots + \alpha_p x_{ip} + \gamma_1 x_{i1}^2 + \ldots + \gamma_p x_{ip}^2 + \ldots \theta_1 x_{i1} x_{i2} + \ldots . + \theta_p x_{ip-1} x_{ip} \qquad \textbf{(2.6)}$$

$H_0 : \forall(\alpha_{i,}\gamma_j, \theta_l) = 0$ (Eş Varyans)

$H_1 : \exists(\alpha_{i,}\gamma_j, \theta_l) \neq 0$

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

$nR^2_{yard.} \approx \chi^2_{k,\alpha}$

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

2.3. DEĞİŞEN VARYANSLILIĞIN SONUÇLARI

Dağılım kuramı, güven aralıkları ve hipotez testleri standart regresyon varsayımlarının geçerliliği halinde anlam taşır. Varsayımlardan sapma olduğunda sonuçlara ilişkin uygulamalar ciddi hatalara neden olmaktadır (Hadi ve Chatterjee, 2006). Basit ve çoklu regresyonda regresyon katsayılarının anlamlılığı ve testlerinin kesinliği eş varyanslılık varsayımına bağlıdır. Regresyon katsayılarının testi için Jackkinfe ve bootstrap gibi sağlam yöntemler kullanılsa bile değişen varyanslılık söz konusu olduğunda sonuçlar çok hatalı olabilir.

Değişen varyanslılık durumunda $\hat{\beta}$ kendi parametresinin yansız bir tahmin edicisidir; ancak standart hatası ($\hat{\sigma}_{\hat{\beta}}$) çok büyük ya da çok küçük olduğundan etkinliğini yitirir. Burada, $\hat{\sigma}_{\hat{\beta}} = \sigma\sqrt{köşegen(X'X)^{-1}}$ dir.

Eğer $\hat{\sigma}_{\hat{\beta}}$ olduğundan çok daha büyük tahmin edilirse $\hat{\beta}$ kendi parametresinin etkin bir tahmin edicisi olmaz ve $t = \hat{\beta} / \hat{\sigma}_{\hat{\beta}}$ istatistiği küçülür. Bu durumda, doğru olan alternatif hipotezin ret olasılığı olan II. Tip hata ortaya çıkar. Eğer $\hat{\sigma}_{\hat{\beta}}$ olduğundan

çok daha küçük tahmin edilirse $\hat{\beta}$ kendi parametresinin etkin bir tahmin edicisi olmaz ve $t = \hat{\beta} / \hat{\sigma}_{\hat{\beta}}$ hesap değeri büyür ve doğru olan bir yokluk hipotezinin ret olasılığı olan I. Tip hata ortaya çıkar (Friel, 2009).

3. VARYANS FONKSİYONLARI ve DEĞİŞEN VARYANSLILIĞIN DENGELENMESİ

Pek çok alanda, örneğin ekonomi, sağlık ve biyoistatistikte doğrusal modeldeki eş varyanslılık varsayımı geçerli değildir ve etkinlik açısından düzeltme gerektirir. Bu konudaki temel yaklaşım, bağımlı değişkenin dönüşümüne dayanan Box-Cox dönüşümüdür (Aitkin, 1987).

Pardoe ve Cook (2007), standart doğrusal ve daha karmaşık regresyon modellerinde tahmin belirsizliğinin doğru değerlendirilmesi için varyans fonksiyonunun belirlenmesinin çok önemli olduğunu belirtmişlerdir. Varyans fonksiyonu tahmini, regresyon analizinde olduğu gibi eğri uydurma ile yapılır. Regresyon analizinin pek çok boyutu olmasına rağmen üç tanesi göze çarpar. İlki, standart grafik tekniği olan saçılım grafiğidir. İkincisi, kanonik tahmin yönteminin varlığı olarak en küçük karelerdir. Aykırı değerlere ve etkili gözlemlere duyarlı olmasına karşın, en küçük kareler hemen hemen tüm regresyon analizlerinde temel olarak kullanılan ve anlaşılması kolay bir yöntemdir. Regresyon analizinin üçüncü bileşeni, doğrusal modeldir.

3.1 VARYANS FONKSİYONU TAHMİNİ

Eş varyanslılık varsayımı geçerli olmadığında varyansları dengelemek için en çok başvurulan yol dönüşümdür. Dönüşüm yeterli olmadığında varyansların modellenmesi gereklidir (Verbyla, 1993). Varyanslar, varyans fonksiyonları ile modellenir. Varyans fonksiyonu tahmininde, değişkenlerin bir fonksiyonu olarak varyansların yapısı anlaşılmaya çalışılır. Varyans fonksiyonu, regresyona benzer şekilde ele alınır. Tıpkı regresyonda olduğu gibi, yapıyı anlamak için basit çizim teknikleri yararlıdır. Burada amaç, ilk olarak konum (location) etkilerine uygun bir fonksiyon uydurduktan sonra artıkları düzgünleştirmektir. Carroll ve Ruppert (1982)'e göre, tahminin standart sapması hakkındaki bilgi artıkça regresyon parametrelerinin etkinliği artar.

Tahminin varyansının yapısını anlamada ilk adım varyansların logaritmalarının modellenmesidir. Modelleme ile ilgili bazı örnekler aşağıdadır:

$\log(e_i^2) = \theta_0 + \theta_1 x + \theta_2 x^{-1}$

ya da

$\log(e_i^2) = \theta_0 + \theta_1 \mu(\beta)$

ya da

$l\,\mathrm{og}(e_i^2) = \theta_0 + \theta_1 \log\left[\mu(\beta)\right]$.

Son model ortalama üslü bir modeldir. Varyansların logaritmalarını doğrusal olarak dikkate almanın avantajı, tahminle elde edilen varyansların da pozitifliğinin garanti olmasıdır. Log-doğrusal varyans modellerine alternatif modellerin çoğunda pozitif varyans garantisi yoktur. Bazı örneklemlerde varyans, sabite ve ortalamaya bağlı iki bileşene sahiptir. Bu durum

$$\sigma^2 g^2\left[\mu_i(\beta), z_i, \theta\right] = \theta_1 + \theta_2\left[\mu_i(\beta)\right]^{\theta_3} \qquad \textbf{(3.1)}$$

şeklinde gösterilir. Alternatif olarak standart sapmanın, bağımsız değişkenlerin karesel bir fonksiyonu biçiminde görsel olarak modellenmesine sıklıkla rastlanabilir:

$$\sigma_i = \sigma g\left[x_i, \theta\right] = 1 + \theta_1 x_i + \theta_2 x_i^2 \qquad \textbf{(3.2)}$$

Eğer burada birden çok bağımsız değişken varsa, model, bu değişkenlerin kendilerinin, karelerinin ve çapraz çarpımlarının yer aldığı bir doğrusal regresyona dönüşebilir.

Değişen varyanslı regresyon modellerinde hata terimlerine ilişkin varyanslar bilinirse, parametre tahminleri ağırlıklı en küçük kareler kullanılarak yapılabilir. Ancak pratikte hata terimlerinin varyansları bilinmez. İstatistik literatüründe ağırlık tahminlerinin kullanıldığı, asimptotik olarak ağırlıklı en küçük kareler tahminlerine denk sonuçlar veren pek çok çalışma vardır (Shick, 1999).

3.2. KARESEL ARTIKLARLA AĞIRLIKLI EN KÜÇÜK KARELER

Bu yöntem için hareket noktası, $E(e_i^2)$'nin yaklaşık olarak varyansa eşit olmasıdır. Bağımlı değişkenin karesel artıklar olduğu ve bu durum için regresyon

fonksiyonunun beklenen değerinin yaklaşık $\sigma^2 g^2\left(\mu(\beta), z, \theta\right)$ olduğu doğrusal olmayan bir regresyon fonksiyonu düşünülebilir. Böylece θ ve σ 'nın tahmini için

$$\sum_{i=1}^{N}\left\{\left[y_i - f(x_i, \hat{\beta}_*)\right]^2 - \sigma^2 g^2\left(\mu_i(\hat{\beta}_*), z_i, \theta\right)\right\}^2 \qquad \textbf{(3.3)}$$

eşitliği minimize edilir. Normal dağılan veri için karesel artıkların varyans fonksiyonu yaklaşık $\sigma^4 g^4\left(\mu_i(\hat{\beta}_*), z_i, \theta\right)$ ile orantılıdır. Bu nedenle genelleştirilmiş en küçük kareler yöntemine başvurulabilir. β 'nın ağırlıklı en küçük kareler ve genelleştirilmiş en küçük kareler tahmini, aşağıda verilen modellemeye paralel olarak elde edilebilir:

Bağımlı Değişken : $\left[y_i - f(x_i, \hat{\beta})\right]^2 \rightarrow e_i^2$ **(3.4)**

Parametreler : (σ, θ)

Regresyon Fonksiyonu : $\sigma^2 g^2\left(\mu_i(\hat{\beta}_*), z_i, \theta\right)$

Varyanslar : $\sigma^4 g^4\left(\mu_i(\hat{\beta}_*), z_i, \theta\right)$

Burada β_* varyans fonksiyonun parametresinin tahmini, z, varyansın fonksiyonu olduğu bağımsız değişken ve θ üs parametresidir. Bu eşitlikten elde edilecek $\widehat{e_i^2}$ tahmin değerleri $\hat{w}_i = 1/\widehat{e_i^2}$ şeklinde ağırlık elde etmek için kullanılır. Karesel artıklarla çalışmanın dezavantajı dağılımlarının oldukça çarpık olması ve bunun sonucunda da normal dağılım özelliklerinin kullanılamamasıdır. Örneğin, artıkların dağılımı yaklaşık normal varsayıldığında karesel artıklar yaklaşık 1 serbestlik dereceli Ki-Kare dağılımına uyar. Bu durumda karesel artıkların küp kökleri, küp kök Ki-Kare (χ^2) rasgele değişkeni dolayısıyla yaklaşık normal dağılımlı değişken olur (Carroll ve Ruppert, 1988).

3.3. MUTLAK ARTIKLARA DAYANAN YÖNTEMLER

En küçük karelerin uygulanmasında, aykırı değerlerin varlığı durumunda, aykırı değerler karesel artıklarda daha etkin olduklarından mutlak artıkların kullanılması önerilmektedir. Pardoe ve Cook (2007), varyans fonksiyonunun karekökünün

standart sapma fonksiyonunun tahmini olarak düşünülebileceğini belirtmişlerdir. Burada hareket noktası,

$$E\left(\left|y_i - f(x_i\hat{\beta})\right|\right) = c_f g(\mu_i(\hat{\beta}), z_i, \theta) \quad \textbf{(3.5)}$$

beklenen değerine sahip mutlak sapma olarak adlandırılan moment varsayımıdır. Burada c_f, x_i tasarım noktasından bağımsız bir orantı sabitidir. Birçok amaç için (2.1) ve (3.5)'deki ortalama ve varyans varsayımları eşanlı olarak sağlandığında (3.5) hataları bağımsız ve türdeş dağılır. Çarpık veri için ortanca, ortalamaya göre daha kolay yorumlanabilir bir ölçü olduğundan ortancanın tercih edilmesi önerilmektedir. Mutlak artıklar, yaklaşık olarak (3.5) ile verilen beklenen değere ve g^2 ile orantılı varyansa sahiptir. Bu, genelleştirilmiş en küçük karelerin mutlak artıklara uygulanmasının bir sonucudur. $\hat{\beta}$ için ağırlıklı en küçük kareler hesaplamaları,

Bağımlı Değişken $= \left|\left(y_i - f(x_i, \hat{\beta}\right)\right|$

Parametreler $= \left(c_f, \theta\right)$

Regresyon Fonksiyonu $= c_f g\left[\mu_i(\hat{\beta}), z_i, \theta\right]$

Varyans Fonksiyonu $= g^2\left[\mu_i(\hat{\beta}), z_i, \theta\right]$

olmak üzere,

$$\left|\left(y_i - f(x_i, \hat{\beta}\right)\right| = c_f g\left[\mu_i(\hat{\beta}), z_i, \theta\right]$$

regresyonundan elde edilen mutlak sapma tahminleri, ağırlıkların elde edilmesinde kullanılır. Öte yandan $\hat{\theta}$ tahmin edicisi elde edildikten sonra $\hat{\beta}$ tahmin edicisi genelleştirilmiş en küçük kareler ile güncelleştirilir. Bu işlem birkaç kez tekrarlanabilir; yeni bir $\hat{\theta}$ tahmin edicisi ve sonrasında $\hat{\beta}$ tahmin edicisi vb. şekilde sürdürülür. Yakınsama garanti değildir. Ancak simetrik dağılan hatalar için $\hat{\theta}$'nın iterasyonla tahmini çok daha etkin olur. Öte yandan Harvey (1976), mutlak artıkların logaritmasının mutlak artıkların yaklaşık beklenen değerlerinin logaritmasında regresyonu önermiştir:

$$\log\left(\left|y_i - f(x_i\hat{\beta})\right|\right) = \log\left[\sigma g\left(\mu_i(\hat{\beta}), z_i, \theta\right]\right. . \quad \textbf{(3.6)}$$

Hataların bağımsız ve türdeş dağıldığı varsayımıyla bu yaklaşık eş varyanslı bir regresyondur. Yürütülmesi yalnızca sıradan doğrusal olmayan en küçük kareler

gerektirdiğinden kolaydır. Buna karşılık artıklar sıfıra çok yakın ise, sıfıra çok yakın değerlerin logaritması büyüyeceğinden, sahte aykırı değerler sorunu çıkar ortaya. Bu potansiyel zorluktan kaçınmak için varyans modellemesi yapılırken birkaç çok küçük artığın silinmesi yoluna gidilmelidir

Carroll ve Davidain (1987), ağırlıklı artık karelerinin etkinliğini ya da θ'nın sözde olabilirlik tahminini ağırlıklı mutlak artıkları kullanarak uygulamışlardır. Eğer varyans fonksiyonu ortalama ile ilgili değilse ya da hatalar simetrik ya da σ küçükse

$$\text{Mutlak artik tahmininin asimptotik bağıl etkinliği} = \frac{[2+\kappa][1-\delta]}{4\delta} \qquad \textbf{(3.7)}$$

olduğunu göstermişlerdir. Burada κ hataların basıklığı, δ ise

$$r_i = \left[y_i - f(x_i,\beta)\right] \Big/ \left[\sigma g(\mu_i), z_i, \theta\right]$$

şeklindeki standardize artıkların varyansıdır.

3.4. ARDIŞIK ARTIKLARA DAYANAN YÖNTEMLER

Temelde en küçük kareler artıkları e_i, $(\mathrm{i} = 1,..,\mathrm{n})$ varyansların tahmini amacıyla kullanılabilir. Alternatif olarak Student-türü artıklar da $r_i = e_i \Big/ \sqrt{1-h_{ii}}$,

$h_{ii} = köşegen\left(X(X'X)^{-1}X'\right)$ ile hesaplanarak kullanılabilir. Student-türü artıklar hala korelasyonlu olabilir, ancak model doğru belirlenmiş ise eş varyanslıdır. Bir başka artıklar sınıfı da ardışık artıklardır. Kianifard ve Swallow (1996), en küçük kareler artıkları ya da Student-türü artıklardan farklı olarak ardışık artıkların eş varyanslı ve bağımsız olduğunu belirtmişlerdir. Model doğru olarak kurulmuş ise ardışık artıklar bağımsız ve türdeş dağılır. Ardışık artıklar, gözlemlerin ilk p baz ile (p parametre sayısıdır) ile başlanarak her adımda bir gözlemin eklenmesi ile hesaplanır. Her adımda hesaplanan regresyon eşitliği bir sonraki adımda eklenecek gözlem için artığın hesaplanmasında kullanılır. Bu süreç son artık hesaplanıncaya kadar devam eder (Rawlings ve diğ., 1998). Ardışık artıklar korelasyonsuz artıklar sınıfına girer ve en küçük kareler artıklarına göre hata varyanslarını daha kesin yansıtırlar.

ω_i, i'inci ardışık artığı göstermek üzere, algoritma aşağıdaki gibidir:

i = p to n-1

$\hat{\beta}_i = (X_i^{'} X_i)^{-1} X_i^{'} Y_i$

$$\hat{y}_{i+1} = x'_{i+1} \hat{\beta}_i \quad \textbf{(3.8)}$$

$$\omega_{i+1} = \frac{y_{i+1} - \hat{y}_{i+1}}{\sqrt{1 + x'_{i+1}(X'X)^{-1} x_{i+1}}} \quad \textbf{(3.9)}$$

end;

Burada, x_{i+1}: X matrisinin (i+1)'inci satırına karşılık gelen vektörüdür. Bu işlem her defasında 1 gözlem eklenerek (n-p) kez tekrarlanır ve (n-p) tane ardışık artık elde edilir. Ardışık artıklardan ilk p tane ardışık artık sıfırdır. Ardışık artıkların varyansı,

$$\hat{\sigma}_{\omega}^2 = \frac{\sum_{i=p+1}^{n} (\omega_i - \hat{\omega})^2}{(n - p - 1)} \quad \textbf{(3.10)}$$

dir. Burada $\hat{\omega} = \sum_{i=p+1}^{n} \omega_i \Big/ (n - p)$ dir. Varyansların tahmini için $\sigma_1^2, .., \sigma_n^2$'nin bilindiği ve ilk iterasyon için en küçük karelerin kullanıldığı göz önünde tutulmalıdır. Daha çok iterasyon için σ_i^2 mevcut tahmini $\hat{\sigma}_i^2$ ile yer değiştirir (Rawlings ve diğ., 1998).

3.5. DEĞİŞTİRİLMİŞ (MODIFIED) ARDIŞIK ARTIKLAR YÖNTEMİ

Bischoff ve diğ. (2006) tarafından önerilen yöntemin algoritması kesim 3.4'de açılanan algoritmayla aynıdır. Ek olarak, simetri için artık dönüşümleri kullanılmaktadır. İlk p gözlem (p, parametre sayısıdır) kullanılarak (p+1)'inci artık tahmin edilir. Daha sonra, her adımda bir gözlem eklenerek bu işlem (n-p) defa tekrarlanır ve (n-p) sayıda bağımsız artık elde edilir. Bu yöntemde, hatalar normal dağılımlı varsayıldığından 'ω' ardışık artık vektörü, "0" ortalamayla normal dağılıma sahip olur. Bundan dolayı karesel artıklar sıfıra çok yakın olduğundan $Ln(\omega^2)$ dönüşümü uygulandığında negatif değerlerin sayısı fazlalaşır ve artık dağılımındaki simetri bozulur. Bu problemden kaçınmak için alternatif Box-Cox dönüşümü kullanılabilir.

T(.) dönüşüm, z, dönüştürülecek öğeyi göstermek üzere,

$$T(z) = \ln(z + M),\ \text{M}=\left(M,....M\right)^T \in R^n; z = \omega^2 \quad \textbf{(3.11)}$$

$$M = \frac{\left(ortc\left(\omega_i^2\right)\right)^2}{\max(\omega_i^2)} \quad \textbf{(3.12)}$$

biçiminde verilir. Örneğin, mutlak ardışık artıklar dönüştürülecekse $z = |\omega|$ karesel ardışık artıklar dönüştürülecekse $z = \omega^2$ olur.

Bu işlemde dikkat edilmesi gereken nokta, varyansların, p sayıda bilinmeyen parametre vektörü $\hat{\beta}$ ile tahmin edildikleridir. Bu işlem, n - q (n >> q) tane ardışık artık ile yapılabileceği gibi n tane standart en küçük kareler artığı ile de yapılabilir. Bu durumda, p<(n-q) parametreye bağlı n tane $\hat{\sigma}_1^2,..,\hat{\sigma}_n^2$ varyans tahmini elde edilir.

Ardışık artıklar ile ilgili en büyük eleştiri, çok hesaplama gerektirmeleri ve başlangıç en küçük kareler artıklarıyla ortak bir taraflarının olmamasıdır.

3.6. YENİDEN AĞIRLIKLANDIRILMIŞ İTERATİF EN KÜÇÜK KARELER YÖNTEMİ

Meloun ve diğ. (2002), tarafından önerilen yöntem, verinin normalden başka bir dağılıma uyması, değişen varyanslılık, aykırı değerler ya da yüksek kaldıraç noktalarının varlığı durumunda kullanılmaktadır.

Regresyon parametreleri, yeniden ağırlıklandırılmış iteratif en küçük kareler yöntemi (YAIEKK) ile aşağıda verildiği gibi hesaplanır:

$$\sum_{i=1}^{n} w_i(r) y_i x_{ij} = \sum_{i=1}^{n} \sum_{k=1}^{m} w_i(r) x_{ij} x_{ik} \hat{\beta}_k, \quad r_i = e_i / \sigma,\ j = 1,.., p \quad \textbf{(3.13)}$$

Ortancaları kullanan sağlam yöntemin aşamaları aşağıdadır:

- En küçük karelerden elde edilen artıklarının ortancası bulunur
- j, iterasyon indisini göstermek üzere,

$$\hat{\sigma}^{(j)} = \frac{ortc\left|\hat{e}^{(j)} - ortc(\hat{e}^{(j)})\right|}{0.6745} \quad \textbf{(3.14)}$$

ya da $\hat{\sigma}^{(j)} = 2.1\ \text{ortc}\left|\hat{e}^{(j)}\right|$ alınır.

Burada,

$e^{(j)}$: j'inci iterasyon için elde edilen artıklar vektörü,

$\hat{\sigma}^{(j)}$: j'inci iterasyon için (3.14)'den hesaplanan standart sapma tahminidir.

0.6745 değeri hataları normal dağılıma sahip büyük örneklemler için $\hat{\sigma}$'yı artık standart sapmasına eşit kılma sabitidir.

- Ağırlıklar, $w_i^{(j)} = köşegen(r_i^{(j)} / \hat{\sigma}^{(j)})$ olarak alındığında, ağırlıklı en küçük kareler regresyonundan,

  $\hat{\beta}^{(j)} = (X'W^{(j)}X)^{-1}X'W^{(j)}Y$ hesaplanır.

  δ çok küçük bir sayı iken, $\hat{\beta}^{(j)}, \hat{\beta}^{(j-1)}$ ardışık iki parametre tahmini olmak üzere,

$\left|\hat{\beta}^{(j)} - \hat{\beta}^{(j-1)}\right| < \delta$ olmadıkça adımlar tekrarlanır.

3.7. CHATTERJEE ve MACHLER TARAFINDAN ÖNERİLEN ITERATİF YÖNTEM

Chatterjee ve Machler (1997) tarafından önerilen bu yöntemde en küçük kareler tahminlerine başvurulmaksızın doğrudan iteratif ağırlıklı en küçük kareler ile başlanır.

- İlk adımda, h_{ii}, $H = X(X'X)^{-1}X'$ matrisinin i'nci köşegen değeri ve

  $\tilde{h} = p / n$ olmak üzere $w_i^{(0)} = 1 / \max(h_{ii}, \tilde{h})$, ağırlık olarak seçilir. Bu başlangıç ağırlıkları ile

  $\hat{\beta}^{(0)} = (X'W^{(0)}X)^{-1}X'W^{(0)}Y$ başlangıç tahminleri bulunur;

  Burada, $W^{(0)} = köşegen(w_1^{(0)}, w_2^{(0)}, .., w_n^{(0)})$ dir.

- İkinci ve daha sonraki adımlar için en son regresyondan elde edilen artıklar kullanılarak yeni ağırlıklar hesaplanır işlemler j'inci iterasyon için aşağıdaki gibidir:

$$e^{(j-1)} = y - \hat{y}^{(j-1)} = y - X\hat{\beta}^{(j-1)} \quad \textbf{(3.15)}$$

$$w_i^{(j)} = (1 - h_{ii})^2 / \max\left\{\left|e_i^{(j-1)}\right|, ortc\left|e_i^{(j-1)}\right|\right\} \quad \textbf{(3.16)}$$

Bu işlem istenen yakınsama gerçekleşinceye kadar tekrarlanır. Yakınsama garantisi yoktur.

3.8. TEKRARLI GÖZLEMLER İÇİN ORTALAMA ve VARYANS FONKSİYONLARI

Amaç öngörü (prediction) olduğunda her $\underline{x}_i'$ bağımsız değişkenler satır vektörü için m_i sayıda y_i değeri elde edilir.

Başka bilgi olmaksızın öngörü değişkenlerinin her bir düzeyindeki örneklem varyanslarının tersi ağırlık alınarak β parametresi ağırlıklı en küçük karelerle tahmin edilir.

i'inci tekrarın örneklem ortalaması $\overline{y}_i$ iken,

örneklem varyansı $s_i^2 = (m_i - 1)^{-1} \sum_{j=1}^{m_i} (y_{i,j} - \overline{y}_i)^2$,

ağırlık tahminleri $\hat{w}_i = 1 / s_i^2$,

ve $\hat{W} = kösegen(w_1, w_2, .., w_n)$ olmak üzere β vektörü,

$$\hat{\beta}_w = (X'\hat{W}X)^{-1} X'\hat{W}Y \qquad \textbf{(3.17)}$$

ile tahmin edilir.

Tekrar sayısı genelde oldukça küçüktür. Tipik olarak 2 ya da 3 olabilir. Yöntemin dezavantajı, tekrar sayısı küçük olduğunda ağırlıkların çok zayıf bir biçimde tahmin edilmesidir (Carroll ve Ruppert, 1988).

3.9. BOOTSTRAP YÖNTEMİ

Regresyon modellerinde eş varyanslı olmayan hatalar durumunda tahmin, yeniden ağırlıklandırılmış en küçük kareler ya da bazı sağlam yöntemler kullanılarak yapılırsa tahmin varyanslarının değerlendirilmesi için standart yöntemler işlemeyebilir. Bootstrap, Efron tarafından bu tip karmaşık durumlarda tahminlerinin kesinliğini sağlamak için geliştirilmiştir. Bootstrap yöntemi, gözlenen örneklemden aynı büyüklükte rastgele iadeli örneklemlerin seçilmesidir. Bu yöntemde aynı gözlemin birkaç kez seçilebilmesi ya da orijinal örneklemde bulunan bazı gözlemlerin seçilmemesi olasıdır.

Basit doğrusal regresyon için orijinal örnekteki (x, y) çiftlerinin bootstrap örneklemesi kullanılır. İadeli bootstrap örnekleme ile n tane (x^*, y^*) çifti elde edilir. Bootstrap örneklemi oluşturulduktan sonra regresyon katsayıları standart yöntemler

kullanılarak bulunur. β parametresi için ilk bootstrap tahmini b_1^* bulunduktan, sonra n örneklem büyüklüğü için, yeteri kadar tekrarlanarak bir örnekleme dağılımı oluşturulur.

Kutner ve diğ. (2005), göre regresyon için bootstrap örneklemesinin ikinci bir yolu da vardır: Orijinal verilerden standart yöntemlerle elde edilen e_i artıkları örneklem kümesi kabul edilerek, bu artıkların örnekleminden rasgele iadeli örneklem seçilir. Örneklemeden elde edilen n sayıda e_1^*, e_2^*,.., e_n^* bootstrap artığı, orijinal $\hat{y}_i$'lere

$$Y_i^* = \hat{y}_i + e_i^* \qquad \textbf{(3.18)}$$

şeklinde eklenerek Y_1^*, Y_2^*,..,Y_n^* bootstrap değerleri elde edilir. Elde edilen bu değerler yeni bağımlı değişken gibi dikkate alınır ve bağımsız değişkenler ile regresyonu yapılarak b^* bootstrap parametre tahmin edicisi elde edilir. Bu süreç yeteri sayıda tekrarlanarak b^*'nın örnekleme dağılımı oluşturulur. Örneklem çekme işlemi bazen çok az sayıda örneğin 50 kez yapılır. Genelde 200-500 sayıda yineleme yeterli olur. Bootstrap tahminlerinin doğruluğu, yineleme sayısı artıkça $S^*\{b^*\}$'nın kararlı hale gelmesiyle gözlenebilir. $S^*\{b^*\}$, tüm bootstrap tahminlerinin standart sapması aynı zamanda $\hat{\beta}$'nın örnekleme dağılımının değişkenliğinin tahmini, yani kesinliğinin ölçüsüdür.

Bootstrap, yaklaşık güven aralıklarının elde edilmesi için de kullanılır. Bootstrap güven aralıklarının oluşturulmasında pek çok yöntem vardır. Basitçe, $\%(1-\alpha)$'lık güven aralığı oluşturmak için kullanılan yöntemlerden biri yansıma (reflection) yöntemidir. Yöntemin her zaman olmasa da genelde iyi sonuçlar verdiği belirtilmektedir. β için yansıma güven aralığı, b^*'ın bootstrap dağılımının $(\alpha/2)100$ ve $(1-\alpha/2)100$ yüzdeliklerine karşılık gelir. Bu yüzdelikler,

$$d_1 = b - b^*(\alpha/2)$$
$$d_2 = b^*(1-\alpha/2) - b$$

biçimde hesaplanarak β için $\%(1-\alpha)$'lık güven aralığı

$$b - d_2 \le \beta \le b + d_1 \qquad \textbf{(3.19)}$$

olarak elde edilir.

3.10. JACKKNIFE YÖNTEMİ

Bazı istatistiksel tahmin problemlerinde, teorik karmaşıklık ya da istatistiksel model ile ilgili bilgi eksikliği, tahminin standart hata ve yanlılık (bias) hesaplanmasını engeller. Jackknife yöntemleri, hem yanın hem de standart hatanın sayısal yaklaşımlarla doğrudan elde edilmesini sağlar ve güven sınırları verir. Yinelemeli örnekleme yöntemlerinden biri olan Jackknife istatistiği θ, n olası örneklem büyüklüğü ile her adımda bir gözlem dışarıda bırakılarak (n-1) gözlem için hesaplanır (Hinkly, 2010). Wu (1986), Jackknife yönteminin düzgün (smooth) tahmin ediciler için varyans ve yanlılık tahmininde iyi işlediğini belirtmiştir. Ancak tutarsız varyans tahminleri veren örneklem ortancası gibi düzgün olmayan tahmin ediciler için Jackknife yöntemi iyi sonuçlar vermez. Yöntemin bir diğer dezavantajı ise orijinal $\hat{\theta}$'dan i. gözlemin silinmesiyle elde edilen $\{\theta_{\{i\}}\}_i^n$'lerin normalize edilmiş histogramı standart normale yaklaşmaz. Bu nedenle de varyans tahminleri yapılmadan geçerlilik aralıkları oluşturulamaz. Klasik $y_i = x_i^{'}\beta + \varepsilon_i$ doğrusal regresyon modelinde, ε_i hata terimleri sıfır ortalama ve σ_i^2 varyansı ile aralarında ilişkisizdir. Tekil olmayan $(X'X)$ matrisi için β'nın en küçük kareler tahmin edicisi $\hat{\beta} = (X'X)^{-1}X'Y$ dir. β'nın tahmin edicisi olan $\hat{\beta}$, (x_i, y_i) çiftinin silindiği örneklemden $\hat{\beta}_{(i)}$ Jackknife tahmin edicisi yeniden hesaplama ile

$$\hat{\beta}_{(i)} = (X_{(i)}^{'}X_{(i)}^{'})^{-1}X_{(i)}^{'}y_{(i)}^{'} \qquad \textbf{(3.20)}$$

formülünden elde edilir. β'nın doğrusal olmayan $\theta = g(\beta)$ şeklindeki fonksiyonunun tahmini $\hat{\theta} = g(\hat{\beta})$ için, Jackknife tahmin edicisi $\hat{\theta}_{(i)} = g(\hat{\beta}_{(i)})$ ve sözde değerler (pseudo value) $p_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}$ dir. θ'nın Jackknife nokta tahmin edicisi $\tilde{\theta} = n^{-1}\sum_1^n p_i$ ve $\hat{\theta}$ için varyans tahmin edicisi,

$$v_j = \frac{1}{n(n-1)}\sum_1^n (p_i - \tilde{\theta})(p_i - \tilde{\theta})' \qquad \textbf{(3.21)}$$

ile elde edilir.

3.11. PARAMETRİK OLMAYAN DÜZLEME

Parametrik yöntemlere bir başka alternatif yaklaşım olarak varyanslar parametrik olmayan regresyon teknikleriyle tahmin edilebilir. Varyansların nasıl belirleneceğinin bilinmediğini ve bunun yerine yalnızca bilinmeyen bir g_0 fonksiyonunun olduğunu ve bu fonksiyon için,

$$st.sapma(y_i) = g_0(x_i) \qquad \textbf{(3.22)}$$

ya da

$$st.sapma(y_i) = g_0\left(\mu_i(\beta)\right) \qquad \textbf{(3.23)}$$

biçiminde bir ilişkinin var olduğu varsayılsın. Sorun, (3.22) ya da (3.23) durumuna ek olarak ağırlıklar bilindiğinde bir β tahmin edicisinin hesaplanmasının olası olup olmadığıdır. Her bir x_i tasarım noktasında gözlem tekrarı sayısı m_i olsun. Pratikte yineleme sayısı genelde küçük olur. İlk olarak veriye ağırlıksız en küçük kareler uygulanır ve mutlak sapmalar,

$$d_{ij}(\hat{\beta}_{ls}) = \left| y_{ij} - f(x_i, \hat{\beta}_{ls}) \right| \qquad \textbf{(3.24)}$$

biçiminde hesaplanırsa, x_i noktasındaki varyans tahmini,

$$\hat{\sigma}_i^2 = m_i^{-1} \sum_{i=1}^{m_i} d_{ij}^2(\hat{\beta}_{LS}) \qquad \textbf{(3.25)}$$

olur. Tahmini ağırlıklar $w_i = 1.0/\hat{\sigma}_i^2$ biçiminde bulunarak genelleştirilmiş en küçük kareler uygulanır. Bu ağırlıkların çok büyük bir bölümü, ağırlıklar küçük serbestlik derecelerine sahip olduklarından güvenilir değildir. β için asimptotik varyanstaki artış, tutarsız fonksiyon kullanımı nedeniyle önemli olabilir. (3.25)'deki varyans tahminleri yeteri kadar düzgün değildir ve β'nın genelleştirilmiş en küçük kareler sonuçlarının zayıf olmasına yol açar. Bu durumda x'e çok "yakın" tasarım noktaları en büyük ağırlığı alır. Varyans fonksiyonu tahmin yöntemlerinden biri,

$$Varyans(y_i) = E\left(y_i^2 \mid x_i\right) - \left[E\left[y_i \mid x_i\right]\right]^2$$

dir. Bu yaklaşım, verilen x_i için y_i^2 ve y_i' nin koşullu beklenen değerinin tahminiyle parametrik regresyon uygulamasını öneren bir yöntemdir. Ortalama fonksiyonunun $f(x_i, \beta)$ olduğuna dair kuşku yoksa o zaman parametrik olmayan bir regresyon uygulaması,

$$Varyans(y_i) = E\left(d_i^2(\beta) \mid x_i\right) \tag{3.26}$$

ilişkisinin kullanılmasıyla daha etkin olur. x_{ij}, p elemanlı bir vektör iken β'nın bilindiği, varyansların (3.26) modeline uyduğu ve $d_i^2(\beta)$ terimlerinin ağırlıklı ortalaması olduğu varsayılsın. O halde, $g_x(0)$ için $\sum_{i=1}^{N} c_{i,n}(x) = 1$ olmak üzere varyansların $d_i^2(\beta)$ kullanılarak tahmini,

$$\hat{g}_0^2(x, \beta) = \sum_{i=1}^{N} c_{i,n}(x) d_i^2(\beta) \tag{3.27}$$

olur.

Burada $c_{i,n}(x)$, $h \to 0$, yerel (local) ortalama miktarını kontrol eden bant genişliği (düzleme parametresi); $a_{i,n}(x)$, K(.) orijinde tek tepeli ve p boyutlu

$$a_{i,n}(x) = K\left(\frac{x - X_{i1}}{h}, \frac{x - X_{i2}}{h}, ..., \frac{x - X_{ip}}{h}\right)$$

şeklinde yaygınlık fonksiyonu olmak üzere,

$$c_{i,n}(x) = \left(a_{i,n}(x) \Big/ \sum_{j=}^{n} a_{j,n}(x) \right) \tag{3.28}$$

dir.

h'nın büyük değerleri için fonksiyon daha çok düzgünleşir ve yanlılık artar bundan dolayı sabit bir n için $h \to \infty$'a ıraksadığında,

$$\hat{g}_0^2(x, \beta) \to n^{-1} \sum_{i=j}^{N} d_j^2(\beta)$$

olur.

Varyans fonksiyonunun tutarlı bir tahminini elde etmenin bir yolu çekirdek fonksiyonlarının kullanımıdır (Carroll ve Ruppert, 1988). Çekirdek fonksiyonları, simetrik olup olasılık fonksiyonları özelliklerini taşırlar. Çizelge 3.1'de verilen bu

fonksiyonlar, en iyi oldukları için değil kolay anlaşılır ve diğer yöntemlere aracılık yapmaları nedeniyle tercih edilirler (Kıroğlu, 2001).

Çizelge 3.1. Çekirdek Fonksiyonları

Çekirdek	Fonksiyon
Epanechnikov	$K = \frac{3}{4}(1\text{-}t^2)\ (I(\lvert t\rvert \le 1))$
İki Ağırlıklı (Biweight)	$K = \frac{15}{16}\left(1-t^2\right)^2\ \ (I(\lvert t\rvert \le 1))$
Üçgensel (Triangular)	$K = (1\text{-}\lvert t\rvert)(I(\lvert t\rvert \le 1))$
Normal (Gaussian)	$K = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$
Uniform (Rectangular)	$K = \frac{1}{2}(I(\lvert t\rvert \le 1))$
Üç Ağırlıklı (Triweight)	$K = \frac{35}{32}\left[1-t^2\right]^3 (I(\lvert t\rvert \le 1))$
Cosinus	$K = \frac{\pi}{4}\cos\left(\frac{\pi}{2}t\right)\ \ (I(\lvert t\rvert \le 1))$

Çizelge 3.1'deki I(.) gösterge fonksiyon olup, koşul gerçeklendiğinde 1, aksi halde 0'dır (Eviews 4.1 User Guide).

Genel olarak bir f(x) fonksiyonunun çekirdek tahmin edicisi,

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

şeklinde tanımlanır.

Bu fonksiyon,

$K(u) \ge 0$ pozitif,

$K(u) = K(-u)$ simetrik,

$\int_{-\infty}^{\infty} K(u)du = 1$ olasılık yoğunluk fonksiyonu,

$\int_{-\infty}^{\infty} K^2(u)du < \infty$ sonlu olma,

$\int_{-\infty}^{\infty} uK(u)du = 0$,

özelliklerine sahiptir.

Bant genişliğinin seçimi oldukça önemlidir. Bant genişliği, fonksiyonun düzleme miktarını kontrol eder. Küçük bant genişlikleri sivri ve pürüzlü histogramlar verirken, büyük bant genişlikleri daha düzgün histogramlar verir (Kıroğlu, 2001).

$\hat{\beta}$, β'nın bir tahmin edicisi olsun. Ağırlıksız en küçük karelerin uygulanması olasıdır ama etkili değerlerden etkilendiğinden tercih edilmez. Değişen varyanslılıktan dolayı, ağırlıksız en küçük kareler yüksek kaldıraç noktalarında aykırı değerleri bulmaya yönelir. Bu durumun önüne geçmenin bir yolu, varyansları çekirdek fonksiyonları ile tahmin etmektir. Çekirdek varyans fonksiyonun biçimi $\hat{g}_0^2(x,\beta)$ ve ağırlık fonksiyonunun biçimi,

$$\hat{w}(x,\hat{\beta}) = \left[\hat{g}_0^2(x,\beta) + \eta_n\right]^{-1} \qquad \textbf{(3.29)}$$

dir. Burada η_n, bant genişliğine göre sıfıra daha yavaş yakınsayan ve işlevi, $\left[\hat{g}_0^2(x,\beta) + \eta_n\right]$ ifadesinin sıfır olmasını engelleyerek sonsuz ağırlıkların oluşmamasını sağlayan pozitif sayılar dizisidir (Carroll ve Ruppert, 1988).

3.12. DÖNÜŞÜMLER

Regresyon analizinde bağımlı değişken dönüşümü, çoğunlukla üç farklı amaç için yapılır: Normal dağılan toplamsal hatalar, artık varyanslarını dengeleme ve bağımsız değişkenlerle temsil edilen basit bir doğrusal model elde etme (Hamasaki ve Goto, 2005). En sık başvurulan dönüşüm, Box ve Cox (1964) tarafından geliştirilen bağımlı değişken dönüşümlerdir. Varyansın bağımlı değişkenin artan değerlerine paralel olarak arttığı durumlarda kullanılan bu dönüşüm, pozitif değerli değişkenlere uygulanabilmektedir. Bu dönüşümler, Y artarken Y'nin varyansı da artıyorsa varyansı dengelemek için; Y'nin hataları sağa çarpıksa Y'nin dağılımını normalleştirmek ya da bağımlı değişken ile bağımsız değişken arasında sürekli artan bir eğim söz konusu olması durumunda ise modeli doğrusallaştırmak için kullanılmaktadır (Albayrak, 2008).

Tek başına değişen varyanslılık, ağırlıklandırma teknikleriyle düzeltilebilir; ancak yalnızca varyansların tersi, hatalar normal dağılmadığında uygun olmaz (Carroll ve Ruppert, 1988). Çarpık verinin modellenmesi konusunda çok iyi kurgulanmış iki yol vardır: İlkinde hata paylarının yoğunluğuna çarpık parametrik yoğunluk fonksiyonu

(örneğin gamma) ile yaklaşılır (McCullagh ve Nelder, 1983). Bu yaklaşım, Y'nin simetrik bir dağılıma dönüştürülebildiğini varsayar.

$h(y,\lambda)$, λ'ya bağlı Y rastlantı değişkeni için dönüşümleri göstersin. Çoğu kez λ bir skalerdir. Bu aile genelde $h(y,\lambda)$'nın Y'de monoton olmasını gerektirir; aksi halde, Y için $h(y,\lambda)$ ile ters dönüşüm olmaz.

En çok kullanılan dönüşüm ailesi, Box ve Cox'un (1964) değiştirilmiş üs (modified power) dönüşümüdür:

$$h(y,\lambda)=\begin{cases}\left(y^{\lambda}-1\right)/\lambda & \lambda\neq 0 \text{ ise}\\ \log y & \lambda=0 \text{ ise}\end{cases} \qquad \textbf{(3.30)}$$

Bir başka olası üs dönüşümü ise,

$$h(y,\lambda,\mu)=\left(y-\mu\right)^{(\lambda)} \qquad \textbf{(3.31)}$$

şekindeki dönüşümüdür.

(3.30)'un kullanımı Y değerlerinin pozitif olmasını varsayar. Eğer Y, bir minimum değere sahipse, o zaman $\left(y-\mu\right)$'nün uygun bir μ seçimiyle pozitif olması sağlanır. Dönüşüm ailesinin seçiminden sonra hataları, 0 ortalama ve σ^2 varyansla normal dağılan toplamsal $Y=f(x,\beta)+\varepsilon$ modeli,

$$h(y,\lambda)=h(f(x,\beta),\lambda)+\varepsilon \qquad \textbf{(3.32)}$$

modeli ile yer değiştirir. Model, Y ve $f(x,\beta)$'ye bazı dönüşümler uygulandıktan sonra artıkların eş varyansla normal dağıldığını ifade eder. (3.32) modelinin hareket noktası, $\varepsilon\equiv 0$ olduğunda hiçbir değişim kaynağı yoksa (3.32) modeli $y=f(x,\beta)$'eşitliğine denktir. $h(y,\lambda)$ bazı λ değerleri için y'de doğrusalsa o zaman (3.32) modeli $y=f(x,\beta)$ modelini özel bir hal olarak kapsar ki veri böyle bir göstergeye sahip olduğunda dönüşüm yapmama seçeneği ortaya çıkar. Örneğin, değiştirilmiş üs dönüşümü ailesi kullanıldığında "dönüşüm ihtiyacı yok" şeklinde ifade edilebilen yokluk hipotezi, $H_0:\lambda=1$ ile test edilir.

Regresyon eşitliğini korumak için, eşitliğin her iki tarafına dönüşüm uygulanabilir. Artık değişimi çarpıklık ya da eş varyanslı olmayan bir manzara sergilediğinde Y'ye basitçe $f(x,\beta)$ ile ağırlıksız olarak eğri uydurmak yerine çift yanlı dönüşüm yöntemleri kullanmak için pek çok neden vardır. İlk olarak normalleştirme ve eş

varyansa dönüşüm yapmak, β parametre tahmin edicilerinin daha etkin olmasını sağlar. Herhangi bir dönüşüm olmaksızın β'nın standart en küçük kareler tahmin edicisi büyük varyansa sahip olur. Dahası çift yanlı dönüşüm modeli, verilen X için Y'nin tam koşullu bir modeline olanak sağlar. Böylece ortalama için güven aralıkları ya da Y'nin çeyrekleri kadar Y için öngörü aralıkları ve verilen X için kalibrasyon aralıkları mevcut olur. $\hat{\beta}$, β'yı çok iyi tahmin etse bile standart en küçük karelere dayalı güven, öngörü ve kalibrasyon aralıkları, artık değişimi normalden uzak ve varyans dengeli değilse çok büyük yanılgılara neden olur.

Dönüşümler normalliği ve eş varyanslılığı sağlayabilse bile verilen herhangi bir veri kümesi için tek bir dönüşüm her ikisini garanti etmez. Çarpıklığı gidermek ve varyansı dengelemek için eş anlı olarak ağırlıklandırma ve dönüşüm gerekebilir. Aykırı değerler, dönüşüm parametreleri tahmini üzerinde çok şiddetli etki yapar; örneğin tek bir aykırı değer normal ve eş varyanslı veri kümesinde çok büyük bir çarpıklığa ve görünür bir değişen varyanslılığa neden olabilir. Karşıt olarak, sağa çarpıklık sol kuyruktaki tek bir aykırı değer tarafından maskelenebilir ve küçük varyansa sahip bir gözlemde aykırı değer ortaya çıkarsa değişen varyanslılık gizlenebilir.

Doğrusallaştırma dönüşümleri en küçük karelerin etkinliğini çok azaltarak asimetriye ya da değişen varyanslılığa neden olur. Kimi zaman doğrusallaştırma şiddetli değişen varyanslılığa ve fiziksel olarak mümkün olmayan bir parametre tahminine neden olur (Carroll ve Ruppert, 1998).

Çift yanlı dönüşümlerin bir örneği, dönüştürülmüş yanıt $y^{(\lambda)}$'nın dönüştürülmüş bağımsız değişken $x^{(\lambda)}$ ile

$$y^{(\lambda)} = \alpha_0 + \alpha_1 x^{(\lambda)} + \varepsilon \qquad \textbf{(3.33)}$$

biçimindeki doğrusal regresyonudur.

Dönüşümler, yalnızca eş varyanslı ya da normal dağılan hata terimleri elde etmek için yapılmaz. Bazı veri kümelerinde bağımlı ve bağımsız değişkenlerin dönüşümleri, veriye çok iyi uyan bir model sonucu verdiği için yapılabilir (Ryan, 1997) .

3.13. AĞIRLIKLANDIRMA ve DÖNÜŞÜMLERİN BİRLEŞTİRİLMESİ

Bazı veri kümeleri için tek bir dönüşümün normalliği sağlayamaması ve varyansı dengeleyememesi çok beklenmedik bir durum değildir. Aslında beklenmedik olan tek bir dönüşümün her iki durumu yerine getirmesidir. Buna rağmen bazı durumlarda eş varyansla normale yakın hatalar elde etmek için hem dönüşüme hem de ağırlıklandırmaya başvurmak gerekebilir. Bu, özellikle varyans bir değişkene bağlı olduğunda doğrudur (Carroll ve Ruppert, 1988).

Bazı veri kümlerinde her ikisi muhtemelen gerekli olmamasına rağmen dönüşüm ya da ağırlıklandırmadan hangisinin tercih edilmesi gerektiği açık değildir. Bu durum varyans ortalamanın monoton bir fonksiyonu olduğunda daha çok geçerlidir. Willet ve Singer (2002) tarafından ifade edildiği üzere ağırlıklandırma için uygun bir yöntem olmadığında ağırlıklandırma stratejisinin uygulanması, ağırlıklı en küçük kareler sonuçlarında çok büyük belirsizliklere neden olur.

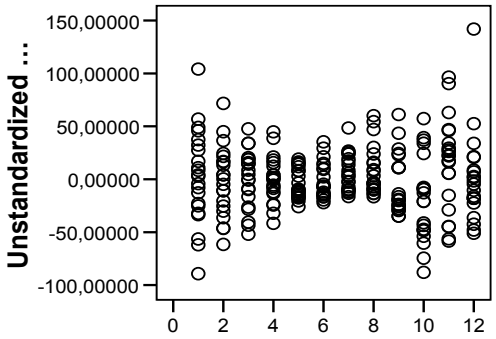
3.14. KELEBEK DAĞILAN ARTIKLAR İÇİN DEĞİŞEN VARYANSLILIĞI DENGELEMEDE ALTERNATİF YÖNTEMLER

Bu bölümde, kelebek dağılan artıklar için değişen varyanslılığın dengelenmesinde iki yönteme ek olarak bir de araç yöntem önerilmiştir. Önerilen yöntemler ve ayrıntıları aşağıda açıklanmıştır.

3.14.1. Dışsal Değişken ve Merkezileştirme

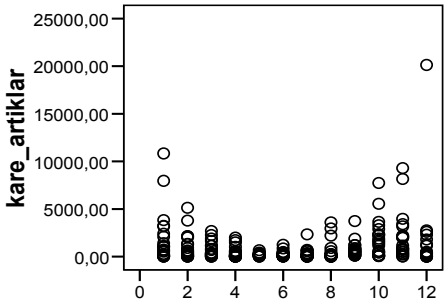
İstatistik literatüründe var olan değişen varyanslılık kalıpları, genel olarak monoton azalan ya da monoton artan varyanslar şeklindedir. Buna koşut olarak varyans dengeleme teknikleri de monoton artan ya da monoton azalan varyansı dengelemeye yöneliktir. Bu çalışmaya konu olan kelebek dağılan artıklar konveks bir yapıya sahiptir. Konveksliği basitçe şu şekilde tanımlarız: y = f(x) fonksiyonu, [a, b] aralığında 2'nci türeve sahip ve $\forall x \in [a,b]$ için f''(x) > 0 ise f(x) fonksiyonu, bu aralıkta konvekstir; $\forall x \in [a,b]$ için f''(x) < 0 ise f(x) fonksiyonu bu aralıkta konkavdır.

Artıklar kelebek şeklinde bir dağılma gösterdiklerinde (Şekil 3.1), artıkların varyans ya da standart sapma fonksiyonları "azalan – artan" şeklinde monoton olmayan bir yapı sergiler.

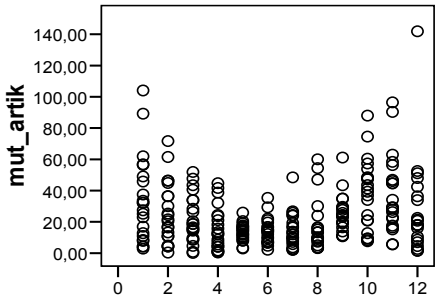


Şekil 3.1.Aylar - Ham Artıklar

Artıklar kelebek şeklinde bir dağılıma sahip olduğunda, varyansların tahminleri olan artıkların kareleri ve standart sapmaların tahminleri olan mutlak artıklar, kolları yukarı doğru olan 2. dereceden bir karesel forma uygun dağılma gösterirler (Şekil 3.2 ve 3.3) .



Şekil 3.2. Aylar - Karesel Artıklar



Şekil 3.3. Aylar - Mutlak Artıklar

Bu bölümde yukarıda belirtilen bilgiler doğrultusunda, ham, mutlak ve karesel artıkların çizimleri dikkate alınarak alternatif yöntemler önerilmiştir.

Kelebek dağılan artıklar için mutlak ya da karesel artık çizimleri dikkate alındığında artık varyanslarının uçlarda büyüdüğü orta bölümde ise küçüldüğü görülmektedir. Öyle bir değişken tanımlansın ki bu değişken uç kısımlarda, orta kısımlara göre daha

büyük değerler alsın ve ağırlıklar tahmin edilirken monoton olmayan varyanslara duyarlı olsun. Aşağıda algoritması verilen yöntem bu algıya dayanmaktadır.

Algoritma:

1. $Y = f(X, \beta) + \varepsilon$ regresyon fonksiyonu için Y_{nx1} bağımlı değişken, X_{nxp} tasarım matrisi ve $\beta_{(p+1)x1}$ boyutlu parametre vektörü olmak üzere, en küçük kareler parametre tahmin edicisi $\hat{\beta} = (X'X)^{-1}(X'Y)$, bağımlı değişkenin tahmin değerleri $\hat{Y} = X\hat{\beta}$ ve artıklar $e_i = y_i - \hat{y}_i$ eşitlikleriyle elde edilir.

2. $d = \{1, 2, .., \mathrm{n}\}$ şeklinde yeni bir sıra değişkeni tanımlanır. Burada n, veri kümesindeki gözlem (satır) sayısıdır. Yeni değişkenin kullanıldığı, mutlak artıklar $|e_i|$ için standart sapma fonksiyonu,

$$\widehat{s_i} = |e_i| = \alpha_0 + \alpha_1 d_i + \alpha_2 d_i^{\,2} \qquad \textbf{(3.34)}$$

ya da karesel artıklar e_i^2 için varyans fonksiyonu

$$\widehat{s_i^2} = e_i^2 = \alpha_0 + \alpha_1 d_i + \alpha_2 d_i^{\,2} \qquad \textbf{(3.35)}$$

olarak yazılır.

Bu fonksiyonların dışsal değişken d 'ye göre 1. dereceden türevi

$$\frac{\partial s}{\partial d} = \alpha_{_1} + 2d\alpha_2$$

dir. Türev fonksiyonunun sıfır olduğu nokta kelebek dağılan artık yüzeyinin en dar noktasına karşılık gelen merkez (boğum noktası) olacaktır. Benzer şekilde galaksi dağılan artıklar söz konusu olduğunda merkez, dağılım yüzeyinin yaygınlığının en fazla olduğu noktaya karşılık gelecektir.

$$m = -0.5\frac{\hat{\alpha}_1}{\hat{\alpha}_2} \qquad \textbf{(3.36)}$$

$\hat{\alpha}_1$ ve $\hat{\alpha}_2$ değerleri (3.34) ya da (3.35) eşitliklerinin en küçük kareler çözümünden elde edildikten sonra (3.36) formülüyle boğum noktası hesaplanır.

3. Boğum noktasına göre dönüştürülmüş yeni merkezi dışsal değişken,

$$o_i = (d_i - m) \qquad \textbf{(3.37)}$$

olur**.**

4. Önceki adımda elde edilen "o_i" merkezi değişkenin bağımsız değişken olduğu ve standart sapma fonksiyonu (Şekil 3.4) için mutlak artıkların bağımlı değişken olduğu,

$$\hat{s}_i = |e_i| = \theta_1 o_i + \theta_2 o_i^2 \qquad \textbf{(3.38)}$$

ya da varyans fonksiyonu için karesel artıkların bağımlı değişken olduğu,

$$\widehat{s_i^2} = e_i^2 = \theta_1 o_i + \theta_1 o_i^2 \qquad \textbf{(3.39)}$$

şeklinde regresyon modelleri kurulur. Bunların çözümünden, $\widehat{|e|}$ 'ya dayalı ağırlık tahminleri,

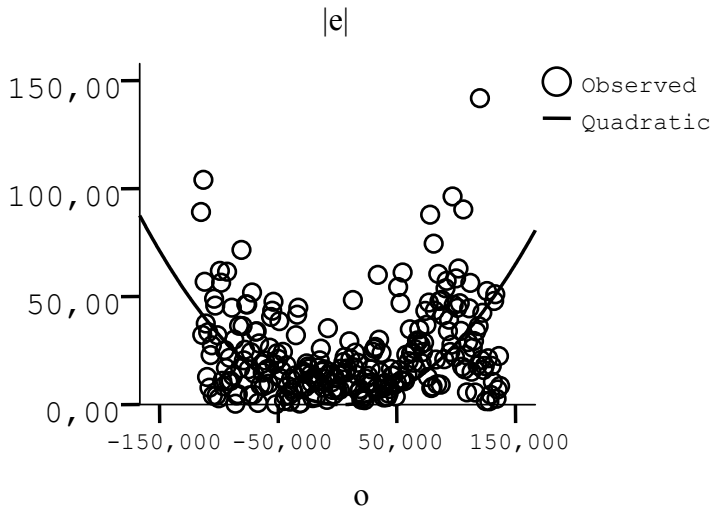
$$\hat{w}_i = 1/\widehat{|\hat{e}_i|}, \qquad \widehat{|\hat{e}_i|} \neq 0 \qquad \textbf{(3.40)}$$

ya da $\widehat{e^2}$ 'ya dayalı ağırlık tahminleri,

$$\hat{w}_i = 1/\widehat{e_i^2}, \qquad \widehat{e_i^2} \neq 0 \qquad \textbf{(3.41)}$$

ile elde edilerek $\hat{\beta}_w = (X'\hat{W}X)^{-1} X'\hat{W}Y$ ağırlıklı en küçük kareler uygulanır.

Ağırlık tahminleri için karesel ya da mutlak artıklardan hangisinin tercih edileceği kararı, görsel olarak grafiklere bakılarak ve ya belirtme katsayısı dikkate alınarak verilebilir.



Şekil 3.4. Mutlak Artıklardan Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu

Böylelikle modelde yer almayan bir değişken ağırlık tahmininde kullanılarak modele dâhil edilmiş olur.

3.14.2. Merkezi Dışsal Değişken ile Doğrudan Ağırlık Tahmini

Uygulamada da görüleceği üzere, verinin boğum noktasının asimetrik olması durumunda, veriye bağlı olarak karesel artıkların ya da mutlak artıkların dışsal değişkenle ağırlık tahminlerinin elde edilmesinde kullanılması, boğum noktası çevresindeki artıkların sıfıra çok yakın olması, bu artıklara karşılık gelen gözlemlerin çok aşırı ağırlık almasına neden olabilir. Böyle bir durumda aykırı değerler sorunu ortaya çıkabilir. Bununla birlikte standart sapmalar ya da varyanslar, ortalamanın ya da bir ya da daha çok bağımsız değişkenin fonksiyonu olarak "azalan" ya da "azalan-artan" bir yapı sergileyebilir. Böyle bir durumda yatay eksende yer alan değişken artan ölçekte olduğundan söz konusu değişkenin mevcut skalası, varyansları değişkenin tersi ya da tersinin üsleriyle dengelemeye izin vermeyebilir. Böyle bir durum için merkezileştirilmiş dışsal değişkenin mutlak değerlerinin doğrudan ağırlık tahminlerinin elde edilmesinde kullanılması, değişen varyanslılığın dengelenmesi konusunda işe yarayabilir. Bu tip durumlarda kesim 3.14.1'de verilen algoritmanın ilk üç adımında elde edilen $o_i = (d_i - m)$ merkezi dışsal değişkenin mutlak değerleri

$$\hat{w}_i = 1/|o_i|, \qquad |o_i| \neq 0 \quad \text{için} \tag{3.42}$$

şeklinde ağırlık tahminlerinin doğrudan elde edilmesinde ikinci bir alternatif olarak kullanılabilir. Buna ilişkin bir uygulama kesim 4.12.6'da verilmiştir.

Verilen iki yönteme ek olarak, merkezi dışsal değişkenin kesim 3.14.1'de verilen algoritmasının ilk iki adımında elde edilen boğum ya da merkez noktası, kesim 4.11'de bir uygulamasına yer verilen veri kümelerinin parçalanmasının gerektiği durumlarda bölme ekseninin elde edilmesinde araç olarak kullanılabilir.

4. UYGULAMA

Bu bölümde, önceki bölümde tanıtılan yöntemlerin kelebek dağılan artıklar için değişen varyanslılığa etkisi, hava kirliliğini etkileyen değişkenler için değerlendirilmiştir. Değerlendirmede artık çizimlerinin yanısıra belirtme katsayısı, denklemin standart hatası, değişim katsayısı gibi model yeterlilik ve karşılaştırma ölçüleri; AIC, BIC, SBC gibi model seçme kriterleri dikkate alınmıştır. Öte yandan önerilen yöntemler hem veri kümesinde hem de farklı örneklem genişliğinde yapay veri kümeleriyle de sınanmıştır.

4.1. HAVA KALİTESİ İNDEKSİ

Hava kalitesi, belli kirletici konsantrasyonlarını kaydeden ölçüm cihazlarından oluşan bir ağ yardımı ile ölçülür. Bu ham ölçüm değerleri, geliştirilen standart formüller kullanılarak hava kalitesi indeksi (HKİ) değerlerine dönüştürülmektedir. HKİ değeri, bölgedeki her bir kirletici için ayrı ayrı hesaplanır (yer seviyesindeki ozon, partiküller madde, karbon monoksit, kükürt dioksit ve azot dioksit). Her bir kirletici için hesaplanan en yüksek HKİ, o güne ait HKİ değerini oluşturur. Örneğin, belli bir alandaki HKİ değerleri, ozon için 90, kükürt dioksit için 88 ise, o güne ait HKİ değeri ozon için hesaplanan 90 değeri olacaktır.

Partiküller madde (PM) terimi, havada bulunan katı partiküller ve sıvı damlacıkları ifade eder. PM, insan faaliyetleri sonucu ve doğal kaynaklardan, doğrudan atmosfere karışır. Atmosferde diğer kirleticiler ile reaksiyona girerek PM oluşturur ve atmosfere verilirler. Katı ve sıvı partiküllerin boyutları geniş bir aralığa yayılır. 2,5 μm den daha küçük partiküller, ince partiküller olarak adlandırılır. İnce partikül kaynakları, tüm yanma prosesleri ve bazı endüstriyel prosesleri içerir. 2,5-10 μm. aralığındaki partiküller, kaba partiküller olarak adlandırılır. Kaba partikül kaynakları ise kırma, öğütme işlemleri ile ortaya çıkan, yollardan kalkan tozlardır (www.rshm.saglik.gov.tr/hki/hkindex/index5.htm).

Uygulama için Sivas iline ilişkin Ocak 1990-Aralık 2010 dönemleri arasında aylık ortalama partiküller madde, kükürt ve sıcaklık ölçüm değerleri kullanılmıştır.

Analizlerde kullanılan SAS 9.2, SPSS 15.0, Eviews 4.1 demo ve S-Plus 8.0 demo programların çıktılarında başvurulan terimlerin açıklamaları aşağıdadır:

Analiz Çizelgelerinde Kullanılan İstatistik ve Kriterler:

Düzeltilmiş çoklu belirtme katsayısı: $\overline{R}^2 = 1 - \frac{n-1}{n-p} R^2$

AIC: Akaike Bilgi Kriteri (Akaike's Information Criterion)

SBC: Schwarz'ın Bayes Kriteri (Schwarz's Bayesian Criterion)

BIC: Bayes Bilgi Kriteri (Bayes Information Criterion)

AIC, SBC ve BIC model değerlendirme kriterleridir. Özellikle değişken seçimi yapılırken başvurulan ölçüleridir. Modele eklenen her bir değişken için ceza ölçüsüdür. Eklenen her parametre bu kriterleri büyütür. Kriter değeri en küçük olan model tercih edilir.

$$AIC = n \ln(AKT / n) + 2p$$

$$SBC = n \ln(AKT / n) + 2p \ln(n)$$

$$BIC = n \ln(AKT / n) + 2n(p+2)\frac{\sigma^2}{AKT} - 2\left(\frac{n\sigma^2}{AKT}\right)^2$$

Collinearity Diagnostics (Çoklubağlantı Tanıları):

Bağımsız değişkenler arasındaki ilişkinin olup olmadığı $(X'X)$ matrisinin öz değerleriyle değerlendirilir. λ_i öz değerleri, X uzayı temel bileşenlerinin yayılma miktarının ölçüleridir. Koşul indeksi (Condition İndex) ve Koşul Sayısı (Condition Number) ölçüleri öz değerlere dayalı ilişki ölçüleridir. Koşul sayısı, X uzayındaki her bir temel bileşene ilişkin bir ölçü; koşul indeksi ise i'inci temel bileşene ilişkin bir ölçüdür: Söz konusu ölçüler aşağıdaki gibi ifade edilirler.

Koşul Sayısı, $CN = \left[\lambda_{\max} / \lambda_{\min}\right]^{1/2}$ ve *Koşul İndeksi* $CI_i = \left[\lambda_{\max} / \lambda_i\right]^{1/2}$.

Koşul indeksinin 10 dolaylarında olması zayıf ilişki, 30 ila 100 arasında olması kuvvetli ilişkiyi ve 100'ün üzerinde olması çok ciddi bir ilişkiyi ifade eder.

Bir diğer ölçü ise varyans büyüme faktörüdür (variance inflation factor -V.I.F). İlişki matrisi tersinin köşegen öğeleri olan bu kriterin, $V.I.F_j > 10$ olması bağımsız değişkenler arasında çoklubağlantıyı ifade eder.

Test of First and Second Moment Specification: Bu test, kesim 2.2.2.3'de verilen, değişen varyanslılığın saptanmasında kullanılan White testine karşılık gelmektedir.

Durbin Watson D İstatistiği: Ardışık artıklar arasındaki otokorelasyonun testidir.

D test istatistiği ile otokorelasyonun olup olmadığı test edilir. Burada $\hat{\rho}$, $e_i - e_{i-1}$'ler arasındaki örneklem korelasyonudur. Artıklar arasındaki otokorelasyon artıkça Durbin Watson-D istatistiği küçülür. D, istatistiğinin 1'den küçük olması otokorelasyonun varlığına; 2'ye yakın olması yokluğuna işarettir

4.2. MODEL SEÇİMİ ve ARTIKLAR

Başlangıçta partiküller madde miktarı bağımlı değişken; taşıt sayısı, sıcaklık, doğal gaz kullanımı ve kükürt miktarı bağımsız değişken alınarak modelleme yapıldı. Sıcaklığa ilişkin ortalama aylık veriler Devlet Meteoroloji Müdürlüğü'nden, kükürt miktarı, taşıt sayısı ve partiküller maddeye ilişkin aylık ortalama veriler ise TUİK kayıtlarından elde edilmiştir (www.tuik.gov.tr/cevredagitimapp/hava.zul). Taşıt sayısına ilişkin TÜİK kayıtlarının başlangıç yılı 1994 Temmuz; diğer değişkenlere ilişkin kayıtların başlangıcı ise Ocak 1990 ayıdır. Doğalgaz değişkeni, doğalgazın bu ilde kullanılmaya başlandığı yıl olan 2004 Nisan ayından öncesi için 0, sonrası için 1 şeklinde kodlanarak modele dâhil edildi. Kükürt ve partiküller madde ölçümlerinde eksik olan gözlemler doğrusal interpolasyonla bulunmuştur.

Alınan başlangıç modeli ve model tahminleri aşağıdadır:

$$Partiküller_Madde = \beta_0 + \beta_1 * Sıcaklık + \beta_2 * Kükürt + \beta_3 * Tasit_sa + \beta_4 * dgaz + \varepsilon$$

Çizelge 4.1. Tam Küme Parametre Tahmin Edicileri ve Model Seçimi

Number of Observations Read 253
Number of Observations Used 199
Number of Observations with Missing Values 54

R-Square 0.6400

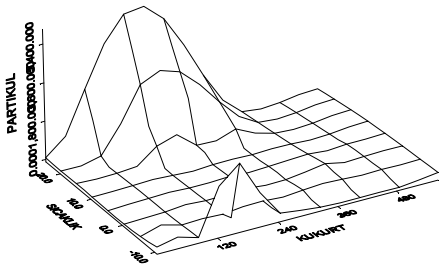
Variable	DF	Estimate	S.E	t Value	Pr > |t|
Intercept	1	66.26640	8.83994	7.50	<.0001
sıcaklık	1	-2.71338	0.30537	-8.89	<.0001
Kükürt	1	0.29889	0.05462	5.47	<.0001
Taşıt Sayısı	1	0.00020002	0.00012476	1.60	0.1105
Dgaz	1	0.41247	4.58087	0.09	0.9283

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.5682	1341.5573	1342.9369	1348.14396	sicaklik
2	0.4649	1384.2607	1384.8293	1390.84730	Kükürt
3	0.0114	1506.4059	1505.1154	1512.99253	dgaz
4	0.0040	1507.9049	1506.5962	1514.49155	tasit_sa
5	**0.6351**	**1310.0485**	**1312.1215**	**1319.92843**	**Kükürt Sıcaklık**
6	0.5813	1337.4362	1338.7008	1347.31614	Sıcaklık dgaz
7	0.5703	1342.6133	1343.7302	1352.49318	Sıcaklık tasit_sa
8	0.4852	1378.5538	1378.6985	1388.43374	Kükürt dgaz
9	0.4751	1382.4287	1382.4744	1392.30864	Kükürt tasit_sa
10	0.0167	1507.3407	1504.8144	1517.22062	dgaz tasit_sa
11	**0.6400**	**1309.4071**	**1311.6116**	**1322.58035**	**Kükürt Sıcaklık tasit_sa**
12	0.6352	1312.0182	1314.1168	1325.19141	Kükürt Sıcaklık dgaz
13	0.5844	1337.9638	1339.0479	1351.13704	Sıcaklık dgaz tasit_sa
14	0.4934	1377.3467	1377.0323	1390.51991	Kükürt dgaz tasit_sa
15	0.6400	1311.3988	1313.6552	1327.86533	Kükürt Sıcaklık dgaz tasit_sa

Çizelge 4.1 incelendiğinde, seçilen başlangıç modelinde taşıt sayısı ve doğalgaz kullanımı istatistiksel olarak anlamlı çıkmamıştır. Diğer yandan ileriye doğru seçim (forward) tekniği ile AIC, BIC eve SBC kriterleri açısından modeller değerlendirildiğinde model 5 ile model 11'in değerleri bir birine çok yakındır. AIC ve BIC kriterleri açısından "Sıcaklık Kükürt ve Taşıt Sayısı" değişkenlerinin yer aldığı model 11; SBC kriteri açısından model 5 daha uygundur. Öte yandan taşıt sayısına ilişkin kayıtlar 1994 Temmuzundan itibaren mevcuttur. Dolayısıyla taşıt sayısı değişkenini içeren model 11 seçilirse 54 tane gözlemin vereceği bilgiden yoksun kalınacaktır. Bu nedenlerle uygulamanın bundan sonraki sürecinde, taşıt sayısı ve doğalgaz kullanımı değişkenleri analiz dışı bırakılarak,

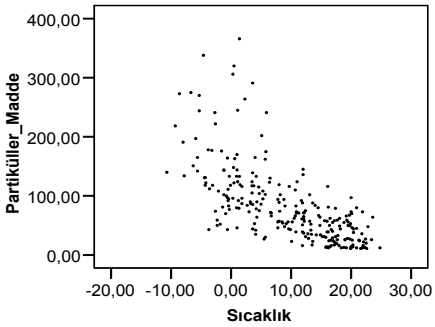
$$Partiküller_Madde = \beta_0 + \beta_1 * Sıcaklık + \beta_2 * Kükürt + \varepsilon$$

modeli ile çalışılmıştır. Partiküller maddenin, sıcaklık ve kükürtle değişim grafikleri (Şekil 4.1) incelendiğinde sıcaklık artışının partiküller madde miktarını azalttığı; kükürt miktarı artışının arttırdığı görülmektedir.

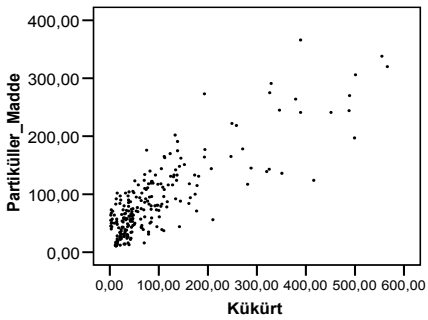


Şekil 4.1. Partiküller Maddenin Kükürt ve Sıcaklıkla Değişimi

Bu durum partiküller maddenin sıcaklık (Şekil 4.2) ve kükürtle değişimi (Şekil 4.3) incelendiğinde daha açık olarak görülmektedir.



Şekil 4.2. Sıcaklık- Partiküller Madde Değişimi



Şekil 4.3. Kükürt - Partiküller Madde Değişimi

Partiküller maddenin bağımlı, sıcaklık ve kükürt ölçümlerinin bağımsız değişken olduğu en küçük kareler analiz sonuçları Çizelge 4.2'de verilmiştir.

Çizelge 4.2. En Küçük Kareler Sonuçları

Dependent Variable: partikul
Number of Observations Read 253
Number of Observations Used 253

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	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 Value	Pr > |t|	95% Conf.	Limits	V.I.F
Intercept	1	70.12801	4.44628	15.77	<.0001	61.37	78.88	
Sıcaklık	1	-2.40700	0.27389	-8.79	<.0001	-2.95	-1.87	1.43
kukurt	1	0.40456	0.02218	18.24	<.0001	0.361	0.448	1.43

R-Square	AIC	BIC	RMSE	SBC	Variables in Model
0.7700	1749.1402	1751.2119	31.52893	1759.74038	sıcaklık Kükürt

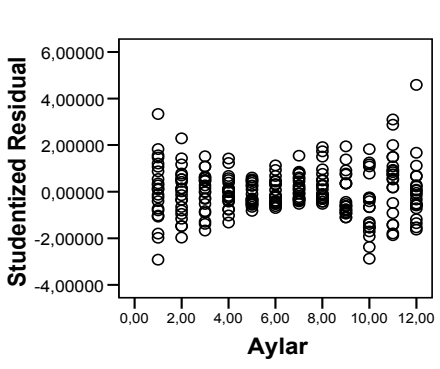
Collinearity Diagnostics

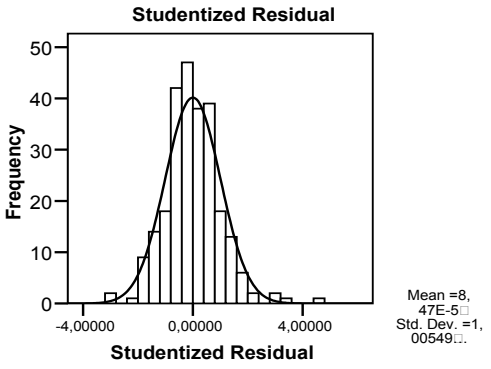
Condition Number	Eigenvalue	Index	Proportion of Variation: Intercept	Sıcaklık	Kükürt
1	2.06480	1.00000	0.04371	0.04705	0.04950
2	0.82188	1.58502	0.00018456	0.16749	0.29037

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
5	14.33	0.0136

Durbin-Watson D **1.249**
1st Order Autocorrelation 0.359





Çizelge 4.2 incelendiğinde kelebek dağılan artıklar dolayısıyla değişen varyanslılık ve aykırı değerler sorunu dışında herhangi bir varsayım bozulmasının olmadığı görülür.

4.3. YENİDEN AĞIRLIKLANDIRILMIŞ İTERATİF EN KÜÇÜK KARELER

Burada, Meloun ve diğ. (2002) tarafından önerilen ve kesim 3.6'da detayları verilen YAIEKK yöntemi için SAS 9.2 programında yazılan kodlar ile elde edilen ağırlıklar kullanılarak ağırlıklı en küçük kareler uygulanmıştır. Analiz sonuçları Çizelge 4.3'de verilmiştir.

Çizelge 4.3. Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler Sonuçları

Number of Observations Read 253
Number of Observations Used 253
Weight: SONAGIRLIK

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2.57687E16	1.288435E16	1.45E13	<.0001
Error	250	221905	887.62096		
Corrected Total	252	2.57687E16			

Dependent Mean	184.80360	Adj R-Sq	1.0000
Coeff Var	16.12142		

R-Square	AIC	BIC	RMSE	SBC	Variables in Model
1.0000	1720.4783	1722.5500	29.79264	1731.07843	sıcaklık Kükürt

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	74.58443	0.00006454	1155717	<.0001	0
sıcaklık	1	-2.90632	0.00010482	-27728	<.0001	191.42976
Kükürt	1	0.36812	0.00000102	361768	<.0001	191.42976

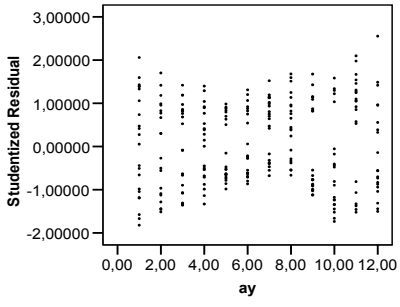
Collinearity Diagnostics

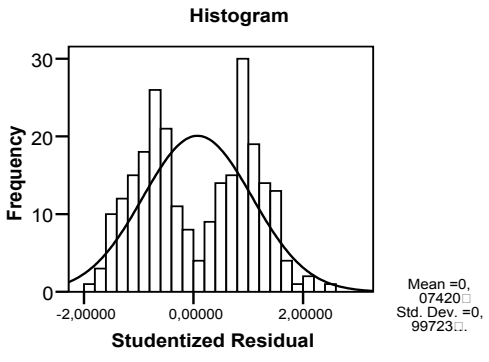
Number	Eigenvalue	Index	Intercept	Sıcaklık	Kükürt
1	2.75532	1.00000	0.00640	0.00019529	0.00024336
2	0.24383	3.36155	0.16415	0.00054762	0.00181
3	0.00084491	57.10589	0.82945	0.99926	0.99794

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
3	3.04	0.3851

Durbin-Watson D 1.023
1st Order Autocorrelation 0.483





38 İterasyon soncunda elde edilen sonuç ağırlıklar kullanılarak elde edilen regresyon analizi sonuçları (Çizelge 4.3) incelendiğinde, çok yüksek bir çoklu belirtme katsayısı ($R^2 = 1$) elde edilmesine rağmen değişen varyanslılığın dengelenemediği görülmektedir. Diğer yandan, söz konusu yöntemin uygulanması sonucunda, hem çoklubağlantı hem de normal dağılımdan sapma durumunun ortaya çıktığı görülebilir.

4.4. CHATTERJEE ve MACHLER TARAFINDAN ÖNERİLEN İTERATİF YÖNTEMİN SONUÇLARI

Chatterjee ve Mahler (1997) tarafından önerilen yöntem için SAS programında yazılan kodlarla elde edilen son ağırlıklarla ağırlıklı en küçük kareler uygulanmış ve Çizelge 4.4'de verilen sonuçlara ulaşılmıştır.

Çizelge 4.4. Chatterjee ve Machler İteratif Yöntemi Sonuçları

Dependent Variable: partikul
Number of Observations Read 253
Number of Observations Used 253
Weight: SON_AGIRLIK

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5452.73794	2726.36897	403.66	<.0001
Error	250	1688.51001	6.75404		
Corrected Total	252	7141.24794			

Dependent Mean	82.40007	Adj R-Sq	0.7617
Coeff Var	3.15395		

R-Square	AIC	BIC	RMSE	SBC	Variables in Model
0.7636	486.2477	488.3194	2.59885	496.84788	Sıcaklık Kükürt

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	70.03857	4.45928	15.71	<.0001	0
Sıcaklıkk	1	-2.40097	0.27359	-8.78	<.0001	1.43066
Kükürt	1	0.40482	0.02277	17.78	<.0001	1.43066

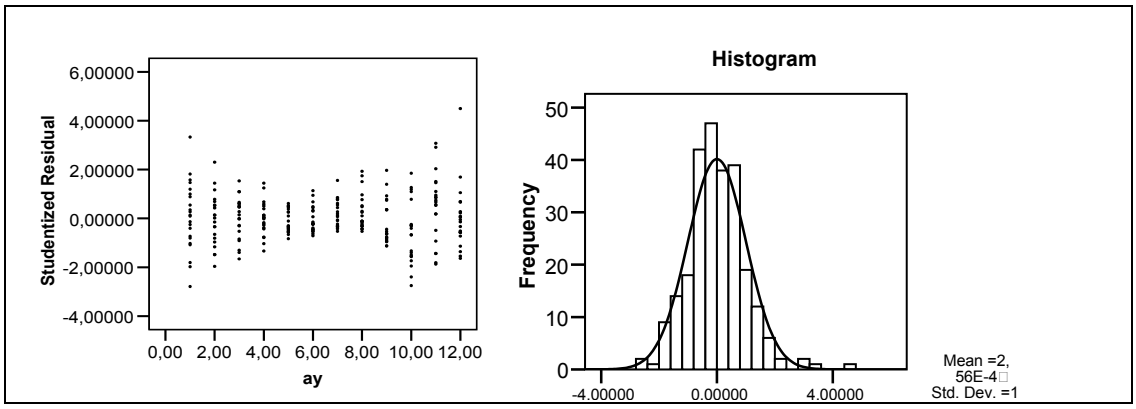
Collinearity Diagnostics

Condition	---------Proportion of Variation---------				
Number	Eigenvalue	Index	Intercept	Sıcaklık	Kükürt
1	2.07422	1.00000	0.04253	0.04625	0.04905
2	0.81441	1.59591	0.00020067	0.16617	0.29102
3	0.11137	4.31555	0.95727	0.78757	0.65992

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 artıkların aylara karşı çiziminden anlaşılmaktadır. Yöntem, diğer kriterler açısından incelendiğinde birkaç aykırı değer dışında herhangi bir sorun göze çarpmamaktadır.

4.5. DEĞİŞTİRİLMİŞ ARDIŞIK ARTIKLAR YÖNTEMİ SONUÇLARI

Bischoff ve diğ. (2006) tarafından önerilen ve işleyişi kesim 3.5'de verilen algoritmaya paralel olarak elde edilen değiştirilmiş ardışık artıklar, ağırlık tahminlerinin belirlenmesinde edilmesinde kullanılmıştır. Elde edilen ağırlıklar ile ağırlıklı en küçük kareler regresyonu gerçekleştirilmiş ve Çizelge 4.5'de özetlenen sonuçlara ulaşılmıştır.

Çizelge 4.5. Değiştirilmiş Ardışık Artıklar ile Ağırlıklı En Küçük Kareler Sonuçları

Dependent Variable: partikull
Number of Observations Read 253
Number of Observations Used 250
Number of Observations with Missing Values 3
Weight: weight

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	557.39853	278.69926	343.98	<.0001
Error	247	202.55270	0.81021		
Corrected Total	249	759.95122			

Dependent Mean	57.62722	Adj R-Sq	0.7313
Coeff Var	1.56197		

R-Square	AIC	BIC	RMSE	SBC	Variables in Model
0.7335	-50.2645	-48.1928	0.90012	39.66435	sıcaklık kükürt

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	70.35585	4.50072	15.63	<.0001	0
sıcaklık	1	-2.46777	0.26653	-9.26	<.0001	1.47551
kükürt	1	0.38588	0.02582	14.95	<.0001	1.47551

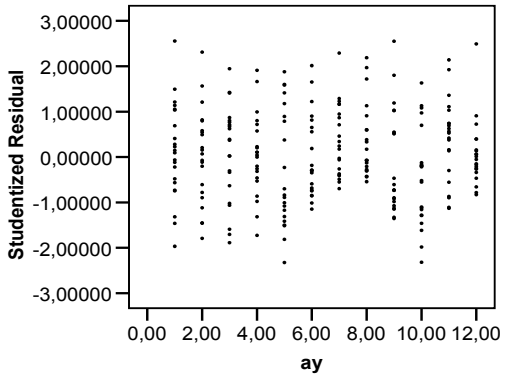
Collinearity Diagnostics

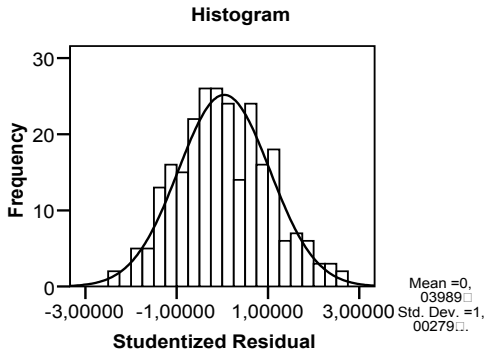
Condition	---------Proportion of Variation---------				
Number	Eigenvalue	Index	Intercept	Sıcaklık	kükürt
1	2.21932	1.00000	0.02008	0.02270	0.04337
2	0.72218	1.75302	0.00250	0.05214	0.44425
3	0.05851	6.15904	0.97741	0.92516	0.51238

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





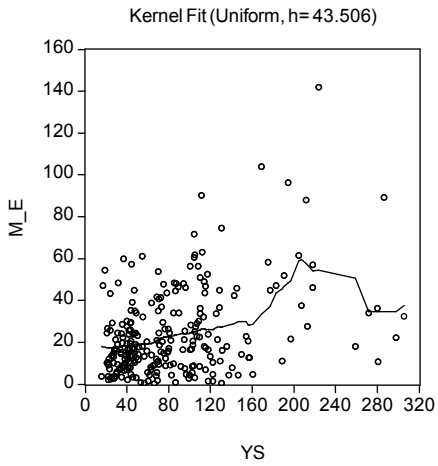
Çizelge 4.5 incelendiğinde, değiştirilmiş ardışık artıklar yöntemiyle değişen varyanslılık sorununun kısmen düzeldiği ve diğer analiz kriterleri açısından herhangi bir sorun olmadığı görülebilir.

4.6. PARAMETRİK OLMAYAN DÜZLEME

Bu kesimde, varyans ve standart sapma fonksiyonlarının tahmini için parametrik olmayan regresyon yöntemleri kullanılmıştır.

4.6.1. Mutlak Artıklarla Parametrik Olmayan Düzleme

Burada, standart sapma fonksiyonu düzlemesinde, mutlak artıklar bağımlı değişken ve $\hat{y}_i$ değerleri, bağımsız değişken alınarak uniform çekirdek fonksiyonu için yerel doğrusal regresyon (local lineer regresyon) uygulanmıştır. Genel olarak, yerel polinomiyel çekirdek regresyonları kullanılarak her x değerinde, artık kareleri toplamlarını minimize eden $\hat{\beta}$'ların seçimiyle $\hat{y}$ tahmin edilir (Eviews 4.1 User Guide).



Şekil 4.4. Mutlak Artıkların Bağımlının Tahmin Değerleriyle Çekirdek Düzlemesi[1]

Mutlak artıkların, bağımlı değişkenin en küçük kareler tahmin değerleriyle yerel doğrusal regresyonundan elde edilen tahmin değerleri (Şekil 4.4'deki kırık çizgi) standart sapma tahminleridir. Bu standart sapma tahmin değerleri, ağırlık tahminleri için kullanılarak, ağırlıklı en küçük kareler uygulanmıştır. Analiz sonuçları Çizelge 4.6'da verilmiştir.

Çizelge 4.6. Mutlak Artıklar İçin Parametrik Olmayan Regresyon ile Ağırlık Tahminleri ve Ağırlıklı En Küçük Kareler Sonuçları

Dependent Variable: partikul
Number of Observations Read 253
Number of Observations Used 253

Weight: $\hat{w}_i = 1/|\widehat{e_i}|$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	876.43324	438.21662	283.16	<.0001
Error	250	386.90253	1.54761		
Corrected Total	252	1263.33576			

Root MSE	1.24403	R-Square	0.6937
Dependent Mean	63.17089	Adj R-Sq	0.6913
Coeff Var	1.96931		

R-Square	AIC	BIC	RMSE	SBC	Variables in Model
0.6913	113.4702	115.5419	1.24403	124.07034	Kükürt sicaklik

Parameter Estimates

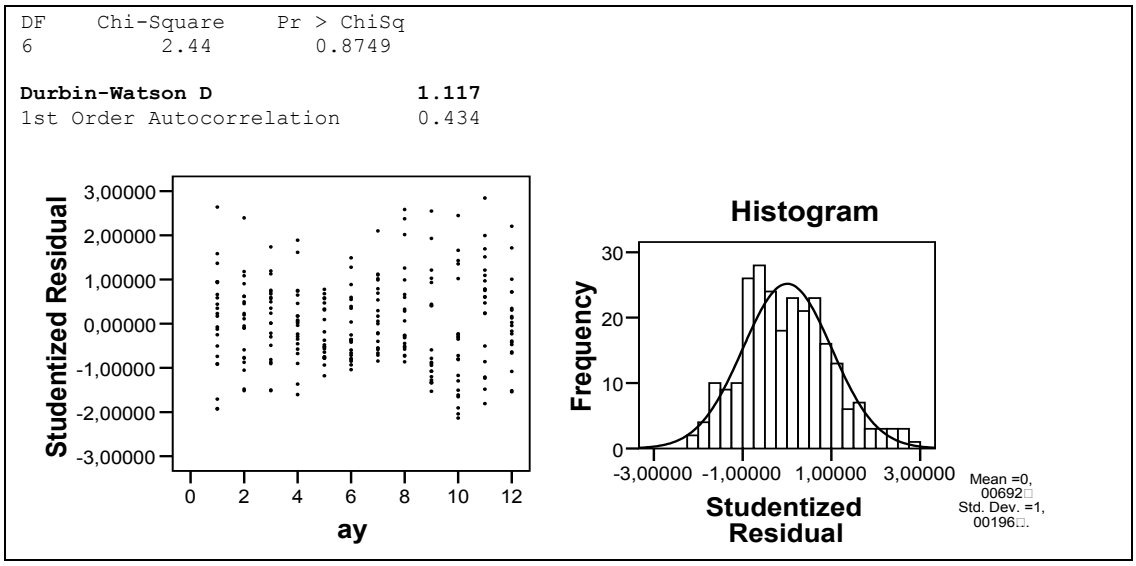
Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	69.03554	4.54418	15.19	<.0001	0
Kükürt	1	0.39100	0.02944	13.28	<.0001	1.46382
Sıcaklık	1	-2.24265	0.25352	-8.85	<.0001	1.46382

Collinearity Diagnostics

Condition		Proportion of Variation		
Number	Eigenvalue	Index	Intercept	Kükürt	Sıcaklık
1	2.19479	1.00000	0.02656	0.04415	0.03088
2	0.72787	1.73648	0.00146	0.37528	0.09329
3	0.07734	5.32702	0.97198	0.58058	0.87583

Test of First and Second Moment Specification

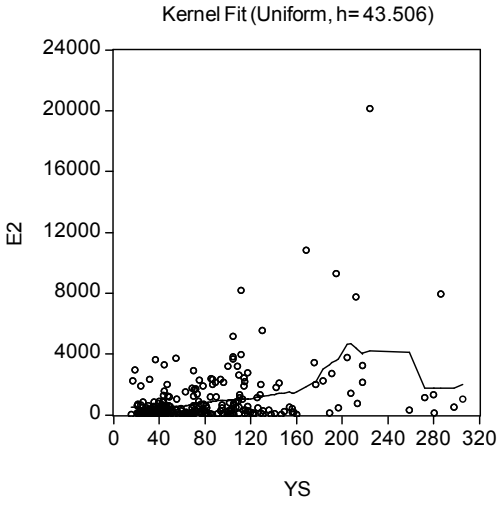
[1] Grafikte M_E, mutlak artıklar ve YS da $\hat{y}$ 'dır.



Mutlak artıkların, bağımlı değişkenin en küçük kareler tahminleriyle düzlemesinden elde edilen mutlak artık tahmin değerlerinin tersinin ağırlık olduğu regresyon sonuçları incelendiğinde skor testinin sonucuna rağmen değişen varyanslılık sorunun ortadan kalkmadığı görülmektedir. Bunun dışında herhangi bir istatistiksel varsayım bozulması gözlenmemektedir.

4.6.2. Karesel Artıklarla Parametrik Olmayan Düzleme

Burada, karesel artıklar bağımlı değişken ve $\hat{y}_i$ değerleri bağımsız değişken alınarak uniform çekirdek fonksiyonu için yerel doğrusal regresyon uygulanmıştır (Şekil 4.5).



Şekil 4.5. Karesel Artıkların Bağımlı Tahminleri ile Çekirdek Düzlemesi[2]

[2]Grafikte, E2 Karesel artıklar, YS de $\hat{y}$ 'dır.

Karesel artıkların, uniform çekirdek fonksiyonu için, bağımlı değişkenin EKK tahmin değerleriyle yerel doğrusal regresyonundan elde edilen $\widehat{e_i^2}$ değerleri (Şekil 4.5'deki kırık çizgi $\widehat{e_i^2}$ değerlerine karşılık gelmektedir.), ağırlık tahmininde kullanılmıştır. Elde edilen ağırlıklar ile ağırlıklı en küçük kareler uygulanmıştır. Analiz sonuçları Çizelge 4.7'de verilmiştir.

Çizelge 4.7. Karesel Artıklar İçin Parametrik Olmayan Regresyon ile Ağırlık Tahminleri ve Ağırlıklı En Küçük Kareler Sonuçları

Dependent Variable: partikul
Number of Observations Read 253
Number of Observations Used 253

Weight $\hat{w}_i = 1/\widehat{e_i^2}$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	571.29498	285.64749	285.86	<.0001
Error	250	249.81587	0.99926		
Corrected Total	252	821.11084			

Root MSE	0.99963	R-Square	0.6958
Dependent Mean	63.06634	Adj R-Sq	0.6933
Coeff Var	1.58505		

R-Square	AIC	BIC	MSE	RMSE	SBC	Variables in Model
0.6958	2.7957	4.8674	0.99926	0.99963	13.39583	Kükürt Sıcaklık

Parameter Estimates

	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	68.87696	4.55123	15.13	<.0001	0
Kükürt	1	0.39316	0.02917	13.48	<.0001	1.45991
Sıcaklık	1	-2.23429	0.25452	-8.78	<.0001	1.45991

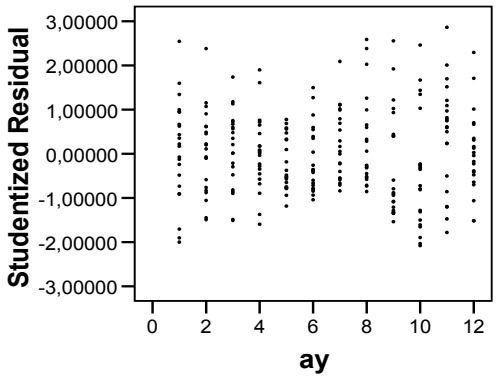
Collinearity Diagnostics

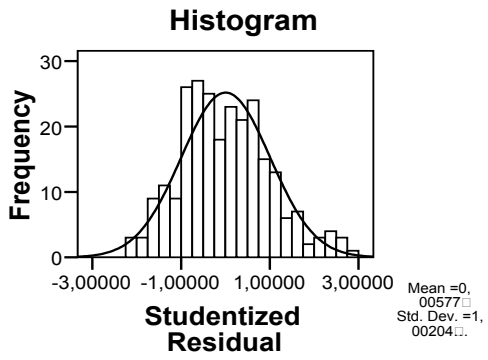
Condition		Proportion of Variation		
Number	Eigenvalue	Index	Intercept	Kükürt	Sıcaklık
1	2.19208	1.00000	0.02663	0.04438	0.03081
2	0.73064	1.73211	0.00153	0.38109	0.09086
3	0.07728	5.32594	0.97184	0.57453	0.87833

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
6	2.11	0.9089

Durbin-Watson D 1.117
Number of Observations 253
1st Order Autocorrelation 0.434





Karesel artıklardan, düzleme ile elde edilen tahminlerin ağırlık olarak kullanılması ile gerçekleştirilen ağırlıklı en küçük kareler sonuçları incelendiğinde skor testinin sonucuna rağmen değişen varyanslılık sorunun ortadan kalkmadığı ve ek olarak artık dağılımında küçük çaplı bir bozulma dışında herhangi bir istatistiksel varsayım bozulması gözlenmemektedir.

4.7. TEKRARLI GÖZLEMLER İÇİN ORTALAMA ve VARYANS FONKSİYONLARI

Bu bölümde, önce bağımlı değişkenin standart sapmaları; sonra da karesel artıkların standart sapmaları ağırlık elde etmede kullanılarak ağırlıklı en küçük kareler uygulanmıştır.

4.7.1. Alt Gruplara Karşılık Gelen Bağımlı Değişken Değerlerinin Standart Sapmaları

Çalışmada öncelikle veri kümesi aylara göre gruplara bölünerek her bir gruba karşılık gelen bağımlı değişken standart sapmaları hesaplanmıştır. Hesaplanan standart sapmalar yeni bir değişken olarak atanarak, bağımlı değişken alt gruplarının standart sapmaları ağırlık değişkeni, partiküller madde miktarı bağımlı değişken; sıcaklık ve kükürt bağımsız değişken olmak üzere $\hat{w}_i = s_{y_{ij}}^{\lambda_i}, \lambda_i \in [-2,2]$ için log olabilirlik (log likelihood) değeri en büyük olan λ_i'ye karşılık gelen üs kullanılarak ağırlıklar elde edilmiştir. Burada, log olabilirlik fonksiyonu $l = -\frac{n}{2}(1+\log(2\pi)+\log(AKT/n))$'dir.

Bu işlem SPSS 15.0 paket programı ağırlık tahmin modülü (weight estimation prosedürü) otomatik olarak yapılmaktadır. En büyük log olabilirlik değeri (-3/2) için $\hat{w}_{ij} = s_{y_{ij}}^{-3/2}$ j= 1 ,.. 1 2 ; i= 1 ,.., n eşitliğinden elde edilen ağırlıklar ile gerçekleştirilen ağırlıklı en küçük kareler analiz sonuçları Çizelge 4.8'de verilmiştir.

Çizelge 4.8. Bağımlı Değişken Standart Sapmaları için AEKK Sonuçları

```
Dependent Variable: Partikuller_madde
Number of Observations Read         253
Number of Observations Used         253

Analysis of Variance
Source                  DF     Sum of Squares   Mean Square   F Value    Pr > F
```

Model	2	2084.43211	1042.21605	280.48	<.0001
Error	250	928.97397	3.71590		
Corrected Total	252	3013.40607			

Dependent Mean	56.64956	Adj R-Sq	0.6893
Coeff Var	3.40279		

R-Square	AIC	BIC	RMSE	SBC	Variables in Model
0.6917	335.0749	337.1466	1.92767	345.67505	Sıcaklık Kükürt

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	71.48031	4.83635	14.78	<.0001	0
Sıcaklık	1	-2.38059	0.27011	-8.81	<.0001	1.41713
Kükürt	1	0.35956	0.02627	13.69	<.0001	1.41713

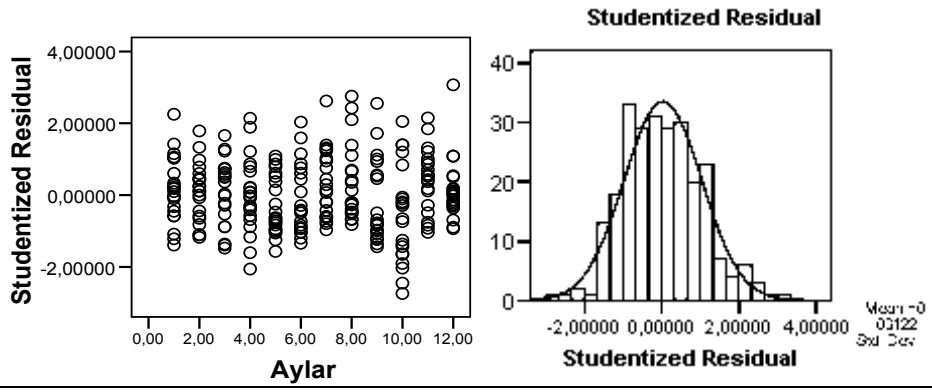
Collinearity Diagnostics

Condition		Proportion of Variation		
Number	Eigenvalue	Index	Intercept	Sıcaklık	Kükürt
1	2.23692	1.00000	0.01942	0.02197	0.04525
2	0.70543	1.78073	0.00305	0.04986	0.47779
3	0.05765	6.22920	0.97752	0.92817	0.47696

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
6	7.60	0.2686

Durbin-Watson D	**1.077**
1st Order Autocorrelation	0.458



Skor testi ve Student-türü artık çizimlerine bakıldığında varyansın kısmen dengelendiği görülmektedir. Diğer sonuçlar incelendiğinde herhangi bir varsayım bozulması sorunun olmadığı görülmektedir.

4.7.2. Karesel Artıklar

Ağırlıksız en küçük karelerden elde edilen artıkların kareleri, bağımsız bir değişken gibi düşünülerek her bir alt gruba karşılık gelen standart sapmaları hesaplanmıştır. Kesim 4.7.1'de bağımlı değişken için yapılan işlemler burada artıkların kareleri için tekrarlanarak ağırlıklar elde edilmiştir. En büyük log olabilirlik değerine karşılık gelen (-1/2) için $\hat{w}_j = 1/s^2_{e_{ij}}$ j=1,..,12; i=1,..,n; eşitliğinden elde edilen ağırlıklar ile gerçekleştirilen ağırlıklı en küçük kareler regresyonunun sonuçları Çizelge 4.9'da verilmiştir.

Çizelge 4.9. Karesel Artık Standart Sapmaları için AEKK Sonuçları

Dependent Variable: Partikuller_madde
Number of Observations Read 253
Number of Observations Used 253

Analysis of Variance

Source	DF	Sum of Squares	Mean square	F Value	Pr > F
Model	2	759.24439	379.62219	370.61	<.0001
Error	250	256.08096	1.02432		
Corrected Total	252	1015.32534			

Dependent Mean	58.16094	Adj R-Sq	0.7458
Coeff Var	1.74015		

R-Square	AIC	BIC	RMSE	SBC	Variables in Model
0.7478	9.0623	11.1341	1.01209	19.66252	Sıcaklık Kükürt

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	67.32753	4.55037	14.80	<.0001	0
Sıcaklık	1	-2.21379	0.25606	-8.65	<.0001	1.51868
Kükürt	1	0.40890	0.02572	15.90	<.0001	1.51868

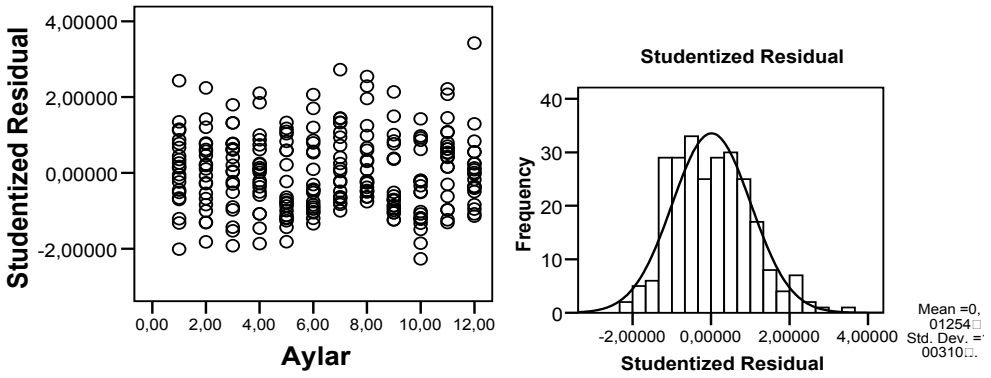
Collinearity Diagnostics

Condition		Proportion of Variation		
Number	Eigenvalue	Index	Intercept	Sıcaklık	Kükürt
1	2.19002	1.00000	0.02142	0.02390	0.04218
2	0.74957	1.70930	0.00207	0.05257	0.42422
3	0.06041	6.02103	0.97651	0.92353	0.53361

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
6	4.52	0.6060

Durbin-Watson D 1.166
1st Order Autocorrelation 0.409



Skor testi ve Student-türü artıkların grafiği incelendiğinde değişen varyanslılığın dengelendiği görülmektedir.

4.8. JACKKNİFE YÖNTEMİ SONUÇLARI

S Plus 8.0 Demo programı kullanılarak elde edilen parametre tahmin edicilerinin ortalamasına karşılık gelen değerler regresyon eşitliğinde yerine yazılarak artıklar elde edilmiştir. Bu artıkların normal dağılım için histogramları, değişen varyanslılık için artık grafikleri ve parametre tahmin edicilerinin örnekleme dağılımları Çizelge 4.10'da verilmiştir.

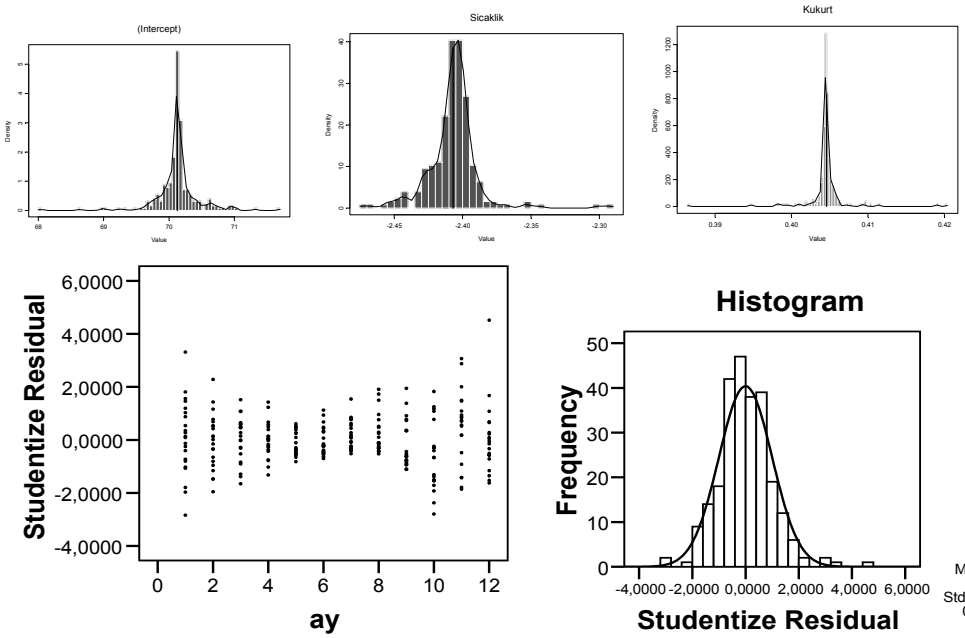
Çizelge 4.10. Jackknife Parametre Tahmin Edicileri ve Artık Çizimleri

```
jackknife(data = siv, statistic = coef(lm(Partikul ~ Sıcaklık + Kükürt,
siv)),
      assign.frame1 = T)
Number of Replications: 253

Summary Statistics:
            Observed       Bias     Mean      S.E
(Intercept)  70.1280 -0.0448217  70.1278 5.30786
   Sıcaklık  -2.4070  0.0030980  -2.4070 0.29637
     Kükürt   0.4046  0.0001199   0.4046 0.03697

Empirical Percentiles:
                  2.5%          5%         95%       97.5%
(Intercept)  69.5195681  69.7074327  70.6555462  70.8988473
   Sıcaklık  -2.4456687  -2.4355400  -2.3874572  -2.3734147
     Kükürt   0.4006256   0.4024015   0.4061704   0.4092484

Correlation of Replicates:
            (Intercept)   Sıcaklık     Kükürt
(Intercept)   1.0000000 -0.9030347 -0.7814282
   Sıcaklık  -0.9030347  1.0000000  0.6162596
     Kükürt  -0.7814282  0.6162596  1.0000000
```



253 yineleme sonucunda gerçekleşen Jackknife parametre tahminleri için elde edilen artık çizimleri incelendiğinde kelebek dağılan artıklar dolayısıyla değişen varyanslılık sorununun devam ettiği gözlenmektedir. Öte yandan tek tek parametre tahmin edicilerinin örnekleme dağılımı için Jackknife yöntemi sonuçlarına bakıldığında parametre tahmin edicilerinin örnekleme dağılımının normal dağılma uymadığı görülmektedir.

4.9. BOOTSTRAP YÖNTEMİ SONUÇLARI

S Plus 8.0 Demo programı kullanılarak elde edilen bootstrap parametre tahmin edicilerinin ortalamalarına karşılık gelen değerler regresyon eşitliğinde yerine yazılarak artıklar elde edilmiştir. Elde edilen artıkların normal dağılım için histogramları, değişen varyanslılık için artık grafikleri ve bootstrap parametre tahmin edicilerinin örnekleme dağılımları Çizelge 4.11'de verilmiştir.

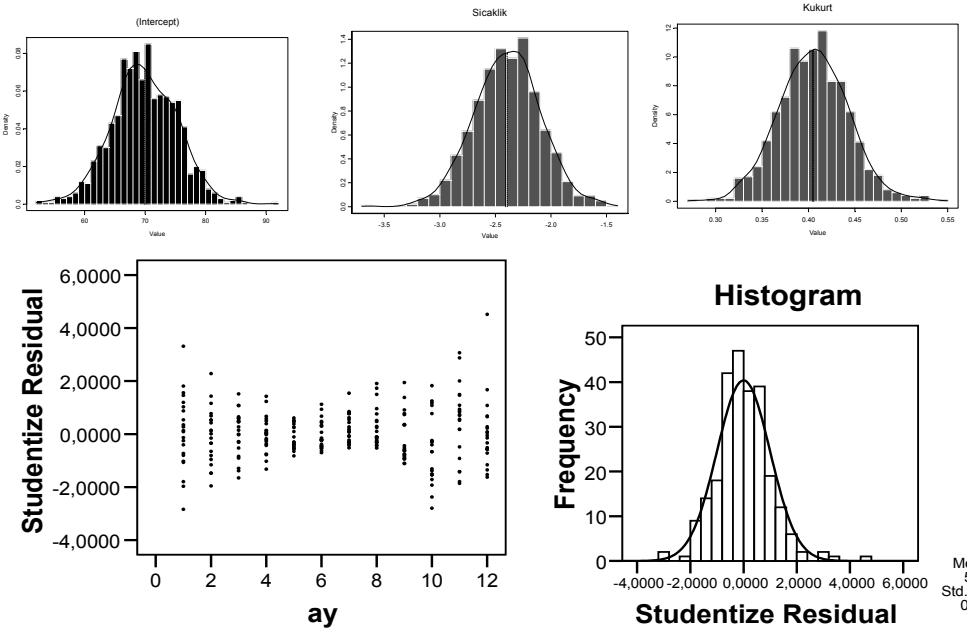
Çizelge 4.11. Bootstrap Yöntemi için Parametre Tahmin Edicileri ve Artık Çizimleri

```
bootstrap(data = siv, statistic = coef(lm(Partikul ~ Sıcaklık + Kükürt,
siv)),
       B = 1000, trace = F, assign.frame1 = F, save.indices = F)
Number of Replications: 1000

Summary Statistics:
            Observed      Bias      Mean      S.E
(Intercept)  70.1280 -0.044473  70.0835 5.12292
   Sıcaklık  -2.4070  0.009035  -2.3980 0.28910
     Kükürt   0.4046 -0.001319   0.4032 0.03562

Empirical Percentiles:
                   2.5%         5%         95%       97.5%
(Intercept)  60.2389990  61.549522  78.5872286  80.1259570
   Sıcaklık  -2.9574931  -2.868212  -1.9123829  -1.8321217
     Kükürt   0.3319762   0.343453   0.4646126   0.4781252


Correlation of Replicates:
            (Intercept)   Sıcaklık     Kükürt
(Intercept)   1.0000000 -0.9000666 -0.7734785
   Sıcaklık  -0.9000666  1.0000000  0.6062647
     Kükürt  -0.7734785  0.6062647  1.0000000
```



1000 tekrar sonucu elde edilen bootstrap için ortalama parametre tahminlerinden elde edilen artık çizimleri incelendiğinde kelebek dağılan artıklar dolayısıyla değişen varyanslılık sorununun devam ettiği gözlenmektedir. Parametre tahmin edicilerinin örnekleme dağılımı incelendiğinde ise Jackknife yöntemine göre sonuçların daha iyi olduğu görülmektedir. Parametre tahmin edicilerin örnekleme dağılımı normal dağılıma uymaktadır.

4.10. BOX-COX DÖNÜŞÜMLERİ

Uygulamanın bu bölümünde, en çok başvurulan bağımlı değişken dönüşümü olan Box-Cox dönüşümleri veri kümesine uygulanmıştır. λ parametresi için $[-2,2]$ aralığında -2'den başlanarak 0.5 artışlarla elde edilen değerleri için bağımlı değişkene dönüşüm uygulanmıştır. Burada, yalnızca $\lambda = 0,5$ ve $\lambda = 0 \rightarrow$ logaritmik dönüşüm değerleri için tam analiz sonuçları verilmiştir.

Bağımlı değişken değerlerinin logaritması alınarak uygulanan regresyon analizi sonuçları Çizelge 4.12'de verilmiştir.

Çizelge 4.12. Box-Cox $\lambda = 0 \rightarrow$ Logaritma Dönüşümü Sonuçları

Dependent Variable: log partikül
Number of Observations Read 253
Number of Observations Used 253

Analysis of Variance

Source	DF	Sum of Squares	MS	F Value	Pr > F
Model	2	19.99065	9.99533	225.85	<.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		

R-Square	AIC	BIC	SBC	Variables in Model
0.5400	-723.1765	-722.1345	-716.10972	Sıcaklık

Parameter Estimates

Variable	DF	Estimate	SE	t Value	Pr > |t|	V.I.F
Intercept	1	70.35585	4.50072	15.63	<.0001	0
Sıcaklık	1	-2.46777	0.26653	-9.26	<.0001	1.47551
kükürt	1	0.38588	0.02582	14.95	<.0001	1.47551

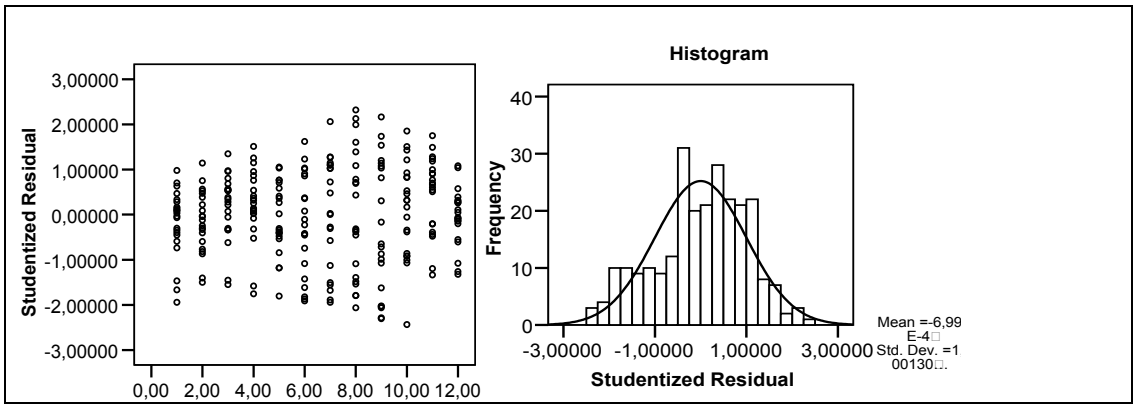
Collinearity Diagnostics

Condition	---------Proportion of Variation---------				
Number	Eigenvalue	Index	Intercept	Sıcaklık	Kükürt
1	2.06480	1.00000	0.04371	0.04705	0.04950
2	0.82188	1.58502	0.00018456	0.16749	0.29037
3	0.11332	4.26856	0.95611	0.78546	0.66013

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 sonucuna bakıldığında istatistiksel olarak logaritma dönüşümünün varyansı dengeleyemediği sonucu ortaya çıkmaktadır. Artık grafikleri incelendiğinde ise kelebek dağılan artıkların yerine, Weisbeg'in (2005), "double outward box" olarak adlandırdığı artık kalıbı ortaya çıkmıştır. Öte yandan dönüşüm ardışık artıklar arasında otokorelasyona neden olmuştur.

Box-Cox'un burada yer verilen , $\lambda = 0.5$ için bağımlı değişkenin dönüşümü sonucu elde edilen analiz sonuçları Çizelge 4.13'de verilmiştir.

Çizelge 4.13. Box-Cox $\lambda = 0.5$ için Dönüşüm Sonuçları

Number of Observations Read 253
Number of Observations Used 253

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	8164.21811	4082.10906	350.07	<.0001
Error	250	2915.23105	11.66092		
Corrected Total	252	11079			

Root MSE	3.41481	R-Square	0.7369
Dependent Mean	15.02736	Adj R-Sq	0.7348
Coeff Var	22.72395		

AIC	BIC	SBC	Variables
624.4117	626.4834	635.01183	Kükürt Sıcaklık

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	15.21774	0.48156	31.60	<.0001	0
KÜKÜRT	1	0.03321	0.00240	13.83	<.0001	1.42876
SICAKLIK	1	-0.33516	0.02966	-11.30	<.0001	1.42876

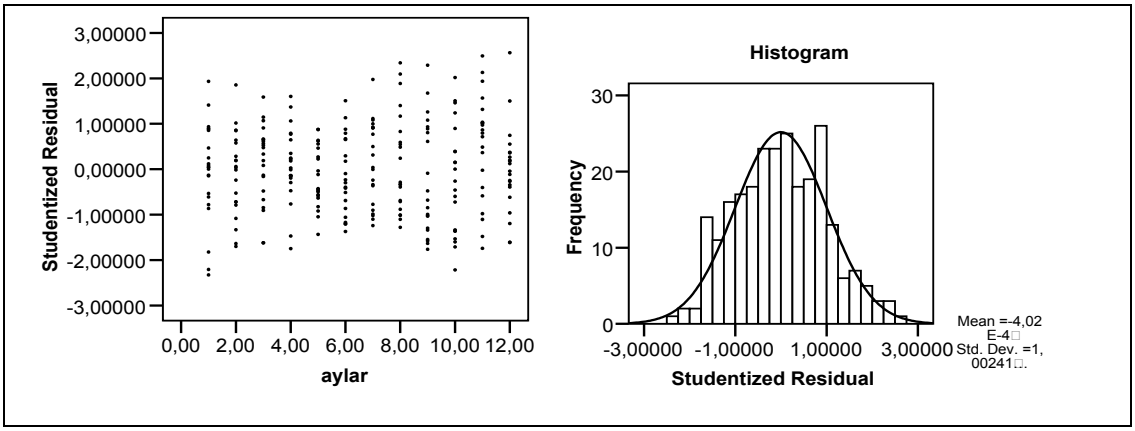
Collinearity Diagnostics

Condition	---------Proportion of Variation---------				
Number	Eigenvalue	Index	Intercept	KÜKÜRT	SICAKLIK
1	2.06480	1.00000	0.04371	0.04950	0.04705
2	0.82188	1.58502	0.00018456	0.29037	0.16749
3	0.11332	4.26856	0.95611	0.66013	0.78546

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
5	1.69	0.8896

Durbin-Watson D	**1.022**
Number of Observations	253
1st Order Autocorrelation	0.479



$\lambda = 0.5$ değeri için değişen varyanslılığın dengelenmediği aylara karşı çizilen Student-türü artıkların çizimlerinden görülmektedir.

Burada yer verilmeyen diğer dönüşüm değerleri için varyansın dengelenmesi bir yana artıkların dağılımında çarpıklaşma ve sivrileşme gözlenmiştir.

4.11. DEĞİŞEN VARYANSLILIĞIN İKİ PARÇALI AĞIRLIKLANDIRMA İLE DENGELENMESİ

Kesim 3.14.1'de verilen algoritmanın ilk üç adımı doğrultusunda veriye en küçük kareler uygulanmıştır. En küçük karelerden elde edilen artıklarının kareleri alınarak dışsal değişken ile regresyonu gerçekleştirilmiştir. Buradan boğum noktası m ,

$$\widehat{e_i^2} = 1860.4478 \; -25.65692 d_i + \; 0.110901 d_i^{\,2} \;,\; m = -0.5 * \left(\frac{-25.65692}{0.110901} \right) \cong 116$$

olarak hesaplanmıştır. Boğum noktasına göre partiküller madde bağımlı, sıcaklık ve kükürttün bağımsız değişkenler olduğu ilk 116 gözlem için ayrı bir regresyon ve kalan 137 gözlem için ayrı bir regresyon uygulanmıştır. Her bir regresyondan elde edilen artıkların kareleri alınmış ve alt gruplar bazında standart sapmaları hesaplanmıştır. Elde edilen standart sapmaların tersi ağırlık tahmini olarak kullanılmıştır. Ağırlıklar birleştirilerek 253 gözlem için ağırlıklı en küçük kareler gerçekleştirilmiştir. Analiz sonuçları Çizelge 4.14'de verilmiştir.

Çizelge 4.14. İkiye Parçalamadan Elde Edilen Ağırlıklar ile AEKK Sonuçları

Dependent Variable: Partiküller Madde
Weight: W
Number of Observations Read 253
Number of Observations Used 253

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	570.26678	285.13339	345.79	<.0001
Error	250	206.14509	0.82458		

Corrected Total	252	776.41187		
Root MSE	0.90806	R-Square	0.7345	
Dependent Mean	55.08142	Adj R-Sq	0.7324	
Coeff Var	1.64859			

AIC	BIC	MSE	RMSE	SBC	Variables
-45.8167	-43.7450	0.82458	0.90806	-35.21657	KÜKÜRT SICAKLIK

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	95% Conf. Lim.		V.I.F
Intercept	1	68.62069	4.57126	15.01	<.0001	59.617	77.623	0
SICAKLIK	1	-2.38378	0.26603	-8.96	<.0001	-2.907	-1.85	1.50
KÜKÜRT	1	0.40000	0.02665	15.01	<.0001	0.3475	0.452	1.50

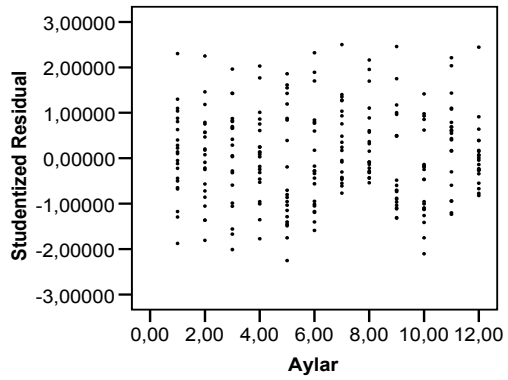
Collinearity Diagnostics

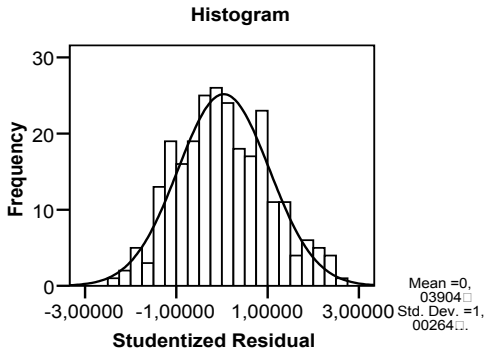
Condition	Proportion of Variation			
Number	Eigenvalue	Index	Intercept	Sıcaklık	Kükürt
1	2.22123	1.00000	0.01843	0.02070	0.04245
2	0.72535	1.74994	0.00253	0.04463	0.44889
3	0.05342	6.44855	0.97904	0.93467	0.50866

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
6	12.85	0.0455

Durbin-Watson D	**1.105**
Number of Observations	253
1st Order Autocorrelation	0.441

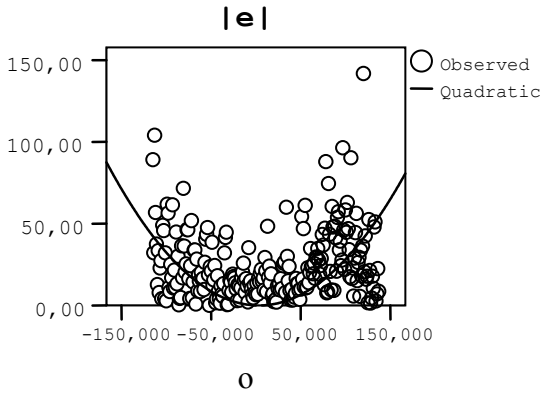




Her bir parça için elde edilen ağırlıkların birleştirilmesi sonucunda elde edilen ağırlıklı en küçük karelerin sonuçları incelendiğinde kelebek dağılan artıklardan kaynaklanan değişen varyanslılık sorununun başarılı bir şekilde ortadan kalktığı, diğer kriterler incelendiğinde ise herhangi ciddi bir varsayım bozulmasının olmadığı görülmektedir.

4.12. DIŞSAL DEĞİŞKEN ve MERKEZİLEŞTİRME SONUÇLARI

Kesim 3.14.1'de algoritması verilen yöntem doğrultusunda merkezileştirilmiş dışsal değişken ile en küçük kareler mutlak artıklarından standart sapma fonksiyonu (Şekil 4.6) tahmin edildikten sonra bu fonksiyondan elde edilen tahmin değerleri ağırlık elde etmekte kullanılarak ağırlıklı en küçük kareler gerçekleştirilmiştir.



Şekil 4.6. Merkezi dışsal değişken ile mutlak artıkların dağılımı

Çizelge 4.15. Merkezi Dışsal Değişken ile Tahmin Edilen Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri

Dependent Variable: |e|

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	b1	b2
Quadratic	,495	122,798	2	251	,000	-,021	,003

The independent variable is o.

Standart sapma fonksiyonu,

$$\widehat{|e_i|} = 0.21o_i + 0.003o_i^2$$

olarak elde edildikten sonra fonksiyonun tahmin değerleri tahmini ağırlıkların elde edilmesinde kullanılmıştır. $\hat{w}_i = 1/\widehat{|e_i|}$ fonksiyonundan elde edilen ağırlıklar ile ağırlıklı en küçük kareler analizi uygulanmıştır. Sonuçlar Çizelge 4.16'de verilmiştir

Çizelge 4.16. Merkezi Dışsal Değişken ile Ağırlık Tahmini ve AEKK Sonuçları

Dependent Variable: Partiküller_madde

Number of Observations Read 253
Number of Observations Used 253

Weight: $\hat{w}_i$

Analysis of Variance

Source	DF	Squares	Square	F Value	Pr > F
Model	2	254192	127096	360.89	<.0001
Error	249	87691	352.17226		
Corrected Total	251	341883			

Root MSE	18.76625	R-Square	0.7435
Dependent Mean	45.18898	Adj R-Sq	0.7414
Coeff Var	41.52838		

R-Square	R-Square	AIC	BIC	SBC	Variables
0.7435	0.7414	1480.7404	1482.8123	1491.32864	Sıcaklık Kükürt

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	95% Conf. Lim		VIF
Intercept	1	65.06665	5.24196	12.41	<.0001	54.74	75.39	0
Sıcaklik	1	-2.32991	0.28001	-8.32	<.0001	-2.88	-1.778	1.59
Kükürt	1	0.42510	0.02803	15.17	<.0001	0.36989	0.4803	1.59

Collinearity Diagnostics

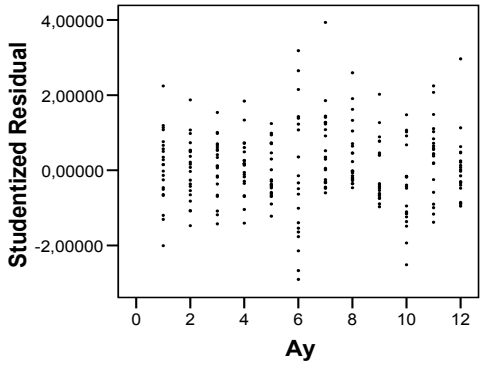
Condition ---------Proportion of Variation---------

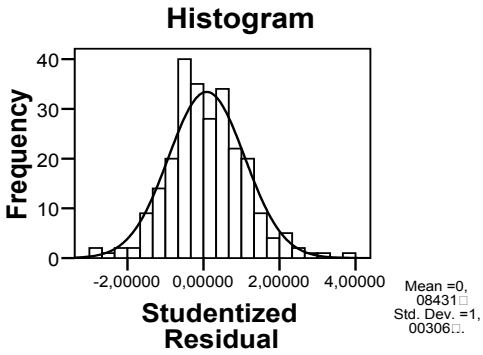
Number	Eigenvalue	Index	Intercept	Sıcaklik	Kükürt
1	2.21441	1.00000	0.01325	0.01444	0.03914
2	0.74844	1.72009	0.00230	0.02355	0.46476
3	0.03714	7.72137	0.98445	0.96201	0.49611

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
6	15.04	0.0199

Durbin-Watson D	1.003
Number of Observations	252
1st Order Autocorrelation	0.491





Merkezi dışsal değişkenle elde edilen standart sapma fonksiyonunun tahmin değerlerinin ağırlık olarak kullanılması sonucu elde edilen ağırlıklı en küçük kareler analizi sonuçları incelendiğinde artık çizimlerindeki kelebek dağılma sorununun ortadan kalktığı ve artıkların normal dağıldığı görülmektedir. Buna karşılık az sayıda da olsa yöntemin aykırı değer ürettiği göze çarpmaktadır.

4.13. YAPAY VERİLER ile ÖNERİLEN YÖNTEMLERİN SINANMASI

Önerilen yöntemlerin geçerliliklerinin sınanması için kelebek dağılan artıkları olan bir regresyon fonksiyonu yaratma gereksinimi ortaya çıktı. 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);
        j = j + 1;
        y[j] = 800 + i + 25*sin^2(2*i*3.14/n)*RAND('NORMAL'));        (4.1)
        x[j] = i;
        d[j] = j
end;
```

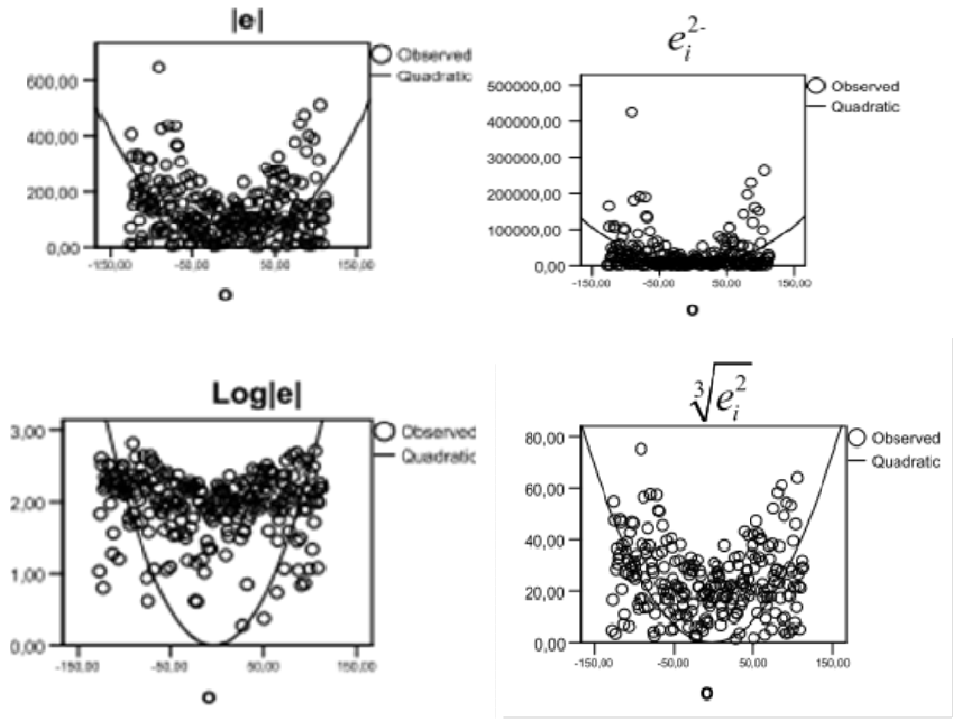
Burada, y[j] bağımlı değişkeni, x[j] bağımsız değişkeni ve d[j] dışsal değişkeni temsil etmektedir.

Bu kısıtlar altında, SAS 9.2 programında yazılan kodlar yardımıyla farklı örneklem büyüklükleri için kelebek dağılan artıkları veren yapay veri kümeleri oluşturularak, kesim 3.14'te önerilen "merkezileştirme ve dışsal değişken" yönteminin geçerliliği sınanmıştır. Kelebek dağılan artıkları elde etmek için fonksiyon, sağdan ve soldan n/5, merkez çevresinde n/10 oranında kırpılmıştır. İzleyen analizlerde belirtilen veri büyüklükleri kırpma işlemi yapıldıktan sonra kalan net veri sayılarıdır. Elde edilen veriler,

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

şeklindeki basit doğrusal regresyon ile önerilen "merkezileştirme ve dışsal değişken" yöntemleri; yaklaşık simetrik, sağa asimetrik ve sola asimetrik kelebek dağılan artıklar için incelenmiştir.

Aynı büyüklükteki veri için en küçük karelerden elde edilen artıkların mutlak değerleri, kareleri, karelerinin küp kökleri ve mutlak artıkların logaritmalarının çizimleri incelenmiş (Şekil 4.7); kelebek dağılan artıklar için mutlak artıklar ile çalışmanın daha iyi standart sapma, dolayısıyla varyans fonksiyonları vereceği sonucuna varılmıştır. Bundan sonraki analizler için standart sapma fonksiyonunun tahmini için mutlak artıkların kullanılması uygun görülmüştür.



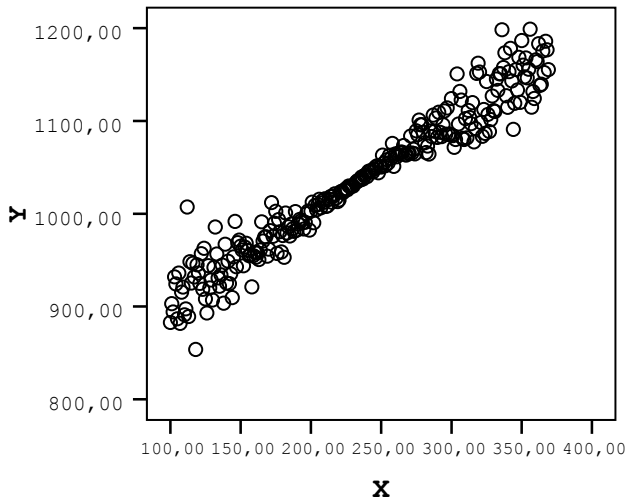
Şekil 4.7. Artık Dönüşümlerinin Merkezi Dışsal Değişken ile Fonksiyonları

Burada normal dağılım gereği sıfır çevresinde yer alması gereken artıkların sayısı fazla olduğundan mutlak artıklar için varyans fonksiyonu tahmin edilirken, merkezi dışsal değişkenin de sıfırı içermesi nedeni ile fonksiyonun orijinden geçmesine olanak tanınarak daha iyi sonuçlara ulaşılması sağlanmıştır. Örneğin, aynı koşullarda orijinden geçmeyen standart sapma fonksiyonu için belirtme katsayısı %20-30 dolaylarında iken orijinden geçen standart sapma fonksiyonu için belirtme katsayısının %50-60 dolayında olduğu gözlenmiştir. Pardoe ve Cook (2007)'da belirtildiği üzere, varyans ya da standart sapma fonksiyonu tahmininin negatif olmaması gerekse bile düzleme sonucunda negatif varyans fonksiyonu tahminleriyle karşılaşılabilir. Bu çalışmada fonksiyonun orijinden geçirilmesi sonucu çok az sayıda da olsa negatif standart sapma tahminleri elde edilmiştir. Bu negatif tahminlerin mutlak değerleri alındıktan sonra ağırlık tahmininde kullanılmıştır.

Bundan sonraki analizlerin tümünde, veriye önce klasik en küçük kareler uygulandıktan sonra önerilen yöntemler uygulanarak sonuçları klasik doğrusal regresyon varsayımları açısından değerlendirilmiştir.

4.13.1. Yaklaşık Simetrik Veri n = 270

Kesim 4.12'de verilen algoritma kullanılarak ve n = 270 örneklem büyüklüğü için Şekil 4.8'de grafiği verilen veri kümesi elde edilmiştir.



Şekil 4.8. n = 270 için Yapay Veri Grafiği

Elde edilen veri kümesine uygulanan en küçük kareler sonuçları Çizelge 4.17'de verilmiştir.

Çizelge 4.17. n = 270 için En Küçük Kareler Sonuçları

Dependent Variable: Y
Number of Observations Read 270
Number of Observations Used 270

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1552745	1552745	4318.49	<.0001
Error	268	96361	359.55774		
Corrected Total	269	1649107			

Root MSE	18.96201	R-Square	0.9416
Dependent Mean	1035.50342	Adj R-Sq	0.9413
Coeff Var	1.83119		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9416	1590.9087	1592.9385	1598.10559	X

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	VIF
Intercept	1	807.33105	3.65889	220.65	<.0001	0
X	1	0.97297	0.01481	65.72	<.0001	1.00000

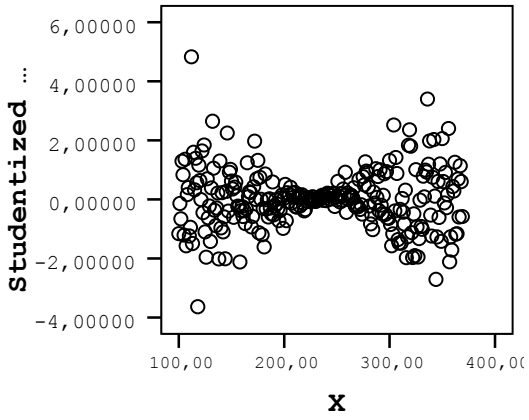
Collinearity Diagnostics

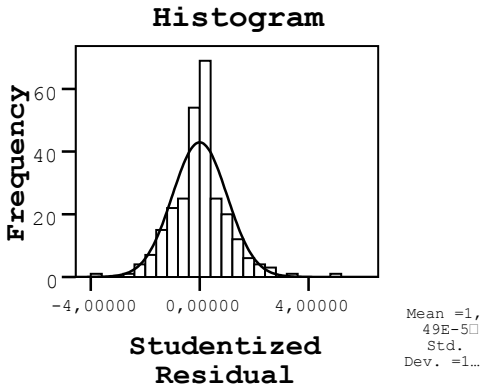
Condition Number	Eigenvalue	Index	--Proportion of Variation- Intercept	X
1	1.94896	1.00000	0.02552	0.02552
2	0.05104	6.17945	0.97448	0.97448

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
2	27.88	<.0001

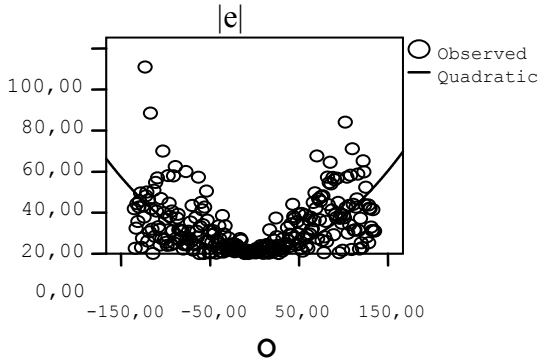
Durbin-Watson D	**2.016**
Number of Observations	270
1st Order Autocorrelation	-0.011





En küçük kareler sonuçları incelendiğinde kelebek dağılan artıklar nedeni ile değişen varyanslılık ve normal dağılımdan sapma göze çarpmaktadır.

En küçük karelerden elde edilen artıkların mutlak değerleri alınarak merkezi dışsal değişken 'o' ile $|e| = \theta_1 o + \theta_2 o^2$ şeklinde standart sapma fonksiyonun (Şekil 4.9) elde edilmesinde kullanılmıştır.



Şekil 4.9. Standart Sapma Fonksiyonu

Elde edilen standart sapma fonksiyonuna ilişkin özet bilgiler Çizelge 4.18'de verilmiştir.

Çizelge 4.18. Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri

Dependent Variable: |e |

	Model Summary					Parameter Estimates	
Equation	R Square	F	df1	df2	Sig.	b1	b2
Quadratic	.568	176.276	2	268	.000	.010	.002

The independent variable is O.

Standart sapma fonksiyonu tahmini,

$|\widehat{e_i}| = 0.010o_i + 0.002o_i^2$

olarak elde edildikten sonra bu fonksiyonun tahmin değerleri ağırlıkların elde edilmesinde kullanılarak ağırlıklı en küçük kareler analizi uygulanmış ve Çizelge 4.19'da verilen sonuçlara ulaşılmıştır.

Çizelge 4.19. n = 270 Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu Tahmini ve AEKK Sonuçları

```
Dependent Variable: Y
Number of Observations Read         270
Number of Observations Used         270

Weight: ŵ_i = 1/|ê_i|,     |ê_i| ≠ 0

Analysis of Variance
Source                   DF     Sum of Squares     Mean Square    F Value    Pr > F
Model                     1          184591          184591      4403.52     <.0001
Error                   268           11234        41.91892
Corrected Total         269          195825
Root MSE              6.47448    R-Square     0.9426
Dependent Mean     1033.46319    Adj R-Sq     0.9424
Coeff Var             0.62648

Model      R-SquareAIC  AIC           BIC          SBC    Variables in Model
1       0.9426          1010.6416     1012.6714     1017.83848    X

Parameter Estimates
```

Variable	DF	Estimate	S.E	t Value	Pr > |t|	VIF
Intercept	1	793.19422	3.62217	218.98	<.0001	0
X	1	1.03554	0.01561	66.36	<.0001	1.00000

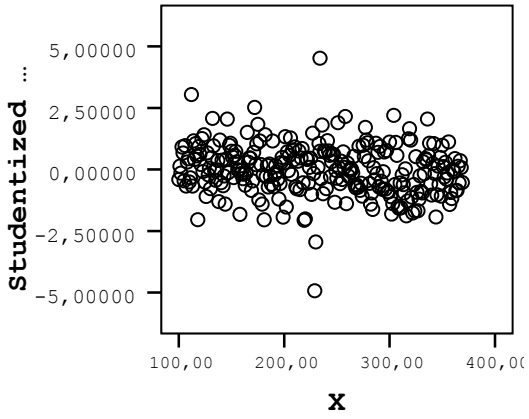
Collinearity Diagnostics

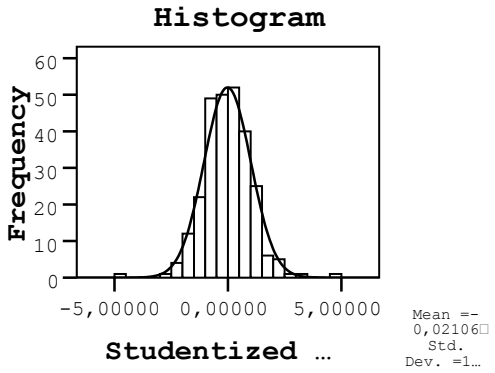
Condition Number	Eigenvalue	Index	--Proportion of Variation- Intercept	X
1	1.99960	1.00000	0.00019774	0.00019774
2	0.00039549	71.10579	0.99980	0.99980

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
3	2.85	0.4159

Durbin-Watson D	**1.849**
Number of Observations	270
1st Order Autocorrelation	0.074

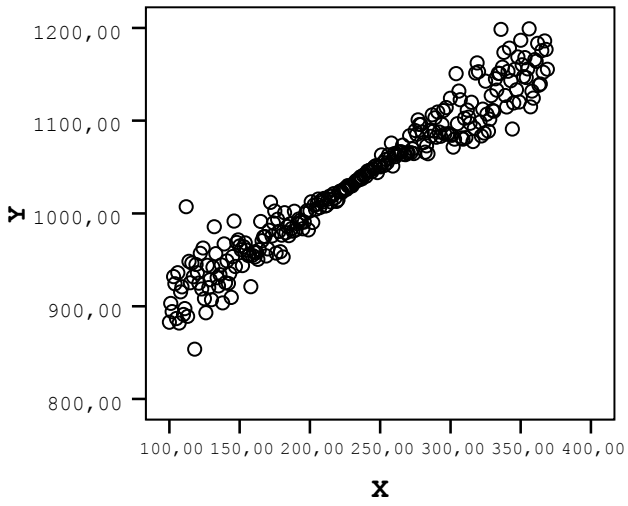




Mutlak artıklar için elde edilen standart sapma fonksiyonu tahmin değerlenin ağırlık olarak kullanımı sonucunda elde edilen regresyon analizi sonuçları incelendiğinde temel olarak odaklandığımız değişen varyanslılık sorununun ortadan kalktığı, artıkların dağılımının normal dağılıma uyduğu, 2 tane aykırı değer dışında herhangi bir varsayım bozulmasının olmadığı ve tüm karşılaştırma kriterlerinde olumlu yönde artışlar olduğu görülmektedir.

4.13.2. Yaklaşık Simetrik Veri n = 1339

Yapay veri üretme süreci örneklem büyüklüğü artırılarak n = 1339 örneklem büyüklüğünde, Şekil 4.10'da grafiği verilen veri kümesi elde edilmiştir.



Şekil 4.10. n = 1339 için Yapay Veri Grafiği

Elde edilen veri kümesine en küçük kareler uygulanarak Çizelge 4.20'de verilen sonuçlara ulaşılmıştır.

Çizelge 4.20. n = 1339 için En Küçük Kareler Sonuçları

Dependent Variable: Y
Number of Observations Read 1339
Number of Observations Used 1339

Analysis of Variance

Source	DF	Sum Squares	Mean Square	F Value	Pr > F
Model	1	204708062	204708062	649592	<.0001
Error	1337	421333	315.13324		
Corrected Total	1338	205129396			

Root MSE	17.75199	R-Square	0.9979
Dependent Mean	1974.75437	Adj R-Sq	0.9979
Coeff Var	0.89895		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9979	7705.2595	7707.2655	7715.65890	X

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	VIF
Intercept	1	800.96967	1.53503	521.79	<.0001	0
X	1	0.99923	0.00124	805.97	<.0001	1.00000

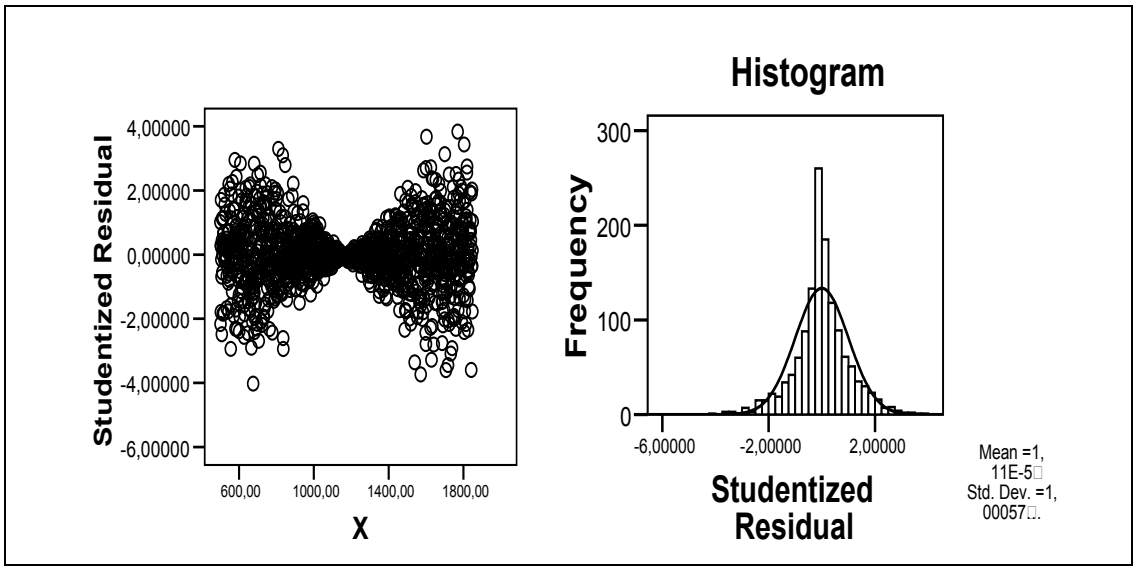
Collinearity Diagnostics

Number	Eigenvalue	Index	Intercept	X
1	1.94875	1.00000	0.02563	0.02563
2	0.05125	6.16618	0.97437	0.97437

Test of First and Second Moment Specification

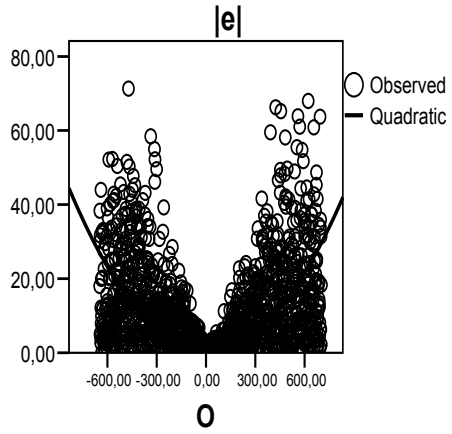
DF	Chi-Square	Pr > ChiSq
2	123.92	<.0001

Durbin-Watson D	1.924
Number of Observations	1339
1st Order Autocorrelation	0.037



n = 1339 örneklem büyüklüğü için en küçük kareler sonuçları incelendiğinde artıklarda kelebek dağılma ve sivrileşme sorunları gözlenmektedir.

n = 1339 için merkezi dışsal değişken ile en küçük kareler mutlak artıklarından standart sapma fonksiyonu (Şekil 4.11) tahmin edilmiştir.



Şekil 4.11. n = 1339 için Standart Sapma Fonksiyonu

Tahmin edilen standart sapma fonksiyona ilişkin özet istatistikler Çizelge 4.21'de verilmiştir.

Çizelge 4.21. Standart Sapma Fonksiyonu İçin Model Özeti ve Parametre Tahminleri

Dependent Variable: |e |

	Model Summary					Parameter Estimates	
Equation	R Square	F	df1	df2	Sig.	b1	b2
Quadratic	.518	718.229	2	1337	.000	-.001	6.21E-005

The independent variable is o.

$\widehat{|e_i|} = -0.001 o_i + 6.2 * 10^{-5} o_i^2$,standart sapma fonksiyonundan elde edilen mutlak artık tahminleri ile ağırlık tahminleri elde edilerek ağırlıklı en küçük kareler uygulanmıştır. Elde edilen sonuçlar Çizelge 4.22'de verilmiştir.

Çizelge 4.22. n = 1339 için Standart Sapma Fonksiyonu ile Ağırlık Tahmini ve AEKK Sonuçları

Dependent Variable: Y
Number of Observations Read 1339
Number of Observations Used 1339

$\hat{w}_i = 1/\widehat{|e_i|}, \quad \widehat{|e_i|} \neq 0$

Analysis of Variance

Source	DF	Sum Squares	Mean Square	F Value	Pr > F
Model	1	21866201	21866201	546575	<.0001
Error	1337	53488	40.00585		
Corrected Total	1338	21919689			

Root MSE	6.32502	R-Square	0.9976
Dependent Mean	1960.46085	Adj R-Sq	0.9976
Coeff Var	0.32263		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9976	4941.6039	4943.6099	4952.00324	x

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	803.02295	1.56686	512.51	<.0001	0
X	1	0.99661	0.00135	739.31	<.0001	1.00000

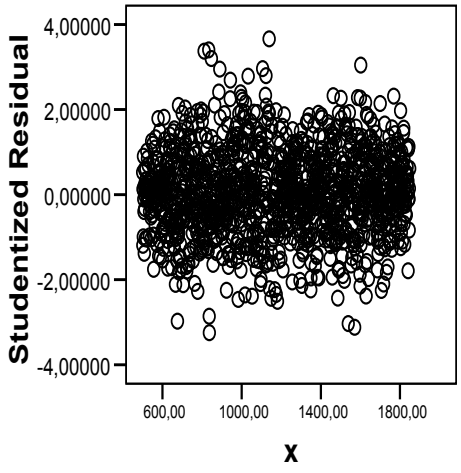
Collinearity Diagnostics

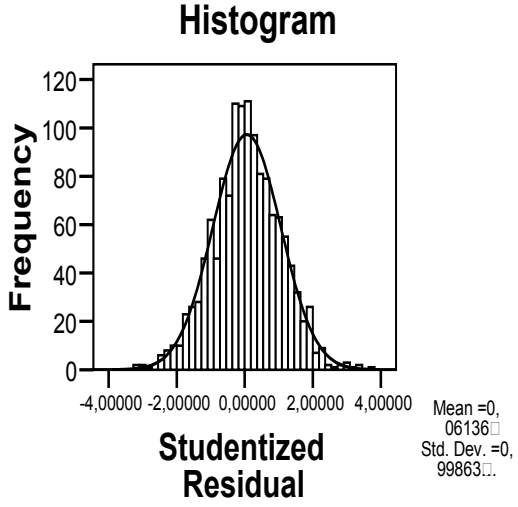
Condition Number	Eigenvalue	--Proportion of Variation- Index	Intercept	X
1	1.99918	1.00000	0.00041079	0.00041079
2	0.00082157	49.32900	0.99959	0.99959

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
3	7.95	0.0471

Durbin-Watson D	**1.914**
Number of Observations	1339
1st Order Autocorrelation	0.043

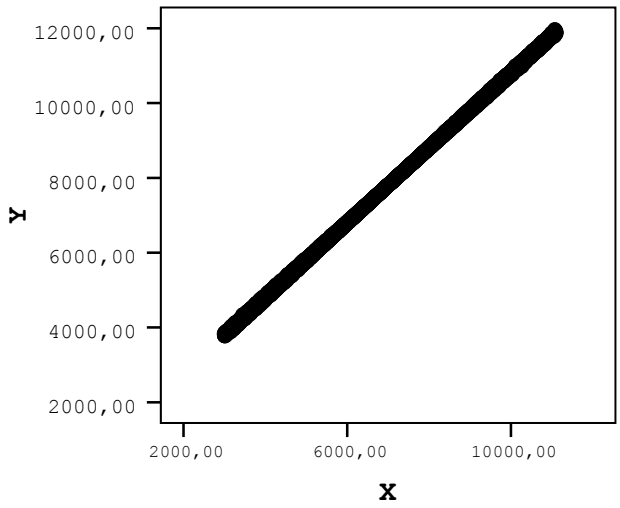




Önerilen yönteminin n = 1339 örneklem büyüklüğü için kelebek dağılan artık sorununu ortadan kaldırdığı ve en küçük kareler artıklarının dağılımındaki sivriliği çok büyük ölçüde düzelttiği görülmektedir. Diğer kriterler incelendiğinde az sayıdaki aykırı değer dışında herhangi bir sorun olmadığı görülmektedir.

4.13.3. Yaklaşık Simetrik Veri n = 8057

Burada, aynı fonksiyon ile veri sayısı bir önceki uygulamaya göre yaklaşık 7 kat büyütülerek önerilen yöntem tekrar sınanmıştır. Veri kümesinin bağımsız değişkeni X ile bağımlı değişkeni Y arasındaki ilişki Şekil 4.12'de verilmiştir.



Şekil 4.12. n = 8057 için Yapay Veri Grafiği

Veri kümesine ağırlıksız en küçük kareler uygulanmış ve Çizelge 4.23'de verilen analiz sonuçları elde edilmiştir.

Çizelge 4.23. n = 8057 için En Küçük Kareler Sonuçları

Dependent Variable: Y
Number of Observations Read 8057
Number of Observations Used 8057

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	44276210162	44276210162	1.361E8	<.0001
Error	8055	2619856	325.24597		
Corrected Total	8056	44278830018			

Root MSE	18.03458	R-Square	0.9999
Dependent Mean	7849.66697	Adj R-Sq	0.9999
Coeff Var	0.22975		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9999	46608.3748	46610.3758	46622	X

Parameter Estimates

Variable	DF	Estimate	SE	t Value	Pr > |t|	VIF
Intercept	1	801.03934	0.63666	1258.20	<.0001	0
X	1	0.99988	0.00008570	11667.5	<.0001	1.00000

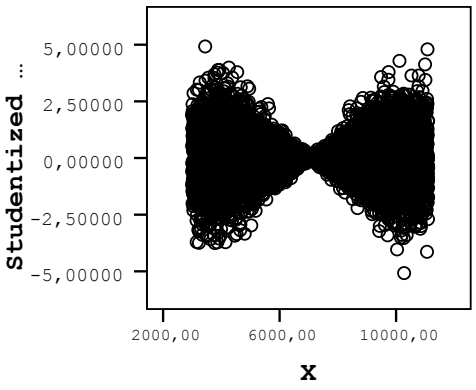
Collinearity Diagnostics

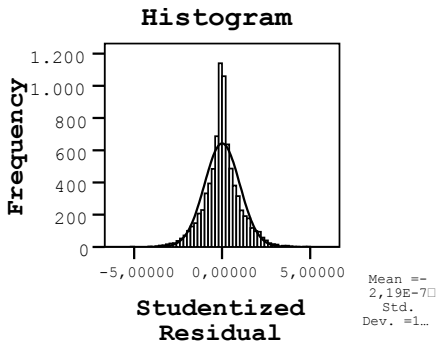
Condition Number	Eigenvalue	Index	Intercept	X
1	1.94890	1.00000	0.02555	0.02555
2	0.05110	6.17555	0.97445	0.97445

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
2	778.63	<.0001

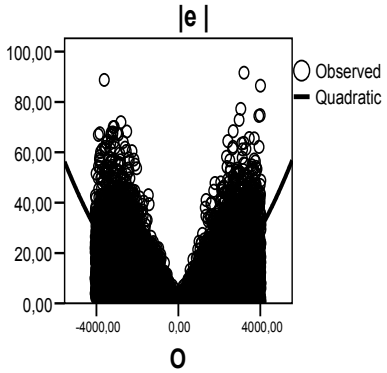
Durbin-Watson D	**1.990**
Number of Observations	8057
1st Order Autocorrelation	0.005





n = 8057 için en küçük kareler sonuçları incelendiğinde kelebek dağılan artıklar dolayısıyla bir değişen varyanslılık sorunu ile birlikte normal dağılımdan sapma da göze çarpmaktadır.

Önerilen yöntem uygulaması için ağırlıksız en küçük kareler artıklarının mutlak değerleri ile merkezi dışsal değişken, standart sapma fonksiyonunun (Şekil 4.13) elde edilmesinde kullanımıştır.



Şekil 4.13. n = 8057 için Standart Sapma Fonksiyonu

Elde edilen standart sapma fonksiyonuna ilişkin istatistikler Çizelge 4.24'de verilmiştir.

Çizelge 4.24. Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri

Dependent Variable: |e|

	Model Summary					Parameter Estimates	
Equation	R Square	F	df1	df2	Sig.	b1	b2
Quadratic	.558	5081.557	2	8055	.000	4.42E-005	1.83E-006

The independent variable is O.

$\widehat{|e_i|} = -4.42*10^{-5} o_i + 1.83*10^{-6} o_i^2$, standart sapma fonksiyonundan elde edilen mutlak artık tahminleri ağırlık elde etmede kullanılarak ağırlıklı en küçük kareler gerçekleştirilmiştir. Analiz sonuçları Çizelge 4.25'te verilmiştir.

Çizelge 4.25. n = 8057 için Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu Tahmini ve AEKK Sonuçları

Dependent Variable: Y

$\hat{w}_i = 1/\left|\widehat{e_i}\right|, \qquad \left|\widehat{e_i}\right| \neq 0$

Number of Observations Read 8057
Number of Observations Used 8057

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4404754406	4404754406	1.124E8	<.0001
Error	8055	315585	39.17880		
Corrected Total	8056	4405069992			

Root MSE	6.25930	R-Square	0.9999
Dependent Mean	7850.51094	Adj R-Sq	0.9999
Coeff Var	0.07973		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9999	29556.1688	29558.1698	29570	X

Parameter Estimates

Variable	DF	Estimate	SE	t Value	Pr > |t|	VIF
Intercept	1	799.16852	0.66524	1201.33	<.0001	0
X	1	1.00009	0.00009432	10603.2	<.0001	1.00000

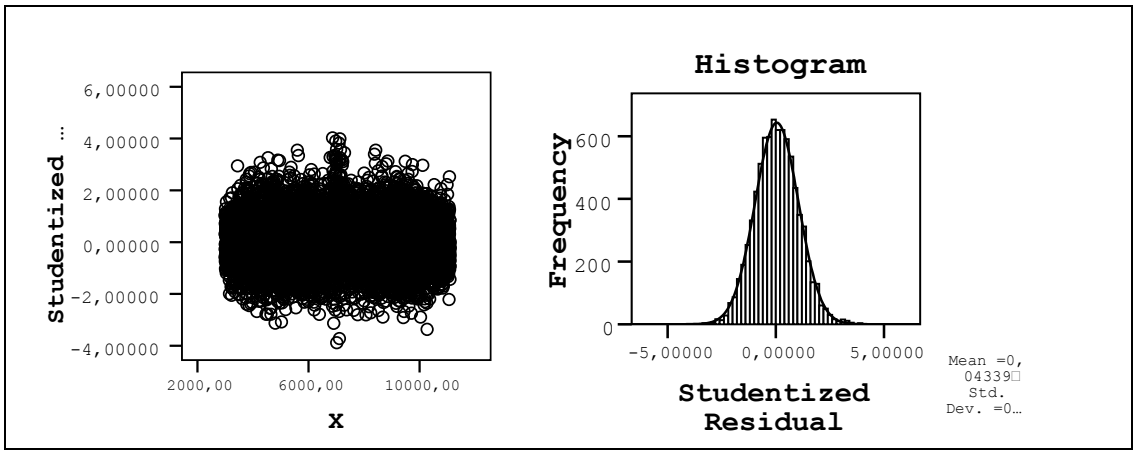
Collinearity Diagnostics

			--Proportion of Variation-	
Number	Eigenvalue	Index	Intercept	X
1	1.99968	1.00000	0.00016062	0.00016062
2	0.00032124	78.89803	0.99984	0.99984

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
3	10.63	0.0139

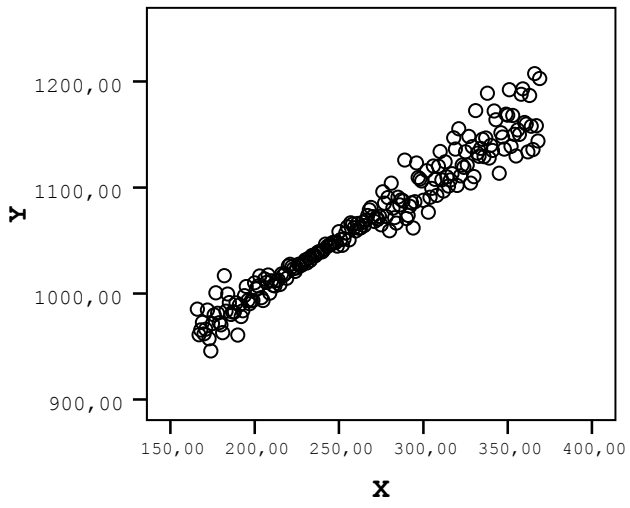
Durbin-Watson D	**1.972**
Number of Observations	8057
1st Order Autocorrelation	0.014



Önerilen yönteminin n = 8057 örneklem büyüklüğü için kelebek dağılan artıklar sorununu ortadan kaldırdığı görülmektedir. Buna ek olarak önerilen yöntem, en küçük kareler artık dağılımındaki normal dağılıma uymama sorununu da çözmüştür. Diğer kriterler incelendiğinde ise merkez çevresinde göze çarpan aykırı değerler dışında herhangi bir sorun olmadığı görülmektedir.

4.13.4. Sağa Asimetrik Veri n = 203

Önerilen yöntem, yaklaşık simetrik farklı örneklem büyüklüğüne sahip veri kümelerindeki kelebek dağılan artıklar sorununu başarılı biçimde çözmüştür. Yöntemin simetrik olmayan kelebek dağılan artıklara etkisi, önce sağa asimetrik (Şekil 4.14) sonrasında ise sola asimetrik (Şekil 4.16) veri kümelerinde de sınanmıştır.



Şekil 4.14. n = 203 için Yapay Veri Grafiği

Bağımsız değişkeni ile bağımlı değişkeninin ilişkisi, Şekil 4.14'te verilen veri kümesine ağırlıksız en küçük kareler uygulanmıştır. Ağırlıksız en küçük kareler analizi sonuçları Çizelge 4.26'da verilmiştir.

Çizelge 4.26. Sağa Asimetrik Veri için En Küçük Kareler Sonuçları

Dependent Variable: Y
Number of Observations Read 203
Number of Observations Used 203

Analysis of Variance

Source	DF	Sum Squares	Mean Square	F Value	Pr > F
Model	1	737497	737497	3768.03	<.0001
Error	201	39341	195.72480		
Corrected Total	202	776838			

Root MSE	13.99017	R-Square	0.9494
Dependent Mean	1067.96943	Adj R-Sq	0.9491
Coeff Var	1.30998		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9494	1073.1621	1075.2017	1079.78854	X

Parameter Estimates

Variable	DF	Estimate	SE	t Value	Pr > |t|		VIF
Intercept	1	794.36523	4.56411	174.05	<.0001	<.0001	0
X	1	1.02547	0.01671	61.38	<.0001	<.0001	1.00000

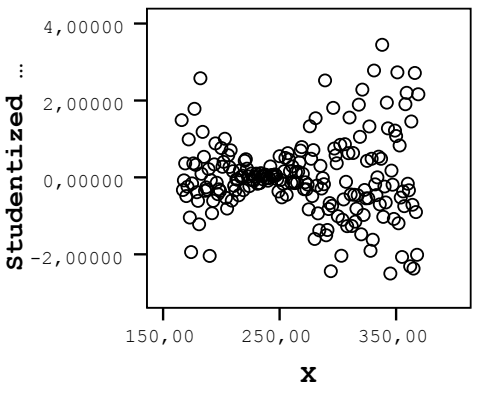
Collinearity Diagnostics

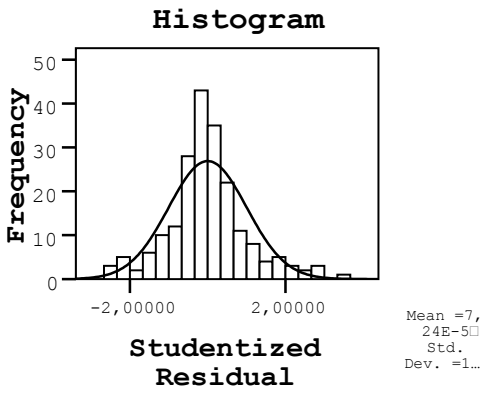
Condition Number	Eigenvalue	Index	Intercept	X
1	1.97658	1.00000	0.01171	0.01171
2	0.02342	9.18748	0.98829	0.98829

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
2	20.03	<.0001

Durbin-Watson D	**2.255**
Number of Observations	203
1st Order Autocorrelation	-0.144

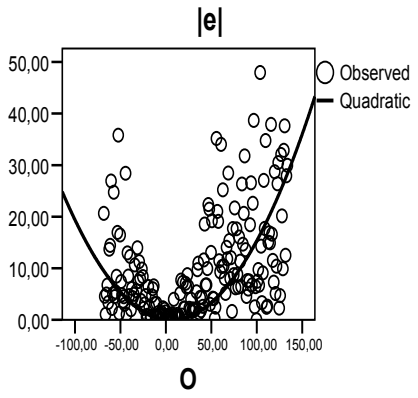




Sağa asimetrik veri için uygulanan en küçük kareler çıktıları incelendiğinde kelebek dağılan artıklar için değişen varyanslılık ve normal dağılımdan sapma gibi sorunların olduğu gözlenmektedir. Diğer kriterler incelendiğinde ise model ve parametrelerin

istatistiksel olarak anlamlı olduğu, ardışık artıklar arasında herhangi bir ilişki olmadığı anlaşılmaktadır.

Sorunun çözümü için ağırlıksız en küçük karelerin mutlak artıkları, merkezi dışsal değişken ile standart sapma fonksiyonun tahmininde (Şekil 4.15) kullanılmıştır.



Şekil 4.15. Sağa Asimetrik Veri için Standart Sapma Fonksiyonu

Elde edilen standart sapma fonksiyonu özeti Çizelge 4.27'de verilmiştir.

Çizelge 4.27. Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri

Dependent Variable | e|

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	b1	b2
Quadratic	.564	130.173	2	201	.000	-.020	.002

The independent variable is O.

Bu standart sapma fonksiyonu tahmininden elde edilen tahmin değerlerinin tersi ağırlık tahminlerinin elde edilmesinde kullanılarak ağırlıklı en küçük kareler analiz gerçekleştirilmiştir. Söz konusu analizin sonuçları Çizelge 4.28'de verilmiştir.

Çizelge 4.28. Sağa Asimetrik Veride Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu Tahmini ve AEKK Sonuçları

Dependent Variable Y
Number of Observations Read 203
Number of Observations Used 203

$$\widehat{|e_i|} = -0.020o_i1 + 0.002o_i^2$$

$$\hat{w}_i = 1/\left|\widehat{e_i}\right|, \quad \left|\widehat{e_i}\right| \neq 0$$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	135450	135450	3580.83	<.0001
Error	201	7603.12846	37.82651		
Corrected Total	202	143053			

Root MSE	6.15033	R-Square	0.9469
Dependent Mean	1040.59765	Adj R-Sq	0.9466
Coeff Var	0.59104		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9469	739.4911	741.5308	746.11756	X

Parameter Estimates

Variable	DF	Estimate	SE	t Value	Pr > |t|	VIF
Intercept	1	794.73784	4.11348	193.20	<.0001	0
X	1	1.02627	0.01715	59.84	<.0001	1.00000

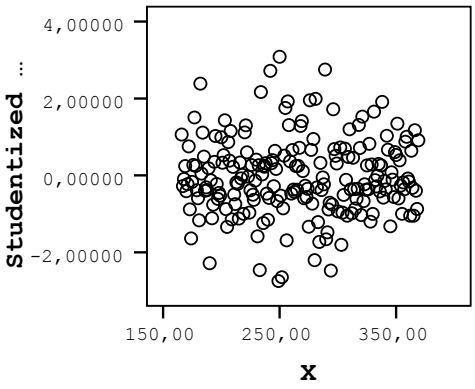
Collinearity Diagnostics

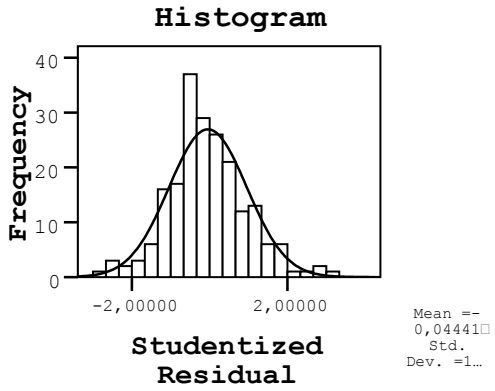
Number	Eigenvalue	Index	Intercept	X
1	1.99882	1.00000	0.00059156	0.00059156
2	0.00118	41.10295	0.99941	0.99941

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
3	1.03	0.7950

Durbin-Watson D	**2.208**
Number of Observations	203
1st Order Autocorrelation	-0.109

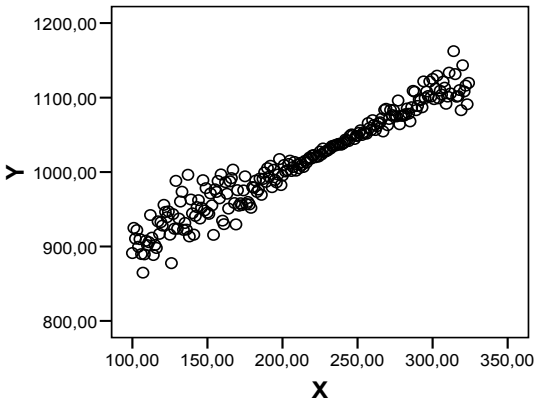




Önerilen yöntemin sağa asimetrik olan verideki değişen varyanslılık ve normal dağılımdan sapma sorunlarını başarılı bir biçimde ortadan kaldırdığı görülmektedir.

4.13.5. Sola Asimetrik Veri n = 223

Önerilen yöntem son olarak 223 örneklem büyüklüğünde sola asimetrik veri için sınanmıştır. Yapay veri üretme fonksiyonu ile elde edilen bağımsız ve bağımlı değişkenin grafiği Şekil 4.16'de verilmiştir.



Şekil 4.16. n = 223 Sola Asimetrik Veri Grafiği

Veriye ağırlıksız en küçük kareler uygulanarak Çizelge 4.29'da verilen sonuçlar elde edilmiştir.

Çizelge 4.29. Sola Asimetrik Veri için EKK Sonuçları

The REG Procedure
Dependent Variable: Y
Number of Observations Read 223
Number of Observations Used 223

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	875363	875363	3572.83	<.0001
Error	221	54146	245.00571		
Corrected Total	222	929509			

Root MSE	15.65266	R-Square	0.9417
Dependent Mean	1013.15760	Adj R-Sq	0.9415
Coeff Var	1.54494		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9417	1228.7768	1230.8128	1235.59110	X

Parameter Estimates

Variable	DF	Estimate	S.E	t Value	Pr > |t|	V.I.F
Intercept	1	807.46421	3.59733	224.46	<.0001	0
X	1	0.97021	0.01623	59.77	<.0001	1.00000

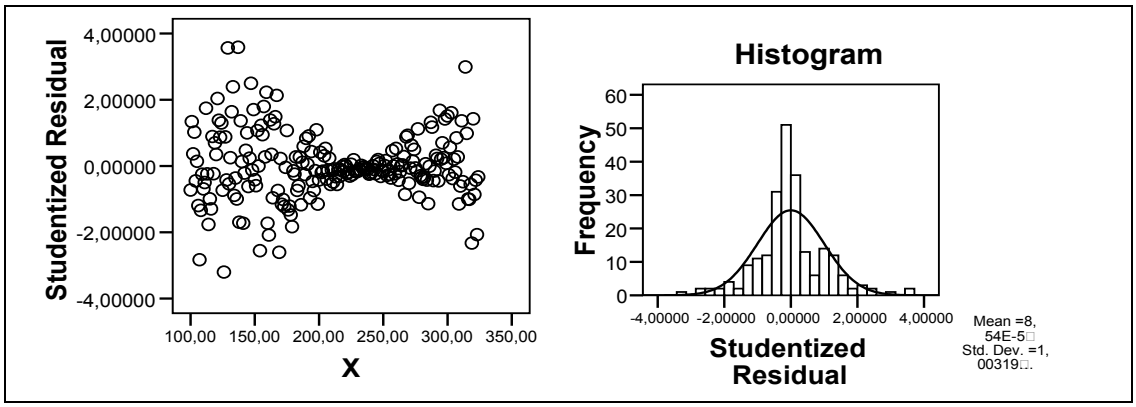
Collinearity Diagnostics

Condition Number	Eigenvalue	Index	Intercept	X
	--Proportion of Variation-			
1	1.95661	1.00000	0.02170	0.02170
2	0.04339	6.71504	0.97830	0.97830

Test of First and Second Moment Specification

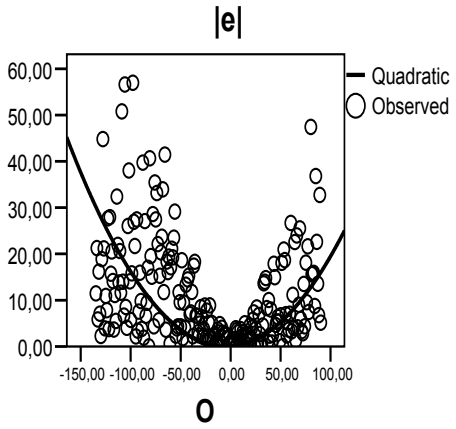
DF	Chi-Square	Pr > ChiSq
2	31.39	<.0001

Durbin-Watson D	**1.798**
Number of Observations	223
1st Order Autocorrelation	0.096



Sola asimetrik veri için en küçük kareler, sola asimetrik kelebek dağılan artıklar ve normal dağılımdan sapmalara neden olmuştur.

Önerilen yöntem ile en küçük kareler mutlak artıklarından standart sapma fonksiyonu (Şekil 4.17) tahmin edilmiştir. Tahmin edilen standart sapma fonksiyonuna ilişkin özet istatistikler Çizelge 4.30'da verilmiştir.



Şekil 4.17. Sola Asimetrik Veri için Standart Sapma Fonksiyonu

Çizelge 4.30. Standart Sapma Fonksiyonu için Model Özeti ve Parametre Tahminleri

Dependent Variable: |e|

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	b1	b2
Quadratic	.510	115.669	2	221	.000	.017	.002

The independent variable is O.

$\widehat{|e_i|} = 0.017o_i + 0.002o_i^2$ standart sapma fonksiyonundan elde tahmin değerlerinin tersleriyle $\hat{w}_i = 1/\widehat{|e_i|}, \quad \widehat{|e_i|} \neq 0$ şeklinde ağırlıklar elde edilerek ağırlıklı en küçük kareler uygulanmıştır. Analizin sonuçları Çizelge 4.31'de verilmiştir.

Çizelge 4.31. Sola Asimetrik Veri için Merkezi Dışsal Değişken ile Standart Sapma Fonksiyonu Tahmini ve AEKK Sonuçları

The REG Procedure
Dependent Variable: Y
Number of Observations Read 223
Number of Observations Used 223

Weight: $\hat{w}_i$

Analysis of Variance

Source	DF	Sum Squares	Mean Square	F Value	Pr > F
Model	1	217.51411	217.51411	3430.18	<.0001
Error	221	14.07742	0.06341		
Corrected Total	222	231.59153			

Root MSE	0.25182	R-Square	0.9392
Dependent Mean	1034.87094	Adj R-Sq	0.9389
Coeff Var	0.02433		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9392	-615.8246	-613.7887	-609.00127	X

Parameter Estimates

Variable	DF	Estimate	SE	t Value	Pr > |t|	VIF
Intercept	1	798.88616	4.03149	198.16	<.0001	0
X	1	1.00784	0.01721	58.57	<.0001	1.00000

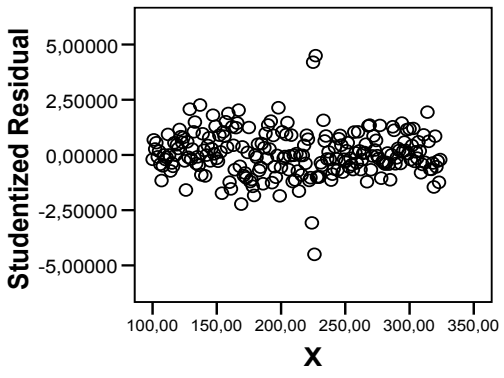
Collinearity Diagnostics

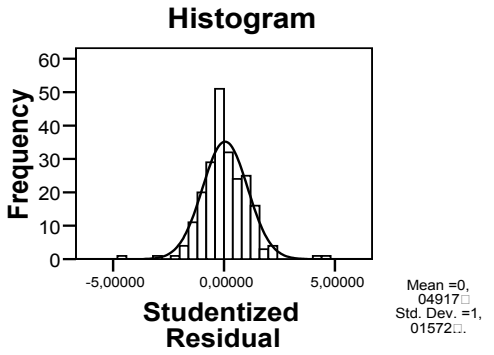
Number	Eigenvalue	Condition Index	Proportion of Variation- Intercept	X
1	1.99945	1.00000	0.00027702	0.00027702
2	0.00055404	60.07360	0.99972	0.99972

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
3	3.64	0.3028

Durbin-Watson D	**1.966**
Number of Observations	223
1st Order Autocorrelation	0.017





Önerilen yöntemin sola asimetrik olan verideki artıkların normal dağılımdan sapma ve kelebek dağılan artıklar için değişen varyanslılığı ortadan kaldırdığı gözlenmektedir. Bununla birlikte az sayıda da aykırı değer ortaya çıkmıştır.

4.13.6. Merkezi Dışsal Değişkenle Doğrudan Ağırlık Tahmini ve Ağırlıklı En Küçük Kareler

Kesim 3.14.2'de belirtildiği üzere merkezileştirilmiş dışsal değişkenin mutlak değerlerinin tersi doğrudan ağırlık tahminlerini elde etmekte kullanılabilir. Yöntem,

kesim 4.12.5'de verilen sola asimetrik veri kümesine uygulanmış ve Çizelge 4.32'de verilen analiz sonuçları elde edilmiştir.

Çizelge 4.32. Mutlak Merkezi Dışsal Değişken ile AEKK Sonuçları

The REG Procedure
Dependent Variable: Y
Number of Observations Read 223
Number of Observations Used 223

$$\hat{w}_i = 1/|o_i|, \quad |o_i| \neq 0$$

Analysis of Variance

Source	DF	Squares	Square	F Value	Pr > F
Model	1	198.46706	198.46706	3699.62	<.0001
Error	221	11.85560	0.05365		
Corrected Total	222	210.32266			

Root MSE	0.23161	R-Square	0.9436
Dependent Mean	1040.09926	Adj R-Sq	0.9434
Coeff Var	0.02227		

Model	R-Square	AIC	BIC	SBC	Variables in Model
1	0.9436	-650.3649	-648.3288	-643.55052	X

Parameter Estimates

Variable	DF	Estimate	SE	t Value	Pr > |t|	VIF
Intercept	1	811.02739	3.76655	215.32	<.0001	0
X	1	0.95398	0.01568	60.82	<.0001	1.00000

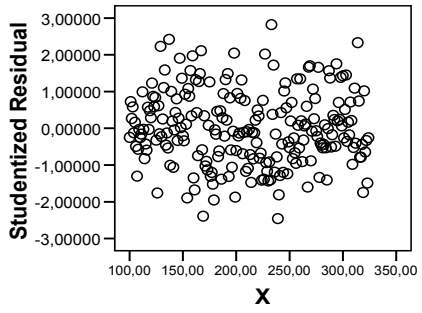
Collinearity Diagnostics

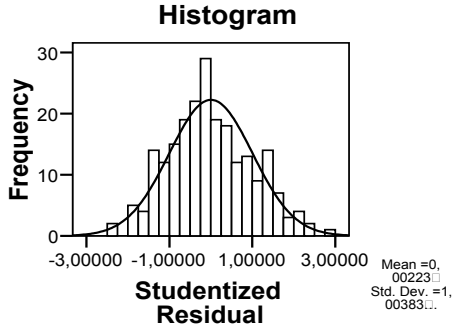
Condition Number	Eigenvalue	Index	Proportion of Variation- Intercept	X
1	1.99988	1.00000	0.00005777	0.00005777
2	0.00011554	131.56151	0.99994	0.99994

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
3	4.70	0.1951

Durbin-Watson D	**1.885**
Number of Observations	223
1st Order Autocorrelation	0.055





Mutlak merkezi dışsal değişkenin doğrudan ağırlık tahmininde kullanılması sonucu kelebek dağılan artıklar için değişen varyanslılığın başarılı bir şekilde dengelendiği görülmektedir. Buna ek olarak yöntemin, Çizelge 4.29'da verilen en küçük kareler artık dağılımındaki bozulmayı giderdiği ve artıkları normal dağılıma uydurduğu görülmektedir. Aynı veri kümesi için değişim katsayıları incelendiğinde ise bu veri

kümesi için en homojen dağılmanın merkezileştirilmiş değişkenin ağırlıklandırmada kullanılması sonucunda elde edildiği gözlenmektedir.

5. SONUÇ ve ÖNERİLER

Bu çalışmada, kelebek dağılan artıkların varlığı durumunda, monoton olmayan değişen varyanslılığın düzeltilmesi için aynı veri kümesine, önerilen yöntemler dışında, 11 başlık altında 8 yöntem uygulanmıştır.

Kelebek dağılan artıklar dolayısıyla değişen varyanslılık sorununu düzeltmek üzere uygulanan, iteratif algoritmalara dayanan Meloun ve diğ., (2002) tarafından önerilen yeniden ağırlıklandırılmış iteratif en küçük kareler, Chatterjjee ve Machler (1997) tarafından önerilen iteratif yöntem, bağımlı değişken dönüşümlerine dayanan Box - Cox (1964) dönüşümleri, yeniden örneklemeye dayanan Jackknife, Bootstrap yöntemleri, karesel ve mutlak artıkların parametrik olmayan düzlemesi istenen sonucu vermemiştir. Adı geçen yöntemlerden Meloun ve diğ. (2002) tarafından önerilen yöntem, iki tepeli bir dağılım sonucuna neden olduğu için normal dağılımı bozmuş ve çoklubağlantı sorunu yaratmıştır.

Dönüşüm ve iteratif algoritmalara dayanan Bischoff ve diğ. (2006) tarafından önerilen değiştirilmiş ardışık artıklar yönteminin yanı sıra, veri kümesinin tekrarlı gözlemler için, karesel ve mutlak artıkların standart sapmalarıyla ağırlıklandırılması, değişen varyanslılık sorununu kısmen ortadan kaldırmıştır. Ancak karesel artıkların standart sapmalarının ağırlık tahmininde kullanılması, bağımlı değişken standart sapmalarının ağırlık tahmininde kullanılmasına göre varyansı dengeleme açısından daha iyi sonuç vermiştir.

Yukarıda sözü edilen yöntemler mevcut pek çok istatistik paket programında yer almamaktadır. Özellikle iteratif olan yöntemler için program kodu yazmak dışında neredeyse seçenek yoktur. Bu bakımdan yöntemlerin kullanımı oldukça zor ve karmaşıktır.

Bu çalışmada önerilen üç yöntemden ilki, ağırlıkların, merkezi dışsal değişken ile karesel ya da mutlak artıklardan varyans ya da standart sapma fonksiyonlarından elde edilmesidir. Bu yöntem, hem uygulama verisi hem de farklı örneklem büyüklüklerinde; yaklaşık simetrik yapay veri kümeleri için kelebek dağılan artıklar

sorununu giderme ve normal dağılıma yaklaştırma konusunda başarılı sonuçlar vermiştir. Söz konusu yöntem, sağa asimetrik ve sola asimetrik yapay veri kümelerine uygulandığında da kelebek dağılan artıklar sorununu giderme ve normal dağılıma yaklaştırma konusunda başarılı sonuçlar vermiştir. Yukarıda sayılan avantajlarına karşılık, merkezi dışsal değişken ile artık dönüşümlerinden, varyans ya da standart sapma fonksiyonları aracılığıyla, ağırlık tahmini elde edilmesinin, görünen tek dezavantajı, boğum noktası çevresinde çok küçük artıklara karşılık gelen gözlemlerin aşırı ağırlık alması dolayısıyla aykırı değerlere neden olabilmesidir.

Önerilen ikinci yöntem olan, merkezi dışsal değişkenin mutlak değerlerinin doğrudan ağırlık tahmini elde etme sürecinde kullanılması, kelebek dağılan artık sorununu giderme konusunda iyi sonuç vermiştir.

Önerilen yöntemlerin bir parçası olan artıklar için merkez ya da boğum noktası elde edilmesi, kelebek ya da galaksi dağılan artıklarla çalışılırken verinin doğru bir biçimde parçalanmasına ve farklı analiz senaryolarına da olanak sağlamıştır.

Bu konuda, merkezi dışsal değişken ile varyans ya da standart sapma fonksiyonları aracılığıyla ağırlık tahminleri elde edilirken, aykırı değerlerin önlenmesi üzerinde çalışılabilir.

KAYNAKLAR

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

Albayrak, A. S., 2008. Değişen Varyans Durumunda En Küçük Kareler Tekniğinin Alternatifi Ağırlıklı Regresyon Analizi ve Bir Uygulama, *Afyon Kocatepe Üniversitesi, İ.İ.B.F. Dergisi.*

Bagnall, A., Whittley, I., Studley, M., Pettipher, M., Tekiner, F., Bull, L., 2006. Variance Statbilizing Regression Ensembles for Enviromental Models, *International Joint Conference on Neural Networks*, Sheraton Vancouver BC, Canada July 16-21.

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

Bischoff, W., **Heck, B., Howind, J., Teusch., A.,** 2006. A procedure 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.

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

Box, G. E. P and Hill, W.J., 1974. Correcting inhomogeneity of variance with power transformation weighting, *Technometrics*, 16, 385-389.

Carroll, R.J and Davidian, M., 1987. Variance Function Estimation, *Journal of the American Statistical Association*, 82, 1079-1091.

Carroll, R. J., 1982. Adapting for Heteroscedasticity in Linear Models, *The Annals of Statistics*, 10, 1224-1233.

Carroll, R.J. and Ruppert, D., 1982. Robust Estimation in Heteroscedastic Linear Models, *The Annals of Statistics,* 10, 429-44.

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

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

Cook, R. D. and Weisberg, S., 1982. Residuals and Influence in Regression, Chapman and Hall.

Darlington, R. B., 1997. Correcting the Standard Errors of Regression Slopes for Heteroscedasticity, http://www.psych.cornell.edu/darlington/heterosc.htm.

Draper, N.R. and H. Smith, 1998. Applied Regression Analysis, 3rd ed. Wiley.

Eviews 4.1 User Guide, 2002. Quantitative Micro Software, LLC.

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

Genceli, M., 2001. Ekonometri ve İstatistik İlkeleri, Filiz Kitapevi.

Gujaratti, D., 2001. Basic Ekonometrics.

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

Hadi, A. S. and Chatterjee, S., 2006. Regression Analysis by Example, Wiley.

Hamasaki, T. and Goto ,M., 2005. Power-Transformatıon Model to Additıvity in Regression, *Behaviormetrika*, Vol.32, No.2, 111–125.

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

Hinkley, D., 2010, Jackknife Methods, *International Statistical Review to Line Intersect. Sampling*, Encyclopedia of Statistical Science Vol.6, Wiley.

Kianifrad, F. and Swallow, W. H., 1996. A Review of the Development and Application of Recursive Residuals in Linear Models, *Journal of the American Statistical Association*, 91, 391-400.

Kıroğlu, G., 2001. Uygulamalı Parametrik Olmayan İstatistiksel Yöntemler, İstanbul.

Kuk, A. Y. C, 1999. Nonparametrically Weighted Least Squares Estimation in Heteroscedastic Linear Regression, *Biometrical Journal* 41, 4: 401-410.

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W., 2005. Applied Linear Statistical Models, McGraw-Hill.

McCullagh, P. and Nelder, J.A., 1983. Generalized Linear Models, Chapman and Hall.

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

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

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

Paulson, D. S., 2006. Handbook of Regression and Modelling, CRC.

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

Rawlings, O. R., Pantula, S. G. and Dickey, D. A., 1998. Applied Regression Analysis, 2nd ed.

Ryan, T. P., 1997. Modern Regression Methods, Wiley.

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

Shick, A., 1999. Improving weighted least squares estimates in heteroscedastic linear regression when the variance is a function of the mean response, *Journal of Statistical Planning and Inference*, 76 127-144.

Silverman, E. and Kot, M., 2000. Rate Estimation for a Simple Movement Model, *Bulletin of Mathematical Biology*, 58, 351-375.

SPSS Trends 10.0, 1999. SPSS Inc.

Verbyla, A.P.,1993. Modelling Variance Heterogeneity Residual Maximum Likelihood and Diagnostics, *Journal of the Royal Statistical Society*, 55, 493-508.

Weisberg, S., 2005. Applied Linear Regression, 3rd ed., Willy.

Willet, J. B. and Singer, J. D., 2002. Ceveats for the Statistical Consumer: A Cautionary Note on Case Weighting as a Method of Implementing Weighted Least-Squares Regression, *Journal of Experimental Educations*, 286-293.

Wu, J., 1986. Jackknife, Bootstrap and Other Resampling Methods in Regresyon Analysis, *The Annals of Statistics*, Vol. 14, No 4.

url_1 http://www.tuik.gov.tr/cevredagitimapp/hava.zul.

url_2 http://www.rshm.saglik.gov.tr/hki/hkindex/index5.htm.

ÖZGEÇMİŞ

Adı Soyadı : Reşit ÇELİK
Doğum Yeri : Sivas
Doğum Tarihi : 29.12.1970
Medeni Hali : Evli ve 1 çocuk sahibi

Eğitim Durumu
İlkokul : Gedikçayır İlkokulu Sivas-Atatürk Çiftliği İlkokulu İstanbul (1981)
Ortaokul : Gazi İlköğretim Okulu İstanbul (1986)
Lise : Pertevniyal Lisesi (1989)
Lisans : Yıldız Teknik Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü (1994)
Yüksek Lisans : Marmara Üniversitesi, Sosyal Bilimleri Enstitüsü, İstatistik A.B.D. (2003)
Yabancı Dil : İngilizce

İş Tecrübesi
1997-1999- : Milli Eğitim Bakanlığı (Öğretmen)
1999- : İstatistik Bölümü Araştırma Görevlisi (Yıldız Teknik Üniversitesi)