

T. C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ PROGRAMI
DOKTORA TEZİ

ROBUST (DAYANIKLI) REGRESYON YÖNTEMLERİ
ÜZERİNE BİR ARAŞTIRMA

Yasin BÜYÜKKÖR

Danışman
Prof. Dr. Ali Kemal ŞEHİRLİOĞLU

İZMİR- 2020

TEZ ONAY SAYFASI



YEMİN METNİ

Doktora tezi olarak sunduğum “Robust (Dayanıklı) Regresyon Yöntemleri Üzerine Bir Araştırma” adlı çalışmanın, tarafımdan, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

.../.../2020

Yasin BÜYÜKKÖR

ÖZET

Doktora Tezi

Robust (Dayanıklı) Regresyon Yöntemleri Üzerine Bir Araştırma

Yasin BÜYÜKKÖR

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Ekonometri Anabilim Dalı

Ekonometri Programı

Birçok regresyon uygulamasında, hataların dağılışı normal olarak varsayılmakta ve parametre tahmini yapılırken En Küçük Kareler (EKK) tahminleme yöntemi kullanılmaktadır. Ancak pratikte hataların dağılışı normal olarak varsayılsa bile, veri seti çözümlendikten sonra elde edilen artıklar genellikle normal dağılışı göstermezler. Eğer veri seti içerisinde uzak gözlem(ler) (sapan) veya sapan olduğundan şüphelenilen gözlem(ler) varsa, normallik varsayımı bozulabilir ve EKK ile yapılan parametre tahminleri sapmalı olabilir. Birçok istatistikçi böyle problemler ortaya çıktığında parametre tahminleri için dayanıklı (robust) yöntemlere başvurmuşlardır. Bu yöntemlerin başında En Çok Olabilirlik (EÇO) Tahminleme yönteminin geliştirilmiş bir versiyonu olan M- Tahminleme Yöntemi gelmektedir. Ancak geleneksel M- Tahminleyiciler eğer veri seti çarpık ve normalden daha fazla basıklık değerine sahipse iyi bir çözüm elde edememektedir. Bu tez çalışmasında Pearson Diferansiyel Denklemi ile Etki Fonksiyonu (IF) arasındaki ilişki kullanılarak Pearson Tip VI (PVI) dağılışıma uyum gösteren veri setleri için M- Tahminleme yöntemi önerilmiştir. Bu yöntemin avantajı, geleneksel M- Tahminleyiciler veri setinin çarpıklık ve basıklık değerlerini dikkate almazken Pearson Diferansiyel Denkleminin bu değerleri dikkate alarak farklı çarpıklık ve basıklık değerleri için dinamik çözümler üretmesini sağlamaktır. PVI dağılışının Olasılık Yoğunluk Fonksiyonu (oyf) kullanılarak Amaç, Etki ve Ağırlık fonksiyonları elde edilmiştir. Ayrıca PVI dağılışının kuyruk özellikleri incelenerek simülasyon çalışmalarındaki davranışı gözlemlenmiştir. Regresyon parametrelerinin tahminini yapabilmek için Ağırlık Fonksiyonu kullanılarak İteratif Olarak Tekrar Ağırlıklandırılan En Küçük Kareler Tahminleme Yöntemi (IRWLS) kullanılmıştır.

Farklı senaryolara sahip simülasyon çalışmaları ve gerçek veri setleri kullanılarak önerilen yöntemin performansı Toplam Mutlak Sapma (TMS) ve Ortalama Karesel Hata (OKH) kriterleri açısından diğer M- Tahminleyiciler ile karşılaştırılmıştır.

Anahtar Kelimeler: M- Tahminleme Yöntemi, Dayanıklı (Robust) Regresyon, Pearson Tip VI Dağılışı, Etki Fonksiyonu, İteratif Olarak Tekrar Ağırlıklandırılan En Küçük Kareler Yöntemi.



ABSTRACT
Doctoral Thesis
Doctor of Philosophy (PhD)
A Research on Robust Regression Methods
Yasin BÜYÜKKÖR

Dokuz Eylul University
Graduate School of Social Sciences
Department of Econometrics
Econometrics Program

In many regression applications, the distribution of the error is assumed to be normal and Least Squares (LS) method is used for parameter estimation. However, in practice, even if the distribution of errors is assumed to be normal, residuals are not generally normally distributed. If the data contains outlier (s) or there are observations (s) which suspected to be outlier, the assumption of normality is violated and parameter estimates which made by using the OLS, will biased. Many statisticians used robust methods for parameter estimates when such problems occur. One of these methods is the M-Estimation Method, a generalized version of the Maximum Likelihood (ML) Estimation method. However, traditional M-Estimators can not achieve a good solution if the data set has skewness and excess kurtosis. In this thesis, using the relationship between Pearson Differential Equation and Influence Function (IF), M-Estimation method is proposed for datasets that are follow Pearson Type VI (PVI) distribution. The advantage of this method is that, while the traditional M-Estimators do not take into account the skewness and kurtosis values of the data set, the Pearson Differential Equation takes into account these values and generate dynamic solutions for different skewness and kurtosis values. Objective, Influence and Weight functions are obtained by using the Probability Density Function (PDF) of the PVI distribution. In addition, the tail properties of PVI distribution are examined and its behavior in simulation studies is observed. By using the Weight Function, the Iteratively Re- Weighted Least Squares Estimation Method (IRWLS) is used to estimate regression parameters. The performance of the proposed method is compared with other M-Estimators in terms of Total Absolute Deviation (TAB)

and Mean Square Error (MSE) criteria by using simulation studies with different scenarios and real data sets.

Keywords: M- Estimation Method, Robust Regression, Pearson Type VI Distribution, Influence Function, Iteratively Re- Weighted Least Squares Method.



ROBUST (DAYANIKLI) REGRESYON YÖNTEMLERİ ÜZERİNE BİR ARAŞTIRMA

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	vi
İÇİNDEKİLER	viii
KISALTMALAR	xi
ŞEKİLLER LİSTESİ	xii
TABLolar LİSTESİ	xiii
EKLER LİSTESİ	xiv
GİRİŞ	1

BİRİNCİ BÖLÜM

ROBUST (DAYANIKLI) İSTATİSTİKSEL YÖNTEMLER

1.1. DAYANIKLILIK ÖLÇME YÖNTEMLERİ	7
1.1.1. Etki Fonksiyonu	7
1.1.2. Kırılma Noktası	9
1.2 M- TAHMİNLEYİCİLER	10
1.2.1. Konum Parametresinin M-Tahminleyicisi	11
1.2.2. Ölçek Parametresinin M-Tahmini	14
1.2.3. Konum ve Ölçek Parametrelerinin M- Tahminleyicileri	15
1.2.4. M- Tahminleyicilerin Etki Fonksiyonu	16
1.3. DAYANIKLI REGRESYON	17
1.3.1. Regresyon M- Tahminleyicisi	21
1.3.1.1. Regresyon Parametrelerinin M- Tahminleyicileri	21
1.3.1.2. Regresyon M- Tahminleyicilerin Asimtotik Dağılışı	24

1.4. HUBER VE TUKEY M- TAHMİNLEME YÖNTEMLERİ	27
1.4.1. Huber M- Tahminleyici	27
1.4.2. Tukey M- Tahminleyici	29

İKİNCİ BÖLÜM

PEARSON DAĞILIŞ SİSTEMİ

2.1. PEARSON DİFERANSİYEL DENKLEMİ	32
2.2. DİFERANSİYEL DENKLEMİN ÇÖZÜMÜNDE PEARSON YAKLAŞIMI	32
2.3. PEARSON DİFERANSİYEL DENKLEMİNİN ÇÖZÜMÜ İÇİN CRAİĞ YAKLAŞIMI	37
2.4. PEARSON DAĞILIŞLARINI TANIMLAYAN KARTLAR	38
2.4.1. Pearson Kartı	38
2.4.2 Craig Kartı	39
2.5. PEARSON ANA TİPLERİ İÇİN GENEL ÇÖZÜM	40
2.5.1. Pearson Tip I (PI) Dağılışı	41
2.5.2 Pearson Tip IV (PIV) Dağılışı	44
2.5.3. Pearson Tip VI (PVI) Dağılışı	46
2.5.3.1. Sağa Çarpık PVI dağılışı	47
2.5.3.2 Sola Çarpık PVI dağılışı	48
2.5.3.3 PVI Dağılışının Momentleri, Modu ve Kümülatif Fonksiyonu	51

ÜÇÜNCÜ BÖLÜM

PEARSON DİFERANSİYEL DENKLEMİNİN ETKİ FONKSİYONU OLARAK KULLANILMASI

3.1. PI DAĞILIŞININ AMAÇ, ETKİ VE AĞIRLIK FONKSİYONU	56
3.2. PIV DAĞILIŞININ AMAÇ, ETKİ VE AĞIRLIK FONKSİYONU	58
3.3. PVI DAĞILIŞININ AMAÇ, ETKİ VE AĞIRLIK FONKSİYONU	60

DÖRDÜNCÜ BÖLÜM

UYGULAMA

4.1 UYGULAMADA KULLANILAN GERÇEK VERİ SETLERİ	68
4.1.1. Eğitim Harcamaları (Education Expenditure) Veri Seti	68
4.1.2 İktisadi Veri Seti	71
4.2. UYGULAMADA KULLANILAN SİMÜLASYON SENARYOLARI	73
 SONUÇ	 87
KAYNAKÇA	88
EKLER	

KISALTMALAR

ARE	Asimtotik Görelî Etkinlik
BP	Kırılma Noktası
EÇO	En Çok Olabilirlik
EKK	En Küçük Kareler
IC	Etki Eğrisi
IF	Etki Fonksiyonu
IRWLS	İteratif Olarak Yeniden Ağırlıklandırılan En Küçük Kareler
LTS	En Küçük Kesilmiş Kareler
LAD	En Küçük Mutlak Sapma
MAD	Medyan Mutlak Sapma
oyf	Olasılık Yoğunluk Fonksiyonu
TMS	Toplam Mutlak Sapma
OKH	Ortalama Karesel Hata
PI	Pearson Tip I Dağılışı
PIV	Pearson Tip IV Dağılışı
PIV	Pearson Tip IV Dağılışı
PVII	Pearson Tip VII Dağılışı

ŞEKİLLER LİSTESİ

Şekil 1: Regresyon Sapanları	s. 20
Şekil 2: Huber M- Tahminleyici Amaç, Etki ve Ağırlık Fonksiyonu	s.29
Şekil 3: Tukey's Bisquare M- Tahminleyici Amaç, Etki ve Ağırlık Fonksiyonu	s.31
Şekil 4: Pearson Kartı	s.39
Şekil 5: Craig Kartı	s.40
Şekil 6: Pearson Tip I Dağılışının Olasılık Yoğunluk Fonksiyonları	s.42
Şekil 7: Pearson Tip IV Dağılışının Olasılık Yoğunluk Fonksiyonları	s.45
Şekil 8: Sağa Çarpık Pearson Tip VI Dağılışının Olasılık Yoğunluk Fonksiyonları	s.48
Şekil 9: Sola Çarpık Pearson Tip VI Dağılışının Olasılık Yoğunluk Fonksiyonları	s.50
Şekil 10: Mod Değeri Orijinde olan Pearson Tip VI Dağılışının Olasılık Yoğunluk Fonksiyonları	s.52
Şekil 11: Pearson Dağılış Sistemi İlişkisi	s.54
Şekil 12: PI Dağılışına ait Amaç Fonksiyonu	s.57
Şekil 13: PI Dağılışına ait Etki Fonksiyonu	s.57
Şekil 14: PI Dağılışına ait Ağırlık Fonksiyonu	s.58
Şekil 15: PIV Dağılışına ait Amaç Fonksiyonu	s.59
Şekil 16: PIV Dağılışına ait Etki Fonksiyonu	s.59
Şekil 17: PIV Dağılışına ait Ağırlık Fonksiyonu	s.60
Şekil 18: PVI Dağılışına ait Amaç Fonksiyonu	s.61
Şekil 19: PVI Dağılışına ait Etki Fonksiyonu	s.62
Şekil 20: PVI Dağılışına ait Ağırlık Fonksiyonu	s.64
Şekil 21: Eğitim Harcamalar Veri Seti için Kutu Grafiği	s.69
Şekil 22: Eğitim Harcamalar Veri Seti için Histogram	s.70
Şekil 23: İktisadi Veri Seti için Kutu Grafiği	s.72
Şekil 24: İktisadi Veri Seti için Histogram	s.72

TABLÖLER LİSTESİ

Tablo 1: Pearson Dağılımlarının Δ , κ ve Köklerine Göre Durumları	s. 36
Tablo 2: Pearson Tip I Dağılımının Biçimleri	s. 42
Tablo 3: Pearson Tip VI Dağılımının Biçimleri	s. 53
Tablo 4: Eğitim Giderleri Verisi Regresyon Analizi Sonuçları	s. 70
Tablo 5: İktisadi Veriler için Regresyon Analizi Sonuçları	s. 73
Tablo 6: Simülasyonda Kullanılan Simetrik Dağılımların TMS ve OKH Sonuçları	s. 81
Tablo 7: Simülasyonda Kullanılan Asimetrik Dağılımların TMS ve OKH Sonuçları	s. 83

EKLER LİSTESİ

Ek 1: Değişmezlik (Invariant) Özelliği	ek s. 1
Ek 2: Medyan Mutlak Sapma (MAD)	ek s. 2
Ek 3: Düzgünlük (Regularity) Koşulları	ek s. 3
Ek 4: Slutsky's Teoremi	ek s. 4
Ek 5: Sandwich Varyans Tahminleyicisi	ek s. 5
Ek 6: Bir Olasılık Dağılışınnın Kuyruk Özellikleri	ek s. 6
Ek 7: M- Tahminleyicilerin Asimtotik Normalliği	ek s. 10
Ek 8: Etki Fonksiyonunun Beklenen Değeri	ek s. 11



GİRİŞ

İstatistik tarihi boyunca birçok araştırmacı, ellerindeki verilerin normal dağıldığını varsayarak analiz yapmışlardır. Bu nedenle verilerde bulunan anormallikler (kalın- uzun kuyruk, aşırı sivrilik, çarpıklık, basıklık, sapan gözlem vb.) araştırmacılar için ayak bağı olarak görülmüştür. Böyle durumların üstesinden gelmek yerine uzun yıllar boyunca bu yapı veya gözlemleri dışlamayı seçmişlerdir.

Veri setinde bulunan anormallikler birçok nedenden kaynaklanabilir. Bu nedenlerin başında ölçüm ve kayıt hataları ya da iki veya daha fazla anakütlenin birbirine karışması gelmektedir. Ancak veri setinde, veri setine ait bir gözlem veya gözlemler anormal gözlem gibi davranabilirler. Araştırmacılar böyle gözlemlerin varlığında veri setini analiz etmekte zorlanmışlar ve böyle durumların üstesinden gelebilmek için anormalliklere ‘dayanıklı’ analiz yöntemlerine yönelmişlerdir.

Dayanıklı (Robust) İstatistik, varsayılan modelden sapmaları ve bu sapmalar gerçekleştiğinde güvenilir ve yeteri kadar etkin istatistiksel prosedürlerin inşa edilmesiyle ilgilenmektedir. İlk defa ‘Robust (Dayanıklı)’ terimi Box (1953) tarafından kullanılmıştır. Ondan önce genellikle istatistikçiler anormalliğe sahip veriler için ‘kötü, kirli’ terimini kullanmışlardır.

Tukey (1960) varsayılan modelden çok küçük sapmaların bile optimal prosedürlerin etkinliğini hızla kaybetmesini sağladığını gözlemlemiştir. Bu nedenle özellikle Tukey (1962), veri analizinin geleceğinin yeni ve daha iyi alternatif prosedürler olması gerektiğini belirtmiştir. Böylece günümüzde kullanılan dayanıklı yöntemlerin temelleri atılmaya başlamış ve daha önce bu yöntemleri kullanmaya sıcak bakmayan istatistikçiler arasında yeni bir tartışma başlatmıştır.

Dayanıklı İstatistik, istatistikçiler arasında artık bir yöntem olarak görülse bile matematiksel bir altyapısı 1960’lı yılların başında henüz oluşmamıştı. Bu eksiklikten yola çıkan Huber (1964), dayanıklı istatistiği, modelin komşuluğunda verilerin dağılımını seçen Doğa ile belirli bir sınıfta bir tahminleyici seçen istatistikçi arasındaki bir oyun olarak görmüştür. Bu yaklaşımın adı Huber’in minimaks yöntemi olarak bilinmektedir. Günümüzde Huber’in minimaks yöntemi herhangi bir dayanıklı yöntem için temel yöntem olarak benimsenmiştir.

İlk defa von Mises (1947) tarafından ortaya atılan istatistiksel fonksiyonellerde süreklilik ve türevlenebilirlik problemleri, dayanıklı istatistiksel yöntemler teorisi oluşmaya başlarken Hampel (1968, 1974)’in yaklaşımıyla yeniden gündeme gelmişlerdir. Özellikle Etki

fonksiyonu (bir fonksiyonelin Gateaux türevi), istatistiksel bir prosedürün dengesini (stability) ölçmede en önemli araçlardan biri haline gelmiş ve yeni dayanıklı yöntemlerin geliştirilmesine olanak sağlamıştır. İstatistiksel fonksiyoneller daha sonra bootstrap ve parametrik olmayan yöntemlerin gelişmesinde önemli rol oynamışlardır.

Huber (1964), dayanıklı istatistiksel yöntemlerin gelişmesinde çok önemli bir yeri olan esnek ve geniş bir tahminleyici sınıfı olan M- tahminleyiciyi geliştirmiştir. Böylece Teorik ve Uygulamalı İstatistik, Ekonometri ve Biyoistatistik gibi alanlarda kullanılabilen bir yapı taşı ortaya çıkmıştır. Geliştirilen dayanıklı yöntemler beraberinde ‘ne kadar dayanıklı?’ sorusunu getirmişlerdir. Bu soruya cevap olarak Etki fonksiyonu (Hampel, 1968-1974) ve Kırılma Noktası verilebilir. Hampel (1968, 1974) tarafından sunulan Kırılma Noktası bir istatistiksel fonksiyonelin yeni eklenen veya değiştirilen bir veri noktasına (point mass) dayanıklılığını ölçmek için kullanılır. Kırılma noktaları dikkate alınarak birçok tahminleyici geliştirilmiştir. Bu tahminleyiciler genel olarak yüksek ve düşük kırılma noktalı tahminleyiciler olarak bilinmektedirler. Ancak düşük kırılma noktasına sahip tahminleyicilerin gelişimi dururken yüksek kırılma noktasına sahip tahminleyiciler için hesaplama ve yeniden örnekleme yöntemleri ilk olarak Rousseeuw (1984) ve Rousseeuw ve Leroy (1987) tarafından geliştirilmiştir.

Dayanıklı yöntemler günümüzde geniş bir uygulama alanına sahiptir. Özellikle Regresyon Analizinde parametre tahmini yapılırken verilerde bulunan aşırılıklar En Küçük Kareler (EKK) tahminleme yönteminin etkinliğini kaybetmesine yol açmaktadır. Böyle verilerin varlığında EKK ile yapılan parametre tahminleri sapmalı olmaktadır (Hampel, 1986). Bu nedenle dayanıklı tahminleme yöntemleri araştırmacıların ilgi odağı haline gelmiştir.

Harwey (1977), dayanıklı regresyon prosedüründe minimum mutlak sapma tahmin edicisini başlangıç çözümü olarak kullanmayı önermiştir. Hinich ve Talvar (1985) ve Andrews (1974) tarafından geliştirilen medyana dayanan M- tahminleyicileri de başlangıç çözümü olarak kullanmış ve etkinliklerini karşılaştırmıştır. Minimum mutlak sapma yönteminin birçok durumda diğer yöntemlere göre daha iyi olduğunu raporlamıştır.

Hogg (1979), veri setinde sapan gözlem olduğunda bu gözlemlerin etkilerini azaltmak için kullanılan robust istatistiksel prosedürleri ele almıştır. Regresyon parametrelerinin tahminlenme süreçlerini incelemiştir. Regresyon parametrelerinin tahminlenmesi için kullanılan yöntem olan IRWLS yöntemi üzerinde durmuş ve asimptotik varyans formülünü ele almıştır. Aynı zamanda L ve R tahminleme yöntemlerini incelemiştir. Andrews (1974) tarafından rapor edilen, Wood ve Gorman (1971) tarafından analiz edilen veri setini kullanmıştır. Buna ek olarak

Zeigler ve Ferris (1973) tarafından incelenen ‘Half-life of Plutonium- 241’, Lenth (1977) tarafından incelenen ‘Splines’ ve Agee ve Turner (1978) tarafından incelenen ‘Automated data reduction’ veri setlerini M- Tahminleyicilerle analiz etmiştir.

Wu (1985), ölçek ve regresyon parametreleri için sık kullanılan M- tahminleyicileri ele almıştır. Bell (1980) tarafından geliştirilen Bell/OLS M- Tahminleyiciler ile Siegel (1982) tarafından geliştirilen yüksek kırılma noktalı Bell/ RM M- tahminleyicileri birçok gerçek veri seti kullanarak karşılaştırmıştır. Tukey’s Bisquare M- tahminleyici ile Bell/OLS M- tahminleyiciler arasındaki benzerlikleri tartışmıştır.

Croux ve Rousseeuw (1992), S_n ve Q_n olmak üzere iki dayanıklı ölçek tahmini geliştirmişlerdir. Geliştirilen tahminleyiciler için kırılma noktaları ve hesaplama algoritmaları üzerinde durmuşlardır. Bu tahminleyicileri hesaplama sürelerine göre karşılaştırmışlardır. Aynı zamanda bu ölçek tahminlerinin regresyon parametre tahmini yaparken kullanmışlardır.

Cantoni ve Ronchetti (2004), çarpık ve kalın kuyruklu verilerde kullanılmak üzere yeni bir dayanıklı yöntem geliştirmişlerdir. Varsayılan modelden sapmalar olduğunda geliştirdikleri yöntemin geleneksel yöntemlere göre daha etkin olduğunu İsviçre’de bir hastanede 100 hastadan elde edilen ‘medical back problems’ verisi ve birçok simülasyon kullanarak kanıtlamışlardır.

Allende ve diğerleri (2006), robust istatistiksel yöntemlerde genel olarak kullanılan simetrik dağılımların eksik yönlerini tartışmış ve G_A^0 dağılımına dayanan asimetric etki fonksiyonlu M- tahminleme yöntemi önermişlerdir. Yeni geliştirilen yöntemi kullanarak uydudan (GPS) elde edilen görüntüleri işlemeye çalışmışlardır.

Mohebbi ve diğerleri (2007), EKK’ya alternatif olan dayanıklı regresyon yöntemleri üzerine inceleme yapmışlardır. Çarpık veri setleri için LAD, Huber ve Parametrik olmayan regresyon yöntemlerini kullanarak karşılaştırma yapmışlardır. Karşılaştırma amacıyla Laplace, Lojistik ve kirletilmiş (contaminated) normal dağılım kullanmışlardır. Çarpık veri seti için Gama dağılımından faydalanmışlardır. Karşılaştırma kriteri olarak Ortalama Karesel Hata ve Toplam Mutlak Sapma kullanmışlardır.

Muthukrishnan ve Radha (2010), veri setinde sapan gözlem olduğunda dayanıklı yöntemlerin kullanılmasını incelemişlerdir. Huber, Tukey ve Hampel M- tahminleme yöntemleri üzerine tartışmışlardır. Birçok simülasyon verisi üzerinde sapan gözlem varlığını ve sapan gözlem ekleyerek yöntemleri araştırmışlardır.

Chen (2013), robust regresyon analizinde parametre tahmini için sıklıkla kullanılan IRWLS tahminleme yöntemini, veri seti çok büyük olduğunda dağıtımli (kümeli) olarak kullanmayı önermiştir.

Rasheed ve diğerleri (2014), veri setinde sapan gözlem veya değişen varyans olduğu durumlarda regresyon parametrelerinin tahminlenmesinde IRWLS kullanıştır. Huber ve Tukey M- Tahminleyicileri için IRWLS ve LTS (En Küçük Kesilmiş Kareler) tahminleme yöntemlerini kullanarak dirençli (resistant) regresyon uygulamaları üzerine araştırma yapmışlardır. Farklı veri setleri kullanarak M- Tahminleyiciler, LTS ve OLS'yi karşılaştırmışlardır.

Khalil ve diğerleri (2016), yeniden azalan M- tahminleyici önermişlerdir. Bu yeni tahminleyiciyi Hampel, Andrews, Tukey ve Qadir M- tahminleyicilerle karşılaştırmıştır. Birçok simülasyon çalışmasına ek olarak 1950-1973 yılları arasında Belçika'dan yapılan uluslararası (Rousseeuw ve Leroy, 1987) telefon görüşmeleri veri setini kullanarak yöntemleri karşılaştırmışlardır.

Bosse ve diğerleri (2016), M- tahminleme yönteminin teorik altyapısını incelemiş ve Robotik çalışmalar için kullanmışlardır.

Sumarni ve diğerleri (2017), normal dağılışa göre daha uzun kuyruklu olan t dağılışını kullanarak, dağılışın yer (location) parametresini dayanıklı (robust) olarak incelemişlerdir. t -dağılışının Amaç, Etki ve Ağırlık fonksiyonlarını elde etmişlerdir. Elde edilen yer parametresi için asimptotik davranışı ve Asimptotik Görelî Etkinliği (ARE) elde etmişlerdir. Farklı serbestlik dereceleri kullanarak ARE'nin nasıl değiştiğini incelemişlerdir. Endonezya Soğan Verimi verisine sapan gözlem ekleyerek parametre tahmini yapmışlardır.

Yulita ve diğerleri (2018), Huber, Hampel, Tukey ve Welsch ağırlık fonksiyonlarını basit ve çok değişkenli regresyon analizi üzerinden karşılaştırmışlardır. Karşılaştırmada birçok simülasyon ve Hindistan Doğu Java Bölgesine ait Human Development Index (HDI) verilerini kullanmışlardır.

Bu tez çalışmasının amacı, verilerde çarpıklık, yüksek basıklık, kalın kuyruklu, uzun kuyruklu, sapan gözlem veya sapan gözlem olduğundan şüphelenilen gözlemler ortaya çıktığında geleneksel M- tahminleyicilerin eksikliklerini gidermektir. M- tahminleyiciler genellikle simetrik ve kalın- uzun kuyruklu veriler için iyi bir çözüm elde ederken anormallikler ortaya çıktığında etkinliklerini kaybetmektedirler. Bu nedenle bu çalışmada veri setinde meydana gelen asimetriyi, yüksek basıklığı ve kalın- uzun kuyruğu dikkate alan PVI dağılışının

ağırlık fonksiyonu kullanılarak regresyon parametre tahminleri yapılmıştır. Pearson Diferansiyel Denklemi ile Etki Fonksiyonu arasında bulunan benzerlikten yola çıkılarak Pearson PVI dağılışının $o(y)$ 'si kullanılarak Amaç, Etki ve Ağırlık Fonksiyonları elde edilmiştir.

Geleneksel M- tahminleme yöntemleri verilerin çarpıklık ve basıklık değerlerini dikkate almazken Pearson Diferansiyel Denklemi bu değerleri dikkate almaktadır. Böylece veri setine ait dağılış tam olarak (tek) belirlenebilmekte ve parametre tahmini yapılırken hatayı en aza indirmektedir.

İlk defa Karl Pearson (1895) tarafından ortaya çıkarılan Pearson Diferansiyel Denklemi, içerisinde bulundurduğu parametrelerin aldıkları farklı değerlere göre farklı olasılık dağılışları türeten bir sistemdir. Bu dağılışların ortaya çıkardığı sisteme Pearson Dağılış Ailesi denilmektedir. İçerisinde 13 farklı dağılış bulundurmaktadır. Bu 13 dağılıştan 3 tanesi Ana Tip geriye kalan 10 tanesi ise geçiş tipleri olarak isimlendirilmektedir. Pearson Dağılış Ailesinin Ana Tipleri;

- Pearson Tip I Dağılışı (PI)
- Pearson Tip IV Dağılışı (PIV)
- Pearson Tip VI Dağılışı (PVI)

olarak yazılabilir. PI dağılışı (Beta Dağılışı) hem sağ hem de soldan sınırlı bir dağılıştır. Bu nedenle kullanım alanı kısıtlıdır. PIV dağılışı ise kökleri sanal olan bir dağılış olmakla birlikte her iki uçtan sınırsızdır. Pearson Dağılış Ailesi içerisinde en zor olarak adlandırılan dağılıştır. PVI dağılışı (İkinci Tip Beta), bir uçtan (sağ veya sol) sınırlı kalın- uzun kuyruklu bir dağılıştır. Dağılışın parametrelerinin aldıkları değerlere göre içerisinde *F*, *Pareto*, *Beta* ve *Gama* Dağılışlarını barındırmaktadır. Ayrıca parametrelerinin yapısı nedeniyle birçok basıklık ve çarpıklık değerinde rahatlıkla kullanılabilir. Başlıca kullanım alanları aşağıda verilmiştir;

- Kayıp Fonksiyonu olarak (Belkama ve Embrechts, 2018),
- Tıpta beyin fonksiyonlarının incelenmesinde (Brascamp ve diğerleri, 2004),
- Salgın hastalıkların modellenmesinde (Tulupyeve ve diğerleri, 2013),
- Meteoroloji ve Hidrolojide (Mielke ve Johnson, 1974),
- Finansal oynaklık verilerinde (Moghaddam ve diğerleri, 2019),
- Gelir Modellemesinde (Ye ve diğerleri, 2012),
- Radardan alınan görüntülerin işlenmesinde (Salazar, 2000) ve
- Güvenilirlik analizinde (Kilany, 2016).

Bu tez çalışması dört bölümden oluşmaktadır. Tezin birinci bölümünde, Dayanıklı İstatistiksel Yöntemlerin dayanıklılıkları ölçmek amacıyla kullanılan Etki Fonksiyonu ve Kırılma noktası ele alınmıştır. Ayrıca genel olarak M- Tahminleyiciler ve regresyon analizinde sıklıkla kullanılan Huber ve Tukey M- Tahminleyiciler üzerinde durulmuştur.

İkinci bölümde, Pearson Diferansiyel denklemi ve bu denklemin Momentler Tahminleme yöntemiyle çözümleri yapılmıştır. Ayrıca PI, PIV ve PVI dağılımları incelenmiş ve PVI dağılımının sağa çarpık ve sola çarpık durumları ele alınmıştır. Buna ek olarak PVI'nın modu, momentleri ve kümülatif dağılım fonksiyonu incelenmiştir.

Üçüncü bölümde, PI, PIV ve PVI dağılımlarının oyfleri kullanılarak, Pearson Diferansiyel Denklemi ile Etki fonksiyonu arasındaki benzerlikten yola çıkılarak mod noktası orijinde olan Pearson ana tipleri için Amaç, Etki ve Ağırlık Fonksiyonları elde edilmiştir. Ayrıca bu tez çalışmasının konusu olan, PVI için Asimtotik Varyans teorik olarak elde edilmiştir.

Dördüncü ve son bölümde, gerçek veri setleri ve simüle edilmiş veri setleri kullanılarak EKK, Huber M- Tahminleme Yöntemi, Tukey M- Tahminleme Yöntemi ve PVI Dağılımının Ağırlık Fonksiyonuna Dayanan Tahminleme Yöntemi Toplam Mutlak Sapma (TMS) ve Ortalama Karesel Hata (OKH) kriterleri açısından karşılaştırılmıştır. Simülasyon çalışması, hata teriminin dağılımının 28 farklı senaryo ve 9 farklı dağılım gösterdiği durumlar için yapılmıştır. Her bir durum için regresyon artıklarının PVI'ya uyum gösterip göstermedikleri katı şartlar altında incelenmiştir. Çalışmada örnek hacimleri 30,50, 100 ve 500 olarak belirlenmiş ve her örnek hacmi için 1000 adet tekrar gerçekleştirilmiştir.

BİRİNCİ BÖLÜM

ROBUST (DAYANIKLI) İSTATİSTİKSEL YÖNTEMLER

Robust (Dayanıklı) terimi ilk defa Box (1953) tarafından kullanılmıştır. O zamana kadar bazı dayanıklı yöntemler nadiren kullanılmışlardır. Ancak Tukey (1960), varsayımlarda bazı küçük sapmaların istatistiklere aşırı etki yaptığını fark eden ve örneklendiren ilk kişi olmuştur. Geleneksel olarak yaygın kullanılan normal modellerin, küçük değişimlere çok duyarlı ve varsayım olarak tehlikeli bir hal aldığını gözlemlemiştir. Huber (1964) ve Hampel (1968) bu konu üzerinde çalışma yapmışlar ve dikkatleri üzerlerine çekmişlerdir.

Çoğu araştırmacıya göre 1960'larda dayanıklı yöntemler daha çok "kötü" olarak görülmekteydi. Bu nedenle öncelikle matematiksel yaklaşımların yapılması gerekmektedir.

Robust istatistik, klasik istatistikte yaygın olarak yapılan birçok varsayımın (normallik, doğrusallık, bağımsızlık vb.) gerçeğe en uygun yaklaşımıdır. Dayanıklı yöntemlerin kullanılmasının nedeni büyük hata (gross error), kopyalama veya aktarma hatalarının veri seti içerisindeki varlıklardır. Genellikle veri setinde sapan (outlier) olarak ortaya çıkar ve birçok geleneksel istatistiksel yöntem için tehlikelidir (Hampel ve diğerleri, 1986:6).

"Robust" kelimesi, bazen tutarsız çağrışımlarla doludur. Göreceli olarak dar bir anlamda, amaçlarımız için robust varsayımların küçük sapmalara duyarsızlıkları olarak ifade edilebilir (Huber, 1981).

1.1. DAYANIKLILIK ÖLÇME YÖNTEMLERİ

Dayanıklı (Robust) istatistiksel yöntemlerin ortaya çıkmasıyla beraber bu yöntemlerin 'Ne kadar dayanıklı?' sorusu gündeme gelmiştir. Bu nedenle yöntemlerin dayanıklılıkları farklı şekillerde ölçülmeye çalışılmıştır. Bu yöntemlerin başında Etki Fonksiyonu (Influence Function) ve Kırılma Noktası (Breakdown Point) gelmektedir.

1.1.1. Etki Fonksiyonu

Bir x noktasındaki sonsuz küçüklüğün (infinitesimal) tahminleyici üzerindeki etkisini ölçmek için kullanılan en etkili yöntemdir. Sonsuz küçüklüğün bir resmidir ve gözlemlerdeki bozulmanın neden olduğu asimptotik sapmayı ölçer. Etki fonksiyonu (IF), orijinal olarak Etki

Eğrisi (IC) olarak adlandırılmış ancak literatürde genellikle IF olarak kullanılmıştır (Hampel 1968, 1974).

Tanım: R , reel sayı uzayı, T ise R 'nin tüm olasılık ölçülerinin bazı alt kümelerinde tanımlı reel değerli bir fonksiyonel ve F , T 'nin tanımlı olduğu R üzerinde bir olasılık ölçüsü olsun. $x \in R$ olmak üzere nokta ağırlığı (point mass) 1 olan bir olasılık ölçüsü δ_x tanımlansın. F ve δ_x 'in karışımı, $0 < t < 1$ olmak şartıyla, $(1-t)F + t\delta_x$ şeklinde yazılabilir. $x \in R$ 'nin her bir noktasında limit tanımlı olmak koşuluyla, tahminleyici T 'nin F dağılışı altındaki Etki Eğrisi (IC);

$$IC_{T,F}(x) = \lim_{t \rightarrow 0} \left\{ T[(1-t)F + t\delta_x] - T(F) \right\} / t \quad (1.1)$$

şeklinde gösterilebilir.

T 'nin tanım aralığı, örnek uzayı (χ) üzerindeki tüm sonlu kümelerin ölçüsünün konveks bir alt kümesi olsun. Tanım aralığındaki bütün G dağılışları için gerçek bir a_1 fonksiyonu ortaya çıkıyorsa, T tahminleyicisi, F dağılışı altında Gateaux anlamında türevlenebilirdir.

$$\lim_{t \rightarrow 0} \frac{T((1-t)F + tG) - T(F)}{t} = \int a_1(x) dG(x)$$

Bu gösterim aynı zamanda;

$$\frac{\partial}{\partial t} [T((1-t)F + tG)]_{t=0} = \int a_1(x) dG(x) \quad (1.2)$$

şeklinde ifade edilebilir (Hampel, 1974: 383). Eşitliğin sol tarafı G yönünde T 'nin F 'deki yönlü türevidir. $G = F$ olursa;

$$\int a_1(x) dF(x) = 0 \quad (1.3)$$

yazılabilir ve $dG(x)$ yerine $d(G - F)(x)$ yazılabilir. Eğer F yakınında bir G dağılışı alınır ve T 'ye F 'de birinci derece von-Mises açılımı (Taylor açılımı) uygulanırsa;

$$T(G) = T(F) + \int IF(x; T, F) d(G - F)(x) + R_i, i = 2, \dots \quad (1.4)$$

elde edilir (von-Mises, 1947: 325).

Asimptotik Varyans ile Etki Fonksiyonu arasında yakın bir ilişki vardır. Eğer gözlemler bağımsız ve aynı F dağılışına sahip olursa ($X_i \stackrel{i.i.d}{\sim} F$), gözlem sayısı n büyük olduğunda ampirik dağılışı F_n , F 'e yaklaşır. Bu nedenle G yerine F_n yazılabilir.

$T_n(X_1, \dots, X_n) = T_n(F_n)$ örnek hacmi büyük olduğunda $T(F_n)$ 'e yeteri kadar yaklaşabilir.

$$\int a_1(x) dF(x) = \int IF(x; T, F) dF(x) = 0$$

alınırsa;

$$T_n(F_n) \approx T(F) + \int IF(x; T, F) dF_n(x) + R_n \quad (1.5)$$

yazılabilir. Eşitlik 1.5 F_n üzerinden integral alınır ve tekrar yazılırsa;

$$\sqrt{n}(T_n - T(F)) \approx \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(X_i; T, F) + R_n \quad (1.6)$$

Merkezi limit teoremine göre Eşitlik 1.6'nın solundaki terim asimptotik normal dağılışa sahiptir.

$n \rightarrow \infty$ iken R_n 0'a yaklaşmaktadır. Bu durumda $\sqrt{n}(T_n - T(F)) \sim N(0, V(T, F))$ olarak gösterilebilir. Burada asimptotik varyans;

$$V(T, F) = \int IF(x; T, F)^2 dF(x) \quad (1.7)$$

olarak yazılır (Hampel ve diğerleri, 1986).

1.1.2. Kırılma Noktası

Bir tahminleyicinin kırılma noktası (breakdown point), gözlemlerdeki en küçük değişimin tahminleyicinin değerinde meydana getirdiği küçük veya büyük değişimdir. Kırılma noktası, bozulmaya karşı direncin bir ölçüsüdür (Wilcox, 2011; Rousseeuw ve Leroy, 1987: 9). İlk defa Hodges (1967) tarafından ortaya atılsa bile Hampel'in doktora tezinde genelleştirilmiştir.

Kırılma noktası küçük örneklem için çok kullanışlı bir dayanıklılık ölçüsüdür ve bir olasılık modeli olmadan tanımlanabilir. Olasılıksal olmayan kırılma noktası için 3 farklı örneklem tanımlanması gerekir.

$X = (x_1, \dots, x_n)$, n örnek hacimli bir örneklem olsun. Bu örneklem 3 farklı yolla bozulursa;

1. **ε - bozulumu:** m adet keyfi $Y = (y_1, \dots, y_m)$ değeri orijinal örneğe eklenebilir. Böylece 'kötü veri' oranı $X' = X \cup Y$ için $\varepsilon = m/(m+n)$ 'dir.
2. **ε - yerine koyma:** m adet veri $y_1, \dots, y_m$ örnek içerisinde keyfi olarak değiştirilir. Böylece 'kötü veri' oranı $\varepsilon = m/n$ olur.
3. **ε - değişiklik:** Ampirik ölçüm uzayı içerisinde bir uzaklık fonksiyonu π olsun. F_n , X için bir olasılık ölçümü ve G_n , X' için bir olasılık ölçümü ve $\pi(F_n, G_n) \leq \varepsilon$ olsun. Bu koşullar dikkate alındığında, T , değerleri Öklid uzayında olan bir

tahminleyici ve $T(X)$, X örneğinde tahminleyicinin değeri olsun. X , örnek uzayında tahminleyici T 'nin kırılma noktası ε^* olur. Burada ε^* , örneklem X' bozulmuş/ yerine konmuş/ değişiklik yapılmış olduğunda ε 'un alacağı en küçük değerdir. Bu nedenle bozulmuş/ yerine konmuş/ değişiklik yapılmış örnekteki maksimum Sapma;

$$Sapma(\varepsilon; X, T) = \sup(T(X) - T(X')) \quad (1.8)$$

olur. Eşitlik 1.8'de supremum bozulmuş/ yerine konmuş/ değişiklik yapılmış bütün örneklemelerin kümesi üzerinden alınmıştır. Sapmadan hareketle kırılma noktası;

$$\varepsilon^*(X, T) = \inf \{ \varepsilon : Sapma(\varepsilon; X, T) = \infty \} \quad (1.9)$$

olur. Bozulmuş/ yerine konmuş/ değişiklik yapılmış örneklemeler için tahminleyicinin değerindeki küçük değişimler önemsizken eğer tahminleyicinin değerindeki değişim sonsuza doğru giderse kırılma noktası bu değerler içerisindeki en küçük sapma değeridir (Donoho ve Huber, 1983).

1.2 M- TAHMİNLEYİCİLER

Geleneksel istatistiksel yöntemler belirli varsayımlara ve kalıplaşmış çözüm yöntemlerine dayanmaktadır. Ancak bu yöntemler gerçek yaşamda ortaya çıkan sorunlara her zaman çözüm olamamışlardır. Dayanıklı yöntemlerin gelişmesi ile birlikte gerçek yaşamda ortaya çıkan problemlerin çözümlenmeleri kolaylaşmış ve yöntemler esnekleşmiştir. Özellikle gerçek yaşamda ortaya çıkan bir problemde dayanıklı yöntemler gerçek modelin yakınına daha hızlı ve daha güvenilir yaklaşmaktadırlar. Böylece dayanıklı istatistiksel yöntemlerin gelişmesiyle birlikte dayanıklı tahminleme yöntemlerine ihtiyaç duyulmuş ve bazı yöntemler geliştirilmiştir. Bu yöntemlerin başında EÇO tahminleme yönteminin genelleştirilmiş bir versiyonu olan M- Tahminleyiciler gelmektedir. M- Tahminleyiciler en esnek, uygulaması (hesaplaması) en kolay ve çok parametrelili yöntemlere çok hızlı ve güvenilir olarak uygulanabilir.

Herhangi bir tahminleyici T_n bir minimizasyon problemi olarak tanımlanırsa Amaç Fonksiyonu;

$$\sum \rho(x_i; T_n) = \min! \quad (1.10)$$

ve minimizasyon probleminin türevi alınır Etki Fonksiyonu;

$$\sum \psi(x_i; T_n) = 0 \quad (1.11)$$

olarak gösterilebilir. Burada $\rho(x_i; \theta)$ her ne kadar keyfi bir fonksiyon olsa bile türevlenebilir bir fonksiyon olması uygun bir seçim olur ve $\psi(x_i; \theta) = \left(\frac{\partial}{\partial \theta}\right) \rho(x_i; \theta)$ M- Tahminleyici (Etki Fonksiyonu) olarak adlandırılır. M- Tahminleyiciler aynı zamanda En Çok Olabilirlik (EÇO) tipi (Maximum Likelihood Estimator, MLE) tahminleyiciler olarak bilinir ve $\rho(\mathbf{x}; \theta) = -\log f(\mathbf{x}; \theta)$ olursa En çok olabilirlik tahminleyicisi olur. Özellikle bir yer ölçüsünün tahmini ile ilgilenilirse;

$$\sum \rho(x_i - T_n) = \min! \quad (1.12)$$

veya türevi;

$$\sum \psi(x_i - T_n) = 0 \quad (1.13)$$

yazılabilir. Eşitlik 1.13'te bir ağırlık fonksiyonu $w_i = \psi(x_i - T_n) / (x_i - T_n)$ olarak tanımlanırsa;

$$\sum w_i (x_i - T_n) = 0 \quad (1.14)$$

olur. Böylece bir yer ölçüsünü ifade eden tahminleyici T_n bir ağırlıklı ortalama olarak yazılabilir;

$$T_n = \frac{\sum w_i x_i}{\sum w_i}. \quad (1.15)$$

Burada ağırlıklar örnekten elde edilmektedir. Eşitlik 1.13 fonksiyonel olarak tanımlanırsa;

$$\int \psi(x; T(F)) dF(x) = 0 \quad (1.16)$$

ile gösterilebilir (Huber, 2011).

1.2.1. Konum Parametresinin M-Tahminleyicisi

M- Konum (location) Tahminleyiciler dikkate alınırken olasılık temelli istatistikler kullanılarak konum parametresinin M-Tahminleyicisi elde edilecektir. Her bir gözlem x_i 'nin değeri bilinmeyen gerçek parametre θ 'nın değerine ve bir şans değişkeni ε_i 'ye (hata terimi) bağlı olsun. Burada hataların eklenebilir olduğu göz önüne alınarak konum modeli;

$$x_i = \theta + \varepsilon_i, \quad (i = 1, \dots, n) \quad (1.17)$$

ile gösterilebilir. Eşitlik 1.17’da hata terimleri $\varepsilon_1, \dots, \varepsilon_n$ şans değişkenleridir. Eğer gözlemler eşit koşullar altında birbirinden bağımsız tekrarlardan elde edildilerse hata terimi için aşağıdaki varsayımlar yapılabilir.

1. $\varepsilon_1, \dots, \varepsilon_n$ birbirleriyle aynı F_0 dağılışına sahiptir ve
2. $\varepsilon_1, \dots, \varepsilon_n$ bağımsızdır.

Böylece gözlemler $x_1, \dots, x_n$ bağımsız ve aynı dağılım fonksiyonuna sahip şans değişkenleridir.

$$F(x) = F_0(x - \theta) \quad (1.18)$$

Eğer sistematik bir hata olmadığı varsayılırsa ε_i ve $-\varepsilon_i$ aynı dağılışa sahip olduğundan $F_0(x) = 1 - F_0(-x)$ olduğu rahatlıkla görülebilir. Konum parametresinin tahminlenmesi için Eşitlik 1.17 dikkate alınırsa, F_0 , ε_i ’nin dağılım fonksiyonu ve $f_0 = F_0'$ olasılık yoğunluk fonksiyonu (oyf) olsun. Gözlemlerin ortak olasılık fonksiyonu (olabilirlik fonksiyonu);

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_0(x_i - \theta) \quad (1.19)$$

olarak yazılabilir. Buradan θ ’nın EÇÖ tahminleyicisi $\hat{\theta}$, olabilirlik fonksiyonunun maksimize edilmesiyle;

$$\hat{\theta} = \hat{\theta}(x_1, \dots, x_n) = \arg \max_{\theta} L(x_1, \dots, x_n; \theta) \quad (1.20)$$

elde edilir. Burada $\arg \max$ terimi değerin maksimize edilmesi anlamında kullanılmıştır. Eğer F_0 kesin olarak biliniyorsa, EÇÖ tahminleme yöntemi, uygun tahmin sınıfları arasında en küçük varyansa sahip optimum tahminleyici sınıfıdır. Ancak F_0 kesin olarak bilinmiyor ancak yaklaşık olarak biliniyorsa tahminleri bulmaktaki amaç;

- a. F_0 Normal ise ‘yaklaşık optimumu’ elde etmek veya
- b. F_0 yaklaşık olarak Normal ise (kirletilmiş normal vs.) ‘yaklaşık optimumu’ elde etmektir.

Eğer $f_0 > 0$ ise logaritması artan bir fonksiyon olacaktır. Bu durumda Eşitlik 1.20;

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n \rho(x_i - \theta) \quad (1.21)$$

olarak yazılabilir. Burada $\rho = -\log f_0$ ’dır ve Amaç Fonksiyonu olarak bilinir. Eğer $F_0 = N(0,1)$ ise;

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (1.22)$$

olur ve sabitler atılırsa $\rho(x) = x^2/2$ olarak yazılabilir. Böylece Eşitlik 1.21;

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n (x_i - \theta)^2 \quad (1.23)$$

olarak yazılabilir. Eğer F_0 çift üstel dağılışı (double exponential) olursa;

$$f_0(x) = \frac{1}{2} e^{-|x|} \quad (1.24)$$

elde edilir ve $\rho(x) = |x|$ olur. Böylece Eşitlik 1.21;

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n |x_i - \theta| \quad (1.25)$$

ile gösterilir. Eşitlik 1.23 ve 1.25 için çözüm yapılırsa örnek ortalaması ve örnek medyanı elde edilir.

Eğer Eşitlik 1.21 θ 'ya göre türevlenirse;

$$\sum_{i=1}^n \psi(x_i - \hat{\theta}) = 0 \quad (1.26)$$

olur. Burada $\psi = \rho'$ 'dır ve Etki Fonksiyonu olarak bilinir. Ancak ψ süreksiz bir fonksiyon ise Eşitlik 1.26'in çözümü bulunamaz. Böyle bir durumda Eşitlik 1.26 yorumlanırken eşitliğin sol tarafı θ 'da işaret değiştirir anlamında yorumlanır. Eğer f_0 simetrik ise ρ çift ve ψ tek fonksiyondur.

Amaç fonksiyonu $\rho(x) = x^2/2$ alınırsa, $\psi(x) = x$ olur ve Eşitlik 1.26;

$$\sum_{i=1}^n (x_i - \hat{\theta}) = 0 \quad (1.27)$$

elde edilir ve bu çözüm örnek ortalaması $\hat{\theta} = \bar{x}$ olarak gösterilir.

Verilen bir ρ fonksiyonu için, Eşitlik 1.21'nin çözümü konum parametresinin M-Tahminleyicisidir. Amaç fonksiyonunun uygun bir şekilde seçilmesi için a ve b amaçları dikkate alınabilir.

Eğer Ağırlık Fonksiyonu;

$$W(x) = \begin{cases} \frac{\psi(x)}{x}, & x \neq 0 \\ \psi'(0), & x = 0 \end{cases} \quad (1.28)$$

alınırsa Eşitlik 1.26;

$$\sum_{i=1}^n (x_i - \hat{\theta}) W(x_i - \hat{\theta}) \quad (1.29)$$

olarak yazılabilir veya;

$$\hat{\theta} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \quad w_i = W(x_i - \hat{\theta}) \quad (1.30)$$

ağırlıklı ortalama olarak yazılabilir. Burada $W(x)$, $|x|$ için artmayan bir fonksiyondur ve ağırlıklar $\hat{\theta}$ 'e bağlıdır. Böylece uzakta bulunan gözlemlere daha küçük ağırlık verilir (Maronna ve diğerleri, 2006; Huber, 1981).

1.2.2. Ölçek Parametresinin M-Tahmini

Ölçek parametresinin M- Tahmininin tek başına bir önemi olmasa bile, regresyon parametrelerinin tahminlenmesi ve çok değişkenli analizlerde yardımcı bir rol oynamaktadır. Buradan yola çıkarak çarpım modeli;

$$x_i = \sigma \varepsilon_i \quad (1.31)$$

ele alınsın. Hata terimleri bağımsız aynı f_0 dağılışına sahip ve $\sigma > 0$ bilinmeyen ölçek parametresi olsun. Gözlemlerin dağılışı bir üstel aile olarak alınırsa dağılışı;

$$\frac{1}{\sigma} f_0\left(\frac{x}{\sigma}\right) \quad (1.32)$$

ile gösterilir (Maronna ve diğerleri, 2006). Ölçek parametresi σ 'nın EÇO tahmin edicisi;

$$\hat{\sigma} = \arg \max_{\sigma} \frac{1}{\sigma^n} \prod_{i=1}^n f_0\left(\frac{x_i}{\sigma}\right) \quad (1.33)$$

olur. Eşitlik 1.33'nin logaritması alınır ve σ 'ya göre türevlenirse;

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{\hat{\sigma}}\right) = 1 \quad (1.34)$$

elde edilir. Burada $\rho(x) = x\psi(x)$ ve $\psi(x) = f_0'(x)/f_0(x)$ 'dir. Eğer f_0 'ın dağılışı $N(0,1)$

ise $\rho(x) = x^2$ ve $\hat{\sigma} = \sqrt{\sum_{i=1}^n x_i^2 / n}$ (Ortalama Karesel Hatanın Karekökü, RMS) olur, ancak

eğer f_0 çift üstel dağılışı ise $\rho(x) = |x|$ ve $\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n |x_i|$ olur. Eğer f_0 çift fonksiyon ise ρ fonksiyonunda çift fonksiyondur ve $\hat{\sigma}$ 'in değeri sadece gözlemlerin mutlak değerine bağlıdır. Ölçek parametresinin M- tahminleyicisi;

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{\hat{\sigma}}\right) = \chi \quad (1.35)$$

şeklinde gösterilebilir. Amaç fonksiyonu kullanılarak elde edilebilir ve burada $\chi > 0$ olmalıdır. Ölçek parametresinin tahminlenmesinde genellikle orijin etrafında ikinci dereceden bir ρ fonksiyonu kullanılır ($\rho(0) = 0$ ve $\rho''(0) > 0$). Böyle durumlarda M- ölçek tahmini bir ağırlıklı ortalama karesel hatanın karekökü şeklinde yazılabilir. Ağırlık fonksiyonu;

$$W(x) = \begin{cases} \frac{\rho(x)}{x^2}, & x \neq 0 \\ \rho''(0), & x = 0 \end{cases} \quad (1.36)$$

olarak alınırsa Eşitlik 1.35;

$$\hat{\sigma} = \frac{1}{n\chi} \sum_{i=1}^n W\left(\frac{x_i}{\hat{\sigma}}\right) x_i^2 \quad (1.37)$$

şeklinde yeniden düzenlenebilir (Maronna ve diğerleri, 2006; Huber, 1981).

1.2.3. Konum ve Ölçek Parametrelerinin M- Tahminleyicileri

Konum ve ölçek parametresi birlikte ele alınarak M- tahminleri elde edilebilir. Bu nedenle konum- ölçek modeli;

$$x_i = \theta + \sigma \varepsilon_i \quad (1.38)$$

ile gösterilebilir. Burada ε_i , f_0 dağılışına sahip ve x_i 'nin dağılışı;

$$f(x) = \frac{1}{\sigma} f_0\left(\frac{x - \theta}{\sigma}\right) \quad (1.39)$$

olur. Eşitlik 1.39'da σ parametresi $\sigma \varepsilon_i$ şans değişkeni için ölçek parametresi ancak x_i şans değişkeni için yayılım parametresi olur. Eşitlik 1.39'daki konum (θ) ve ölçek (σ) parametresi için EÇÖ tahmini;

$$(\hat{\theta}, \hat{\sigma}) = \arg \max_{\theta, \sigma} \frac{1}{\sigma^n} \prod_{i=1}^n f_0\left(\frac{x_i - \theta}{\sigma}\right) \quad (1.40)$$

veya

$$(\hat{\theta}, \hat{\sigma}) = \arg \min_{\theta, \sigma} \left\{ \frac{1}{n} \sum_{i=1}^n \rho_0 \left(\frac{x_i - \theta}{\sigma} \right) + \log \sigma \right\} \quad (1.41)$$

yazılabilir. Eşitlik 1.26 ve 1.35 dikkate alınırsa yer ve ölçek parametrelerinin EÇO tahminleyicileri;

$$\sum_{i=1}^n \psi \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) = 0 \quad (1.42)$$

$$\frac{1}{n} \sum_{i=1}^n \rho_{ölçek} \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) = \chi \quad (1.43)$$

denklem sistemlerinin çözümünden elde edilir. Burada $\psi(x) = -\rho'_0$, $\rho_{ölçek}(x) = x\psi(x)$ ve $\chi = 1$ 'dir. Eğer F dağılışı simetrik ve $\theta \sim N(\theta, v/n)$ ise Eşitlik 1.42 ve 1.43;

$$E \left[\psi \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) \right] = 0 \quad (1.44)$$

$$E \left[\rho_{ölçek} \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) \right] = \chi \quad (1.45)$$

olur ve $v = \sigma^2 \frac{E \left[\psi((x - \theta)/\sigma)^2 \right]}{[E(\psi'((x - \theta)/\sigma))]^2}$ olarak gösterilir (Maronna ve diğerleri, 2006; Huber, 1981:1967).

1.2.4. M- Tahminleyicilerin Etki Fonksiyonu

Eşitlik 1.16'da eğer $F_t = (1-t)F + t\delta_x$ ve $T_t = T(F_t)$ yazılırsa $E_{F_t}[\psi(X; T_t)] = 0$ için;

$$(1-t)E_F[\psi(X; T_t)] + t\psi(x; T_t) = 0 \quad (1.46)$$

elde edilir (bkz. EK 8). Buradan t 'ye göre türev alınırsa;

$$-E_F[\psi(X; T_t)] + (1-t)\frac{\delta}{\delta t}E_F[\psi(X; T_t)] + \psi(x; T_t) + t\frac{\delta}{\delta t}\psi(x; T_t) = 0 \quad (1.47)$$

olur. Burada $\psi'(x, \theta) = \frac{\partial \psi(x, \theta)}{\partial \theta}$ alınırsa zincir kuralı kullanılarak;

$$E[\psi(X; T_t)] + (1-t)E_F[\psi'(X; T_t)] \frac{\delta T_t}{\delta t} + \psi(x; T_t) + t\psi'(x; T_t) \frac{\delta T_t}{\delta t} = 0 \quad (1.48)$$

elde edilir. Eğer örnek hacmi yeterince büyük olursa $T(F_t) = T(F)$ olur ve $t = 0$ civarında;

$$\begin{aligned} -E_F[\psi'(X; T_t)]IF(x, T(F)) + \psi(x; T_t) &= 0 \\ IF(x, T(F)) &= \frac{\psi(x; T_t)}{-E_F[\psi'(X; T_t)]} \end{aligned} \quad (1.49)$$

elde edilir. Eşitlik 1.49'a bakılacak olunursa M- Tahminleyicilerin Etki Fonksiyonları $\psi(\cdot)$ fonksiyonu ile oransaldır. Dolayısıyla $\psi(\cdot)$ fonksiyonu Etki Fonksiyonu olarak kullanılabilir (Huber, 2011). Ayrıca M- Tahminleyicilerin asimtotik dağılışı Ek 7'de verilmiştir.

1.3. DAYANIKLI REGRESYON

İstatistiksel analiz yöntemlerinin arasında Regresyon Analizi önemli bir yer tutmakta ve neredeyse istatistik tarihi kadar eskiye dayanmaktadır. Özellikle 1960'lı yılların başına kadar geleneksel analiz yöntemleri yaygın olarak kullanılmıştır. Ancak, dayanıklı yöntemlerin gelişmesi diğer tüm istatistiksel yöntemlerde olduğu gibi regresyon analizi içinde önemli gelişmelere sebep olmuştur. Özellikle verilerde sapan gözlem ortaya çıktığında dayanıklı yöntemlerin kullanılması kaçınılmaz hale gelmiştir.

Veri seti içerisinde bulunan bir gözlemi sapan olarak tanımlamak araştırmacılar için her zaman çok zor olmuştur. Bu nedenle tam olarak sapan tanımı yapmak zordur. Barnett ve Lewis (1979:4)'te sapan gözlem için "veri setinin geri kalanı için tutarsız gözlemler" olarak belirtmişlerdir. Judge ve diğerleri (1988:889), regresyon artıklarındaki büyük değerlere sapan gözlem adını vermişlerdir. Davies ve Gather (1993)'te sapan gözlem için teorik bir tanım geliştirmişlerdir. Bu tanıma göre F içerisinde sapan gözlem bulunmayan ortalaması μ ve standart sapması σ olan bir dağılış olmak üzere $0 < \alpha < 1$ aralığı için;

$$sapan(\alpha, \mu, \sigma) = \{x; |x - \mu| > z_{1-\alpha/2}\sigma\} \quad (1.50)$$

olursa x_i gözlemi F dağılışı için α - sapanıdır. Ancak bu durumda F dağılışına ait bir gözlem doğal olarak bir α - sapan olarak tanımlanabilir. Burada x_i herhangi bir reel değerdir ve bir gözleme denk gelmek zorunda da değildir. Eşitlik 1.50 için önemli bir özellik ise μ ve σ 'nın

nokta tahminleyicileri değil gerçek değerleri kullanılmıştır. Bu nedenle pratikte sapan gözlemler için bir sınır bölgesi kullanılamaz. Bu yöntemin avantajı, tahminlenmiş parametreler yerine gerçek parametre değerlerinin kullanılması, sapanları tanımlamak için sapanları belirlemede kullanılan yöntemlerden bağımsız olmasıdır. Örneğin μ ve σ yerine $\hat{\mu}$ ve $\hat{\sigma}$ 'in kullanılırsa;

$$x_i \in \{x_i; |x - \hat{\mu}| > z_{1-\alpha/2} \hat{\sigma}\} \quad (1.51)$$

olur. Ancak araştırmacı μ ve σ yerine medyan ve MAD kullanırsa farklı gözlemleri sapan olarak belirleyebilir. Gerçek parametrelerin kullanılması “gerçek sapan” ile diğer yöntemlerden elde edilen sapanları birbirinden ayırır.

Diğer bir sapan gözlem tanımı ise, sapan gözlemlerin sapan gözlem türeten bir kirletilmiş (contaminated) bir dağılıştan gelmesidir. Gözlemlerin bir kısmı p olasılığı ile F dağılışı ve $1-p$ olasılığı ile G dağılışıdan geliyor olsun. Burada G dağılışı kirletilmiş $G_1, G_2, \dots, G_k$ dağılışların bir karışımıdır. Böylece eğer gözlemler F dağılışıdan geliyor ise düzgün, G dağılışıdan geliyor ise sapan (kirli) gözlemler olarak adlandırılır. Bu durumda sapan gözlemlerin belirlenmesi kirli gözlemlerin belirlenmesiyle aynı anlama gelmektedir. Ancak F dağılışıdan gelen gözlemlerin bazıları F 'e göre α - sapan olabilirler. Bir örnek üzerinden açıklırsak, F standart normal dağılışı ve G standart Cauchy dağılışı olsun. Burada F dağılışıdan seçilen bir gözlemin değeri çok küçük bir olasılıkla ($\alpha = 0.01$) 5 olsun. Bu gözlem F dağılışı için bir α - sapandır. Buna ek olarak G dağılışıdan seçilen bir gözlem ise $[-1, 1]$ aralığında olursa F dağılışı için bir sapan gözlem olmaz.

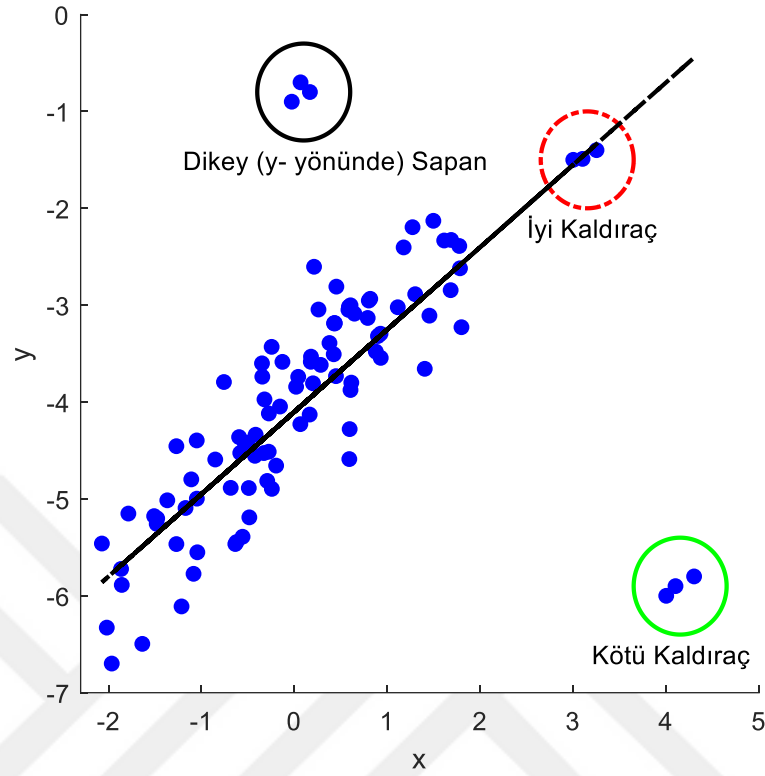
Hampel ve diğerleri (1986) ve Krasker ve diğerleri (1983)'e göre sapan gözlemler büyük hata (gross error) ve model hatası olarak iki gruba ayrılırlar.

Büyük hatalar, kayıt, yazım, ölçüm ekipmanının yetersizliğinden kaynaklanan, birim değişiminden kaynaklanan veya yanlış yorumlamadan kaynaklanan hatalardır. Büyük hatalar genellikle araştırmacı tarafından veri setinden silinirler. Ancak bu hataların belirlenmesi çok zordur. Hampel (1986:25)'e göre bu hataların ortaya çıkması bir istisnai durum değil normal bir süreçtir. Veri setinde çok az sayıda bile büyük hata olsa geleneksel EKK tahminleyicilerinde büyük değişime yol açarlar. Böyle durumların varlığında ise dayanıklı (robust) istatistiksel yöntemlerin kullanılması büyük önem taşır.

Model hatası ise, istatistiksel/ekonometrik modelin yapısından kaynaklanabilir. Böyle durumlarda ekonometrik model kurulurken genellikle kukla değişken kullanılır. Kukla değişken, tek bir veri noktası için 1'e eşit olur ve veri dışlamayı kolaylaştırır. Böylece model

kurulurken varsayılan model bütün gözlemleri tanımlamaz. Bundan daha iyi bir durum ise gözlemlerin çoğunu veya hepsini dikkate alan dayanıklı yöntemlerin kullanılmasıdır. Modelin yaklaşık olarak kurulması farklı formlarda mümkündür. Bu formlar incelenirse, ilk olarak, zaman serisi analizinde basit istatistiksel modeller tarafından yakalanamayacak önemli olay veya durumlar ortaya çıkabilir. Bu duruma örnek olarak 1973 petrol krizi verilebilir. İkincisi, yatay kesit çalışmalarında, gerçek yaşamın tüm karmaşıklığını tek bir modelde yakalamak zor olduğundan sapan gözlemler kolayca ortaya çıkabilmektedir. Örneğin, bir işyerinde çalışan kişilere şimdiki ve altı ay sonraki beklenen maaşları sorulursa çoğu çalışan için bu iki değer arasındaki fark çok az olacaktır. Bu çalışanlardan birinin doktora öğrencisi ve yaklaşık dört ay sonra mezun olacağı düşünülürse altı ay sonra çok daha yüksek bir maaşa sahip olacaktır. Böyle bir durum ortaya çıkarsa sıradan istatistiksel modeller için bu doktora öğrencisi bir sapan gözlem olacaktır. Üçüncü olarak, doğrusal bir model kurulursa, doğrusal olmayan bir ilişki sapan gözlem olabilir. Örneğin, orta gelir grubunda gelir-gider arasındaki ilişki doğrusal olabilir. Ancak yüksek gelir grubunda bu ilişki doğrusal olmayabilir. Eğer hepsi için doğrusal bir model kullanılırsa yüksek gelir grubu büyük ihtimalle sapan olacaktır. Dördüncü ve son olarak, değişkenlerin modelden çıkarılması/ atılması yüksek sapanlara neden olabilir. Örneğin gerekli olan değişkenler modelde yoksa veya araştırmacının bu değişkenlerden haberi (ilişkisi hakkında bilgisi) yoksa böyle durum sapan gözlem ortaya çıkabilir. Örneğin hane halkı harcamaları ele alınırsa, soğuk iklimde yaşayan insanlar kış aylarında yüksek ısıtma ve izolasyon masrafı yaparken eğer modele sıcaklık değişkeni alınmazsa bu ailelerin harcamaları sapan olabilir.

Şekil 1: Regresyon Sapanları



Bu örnek durumlar incelendikten sonra regresyon sapanları için, Şekil 1 ‘de gösterilen doğrusal regresyon modeli $y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$ ele alınırsa;

- Siyah çember içerisindeki gözlemler dikey (y- yönünde) sapanlardır. Bu gözlemlerin x_i değerleri diğer gözlemlere yakındır. Ancak verinin çoğunun sahip olduğu doğrusal ilişkiye aykırı değerlerdir. Daha önce verilen hane halkı harcamaları örnek olarak gösterilebilir.
- Kırmızı çember içerisindeki gözlemler “iyi kaldıraç (good leverage)” noktalarıdır. Verilerin çoğunun sahip olduğu doğrusal ilişkiye sahip ancak büyük x_i değerlerine sahiptirler. Böyle gözlemler EKK’nın etkinliğini artırma eğilimindedirler. Adının iyi olmasının tersine EKK tahminleyicileri üzerinde büyük etkiye sahiptirler. Onları dışlamak regresyon doğrusunu büyük ölçüde değiştirebilir. Buna ek olarak daha uzağa koyulurlarsa “kötü kaldıraç (bad leverage)” olurlar.
- Yeşil çember içerisindeki gözlemler “kötü kaldıraç (bad leverage)” noktalarıdır. Büyük x_i değerlerine sahiptirler ve veri setinin çoğuna uyum göstermezler. EKK tahminleyicileri üzerinde çok büyük etkiye sahiptirler. Büyük hata olabilirler. Daha önce verilen bir doktora öğrencisi buna örnek olarak verilebilir.

Büyük hata veya model hatası gibi durumlarda dayanıklı yöntemlerin kullanılması bu hataların etkilerini en aza indirmek için önemlidir. Bu nedenle bu tez çalışmasında dayanıklı regresyon yöntemleri üzerinde durulmuştur. Dayanıklı regresyon tahmin edicilerinden beklenen bazı özellikler;

- Eğer verilerde sapan gözlem yoksa ve verilerin dağılışı normal ise EKK kadar iyi performansı olmalıdır.
- Birinci durumdaki şartlar sağlanmadığında EKK'dan daha iyi performansı olmalıdır.
- Teorisinin anlaşılması en az EKK yöntemi kadar kolay olmalıdır.
- Verilerdeki küçük değişimlere karşı duyarsız olmalıdır.
- Kolayca hesaplanabilmelidir (Ryan, 2008; Staudte ve Sheather, 2011).

1.3.1. Regresyon M- Tahminleyicisi

Regresyon parametrelerinin M- tahminleme yöntemiyle tahminlenmesi ilk defa Huber (1973) tarafından ortaya atılmış ve M- tahminleme yönteminin teorisinin gelişmeye başlamasıyla önemli bir tahminleme yöntemi haline gelmiştir. Kullanılmaya başladığı ilk yıllarda dirençli (resistant) tahminleme yöntemleri ile sıkça karıştırılmıştır. Ancak Regresyon M- Tahminleyicileri hem hesaplama hem de yorumlama açısından dirençli regresyon yöntemlerine göre çok daha gelişmiş olduklarından hızla dirençli yöntemlerin yerini almışlardır.

1.3.1.1. Regresyon Parametrelerinin M- Tahminleyicileri

Hata terimlerinin dağılışlarının kalın kuyruklu olması veya elde edilen artıklarda sapan gözlem olması durumunda regresyon parametrelerinin EKK yöntemi ile tahminlenmesi sonuçların sapmalı olmasına yol açmaktadır (Hampel ve diğerleri, 1986). Böyle durumların üstesinden gelebilmek için araştırmacılar veri setindeki anormalliklere daha dayanıklı (robust) yöntemlere yönelmişlerdir. Bu dayanıklı yöntemlerin başında Huber (1964) tarafından geliştirilen M- Tahminleyiciler bulunmaktadır. M- Tahminleyici ismi, bu tahminleyiciler EÇO (MLE) tahminleyicilerin genişletilmiş ve geliştirilmiş versiyonu olmasından gelmektedir (Stuart, 2011; Rousseeuw and Leroy, 1987; Andersen 2008).

Doğrusal regresyon modeli ele alınırsa;

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_k x_{ik} + \varepsilon_i, \quad i = 1, \dots, n, \quad j = 1, \dots, k \quad (1.52)$$

ile gösterilir ve burada $k + 1 = p$ olarak yazılabilir. Eşitlik 1.52 matris notasyonu ile yazılırsa;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (1.53)$$

ile verilebilir. Burada $\mathbf{y}$, $n \times 1$ boyutlu açıklanan (cevap) gözlem vektörü, $\boldsymbol{\theta}$, $p \times 1$ boyutlu bilinmeyen regresyon parametreleri, $\mathbf{X}$, $n \times p$ boyutlu açıklayıcı değişken (dizayn) matrisi ve $(\mathbf{X}^T \mathbf{X})^{-1}$ matrisi tam ranklı ve $\boldsymbol{\varepsilon}$, $n \times 1$ boyutlu hata vektörünü göstermektedir.

Klasik EKK yönteminde, hataların kareler toplamaları minimize edilecek olursa;

$$\sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\theta})^2 = \min! \quad (1.54)$$

olur. Eşitlik 1.54, bilinmeyen parametreye göre türevlenirse;

$$\sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\theta}) \mathbf{x}_{ij} = 0 \quad (1.55)$$

p adet denklemden oluşan sistem elde edilir. Bu denklem sistemin çözülmesiyle regresyon parametrelerinin EKK tahminleri bulunur. Eğer Eşitlik 1.55 matris gösterimiyle yazılırsa;

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\theta} = \mathbf{X}^T \mathbf{y} \quad (1.56)$$

elde edilir. Eşitlik 1.56'te bilinmeyen parametre vektörü yalnız bırakıldığında;

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (1.57)$$

bulunur. Tahminlenmiş açıklanan değişken değerleri matris gösterimiyle yazılırsa;

$$\hat{\mathbf{y}} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (1.58)$$

elde edilir. Burada $\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ matrisi 'İzdüşüm Matrisi veya Şapka Matris' olarak adlandırılmakta ve $\mathbf{H}$ ile gösterilmektedir. $\mathbf{H}$ matrisi $n \times n$ boyutlu, simetrik ($\mathbf{H}^T = \mathbf{H}$) ve idempotent ($\mathbf{H}^2 = \mathbf{H}$) bir matristir. $\mathbf{H}$ matrisi, açıklanan değişken değerlerinin tahminlenmiş açıklanan değişken değerleri üzerindeki etkisini ölçmek için kullanılır. Bu matrisin köşegen elemanları h_i ile gösterilmekte ve $0 \leq h_i \leq 1$ arasında değerler almaktadırlar. Ayrıca elde edilen h_i değerleri sapan gözlemlerin ve yüksek etki noktalarının tespit edilmesinde önemli bir rol oynamaktadırlar.

Dayanıklı Regresyon Analizinde, hataların kareler toplamını minimize etmek yerine, yine hataların farklı fonksiyon veya dağılımları minimize edilir. Böylece minimize edilecek fonksiyon (Amaç Fonksiyonu);

$$\sum_{i=1}^n \rho(y_i - \mathbf{x}_i^T \boldsymbol{\theta}) = \min! \quad (1.59)$$

ile gösterilebilir (Susanti ve Pratiwi, 2014: 352). Eşitlik 1.59'da parametrelere göre türev alınır;

$$\sum_{i=1}^n \psi(y_i - \mathbf{x}_i^T \boldsymbol{\theta}) x_{ij} = 0 \quad (1.60)$$

elde edilir. Burada $\psi(\cdot)$ fonksiyonu Etki veya Skor Fonksiyonu olarak adlandırılır. Eşitlik 1.60 çözüldükten sonra elde edilen i -inci artık değerleri;

$$e_i = e_i(\boldsymbol{\theta}) = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\theta}} \quad (1.61)$$

bulunur. Minimize edilecek fonksiyon artıklar üzerinden tekrar yazılırsa;

$$\min \sum_{i=1}^n \rho\left(\frac{e_i}{s}\right) \quad (1.62)$$

elde edilir. Eşitlik 1.62'da bilinmeyen parametreye göre türev alınır ve sıfıra eşitlenirse;

$$\sum_{i=1}^n \psi(r_i) x_{ij} = 0 \quad (1.63)$$

elde edilir. Eşitlik 1.63'da $r_i = e_i/s$ 'dir ve s ölçek değişmezliğinin (scale equaverient) sağlanması için kullanılan standart sapmanın bir tahminidir (Bkz. Ek 1). Birçok farklı s tahmini olsa bile sapan gözlemlerden etkilenmeyen Medyan Mutlak Sapma (MAD) (Bkz. Ek 2) değeri en yaygın olarak kullanılan ölçek tahminidir (Draper and Smith, 2014).

Eşitlik 1.63 tekrar düzenlenir ve ağırlık fonksiyonu $w_i = \psi(r_i)/r_i$ alınır;

$$\begin{aligned} \sum_{i=1}^n w_i r_i x_{ij} &= 0 \\ \sum_{i=1}^n x_{ij} w_i (y_i - \mathbf{x}_i^T \boldsymbol{\theta}) &= 0 \\ \sum_{i=1}^n x_{ij} w_i \mathbf{x}_i \boldsymbol{\theta} &= \sum_{i=1}^n x_{ij} w_i y_i \end{aligned} \quad (1.64)$$

elde edilir. Eşitlik 1.64 bilinmeyen regresyon parametresi $\boldsymbol{\theta}$ 'ya göre çözümlerse ve matris notasyonunda yazılırsa;

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \boldsymbol{\theta} = \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (1.65)$$

bulunur. Böylece $\boldsymbol{\theta}$ parametresinin tahmini;

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (1.66)$$

bulunur. Burada $\mathbf{W} = \text{diag} \{w_i, i = 1, \dots, n\}$ ile gösterilen $n \times n$ ağırlık matrisidir (Huber, 1981). Burada çözüm yapılırken Ağırlıklı En Küçük Kareler yöntemi kullanılır. Ancak ağırlıkların hesaplanabilmesi için ilk çözümün uygun bir yöntemle (genellikle EKK) yapılması ve elde edilen artıklar kullanılarak seçilen M- Tahminleyicilerin ağırlık fonksiyonuyla ağırlıkların hesaplanması gerekmektedir

Regresyon analizinde, nokta tahminlemede EKK'nın varsayımlarından biri olmasa bile hataların dağılışı genellikle normal olarak kabul edilmektedir. EKK'da normal dağılış varsayımı istatistiksel yorumlama için gereklidir. Eğer hataların dağılışı gerçekten normal ise EKK ve EÇO yöntemleri birbirleriyle aynıdır denebilir. Buradan yola çıkarak M- Tahminleyiciler, hataların dağılışı normalden farklı olduğunda (çarpık, uzun kuyruklu, yüksek basıklık vb.) farklı bir fonksiyon veya dağılış kullanarak parametre tahmini yaparlar. Bu yönden M- Tahminleyiciler ve EKK yöntemi EÇO tipi tahminleyicilerdir denilebilir (Stuart, 2011; Rousseeuw and Leroy, 1987; Andersen 2008).

1.3.1.2. Regresyon M- Tahminleyicilerin Asimtotik Dağılışı

Robust regresyon yaklaşımında parametrelerin dağılışlarını elde etmek için birçok araştırmacı Etki Fonksiyonu ve Asimtotik Varyans arasındaki ilişkiyi dikkate alarak Etki Fonksiyonu için beklenen değer ve varyansları elde etmişlerdir. Etki fonksiyonu $\psi(x_i, \hat{\theta})$ ele alınır ve gerçek parametre vektörü θ_0 etrafında Taylor açılımı uygulanırsa;

$$\psi(x_i, \hat{\theta}) = \psi(x_i, \theta_0) + \psi'(x_i, \theta_0)(\hat{\theta} - \theta_0) + R_n \quad (1.67)$$

elde edilir. Eşitlik 1.67'te her iki tarafın toplamı alınırsa düzgünlük koşulları (Bkz. Ek 3) altında

$n \rightarrow \infty$ iken $\sqrt{n}R_n \xrightarrow{p} 0$ olur ve;

$$\sum_{i=1}^n \psi(x_i, \hat{\theta}) = \sum_{i=1}^n \psi(x_i, \theta_0) + (\hat{\theta} - \theta_0) \sum_{i=1}^n \psi'(x_i, \theta_0) \quad (1.68)$$

bulunur. Burada $\psi'(x_i, \theta_0) = \frac{d\psi(x_i, \theta_0)}{d\theta^T}$ ve Eşitlik 1.11 dikkate alınarak $E[\psi(x_i, \theta_0)] = 0$

için;

$$\sum_{i=1}^n \frac{\psi(x_i, \theta_0)}{n} + (\hat{\theta} - \theta_0) \sum_{i=1}^n \frac{\psi'(x_i, \theta_0)}{n} = 0 \quad (1.69)$$

elde edilir. $(\hat{\theta} - \theta_0)$ terimi yalnız bırakılırsa;

$$\begin{aligned} \sum_{i=1}^n \frac{\psi(x_i, \theta_0)}{n} &= -(\hat{\theta} - \theta_0) \sum_{i=1}^n \frac{\psi'(x_i, \theta_0)}{n} \\ (\hat{\theta} - \theta_0) &= -\frac{\sum_{i=1}^n \frac{\psi(x_i, \theta_0)}{n}}{\sum_{i=1}^n \frac{\psi'(x_i, \theta_0)}{n}} \\ (\hat{\theta} - \theta_0) &= \frac{\sum_{i=1}^n \psi(x_i, \theta_0)}{-\sum_{i=1}^n \psi'(x_i, \theta_0)} \end{aligned} \quad (1.70)$$

olur. Eşitlik 1.70'in her iki tarafı $\sqrt{n}$ ile çarpılırsa;

$$\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n}(\psi(x_i, \theta_0)) \left[(-\psi'(x_i, \theta_0)) \right]^{-1} \quad (1.71)$$

bulunur. Düzgünlük koşulları (Bkz. Ek 3) altında $n \rightarrow \infty$ için Slutsky Teoremi (Bkz. Ek 4) kullanarak iki rassal değişkenin toplamının dağılışından Eşitlik 1.68;

$$\hat{\theta} \sim N(\theta_0, V(\theta_0)) \quad (1.72)$$

yazılabilir ve Sandviç Varyans Teoremi (Bkz. Ek 5) kullanılarak Robust Varyans- Kovaryans matrisi $V(\theta_0) = B(\theta_0)^{-1} A(\theta_0) \{B(\theta_0)^{-1}\}^T$ formülü ile elde edilir. Burada

$A(\theta_0) = E[\psi(x_i, \theta_0) \psi(x_i, \theta_0)^T]$ ve $B(\theta_0) = E[-\psi'(x_i, \theta_0)]$ yazılabilir. Regresyon parametrelerinin Varyans Kovaryans matrisi için $\psi(x_i, \theta_0)$ yerine $\psi(y_i - \mathbf{x}_i^T \theta)$ yazılırsa;

$$\begin{aligned} A(\mathbf{X}, \theta_0) &= \frac{1}{n} \sum_{i=1}^n \psi(\varepsilon_i)^2 \mathbf{x}_i \mathbf{x}_i^T \\ &= E[\psi(\varepsilon_i)^2] \mathbf{x}_i \mathbf{x}_i^T \end{aligned} \quad (1.73)$$

$$\begin{aligned} B(\mathbf{X}, \theta_0) &= \frac{1}{n} \sum_{i=1}^n [-\psi'(y_i, x_i, \theta_0)] \\ &= \frac{1}{n} \sum_{i=1}^n \psi'(\varepsilon_i) \mathbf{x}_i \mathbf{x}_i^T \\ &= E[\psi'(\varepsilon_i)] \mathbf{x}_i \mathbf{x}_i^T \end{aligned} \quad (1.74)$$

elde edilir. Sandviç varyans yöntemi kullanılarak;

$$\mathbf{V}(\mathbf{X}, \boldsymbol{\theta}_0) = (\mathbf{X}^T \mathbf{X} / n)^{-1} \frac{E[\psi(\varepsilon_i)^2]}{[E(\psi'(\varepsilon_i))]^2} \quad (1.75)$$

elde edilir (Huber, 1973:812; Hogg, 1979:111; Stefanski ve Boss, 2002:36; Schrader ve Hettmansperger, 1980: 95; Gross, 1977: 343; Bickel ve diğerleri, 1976: 155).

Holland ve Welsch (1977), Robust regresyon çözümlemesi yaparken IRWLS kullanmışlardır. Street ve diğerleri (1988) Ağırlıklı EKK (AEKK) için varyans tahminini $\mathbf{V}_{AEKK}(\hat{\boldsymbol{\theta}}) = s^2 (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$ olarak göstermişlerdir. Ancak robust regresyon parametre vektörü için varyans tahmini düzeltilirse;

$$\mathbf{V}(\hat{\boldsymbol{\theta}}) = s^2 \frac{E[\psi(r_i)^2]}{(E[\psi'(r_i)])^2} (\mathbf{X}^T \mathbf{X})^{-1} \quad (1.76)$$

şeklinde yazılabilir (Chatterjee ve Mächler, 1997: 1392).

Huber (1981: 163), regresyon parametre vektörünün asimtotik varyans- kovaryans matrisi için üç farklı sapmasız formülasyon önermiştir. Bu formüller birbirleri yerine rahatlıkla kullanılabilirler. Bu formüller sırasıyla;

$$\mathbf{V}(\hat{\boldsymbol{\theta}}) = K_h^2 \frac{[1/(n-p)] \sum_{i=1}^n \psi(r_i)^2}{\left[(1/n) \sum_{i=1}^n \psi'(r_i) \right]^2} (\mathbf{X}^T \mathbf{X})^{-1} \quad (1.77)$$

$$\mathbf{V}(\hat{\boldsymbol{\theta}}) = K_h \frac{[1/(n-p)] \sum_{i=1}^n \psi(r_i)^2}{\left[(1/n) \sum_{i=1}^n \psi'(r_i) \right]^2} \mathbf{W}^{-1}$$

$$\mathbf{V}(\hat{\boldsymbol{\theta}}) = K_h^{-1} \frac{1}{n-p} \sum_{i=1}^n \psi(r_i)^2 \mathbf{W}^{-1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{W}^{-1}$$

ile gösterilebilir. Burada düzeltme faktörleri $K_h = 1 + \frac{p}{n} \frac{\text{var}(\psi')}{(E(\psi'))^2}$ ve $\mathbf{W} = \sum_{i=1}^n \psi'(r_i) \mathbf{x}_{ij} \mathbf{x}_{ik}$,

$k = 1, \dots, p$ olarak gösterilir. Bu çalışmanın uygulama bölümünde (Bölüm 4) hesaplama kolaylığı açısından varyans- kovaryans matrisleri Eşitlik 1.77 kullanılarak elde edilmiştir.

1.4. HUBER VE TUKEY M- TAHMİNLEME YÖNTEMLERİ

M- Tahminleyiciler içerisinde en çok bilinen ve en çok kullanılan tahminleme fonksiyonları Huber (1964) ve Tukey (1974) tarafından geliştirilen M- Tahminleyicilerdir.

1.4.1. Huber M- Tahminleyici

Huber (1964) tarafından ilk defa ortaya atılan Amaç ve Etki fonksiyonudur. Bu fonksiyon birden fazla fonksiyonun birleşimiyle oluşan bir fonksiyonudur. En önemli özelliği amaç fonksiyonunun merkezinde normal dağılışı gibi merkezden uzaklaştıkça (önceden belirlenen bir sabit noktadan sonra) Çift Üstel (Laplace) dağılışı gibi davranmasıdır (Hogg, 1979: 109). Önerilen tahminleme ailesi için amaç fonksiyonu;

$$\rho(r) = \begin{cases} \frac{1}{2}r^2 & , |r| < k \\ k|r| - \frac{1}{2}k^2 & , |r| \geq k \end{cases} \quad (1.78)$$

şeklinde tanımlanmaktadır. Burada bulunan ayarlama sabiti (tuning constant) k , eğer dağılışı gerçekten normal ise $\hat{\theta}$ 'nın etkinliğini en yüksek yapmaktadır. Huber (1964), k için farklı değerler üzerinde çalışma yapmış ve bu değer 1.5'e yakın bir değer olması gerektiği belirtilmiştir. Bu değer literatürde Princeton Dayanıklılık Çalıştayı (Princeton Robustness Study) sonrasında $H15$ adını almıştır (Andrews ve diğerleri, 1972:12). Ancak bu sabit, standart sapmanın bir fonksiyonudur. Eğer verilerde sapan veya sapan olduğundan şüphelenilen bir gözlem varsa standart sapma bu gözlemlerden çok etkileneceği için tahmin yapmak hatalı olacaktır. Bu nedenle bu gözlemlerden etkilenmeyecek bir ölçek tahmini yapmak daha doğru olacaktır. Verilerin dağılışı standart normal olduğunda %95 etkinlik için $k = 1.345$ olarak kullanılmaktadır. Ölçek tahmini için MAD kullanıldığında $k = \frac{MAD}{0.6745}$ olmakta ve

$E(MAD) = 0.6745$ (Bkz. Ek 2) için $k \equiv \sigma$ elde edilmektedir. $k \rightarrow \infty$ için M- tahminleyici ortalamaya, $k \rightarrow 0$ için medyana dönüşmektedir.

Huber amaç fonksiyonu incelendiğinde, parametre tahmini yapılırken dağılışın orta kısmında ($-k < r < k$) normal dağılışı (EKK yöntemi) gibi davranırken uç kısımlarda Laplace dağılışı (En Küçük Medyan Kareler yöntemi) gibi davranmaktadır.

Amaç fonksiyonundan yola çıkarak Huber Etki Fonksiyonu;

$$\psi(r) = \begin{cases} r & , |r| < k \\ k \operatorname{sign}(r) & , |r| \geq k \end{cases} \quad (1.79)$$

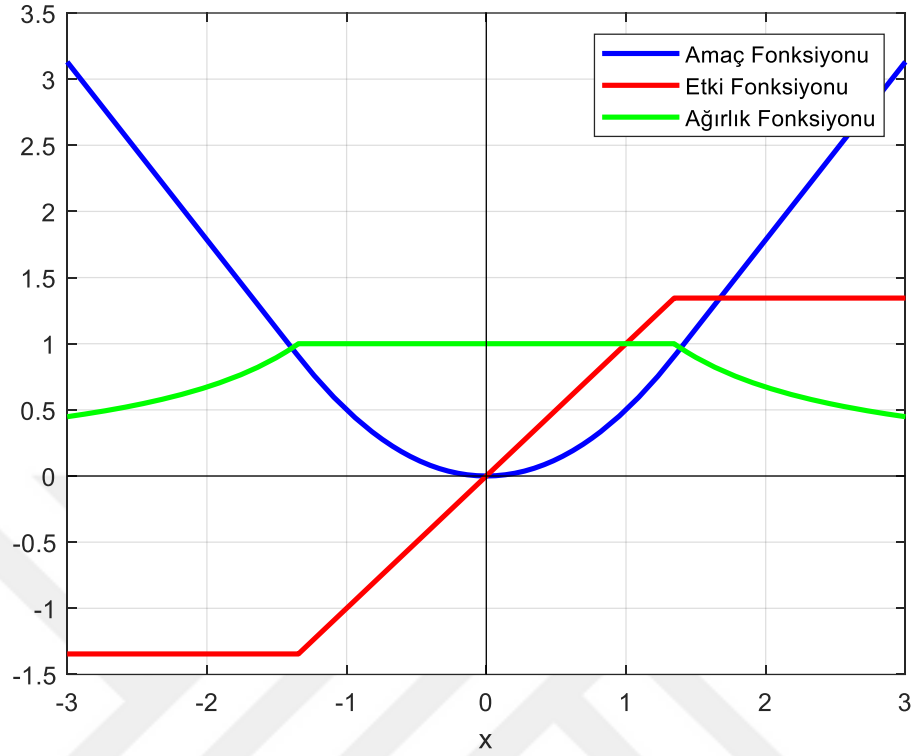
elde edilir. Etki fonksiyonuna göre gözlemler dağılışın orta kısmında kendi değerleri kadar etki yaparken uç kısımlarda maksimum ayarlama sabitinin değeri ile gözlemin işaretinin çarpımı kadar etki yapmaktadırlar. Ayarlama sabitine kadar fonksiyon doğrusal artan (azalan) değerler alırken ayarlama sabitinden sonra fonksiyonun değeri sabit olmaktadır.

Huber M- tahminleyicisine ait Ağırlık Fonksiyonu ise;

$$w(r) = \begin{cases} 1 & , |r| < k \\ \frac{k}{|r|} & , |r| \geq k \end{cases} \quad (1.80)$$

olarak ede edilir. Bu ağırlık fonksiyonuna göre dağılışın merkezinde bulunan gözlemlere eşit ve 1 ağırlığı verilirken merkezden uzaklaştıkça gözlemlerin mutlak değeriyle ters orantılı ağırlıklar verilmektedir. Burada en önemli nokta ise Huber M- tahminleyici için ağırlık fonksiyonunun hiçbir zaman 0'a eşit olmamasıdır (Fox ve Weisberg, 2002:3). Şekil 2'de Huber M- Tahminleyicinin Amaç, Etki ve Ağırlık Fonksiyonları verilmiştir.

Şekil 2: Huber M- Tahminleyici Amaç, Etki ve Ağırlık Fonksiyonu



1.4.2. Tukey M- Tahminleyici

Beaton ve Tukey (1974) tarafından önerilen ağırlık fonksiyonuna dayanan tahminleme yöntemidir. Tukey's Bi-Weight (Bisquare) adıyla bilinmektedir. Bisquare amaç (kayıp) fonksiyonu;

$$\rho(r) = \begin{cases} \frac{k^2}{6} \left\{ 1 - \left[1 - \left(\frac{r}{k} \right)^2 \right]^3 \right\} & , |r| < k \\ \frac{k^2}{6} & , |r| \geq k \end{cases} \quad (1.81)$$

şeklinde tanımlanmaktadır. Huber M- tahminleyicide olduğu gibi bir ayarlama sabiti bulunmaktadır ve standart normal dağılış altında %95 etkinlik için bu sabit $k = 4.685$ olarak kullanılmaktadır. Bisquare Etki Fonksiyonu;

$$\psi(r) = \begin{cases} r \left[1 - \left(\frac{r}{k} \right)^2 \right]^2 & , |r| < k \\ 0 & , |r| \geq k \end{cases} \quad (1.82)$$

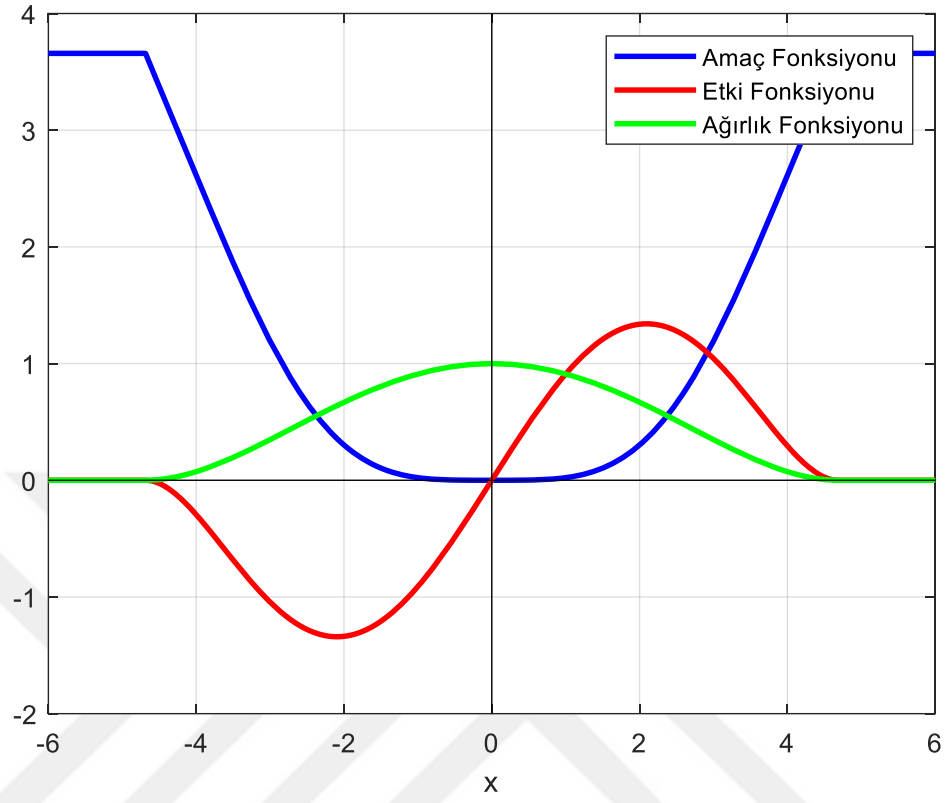
elde edilir. Etki fonksiyonunun en önemli özelliği veri setinde ayarlama sabitinden daha büyük veya eşit gözlemler olduğunda tahminleyici üzerinde herhangi bir etki yapmamasıdır. Bu yöntem sapan gözlemlere karşı dayanıklı olsa bile sapan gözlem olmayan ama sapan olduğundan şüphelenilen (veya sapan olmadığı bilinen) gözlemlerin etkisini ortadan kaldırmaktadır. Bu yöntemde de aynı Huber yönteminde olduğu gibi standart sapmayı hesaplamak sapan gözlem olduğunda hatalı olacaktır. Bu nedenle farklı ölçek tahminleri yapılabilir.

Tukey Bisquare M- Tahminleyicisine ait Ağırlık Fonksiyonu;

$$w(r) = \begin{cases} \left[1 - \left(\frac{r}{k} \right)^2 \right]^2 & , |r| < k \\ 0 & , |r| \geq k \end{cases} \quad (1.83)$$

ile gösterilir. Bu ağırlık fonksiyonuna göre dağılım merkezinde bulunan gözlemler 0 ile 1 arasında ağırlık değerleri alırken dağılım merkezinden uzaklaştığında gözlemlerin ağırlıkları 0 değerini almaktadırlar. Tukey's Bisquare M- Tahminleyiciye ait Amaç, Etki ve Ağırlık Fonksiyonu Şekil 3'de verilmiştir.

Şekil 3: Tukey's Bisquare M- Tahminleyici Amaç, Etki ve Ağırlık Fonksiyonu



İKİNCİ BÖLÜM

PEARSON DAĞILIŞ SİSTEMİ

2.1. PEARSON DİFERANSİYEL DENKLEMİ

İlk defa Karl Pearson (1895) tarafından ortaya atılan ve kendi adıyla anılan Pearson Diferansiyel denklemi;

$$\frac{1}{f(x)} \frac{df(x)}{dx} = \frac{d \ln f(x)}{dx} = \frac{f'(x)}{f(x)} = \frac{x-a}{c_0 + c_1x + c_2x^2} \quad (2.1)$$

ile gösterilir. Bu diferansiyel denklemin çözümleri Pearson Dağılış Ailesini oluşturmaktadır. Burada diferansiyel denklemin paydasında bulunan $C(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \dots$ ifadesi genişletilebilir. Diferansiyel denklem içerisinde bulunan a değeri mod değerini başka bir deyişle;

$$\left. \frac{df(x)}{dx} \right|_{x=a} = 0 \quad (2.2)$$

koşulunu sağlayan değerdir. Paydada bulunan $c_0, c_1, c_3, \dots$ değerleri ise $C(x)$ fonksiyonunun parametreleridir. Dağılışın biçimi bu diferansiyel denklemin parametrelerine göre $a, c_0, c_1, c_3, \dots$ göre belirlenmektedir. $C(x)$ fonksiyonunun bir polinom olarak gösterilmesi diferansiyel denklemin bilinmeyen parametrelerinin tahminlenmesinde momentler yönteminin kullanılmasına olanak sağlar (Şehirlioğlu ve Dündar, 2014:8).

2.2. DİFERANSİYEL DENKLEMİN ÇÖZÜMÜNDE PEARSON YAKLAŞIMI

Eşitlik 2.1'deki diferansiyel denklem yeniden düzenlenirse;

$$\frac{df(x)}{dx} = \frac{(x-a) f(x)}{c_0 + c_1x + c_2x^2} \quad (2.3)$$

$$(c_0 + c_1x + c_2x^2) \frac{df(x)}{dx} = (x-a) f(x) \quad (2.4)$$

ve x şans değişkeninin sınırları r_1, r_2 olarak kabul edilirse yukarıdaki eşitliğin her iki tarafı x^n ile çarpılarak integrali alınır;

$$\int_{r_2}^{r_1} x^n (c_0 + c_1 x + c_2 x^2) \left(\frac{df(x)}{dx} \right) dx = \int_{r_2}^{r_1} x^n (x - a) f(x) dx \quad (2.5)$$

elde edilir. Eşitlik 2.5 kısmi integrasyon yöntemiyle çözülürse;

$$x^n (c_0 + c_1 x + c_2 x^2) f(x) \Big|_{r_2}^{r_1} - \int_{r_2}^{r_1} f(x) [nc_0 x^{n-1} + (n+1)c_1 x^n + (n+2)c_2 x^{n+1}] dx$$

elde edilir. Denklemin sol tarafı ele alınırsa;

$$\begin{aligned} & \int_{r_2}^{r_1} x^n (c_0 + c_1 x + c_2 x^2) \left(\frac{df(x)}{dx} \right) dx \\ &= - \int_{r_2}^{r_1} [nc_0 x^{n-1} f(x) + (n+1)c_1 x^n f(x) + (n+2)c_2 x^{n+1} f(x)] dx \end{aligned} \quad (2.6)$$

bulunur. Burada x eksenini ile uç noktalarda bir temas olduğu varsayımı ile $\lim_{x \rightarrow r_1} f(x) = 0$ ve

$\lim_{x \rightarrow r_2} f(x) = 0$ olduğu varsayılmaktadır. Eşitlik 2.5'in sağ tarafı ise;

$$\begin{aligned} & \int_{r_2}^{r_1} x^n (x - a) f(x) dx \\ &= \int_{r_2}^{r_1} x^{n+1} f(x) dx - \int_{r_2}^{r_1} ax^n f(x) dx \end{aligned} \quad (2.7)$$

gösterilir. Eşitlik 2.6 ve Eşitlik 2.7 birlikte değerlendirilir ve çözülürse ardışık moment denklemi;

$$\begin{aligned} & -nc_0 \mu'_{n-1} - (n+1)c_1 \mu'_n - (n+2)c_2 \mu'_{n+1} = \mu'_{n+1} - a \mu'_n \\ & nc_0 \mu'_{n-1} - \{(n+1)c_1 - a\} \mu'_n - \{(n+2)c_2 + 1\} \mu'_{n+1} = 0 \end{aligned} \quad (2.8)$$

elde edilir. Ardışık moment denkleminde $n = 0, 1, 2, 3$ değerleri yerine koyularak ve gerekli tüm momentlerin mevcut olduğu varsayılarak a, c_0, c_1, c_2 parametreleri;

$$\begin{aligned} a &= - \frac{(20\mu'_2 \mu'_3 \mu'^2_2 + 13\mu'_1 \mu'_2 \mu'_4 - 12\mu'_4 \mu'^3_1 - 9\mu'_1 \mu'^3_2 - 8\mu'_1 \mu'^2_3 - 3\mu'^2_2 \mu'_3 - \mu'_3 \mu'_4)}{Z} \\ c_0 &= - \frac{(4\mu'^2_2 \mu'_4 + 4\mu'^2_1 \mu'^2_3 - 3\mu'_2 \mu'^2_3 - \mu'_1 \mu'^2_2 \mu'_3 - 3\mu'^2_1 \mu'_2 \mu'_4 - \mu'_1 \mu'_3 \mu'_4)}{Z} \\ c_1 &= - \frac{(8\mu'_2 \mu'_3 \mu'^2_1 + 7\mu'_1 \mu'_2 \mu'_4 - 6\mu'_4 \mu'^3_1 - 3\mu'_1 \mu'^3_2 - 2\mu'_1 \mu'^2_3 - 3\mu'^2_2 \mu'_3 - \mu'_3 \mu'_4)}{Z} \\ c_2 &= - \frac{(3\mu'^2_3 + 6\mu'^3_2 + 4\mu'^3_1 \mu'_3 + 2\mu'^2_1 \mu'_4 - 3\mu'^2_1 \mu'^2_2 - 10\mu'_1 \mu'_2 \mu'_3 - 2\mu'_2 \mu'_4)}{Z} \end{aligned} \quad (2.9)$$

elde edilir. Burada $\mu'_0 = 1$ ve ardışık moment denklemlerinin paydasında bulunan ifade ortak olup;

$$Z = 18\mu_2'^3 - 6\mu_1'^2\mu_1'^2 - 10\mu_2'\mu_4' - 32\mu_1'\mu_2'\mu_3' + 10\mu_1'^2\mu_4' + 8\mu_1'^3\mu_3' + 12\mu_3'^2$$

dir. $\mu_1' = 0$ yapılırsa (dağılımın ortalaması 0'a çekilirse);

$$\begin{aligned} c_1 - a &= 0 \\ c_0 + (3c_2 + 1)\mu_2 &= 0 \\ (3c_1 - a)\mu_2 + (4c_2 + 1)\mu_3 &= 0 \\ 3c_0\mu_2 + (4c_1 - a)\mu_3 + (5c_2 + 1)\mu_4 &= 0 \end{aligned} \quad (2.10)$$

denklemleri elde edilir. $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ (Çarpıklık) ve $\beta_2 = \frac{\mu_4}{\mu_2^2}$ (basıklık) olarak alınırsa diferansiyel

denklemin parametreleri;

$$\begin{aligned} c_0 &= -\frac{\mu_2(4\mu_2\mu_4 - 3\mu_2^2)}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3} = -\frac{\sigma^2(4\beta_2 - 3\beta_1)}{10\beta_2 - 12\beta_1 - 18} \\ c_1 = a &= -\frac{\mu_3(\mu_4 - 3\mu_3^2)}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3} = -\frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18} \\ c_2 &= -\frac{2\mu_4\mu_2 - 3\mu_3^2 - 6\mu_2^3}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3} = -\frac{2\beta_2 - 3\beta_1 - 6}{10\beta_2 - 12\beta_1 - 18} \end{aligned} \quad (2.11)$$

elde edilir (Kotz ve diğerleri, 1972:22). Elde edilen parametrelere göre diferansiyel denklem tekrar yazılırsa;

$$\begin{aligned} &\frac{1}{f(x)} \frac{df(x)}{dx} \\ &= -\frac{x + \frac{\mu_3(\mu_4 - 3\mu_3^2)}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3}}{\frac{\mu_2(4\mu_2\mu_4 - 3\mu_3^2)}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3} + \left(\frac{\mu_3(\mu_4 - 3\mu_3^2)}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3}\right)x + \left(\frac{2\mu_4\mu_2 - 3\mu_3^2 - 6\mu_2^3}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3}\right)x^2} \end{aligned} \quad (2.12)$$

veya

$$\begin{aligned} &\frac{1}{f(x)} \frac{df(x)}{dx} \\ &= -\frac{x + \frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18}}{\frac{\sigma^2(4\beta_2 - 3\beta_1)}{10\beta_2 - 12\beta_1 - 18} + \left(\frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18}\right)x + \left(\frac{2\beta_2 - 3\beta_1 - 6}{10\beta_2 - 12\beta_1 - 18}\right)x^2} \end{aligned} \quad (2.13)$$

ile gösterilebilir. Eşitlik 2.12 ve Eşitlik 2.13'te gösterilen momentlerin König teoremi kullanılarak merkezi momentler cinsinden gösterimi;

$$\begin{aligned}\mu_2 &= \mu_2' - \mu_1'^2 \\ \mu_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 \\ \mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'^2\mu_1' - 3\mu_1'^4\end{aligned}\tag{2.14}$$

yazılır.

Pearson dağılış sistemindeki dağılışların tipleri $C(x)$ fonksiyonunun köklerine göre tanımlanmıştır. Fonksiyonun iki köke sahip olduğu durumlar dağılış sisteminin ana tiplerini, tek kök veya katlı kök olduğu durumlar geçiş tiplerini tanımlar (Şehirlioğlu ve Dünder, 2014:15). Pearson dağılış sisteminin 3 ana ve 10 geçiş (ara) tipi vardır. Bu tipler;

1. Köklerin gerçek (reel) ve farklı işaretli olduğu durum Tip I (PI),
2. Köklerin gerçek (reel) ve aynı işaretli olduğu durum Tip VI (PVI),
3. Köklerin karmaşık olduğu durum Tip IV (PIV)'tür.

Tipleri belirlemek amacıyla kullanılabilecek diğer bir yöntem ise kappa (κ) kriteridir. Bu kriter $C(x)$ fonksiyonunun diskriminantının kullanılmasıyla elde edilen bir istatistiktir (Elderton, 1906:42; Hald, 2008:110; Friori ve Zenga, 2009:43; Nagahara, 2008:176). Denklemin diskriminantı;

$$\Delta = c_1^2 - 4c_0c_2\tag{2.15}$$

olarak yazılır ve diskriminantın işaretine göre κ belirlenirse;

$$\kappa = \frac{c_1^2}{4c_0c_2} = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)}\tag{2.16}$$

yazılır (Pearson, 1901). Bu istatistiğe göre dağılışlar Tablo 1'de verilmiştir.

Tablo 1: Pearson Dağılımlarının Δ , κ ve Köklerine Göre Durumları

Δ	κ İstatistiği	c_0	c_1	c_2	Köklerin Durumu	Tip
$\Delta > 0$	$\kappa < 0$	$c_0 \neq 0$	$c_1 \neq 0$	$c_2 \neq 0$	Reel	Tip I (Ana Tip)
$\Delta > 0$	$\kappa > 1$	$c_0 \neq 0$	$c_1 \neq 0$	$c_2 \neq 0$	Reel	Tip VI (Ana Tip)
$\Delta < 0$	$0 < \kappa < 1$	$c_0 \neq 0$	$c_1 \neq 0$	$c_2 \neq 0$	Karmaşık	Tip IV (Ana Tip)
$\Delta = 0$	$\kappa = 1$	$c_0 \neq 0$	$c_1 \neq 0$	$c_2 \neq 0$	Reel	Tip V (Ara tip)
$\Delta < 0$	$\kappa = 0$	$c_0 \neq 0$	$c_1 = 0$	$c_2 \neq 0$	Reel	Tip II ve Tip VII (Ara Tipler)
$\Delta > 0$	$\kappa \rightarrow \pm\infty$	$c_0 \neq 0$	$c_1 \neq 0$	$c_2 = 0$	Reel	Tip III ve Tip X (Ara Tip)
$\Delta = 0$	$\kappa = 0$	$c_0 \neq 0$	$c_1 = 0$	$c_2 = 0$	Reel	Normal Dağılım

(Şehirlioğlu ve Dündar, 2014:16).

Pearson dağılımları için önemli bir durum dağılımın modunun 0'a çekildiği durumdur ($a = 0$).

Bu durumu tanımlama için $X = x - a$ dönüşümü yapılırsa $dX = dx$ olur ve diferansiyel denklem;

$$\frac{df(X)}{dX} = \frac{Xf(X)}{C_0 + C_1X + C_2X^2} = \frac{(x-a)f(x)}{c_0 + c_1(x-a) + c_2(x-a)^2} \quad (2.17)$$

yazılarak paydanın eşitliğinden;

$$C_0 + C_1X + C_2X^2 = c_0 + c_1(x-a) + c_2(x-a)^2 \quad (2.18)$$

özdeşliği kullanılır ve $c_1 = a$ alınır;

$$\begin{aligned} c_2 &= C_2 \\ 2ac_2 + c_1 &= a(2c_2 + 1) = C_1 \\ c_2a^2 + c_1a + c_0 &= c_0 + a^2(1 + c_2) = C_0 \end{aligned} \quad (2.19)$$

olur ve elde edilen parametreler için dağılımın mod değeri 0 olur.

2.3. PEARSON DİFERANSİYEL DENKLEMİNİN ÇÖZÜMÜ İÇİN CRAIG YAKLAŞIMI

Craig (1936), Pearson diferansiyel denkleminde orijinal değişkenler yerine standart değişkenleri kullanarak alternatif bir çözüm üretmiştir. Bu yöntemle göre z standart değişken olmak üzere diferansiyel denklem;

$$\frac{1}{f(z)} \frac{df(z)}{dz} = \frac{a-z}{c_0 + c_1 z + c_2 z^2} \quad (2.20)$$

olarak yazılabilir. Burada ardışık moment denklemi;

$$-nc_0\mu_{n-1} - (n+1)c_1\mu_n - (n+2)c_2\mu_{n+1} = a\mu_n - \mu_{n+1} \quad (2.21)$$

olup $\mu_0 = 1$ ve $\mu_1 = 0$ alınarak $n = 0, 1, 2, 3$ için çözüm yapılır. Burada diferansiyel denklemin parametreleri;

$$\begin{aligned} a &= \frac{-\sigma\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18} \\ c_0 &= \frac{\sigma^2(4\beta_2 + 3\beta_1)}{10\beta_2 - 12\beta_1 - 18} \\ c_1 &= \frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18} \\ c_2 &= \frac{(2\beta_2 - 3\beta_1 - 6)}{10\beta_2 - 12\beta_1 - 18} \end{aligned} \quad (2.22)$$

elde edilir. Standart değişken kullanıldığı için $\mu'_1 = 0$, $\mu'_2 = \mu_2 = 1$, $\mu_3 = \gamma_1 = \sqrt{\beta_1}$ ve $\mu_4 = \beta_2$ olur. Craig yeni bir parametre tanımlamış ve diferansiyel denklemin parametreleri için yeniden çözüm yapmıştır. Bu parametre;

$$\delta = \frac{(2\beta_2 - 3\beta_1 - 6)}{\beta_2 + 3} \quad (2.23)$$

olarak gösterilmiştir. Bu parametre dikkate alınarak çözümler;

$$\begin{aligned}
a &= \frac{-\gamma_1}{2(1+2\delta)} \\
c_0 &= \frac{2+\delta}{2(1+2\delta)} \\
c_1 &= \frac{\gamma_1}{2(1+2\delta)} \\
c_2 &= \frac{\delta}{2(1+2\delta)}
\end{aligned} \tag{2.24}$$

olarak elde edilir.

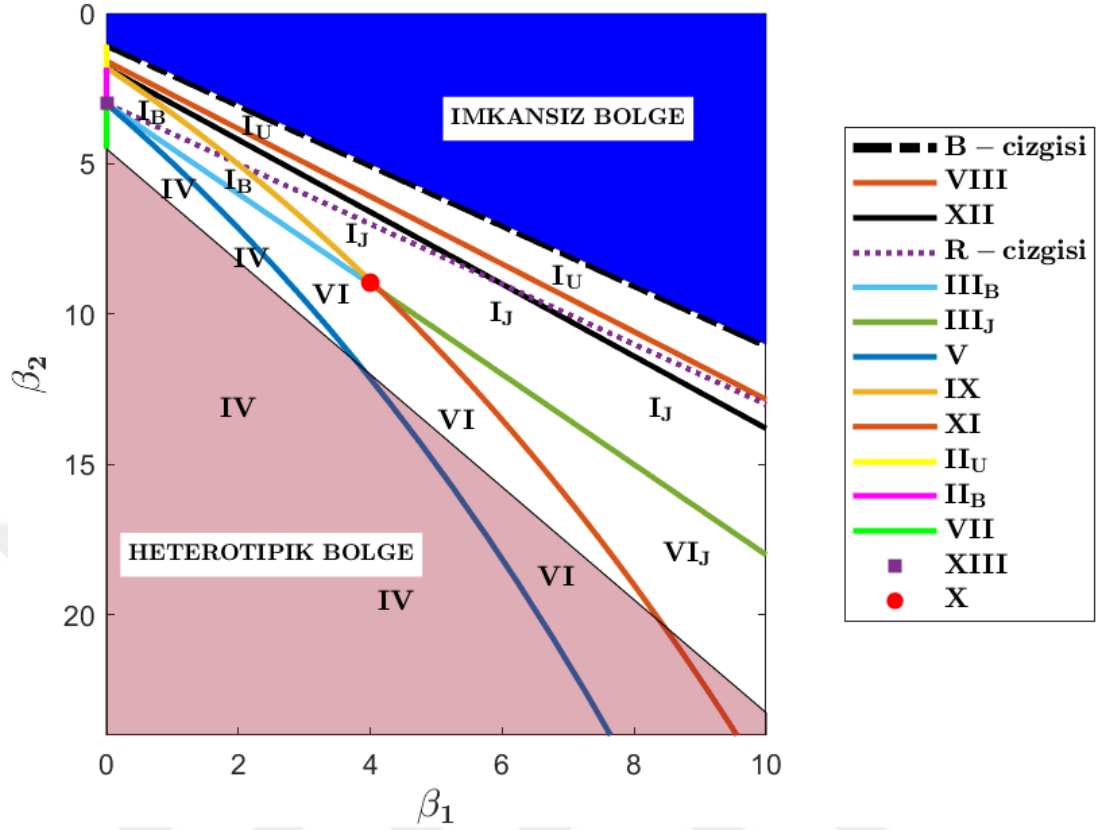
2.4. PEARSON DAĞILIŞLARINI TANIMLAYAN KARTLAR

Gözlenmiş veri setinin dağılışını bulabilmek amacıyla kullanılan yöntemler arasında Pearson (1914) ve Craig (1936) tarafından geliştirilen kartlar bulunmaktadır. Bu kartlar verilerin ilk dört momentine dayanmaktadır.

2.4.1. Pearson Kartı

Pearson (1914) tarafından geliştirilen kart (grafik) verilere uyum yapabilmek için ilk dört momenti kullanır. Bu momentlerin kullanılmasıyla hesaplanan $\beta_1 = \mu_3^2 / \mu_2^3$ yatay eksen ve $\beta_2 = \mu_4 / \mu_2^2$ dikey eksen Şekil 4'teki karta işaretlenir. Böylece verilere uygun bir dağılış kesin olarak belirlenebilir (Şehirlioğlu ve Dünder, 2014: 33). Kart üzerindeki İmkansız Bölgede dağılış ortaya çıkmazken Heterotipik Bölgede sekizinci derece ve üzeri momentlerin sonsuz olduğu dağılışlar yer alır.

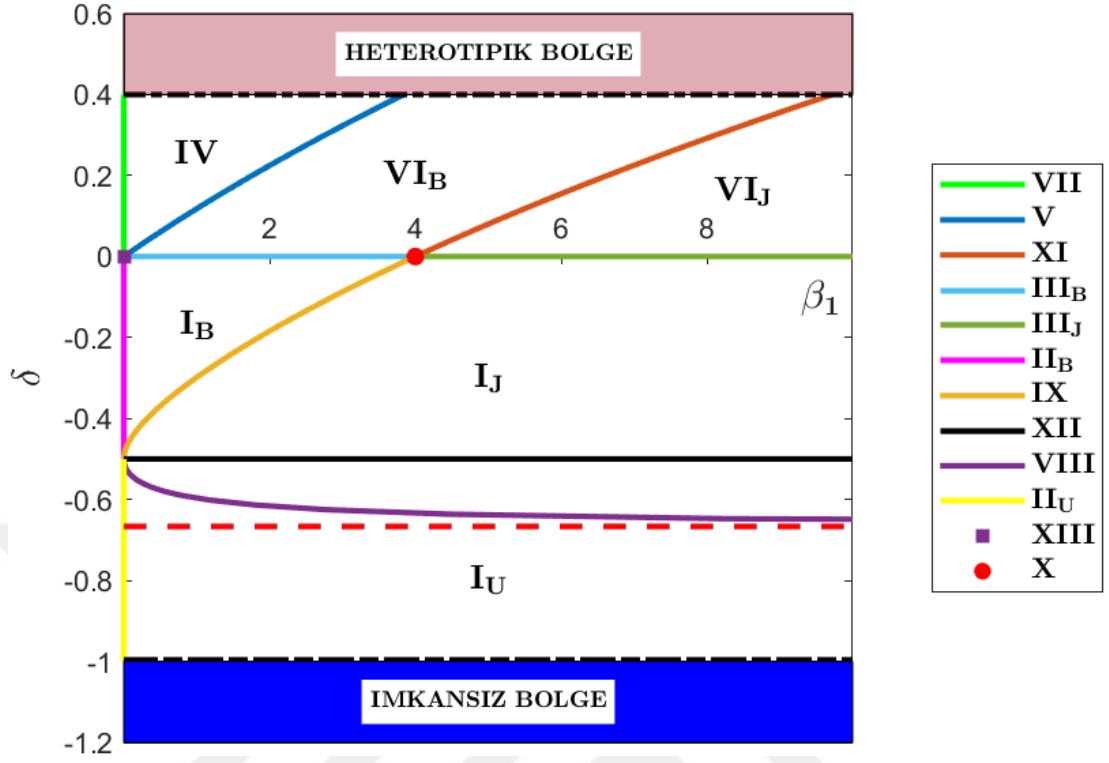
Şekil 4: Pearson Kartı



2.4.2 Craig Kartı

Craig (1936) tarafından geliştirilen ve Pearson dağılışlarını belirlemede kullanılan diğer bir karttır. Şekil 5'te dikey ekseninde δ ve yatay ekseninde β_1 değeri bulunmaktadır. Kartta $\delta = -1$ eksenin altı İmkansız Bölge olarak adlandırılmakta ve bu bölgede herhangi bir dağılış ortaya çıkmamaktadır. $\delta = 0.4$ 'ün üzerinde kalan bölge Heterotipik Bölge olarak adlandırılır.

Şekil 5: Craig Kartı



2.5. PEARSON ANA TİPLERİ İÇİN GENEL ÇÖZÜM

Pearson dağılışı sistemin ana tipleri PI, PIV ve PVI için genel çözüm için öncelikle diferansiyel denklemin paydası $c_2 (x \pm r_1)(x \pm r_2)$ şeklinde çarpanlarına ayrılırsa daha genel bir ifadeyle;

$$c_0 + c_1x + c_2x^2 = c_2 \left[x - \frac{-c_1 + \sqrt{\Delta}}{2c_2} \right] \left[x - \frac{-c_1 - \sqrt{\Delta}}{2c_2} \right] \quad (2.25)$$

yazılabilir. Burada Eşitlik 2.15'ten;

$$r_1 = \frac{-c_1 + \sqrt{\Delta}}{2c_2} \quad (2.26)$$

$$r_2 = \frac{-c_1 - \sqrt{\Delta}}{2c_2}$$

elde edilir. Burada r_1 ve r_2 diferansiyel denklemin paydasının kökleridir ve parametreler cinsinden yazılmıştır. Eşitlik 2.25 ve 2.26 birlikte dikkate alınarak diferansiyel denklem;

$$\begin{aligned}\frac{d \ln y}{dx} &= \frac{x-a}{c_0 + c_1 x + c_2 x^2} \\ &= \frac{x-a}{c_2 (x \pm r_1)(x \pm r_2)}\end{aligned}\quad (2.27)$$

yazılabilir. Diferansiyel denklemin farklı yazılış tipleri için Friori ve Zenga (2009), Nagahara (2008), Elderton, (1953) çalışmaları incelenebilir. Köklerin, parametrelerin ve Δ değerinin durumuna göre dağılışın tipi farklılık gösterir. Bu durumlar aşağıda verilmiştir;

1. Eğer $\Delta > 0$, $\kappa < 0$, $r_2 < 0 < r_1$ ve $c_2 > 0$ ise PI
2. Eğer $\Delta < 0$, $0 < \kappa < 1$, r_1 ve r_2 karmaşık ise PIV
3. Eğer $\Delta > 0$, $\kappa > 1$, $0 < |r_1| < |r_2|$ ve $c_2 < 0$ ise PVI (Pearson, 1895:367; Elderton, 1906:42; Şehirlioğlu ve Dündar, 2014:21; Kotz ve diğerleri, 1972:17).

2.5.1. Pearson Tip I (PI) Dağılışı

PI dağılışı aynı zamanda *Beta Dağılışı* olarak bilinmektedir. Dağılışa ait 4 parametre vardır. Bu parametreler r_2 yer, r_1 ölçek ve m_1, m_2 çarpıklık ve basıklık parametreleridir. Kökler ters işaretli ve dağılışın ortaya çıkabilmesi için $c_2 > 0$ ve $c_0 < 0$ olmalıdır. Bu koşullar sağlandığında diferansiyel denklem;

$$\int d \ln f(x) = \frac{1}{c_2} \int \frac{x-a}{(x-r_1)(x+r_2)} dx = \frac{1}{c_2} \int \left[\frac{A_1}{(x-r_1)} + \frac{A_2}{(x+r_2)} \right] dx \quad (2.28)$$

ifadesi kısmi kesirler yöntemiyle çözülürse;

$$\begin{aligned}A_1 &= \frac{r_1 - a}{r_1 + r_2} \\ A_2 &= \frac{a + r_2}{r_1 + r_2}\end{aligned}\quad (2.29)$$

elde edilir ve integrali alınırsa;

$$\ln f(x) = \frac{A_1}{c_2} \ln |x - r_1| + \frac{A_2}{c_2} \ln |x + r_2| + \ln K \quad (2.30)$$

eşitliği bulunur. Eşitlik 2.30'daki mutlak ifadeleri açılırsa;

$$f(x) = K (r_1 - x)^{\frac{A_1}{c_2}} (x + r_2)^{\frac{A_2}{c_2}} = K (r_1 - x)^{m_1} (x + r_2)^{m_2}, \quad r_2 < x < r_1 \quad (2.31)$$

ve $\int f(x) = 1$ olabilmesi için K integral sabiti;

$$K = \frac{1}{B(m_1 + 1; m_2 + 1)(r_1 + r_2)^{m_1 + m_2 + 1}} \quad (2.32)$$

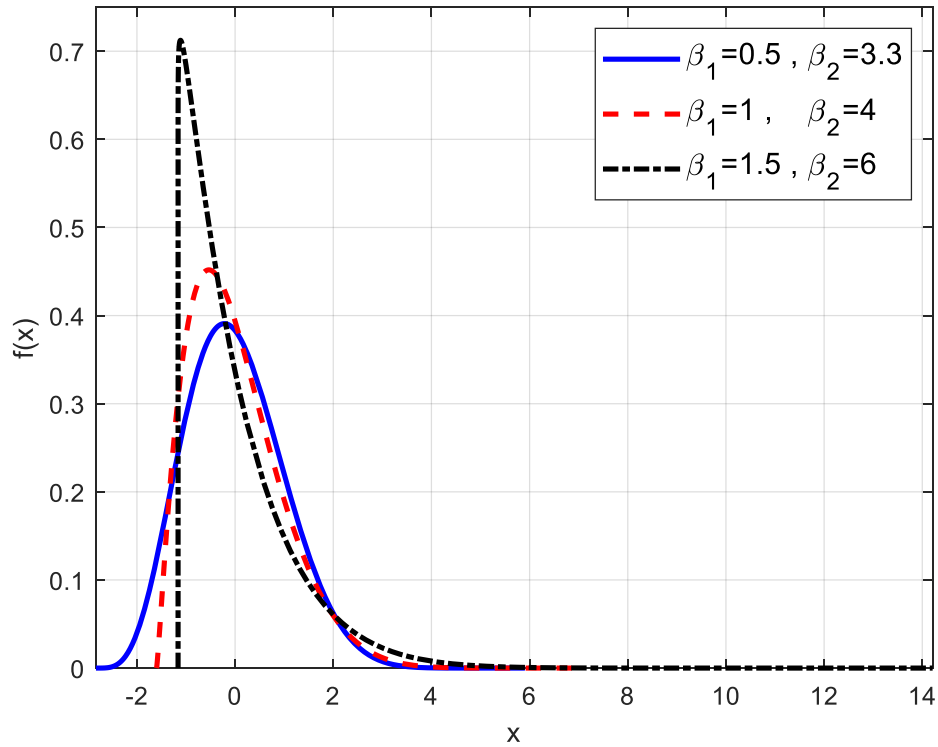
olarak belirlenir. Burada $B(x, y)$, Beta fonksiyonudur. Köklerin ve üs parametrelerinin durumlarına göre ortaya çıkan PI dağılışının biçimleri Tablo 2’de verilmiştir.

Tablo 2: Pearson Tip I Dağılışının Biçimleri

<i>Koşul 1</i>	<i>Koşul 2</i>	<i>Dağılım Biçimi</i>
$ r_2 > a$	$m_1 > 0$ ve $m_2 > 0$	Sağa çarpık B- Biçimli
$r_1 > a$	$m_1 > 0$ ve $m_2 > 0$	Sola çarpık B- Biçimli
$ r_2 < a$	$m_1 > 0$ ve $m_2 < 0$	Sağa çarpık J- Biçimli
$r_1 < a$	$m_1 < 0$ ve $m_2 > 0$	Sola çarpık J- Biçimli
-	$m_1 < 0$ ve $m_2 < 0$	U- Biçimli
-	$m_1 = m_2$	Simetrik (Tip II)

PI dağılışının farklı çarpıklık ve basıklık değerlerine göre oyları Şekil 6’da verilmiştir.

Şekil 6: Pearson Tip I Dağılışının Olasılık Yoğunluk Fonksiyonları



PI dağılışının mod değeri orijine çekilirse $X = x - a$ dönüşümü yapılarak $dX = dx$ elde edilir. Buradan diferansiyel denklem;

$$\int d \ln f(x) = \frac{1}{C_2} \int \left[\frac{A_1^*}{X - R_1} + \frac{A_2^*}{X + R_2} \right] dX \quad (2.33)$$

olur. Kısmi kesirler yöntemiyle;

$$\begin{aligned} A_1^* &= \frac{R_1}{R_1 + R_2} \\ A_2^* &= \frac{R_2}{R_1 + R_2} \end{aligned} \quad (2.34)$$

elde edilir. Burada $v = 1/(C_2(R_1 + R_2))$, $M_1 = vR_1$, $M_2 = vR_2$ olarak alınırsa $-R_2 < X < R_1$ aralığında modu orijinde olan PI dağılışı;

$$f(X) = K^* (R_1 - X)^{M_1} (X + R_2)^{M_2} \quad (2.35)$$

elde edilir. Burada K^* değeri integral sabitidir.

PI dağılışı, parametrelerinin aldıkları değerlere göre farklı dağılışları üretir. Bu dağılışlardan bazıları;

1. $m_2 = 0$ ve $z = x - r_1$ dönüşümü yapılırsa $m_1 + 1 = \lambda$ ve $r_2 - r_1 = k$ için fonksiyon Güç (Power) dağılışı olur.

$$f(z) = \lambda k^{-\lambda} z^{\lambda}, \quad 0 < z < k$$

2. $m_1 = 0$ ve $m_2 = 0$ alınırsa Üiform (Düzgün) dağılış elde edilir.

$$f(z) = \frac{1}{r_2 - r_1}, \quad r_1 < z < r_2$$

3. $p = m_1 + 1$ ve $q = m_2 + 1$ alınırsa ($r_2 = 0$ ve $r_1 = 1$) Standart Beta Dağılışı elde edilir.

$$f(z) = \frac{1}{B(p; q)} z^{p-1} (1-z)^{q-1}, \quad 0 < z < 1 \quad (2.36)$$

4. Eşitlik 2.36'da $p = 1/2$ ve $q = 1/2$ olarak alınırsa Arcsine Dağılışı elde edilir.

$$f(z) = \frac{1}{\pi} \frac{1}{\sqrt{z(1-z)}}, \quad 0 < z < 1$$

2.5.2 Pearson Tip IV (PIV) Dağılışı

Pearson dağılışı sistemi içerisinde en az kullanılan ve matematiksel olarak en zor olan dağılıştır. Diğer ana tiplerden ayrılan noktası dağılışın kökleri karmaşık olmasıdır ve $-\infty$ ile ∞ aralığında tanımlıdır. Dağılışı farklı çarpıklık ve basıklık değerlerine sahip *B- biçimli* bir dağılıştır.

Diferansiyel denklemin paydası $C(x)$ 'in kökleri $r_1 = -r + is$ ve $r_2 = -r - is$ olduğunda diferansiyel denklem;

$$\frac{1}{f(x)} \frac{df(x)}{dx} = \frac{1}{c_2} \left[\frac{A_1}{(x+r-is)} + \frac{A_2}{(x+r+is)} \right] \quad (2.37)$$

yazılabilir. Burada kısmi kesirler yöntemiyle çözüldüğünde;

$$\begin{aligned} A_1 &= \frac{1+hi}{2} \\ A_2 &= \frac{1-hi}{2} \end{aligned} \quad (2.38)$$

elde edilir. Burada $h = \frac{a+r}{s}$ 'dir. Eşitlik 2.37'de her iki tarafın integrali alınır;

$$\begin{aligned} \ln f(x) &= \frac{1}{c_2} \int \left[\frac{A_1}{(x+r-is)} + \frac{A_2}{(x+r+is)} \right] dx \\ \ln f(x) &= \frac{1}{c_2} \left(A_1 \ln((x+r-is)) + A_2 \ln((x+r+is)) \right) + \ln K \\ f(x) &= K (x+r-is)^{\frac{h}{c_2}} (x+r+is)^{\frac{1}{2c_2}} \end{aligned} \quad (2.39)$$

yazılabilir. Burada $v = h/c_2$ ve $m = 1/2c_2$ olarak yazılırsa PIV oyl'si;

$$f(x) = K \left[(x+r)^2 + s^2 \right]^m e^{\text{varctan } \tau}, \quad -\infty < x < \infty \quad (2.40)$$

elde edilir (Pearson, 1895, 1914, 1963; Elderton, 1906:44; Kotz ve diğerleri, 1972:18; Nagahara, 1999: 252). Eşitlik 2.40'da $\tau = \frac{x+r}{s}$ 'dir. İntegral sabiti K 'yı bulmak amacıyla

$\frac{x+r}{s} = \tan \theta$ dönüşümü yapılırsa K ;

$$K = \frac{s^{-2m-1}}{\exp\left(\frac{v\pi}{2}\right) \int_{-\pi/2}^{\pi/2} (\cos \theta)^{-2m-2} \exp(-v\theta) d\theta}$$

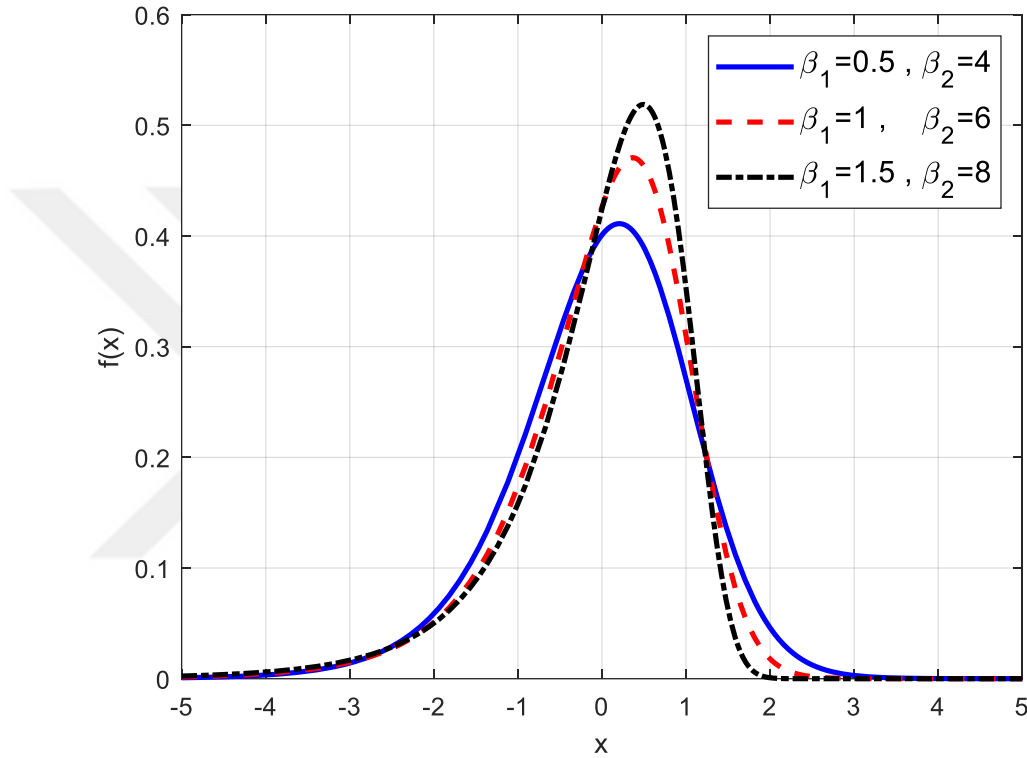
bulunur. Bu eşitlik farklı bir şekilde yazılırsa;

$$K = \frac{s^{-2m-1}}{G(-2m-2, v)}$$

yazılır (Şehirlioğlu ve Dündar, 2014:88). Burada $G(x; y) = \int_0^\pi \sin^x \varphi e^{-y\varphi} d\varphi$ 'dir (Pearson,

1895). Farklı çarpıklık ve basıklık değerlerine göre PIV dağılışının oyları Şekil 7'de verilmiştir.

Şekil 7: Pearson Tip IV Dağılışının Olasılık Yoğunluk Fonksiyonları



PIV dağılışının modunun orijine çekilmesi için $X = x - a$ dönüşümü yapılırsa $dX = dx$ olur.

Dağılım kökleri;

$$\begin{aligned} R_1 &= -R + iS \\ R_2 &= -R - iS \end{aligned} \quad (2.41)$$

yazılabilir. Buradan diferansiyel denklem;

$$\int d \ln f(x) = \frac{1}{C_2} \int \left[\frac{A_1^*}{X - R - iS} + \frac{A_2^*}{X + R + iS} \right] dX \quad (2.42)$$

elde edilir. Kısmi kesirler yöntemiyle $H = \frac{R}{S}$ için $A_1^* = \frac{1+iH}{2}$ ve $A_2^* = \frac{1-iH}{2}$ olur. Böylece

modu orijinde olan PIV dağılışı;

$$f(X) = K^* \left[(X + R)^2 + S^2 \right]^M e^{-V \arctan\left(\frac{X+R}{S}\right)}, \quad -\infty < X < \infty \quad (2.43)$$

olur. Burada $V = H/C_2$ ve $M = 1/(2C_2)$ 'dir (Şehirlioğlu ve Dündar, 2014:88).

PI dağılışında olduğu gibi PIV dağılışı da parametrelerinin aldıkları değerlere göre farklı dağılışları türetirler. Bu dağılışlar;

1. $a = c_1 = v = 0$, $c_2 > 0$ ve $c_0 > 0$ olduğunda Pearson Tip VII (Student t) dağılışı ortaya çıkar. Simetrik ve B-biçimlidir. PIV dağılışının kökleri pür imajiner olduğunda elde edilir. PVII dağılışının mod değeri orijinde olduğundan diferansiyel denklem;

$$\frac{d \ln f(x)}{dx} = \frac{X}{C_0 + C_2 X^2} = \frac{1}{c_2} \left[\frac{A_1^*}{X - iS} + \frac{A_2^*}{X + iS} \right] \quad (2.44)$$

Kısmi kesirler yardımıyla alınırsa $A_1^* = A_2^* = 1/2$ elde edilir. Her iki tarafın integrali alınırsa;

$$\ln f(X) = \frac{1}{2C_2} [\ln |X - iS| + \ln |X + iS|] + \ln K^*$$

olur. Buradan PVII için oyf;

$$f(X) = K^* [(X - iS)(X + iS)]^M, \quad -\infty < X < \infty$$

elde edilir. Burada $M = 1/(2C_2)$ 'dir. Böylelikle oyf açık haliyle yazılırsa;

$$f(X) = K^* [X^2 + S^2]^M, \quad -\infty < X < \infty \quad (2.45)$$

olur (Şehirlioğlu ve Dündar, 2014:144).

2. Standart değişken için tanımlanan Pearson Tip VII dağılışında $m = 2$ alınırsa Breit-Wigner (Cauchy) dağılışı ortaya çıkar (Crooks, 2010:63).

$$f(x) = \frac{2}{s\pi} \left(1 + \frac{x^2}{s^2} \right)^{-m}, \quad -\infty < x < \infty$$

2.5.3. Pearson Tip VI (PVI) Dağılışı

PI ve PIV dağılışlarından farklı olarak PVI (*İkinci Tip Beta*) dağılışı, dağılışın sadece bir tarafında sınırlıdır. Hem sağa çarpık hem sola çarpık olmak üzere *B* ve *J- Biçimli* durumları vardır. En önemli ayırt edici özelliği köklerin her ikisinin birden aynı işaretli olmasıdır. Hem sağa çarpık hem de sola çarpık PVI dikkate alınacaktır.

2.5.3.1. Sağa Çarpık PVI dağılışı

Dağılımın köklerinin sıralaması $-r_2 < -r_1 < 0$ olduğu durum ele alınırsa sağa çarpık PVI dağılımının ortaya çıkabilmesi için;

- $c_2 < 0$
- $c_1 > \sqrt{\Delta}$
- $a < 0$

koşulları sağlanmalıdır. İkinci koşul dikkate alındığında $c_0 < 0$ olmalıdır. Böylece diferansiyel denklemin paydasında bulunan parametrelerin hepsi 0'dan küçük olacaktır. Diferansiyel denklemin çözümü;

$$\int d \ln f(x) = \frac{1}{c_2} \int \frac{x-a}{(x+r_1)(x+r_2)} dx = \frac{1}{c_2} \int \left[\frac{A_1}{(x+r_1)} + \frac{A_2}{(x+r_2)} \right] dx \quad (2.46)$$

Kısmi kesirler yöntemiyle $(A_1 + A_2)x + A_2r_1 + A_1r_2 = x - a$ elde edilir. Bu denklemin çözümünden;

$$\begin{aligned} A_1 &= \frac{-r_1 - a}{r_2 - r_1} \\ A_2 &= \frac{r_2 + a}{r_2 - r_1} \end{aligned} \quad (2.47)$$

bulunur. İntegral alınırsa;

$$\ln f(x) = \frac{A_1}{c_2} \ln|x+r_1| + \frac{A_2}{c_2} \ln|x+r_2| + \ln K \quad (2.48)$$

olur. Mutlak değerli ifadeler tanım aralığı dikkate alınarak açılırsa;

$$f(x) = K (x+r_1)^{m_1} (x+r_2)^{m_2}, \quad -r_1 < x < \infty \quad (2.49)$$

elde edilir. Burada $A_1/c_2 = m_1$ ve $A_2/c_2 = m_2$ 'dir. İntegral sabiti K 'yı belirlemek için

$\int f(x) = 1$ dikkate alınırsa;

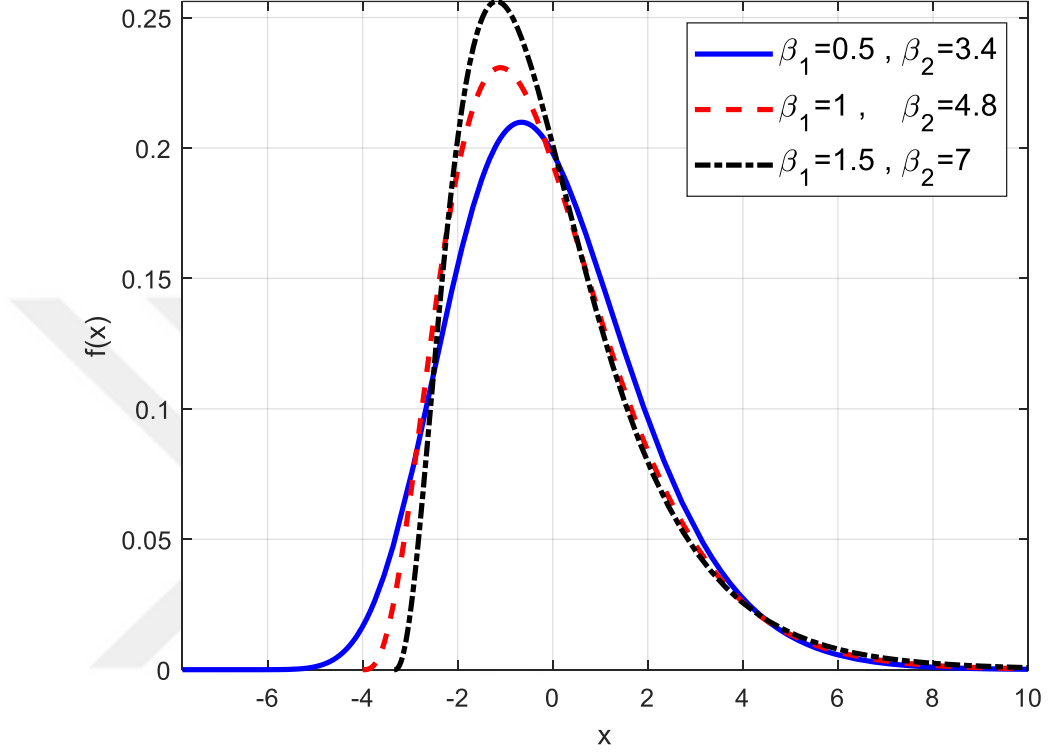
$$K = \frac{1}{B(-m_1 - m_2 - 1; m_1 + 1)(r_2 - r_1)^{m_1 + m_2 + 1}} \quad (2.50)$$

elde edilir. İntegral sabiti yerine koyulursa Sağa Çarpık PVI dağılışı;

$$f(x) = \frac{1}{B(-m_1 - m_2 - 1; m_1 + 1)(r_2 - r_1)^{m_1 + m_2 + 1}} (x+r_1)^{m_1} (x+r_2)^{m_2}, \quad -r_1 < x < \infty \quad (2.51)$$

yazılabilir. Sağa çarpık PVI dağılışının farklı çarpıklık ve basıklık değerleri için oyfleri Şekil 8’de verilmiştir.

Şekil 8: Sağa Çarpık Pearson Tip VI Dağılışının Olasılık Yoğunluk Fonksiyonları



2.5.3.2 Sola Çarpık PVI dağılışı

Dağılışı köklerinin sıralaması $0 < r_2 < r_1$ olduğu durum ele alınırsa sola çarpık PVI dağılışı ortaya çıkabilmesi için;

- $c_2 < 0$
- $c_1 > \sqrt{\Delta}$
- $a > 0$

olmalıdır. Böylece diferansiyel denklemin paydasında bulunan parametrelerin hepsi 0’dan küçük olacaktır. Sola çarpık PVI dağılışı için Sağa çarpık PVI dağılışı koşulları aynı olup sadece kökler yer değiştirmektedir. Bu değişiklik dikkate alınarak diferansiyel denklemin çözümü;

$$\int d \ln f(x) = \frac{1}{c_2} \int \frac{x-a}{(x-r_1)(x-r_2)} dx = \frac{1}{c_2} \int \left[\frac{A_1}{(x-r_1)} + \frac{A_2}{(x-r_2)} \right] dx \quad (2.52)$$

bulunur. Burada;

$$\begin{aligned} A_1 &= \frac{r_1 - a}{r_1 - r_2} \\ A_2 &= \frac{a - r_2}{r_1 - r_2} \end{aligned} \quad (2.53)$$

elde edilir. İntegral alınırsa;

$$\ln f(x) = \frac{A_1}{c_2} \ln|x-r_1| + \frac{A_2}{c_2} \ln|x-r_2| + \ln K \quad (2.54)$$

İntegral sabiti K belirlendikten sonra Sola Çarpık Pearson PVI dağılışı;

$$f(x) = \frac{1}{B(-m_1 - m_2 - 1; m_2 + 1)(r_1 - r_2)^{m_1 + m_2 + 1}} (r_1 - x)^{m_1} (r_2 - x)^{m_2}, \quad -\infty < x < r_2 \quad (2.55)$$

elde edilir (Şehirlioğlu ve Dündar, 2014:111).

PVI dağılışı, Craig Yöntemiyle çözümlürse;

$$\begin{aligned} - \delta &> 0 \\ - \gamma_1 &\neq 0 \\ - \beta_1 &> 4\delta(\delta + 2) \\ - \beta_1 &\neq \frac{[4(1 + 2\delta)^2(2 + \delta)]}{2 + 3\delta} \end{aligned}$$

koşulları sağlandığı zaman ortaya çıkar. Son koşulda eşitlik oluşursa Pearson Tip XI ortaya çıkar. Sağa çarpık ($\gamma_1 > 0$) için $\sqrt{\Delta} < |\gamma_1|$ olduğundan r_1 ve r_2 , γ_1 'den farklı işarete sahiptir.

PVI için diferansiyel denklem;

$$\int d \ln f(z) = \frac{1}{c_2} \int \left[\frac{\pi_1}{z-r_1} + \frac{\pi_2}{z-r_2} \right] dz \quad (2.56)$$

yazılır. Burada $\pi_1 = (c_1 + r_1)/(r_2 - r_1)$ ve $\pi_2 = (-c_1 - r_2)/(r_2 - r_1)$ 'dir. İntegral alınır ve K sabiti hesaplanırsa;

$$f(z) = \frac{1}{B(-m_1 - m_2 - 1; m_2 + 1)(r_1 - r_2)^{m_1 + m_2 + 1}} (z - r_1)^{m_1} (z - r_2)^{m_2}, \quad r_1 < x < \infty \quad (2.57)$$

elde edilir. Eşitlik 2.57’de parametreler yeniden düzenlenir $t = \frac{(z - r_1)}{r_1 - r_2}$ alınırsa PVI (İkinci Tip

Beta) dağılışı;

$$f(t) = \frac{1}{B(-m_1 - m_2 - 1; m_2 + 1)} \left[\frac{t^{m_1}}{(1+t)^{-m_2}} \right], \quad 0 < t < \infty \quad (2.58)$$

bulunur. Burada $q = -m_1 - m_2 - 1$ ve $p = m_1 + 1$ yazılırsa sağa çarpık dağılışı;

$$f(t) = \frac{1}{B(p; q)} \left[\frac{t^{p-1}}{(1+t)^{p+q}} \right], \quad 0 < t < \infty \quad (2.59)$$

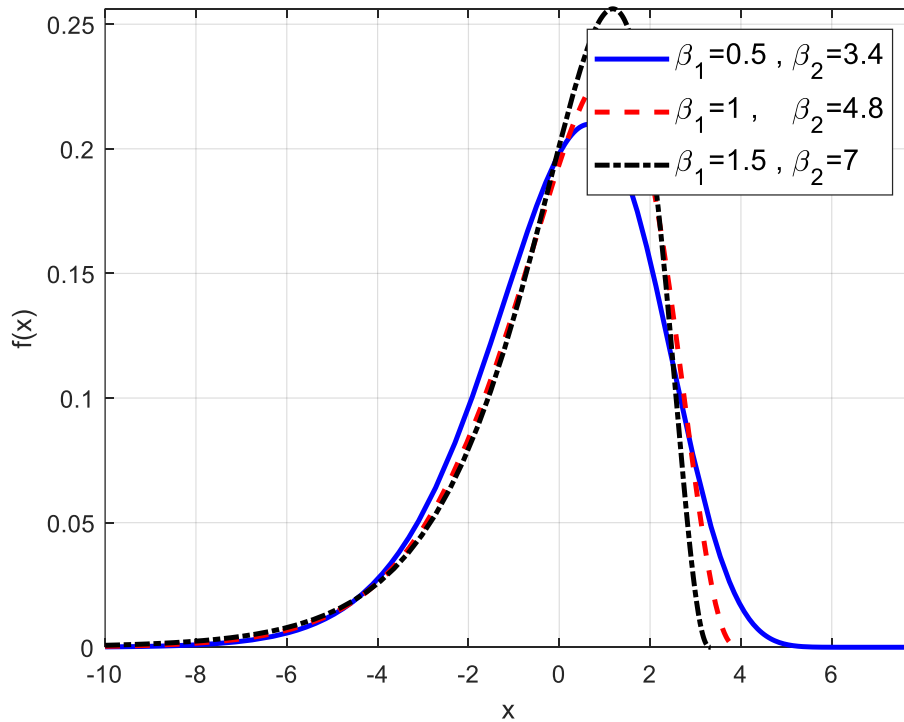
elde edilir.

Sola çarpık PVI dağılışı ($\gamma_1 < 0$) ele alınırsa r_1 ve r_2 , γ_1 ’den farklı işarete sahiptir ve $m_1 < 0$ olmalıdır. Köklerin işaretleri dikkate alınırsa Sola Çarpık PVI Dağılışı;

$$f(t) = \frac{1}{B(|m_1| - m_2 - 2; m_2 + 1)} \left[\frac{t^{|m_1| - m_2 - 2}}{(1+t)^{|m_1|}} \right], \quad 0 < t < \infty \quad (2.60)$$

elde edilir. Sola çarpık PVI dağılışının farklı çarpıklık ve basıklık değerleri için oyfleri Şekil 9’de verilmiştir.

Şekil 9: Sola Çarpık Pearson Tip VI Dağılışının Olasılık Yoğunluk Fonksiyonları



2.5.3.3 PVI Dağılışının Momentleri, Modu ve Kümülatif Fonksiyonu

Eşitlik 2.58 dikkate alınarak PDVI'nın r . momentinin var olabilmesi $r < m_1 + 1$ veya Eşitlik 2.58 dikkate alınırsa $r < q$ olması gereklidir. Orijine göre r -inci moment;

$$\mu_r' = \frac{\Gamma(-m_1 - m_2 - 1 + r) \Gamma(m_1 + 1 - r)}{\Gamma(-m_1 - m_2 - 1) \Gamma(m_1 + 1)} \quad (2.61)$$

elde edilir. Eşitlik 2.59 için r -inci moment;

$$\mu_r' = \frac{\Gamma(p + r) \Gamma(q - r)}{\Gamma(p) \Gamma(q)} \quad (2.62)$$

olarak bulunur. Dağılışı modunun bulunması amacıyla $f(x) = K(x + r_1)^{m_1} (x + r_2)^{m_2}$ fonksiyonu türevlenirse;

$$\frac{df(x)}{dx} = m_2(x + r_1) + m_1(x + r_2) = 0$$

ifadesi elde edilir. İfade x 'e göre çözülürse fonksiyonun Mod noktası;

$$a = \frac{-(m_1 r_2 + m_2 r_1)}{m_1 + m_2} \quad (2.63)$$

elde edilir. Eşitlik 2.58 kullanılarak PVI için birikimli dağılış fonksiyonu;

$$\int_0^x \frac{t^{m_1}}{(1+t)^{-m_2}} dt = \frac{t^{m_1+1}}{m_1+1} {}_2F_1 \left[\begin{matrix} m_1+1 & -m_2 \\ m_1+2 \end{matrix}; -t \right]$$

eşitliği kullanılarak birikimli dağılış fonksiyonu;

$$F(t) = \frac{t^{m_1+1} {}_2F_1 \left[\begin{matrix} m_1+1 & -m_2 \\ m_1+2 \end{matrix}; -t \right]}{(m_1+1) B(-m_1 - m_2 - 1; m_2 + 1)} \quad (2.64)$$

elde edilir. Burada ${}_2F_1 \left[\begin{matrix} a & b \\ c \end{matrix}; z \right] = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!} = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n!}$

ile gösterilen Gauss Hipergeometrik Fonksiyondur (Abramowitz ve Stegun, 1965:556).

Modu orijinde olan PVI dağılışı için $X = x - a$ dönüşümü yapılırsa $dX = dx$ olur ve diferansiyel denklem $a < 0$ için;

$$\frac{d \ln f(X)}{dX} = \frac{X}{C_0 + C_1 X + C_2 X^2} = \frac{1}{C_2} \frac{X}{(X + R_1)(X + R_2)} \quad (2.65)$$

$$= \frac{1}{C_2} \left[\frac{A_1^*}{X + R_1} + \frac{A_2^*}{X + R_2} \right] \quad (2.66)$$

bulunur. Burada

$$A_1^* = \frac{-R_1}{R_2 - R_1} \quad (2.67)$$

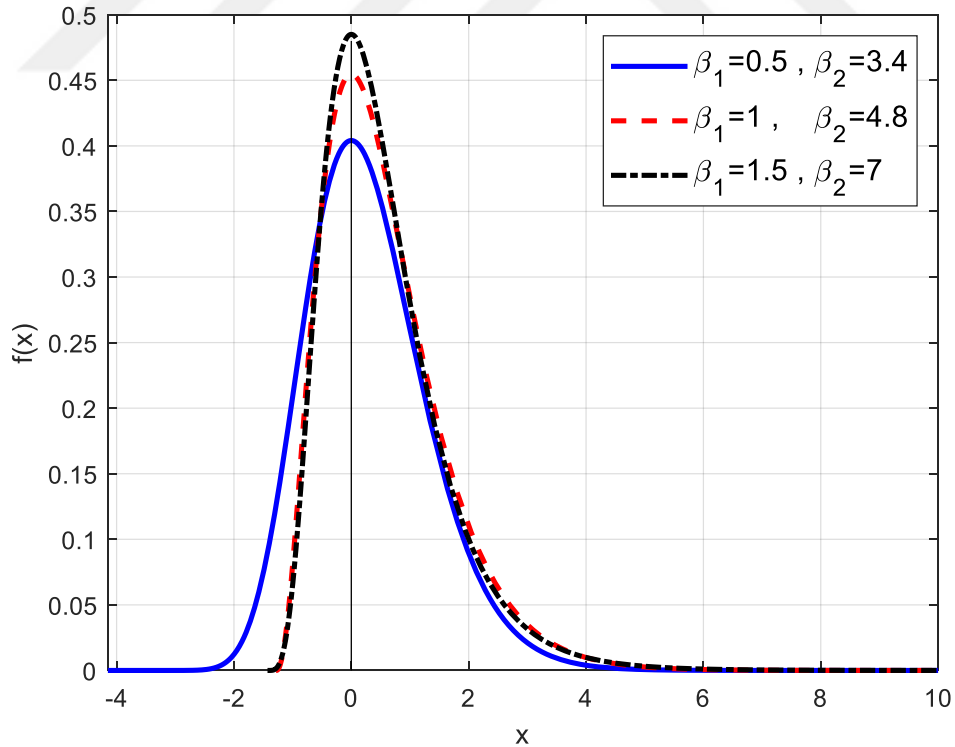
$$A_2^* = \frac{R_2}{R_2 - R_1}$$

olur. Böylece modu 0 olan PVI dağılışı $M_1 = \frac{A_1^*}{C_2}$ ve $M_2 = \frac{A_2^*}{C_2}$ olmak üzere;

$$f(X) = K^* (X + R_1)^{M_1} (X + R_2)^{M_2}, \quad R_1 < R_2 < X < \infty \quad (2.68)$$

elde edilir. Burada K^* integral sabitidir. Mod değeri orijinde olan PVI dağılışının farklı çarpıklık ve basıklık değerleri için oyf'leri Şekil 10'da verilmiştir.

Şekil 10: Mod Değeri Orijinde olan Pearson Tip VI Dağılışının Olasılık Yoğunluk Fonksiyonları



Dağılımın köklerinin ve üs parametrelerinin aldıkları değerlere göre dağılımın biçimleri Tablo 3'te verilmiştir.

Tablo 3: Pearson Tip VI Dağılısının Biçimleri

<i>Koşul 1</i>	<i>Koşul 2</i>	<i>Dağılısın Biçimi</i>
$r_1 < 0$ ve $r_2 < 0$	$m_1 > 0$ ve $m_2 < 0$	Sağa Çarpık B- Biçimli
$r_1 > 0$ ve $r_2 > 0$	$m_1 < 0$ ve $m_2 > 0$	Sola Çarpık B- Biçimli
$r_1 < 0$ ve $r_2 < 0$	$m_1 < 0$ ve $m_2 < 0$	Sağa Çarpık J- Biçimli
$r_1 > 0$ ve $r_2 > 0$	$m_1 < 0$ ve $m_2 < 0$	Sola Çarpık J- Biçimli

PI ve PIV dağılışlarında olduğu gibi PVI dağılışı da parametrelerinin aldıkları değerlere göre farklı dağılışları türetirler. Bu dağılışlar;

1. Eşitlik 2.57’de $f = \left(\frac{v_2}{v_1}\right)t$ dönüşümü yapılırsa $dt = \left(\frac{v_1}{v_2}\right)df$ olur ve f değişkeni

için sınırlar $0 < f < \infty$ olur. Parametrelere $-m_1 - m_2 - 2 = \frac{v_1}{2} - 1$ ve

$-m_2 = \frac{v_1 + v_2}{2}$ dönüşümleri yapılırsa F dağılışı;

$$f(f) = \frac{v_1^{\frac{v_1}{2}} v_2^{\frac{v_2}{2}} f^{\frac{v_1}{2}-1}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right) (v_1 + v_2 f)^{\frac{(v_1+v_2)}{2}}}, \quad 0 < f < \infty \quad (2.69)$$

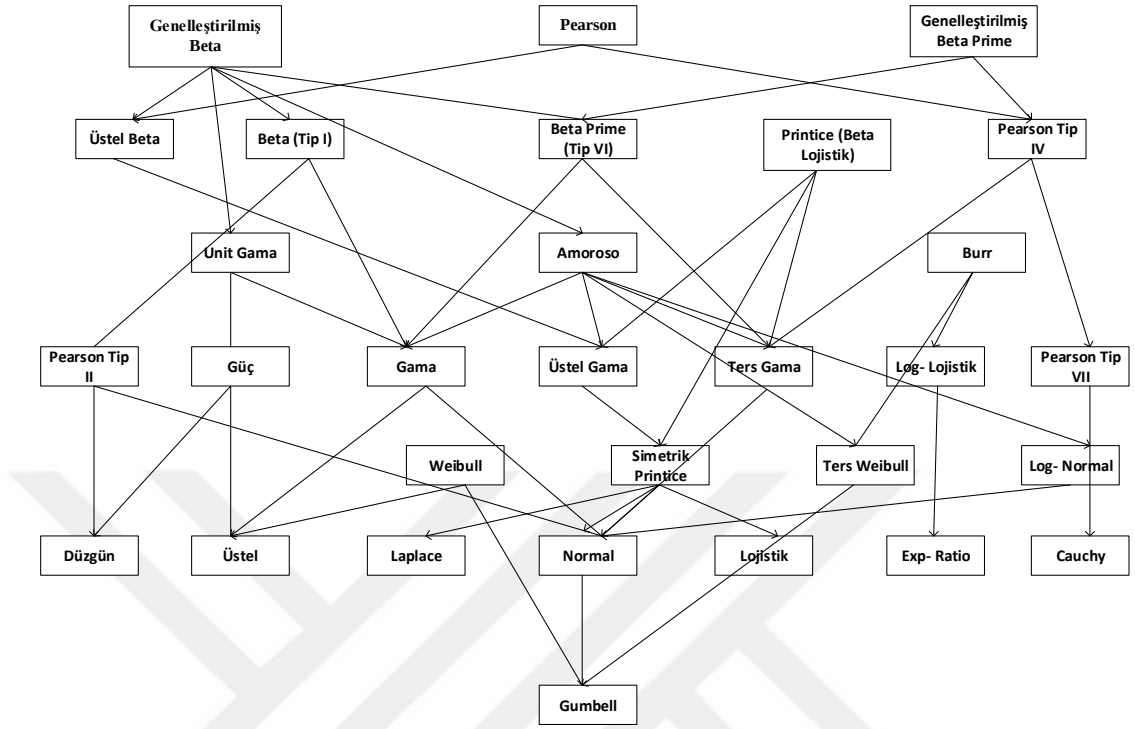
elde edilir.

2. Eşitlik 2.58’de $z = r_2$ olursa (orijin r_2 noktasına kayarsa) $u = z - r_2$ dönüşümü yapılarak yeni sınırlar $r_1 - r_2 = k$ için $k < u < \infty$ olur. Burada $m_1 = 0$ ve $m_2 = -(a+1)$ alınırsa Pareto dağılışı;

$$f(u) = ak^a \frac{1}{u^{a+1}}, \quad k < u < \infty \quad (2.70)$$

elde edilir.

Şekil 11: Pearson Dağılış Sistemi İlişkisi



Kaynak: Crooks, 2010:19.

ÜÇÜNCÜ BÖLÜM

PEARSON DİFERANSİYEL DENKLEMİNİN ETKİ FONKSİYONU OLARAK KULLANILMASI

Etki fonksiyonunun genel tanımı ile Pearson Diferansiyel denkleminin yapısı arasındaki benzerlikten yola çıkarak, Pearson diferansiyel denkleminin parametrelerinin aldığı farklı değerler için farklı yapıdaki Etki Fonksiyonları tanımlanabilir. Bu tanım kullanılarak elde edilecek Ağırlık fonksiyonu regresyon parametrelerinin tahmini yapılırken kullanılabilir. Pearson Diferansiyel Denklemi ve Etki Fonksiyonu arasındaki ilişki;

$$\psi(x) = \frac{d\rho(x)}{dx} = \frac{d[-\ln f(x)]}{dx} = \frac{f'(x)}{f(x)} = \frac{x-a}{c_0 + c_1x + c_2x^2} \quad (3.1)$$

ile gösterilebilir (Dzhun', 2011:261). Böylece Diferansiyel denklemin farklı parametreleri farklı dağılımlar ortaya çıkaracağından her bir dağılım için Etki Fonksiyonu, Diferansiyel Denklem kullanılarak rahatlıkla elde edilebilir (Wiśniewski, 2014:947). Elde edilen Etki fonksiyonunun yardımıyla Ağırlık fonksiyonunun tanımından;

$$w(x) = \frac{\psi(x)}{x} = \frac{x-a}{x(c_0 + c_1x + c_2x^2)} \quad (3.2)$$

yazılabilir. Normal dağılım için ($\beta_1 = 0$ ve $\beta_2 = 3$) $c_0 = \sigma^2$, $c_1 = c_2 = 0$ olduğunda ağırlık fonksiyonu;

$$w(x) = \frac{1}{\sigma^2} \quad (3.3)$$

elde edilir. Simetrik fonksiyonlar için $\beta_1 = 0$ ve $c_1 = 0$ olduğunda;

$$w(x) = \frac{1}{c_0 + c_2x^2} \quad (3.4)$$

olur (Dzhun', 2011:262). Ağırlık fonksiyonunun değeri genellikle dağılımın mod noktasıyla yakından ilişkilidir. Eğer veriler asimetrik ve normalden daha yüksek basıklık değerine sahipse Ağırlık Fonksiyonu, verilerin en fazla yoğunlaştığı mod noktasında maksimum değerini almalıdır. Böyle durumlarda işlemleri kolaylaştırmak amacıyla mod noktası orijine çekilerek işlem yapılabilir. Modun orijinde olduğu Etki Fonksiyonu;

$$\psi(X) = \frac{X}{C_0 + C_1 X + C_2 X^2} \quad (3.5)$$

şeklinde yazılabilir. Böylece elde edilecek ağırlık fonksiyonu ise;

$$w(X) = \frac{1}{C_0 + C_1 X + C_2 X^2} \quad (3.6)$$

elde edilir. Farklı Pearson dağılımları için farklı Amaç, Etki ve Ağırlık fonksiyonları elde edilebilir. Bu bölümde Pearson Dağılım Ailesinin ana tipleri için Amaç, Etki ve Ağırlık fonksiyonları elde edilecektir.

3.1. PI DAĞILIŞININ AMAÇ, ETKİ VE AĞIRLIK FONKSİYONU

Modu orijinde olan PI dağılımının oyf'si ele alındığında sabitler atılırsa Amaç fonksiyonu;

$$\begin{aligned} \rho(X) &= -\ln f(X) \\ &\propto -M_1 \ln(R_1 - X) - M_2 \ln(X + R_2) \end{aligned} \quad (3.7)$$

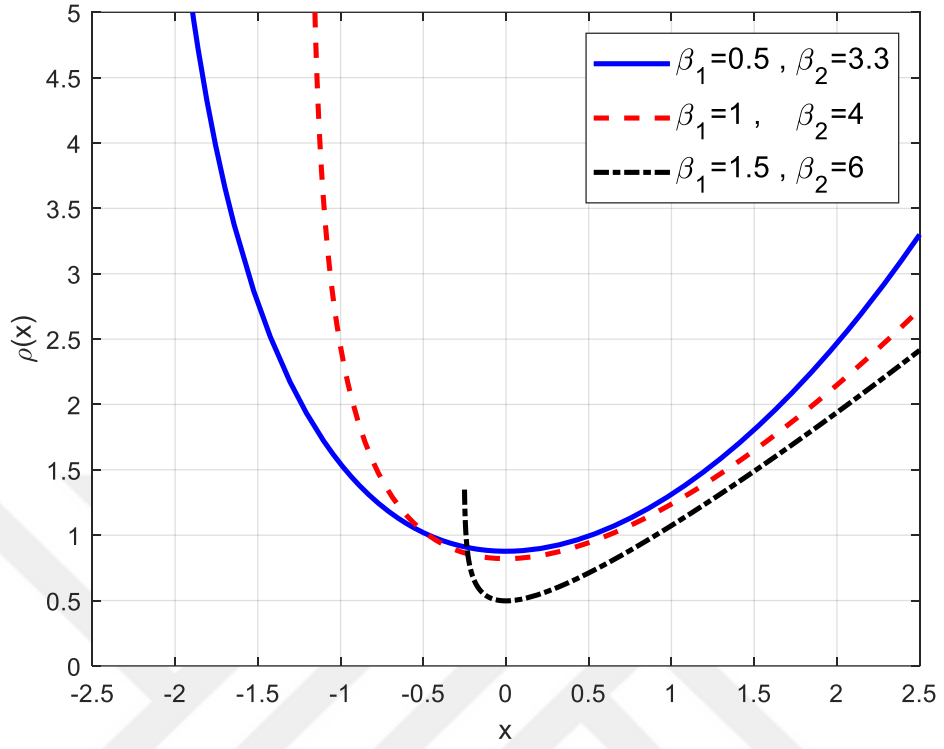
olur. Etki fonksiyonu ve Ağırlık Fonksiyonu;

$$\begin{aligned} \psi(X) &= \frac{d\rho(X)}{dX} \\ &= \frac{-M_1}{R_1 - X} - \frac{M_2}{X + R_2} \end{aligned} \quad (3.8)$$

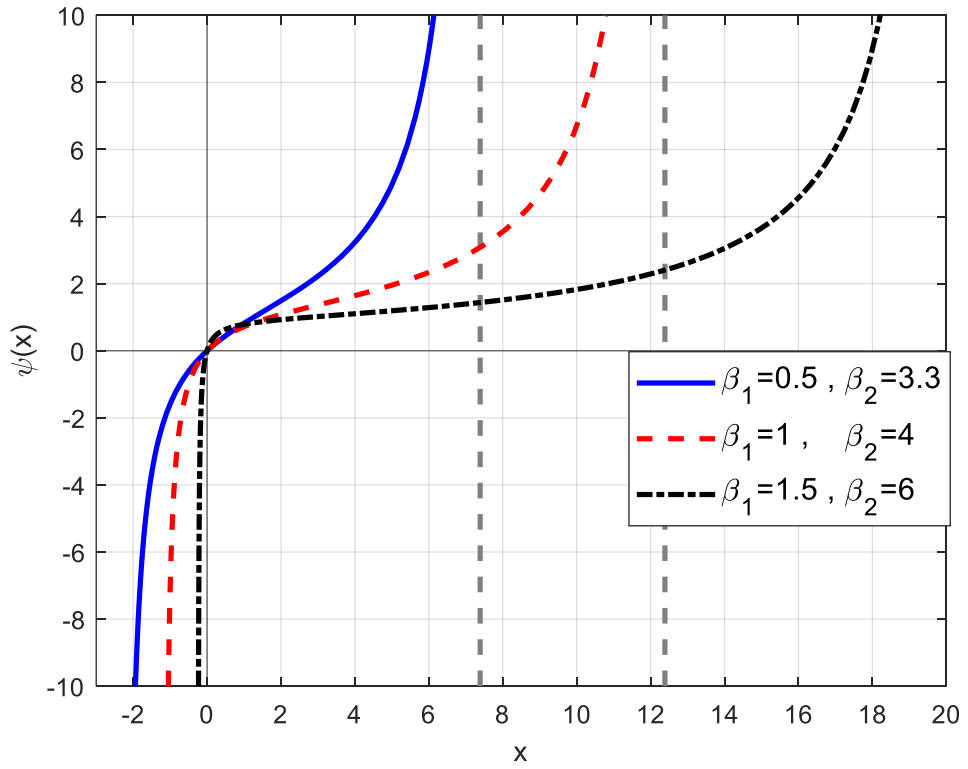
$$\begin{aligned} w(X) &= \frac{\psi(X)}{X} \\ &= \frac{M_2 R_1 - M_1 R_2 - X(M_1 + M_2)}{X(R_1 - X)(X + R_2)} \end{aligned} \quad (3.9)$$

elde edilir. Ancak PI dağılışı için en büyük problem dağılım $r_2 < x < r_1$ aralığında sınırlı olmasıdır. Bu nedenle ağırlık fonksiyonu U- şekillidir (Wiśniewski, 2014:943). PI'e ait Amaç, Etki ve Ağırlık Fonksiyonları Şekil 12, Şekil 13 ve Şekil 14'te verilmiştir.

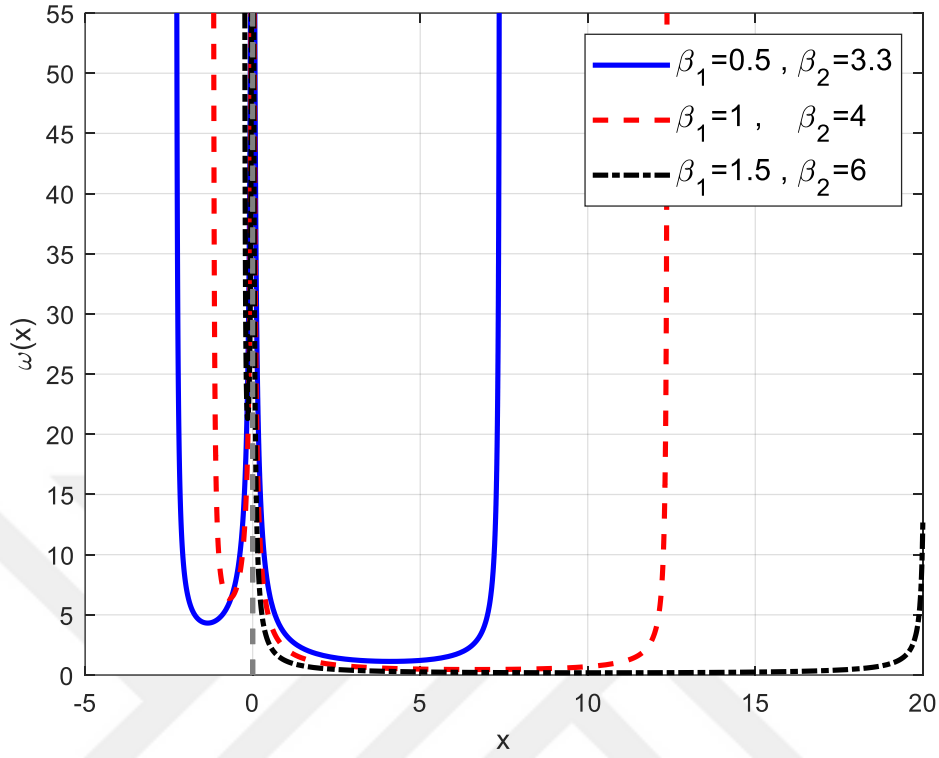
Şekil 12: PI Dağılışına ait Amaç Fonksiyonu



Şekil 13: PI Dağılışına ait Etki Fonksiyonu



Şekil 14: PI Dağılışına ait Ağırlık Fonksiyonu



3.2. PIV DAĞILIŞININ AMAÇ, ETKİ VE AĞIRLIK FONKSİYONU

Modu orijinde olan PIV dağılışının oyf'si ele alındığında sabitler atılırsa Amaç fonksiyonu;

$$\begin{aligned} \rho(X) &= -\ln f(X) \\ &\propto M \ln((X+R)^2 + S^2) + V \arctan \frac{(X+R)}{S} \end{aligned} \quad (3.10)$$

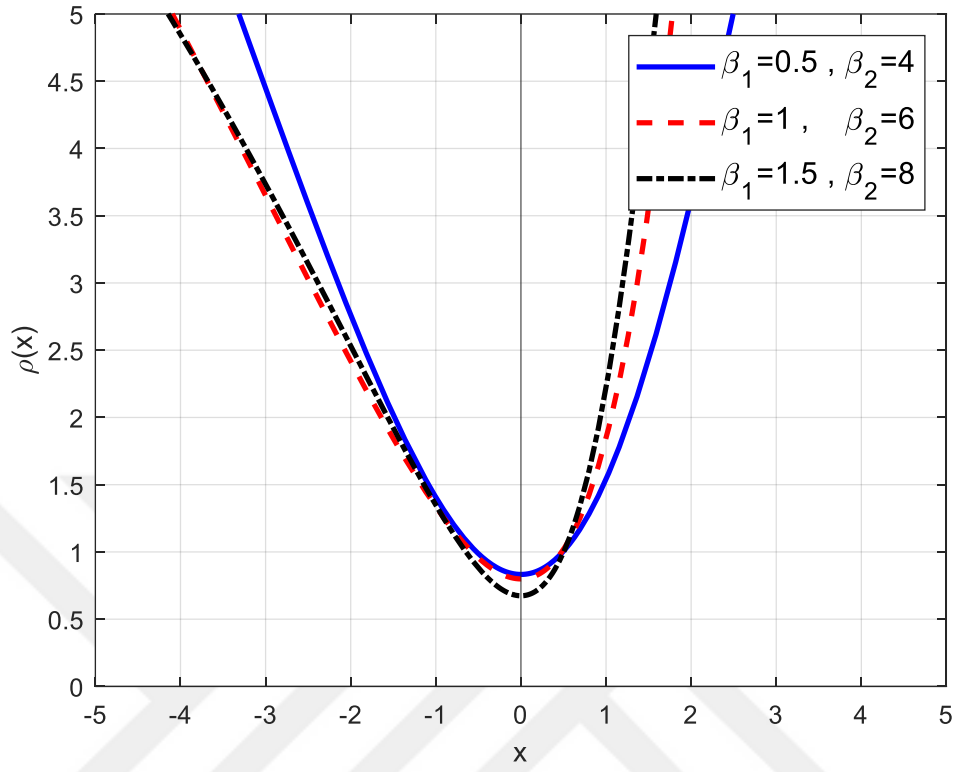
olur. Buradan hareketle Etki ve Ağırlık Fonksiyonunun tanımından;

$$\psi(X) = \frac{2M(X+R) + VS}{(X+R)^2 + S^2} \quad (3.11)$$

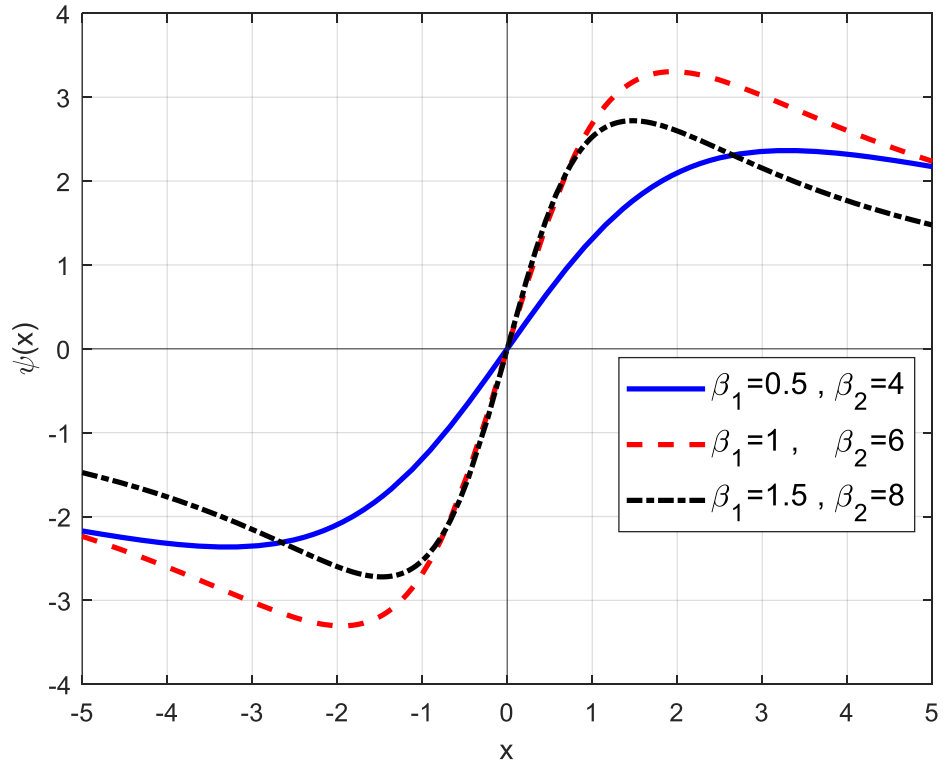
$$w(X) = \frac{2M(X+R) + VS}{X((X+R)^2 + S^2)} \quad (3.12)$$

elde edilir (Wiśniewski, 2014:948). Farklı çarpıklık ve basıklık değerleri için PIV Amaç, Etki ve Ağırlık fonksiyonlarının grafikleri Şekil 15, Şekil 16 ve Şekil 17'de verilmiştir.

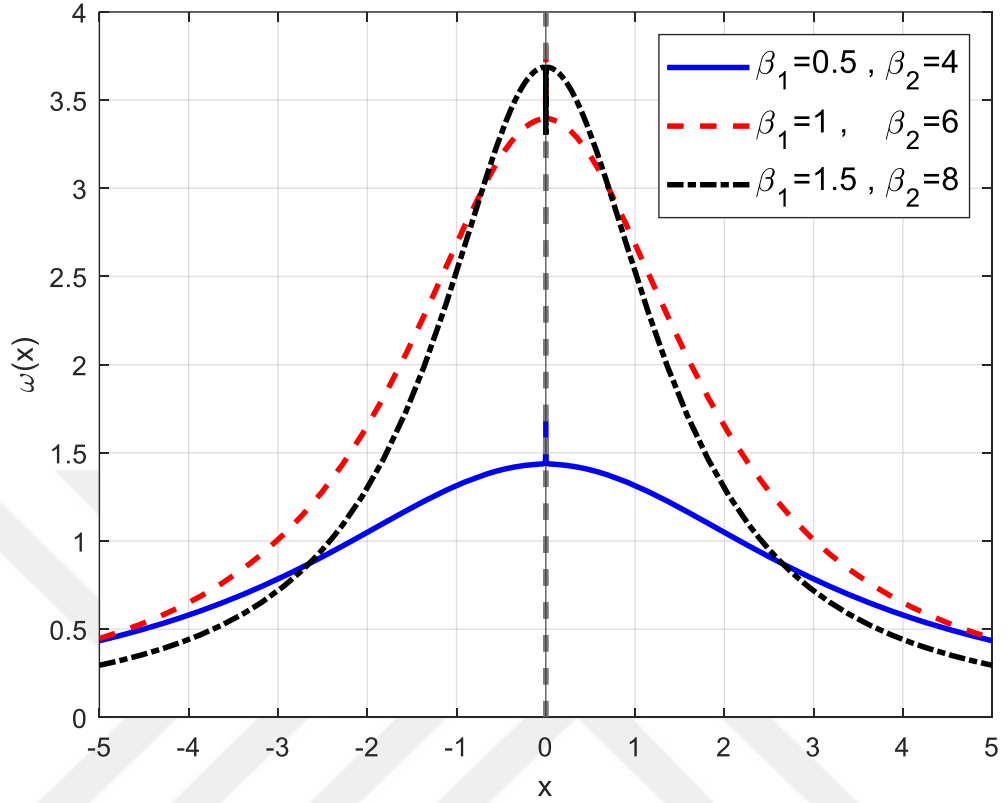
Şekil 15: PIV Dağılışına ait Amaç Fonksiyonu



Şekil 16: PIV Dağılışına ait Etki Fonksiyonu



Şekil 17: PIV Dağılışına ait Ağırlık Fonksiyonu



3.3. PVI DAĞILIŞININ AMAÇ, ETKİ VE AĞIRLIK FONKSİYONU

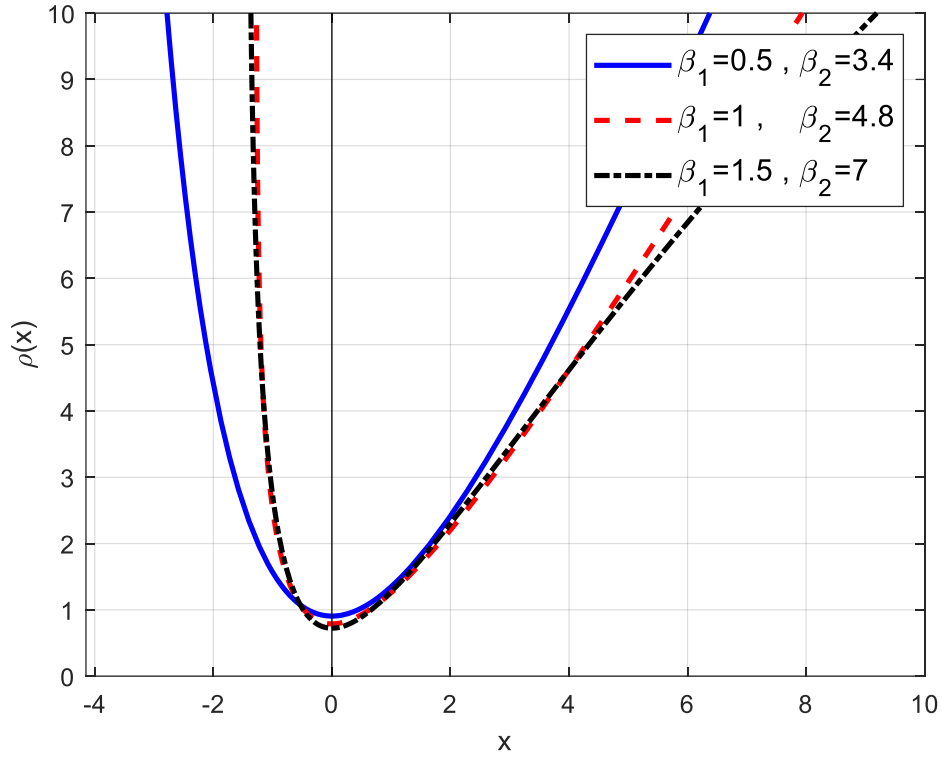
Modu orijinde olan PVI dağılışının oyf'si ele alındığında sabitler atılırsa Amaç fonksiyonu;

$$f(X) \propto (X + R_1)^{M_1} (X + R_2)^{M_2} \quad (3.13)$$

$$\rho(X) = -\ln f(X) = -M_1 \ln(X + R_1) - M_2 \ln(X + R_2) \quad (3.14)$$

elde edilir. PVI amaç fonksiyonu, diğer dayanıklı yöntemlere göre içerisinde barındırdığı yer, ölçek, çarpıklık ve basıklık parametreleri sayesinde minimize edilecek fonksiyonlar açısından esneklik sağlamaktadır. Şekil 18 incelendiğinde farklı çarpıklık ve basıklık değerleri olduğunda PVI Amaç fonksiyonunun aldığı şekiller ortaya çıkmaktadır.

Şekil 18: PVI Dağılışına ait Amaç Fonksiyonu



Bu amaç fonksiyonundan hareketle Etki Fonksiyonu;

$$\psi(X) = \frac{d\rho(X)}{dX} = \frac{d[-\ln f(X)]}{dX} = -\frac{M_1}{X + R_1} - \frac{M_2}{X + R_2} \quad (3.15)$$

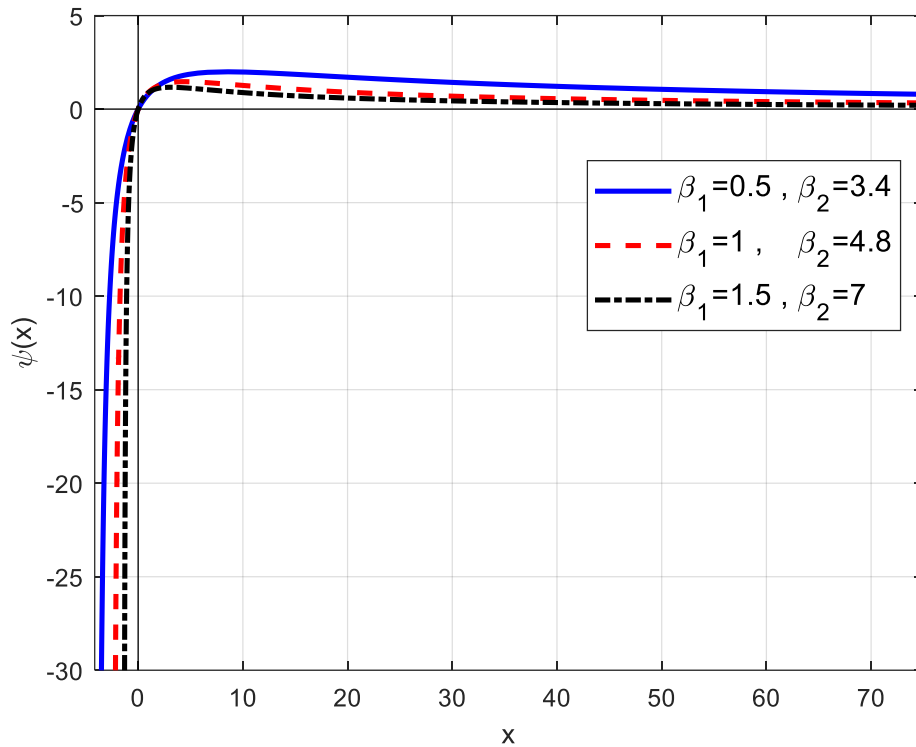
bulunur. PVI dağılışı için Etki fonksiyonu dağılışın aldığı her parametre için değişiklik gösterir ve sonsuz adet oluşabilir. Farklı çarpıklık ve basıklık değerleri için Etki fonksiyonu incelenirse;

- Çarpıklık ($\beta_1 = 0.5$) ve basıklık ($\beta_2 = 3.4$) değerleri diğer durumlara göre düşük olduğunda, veri setinin sol tarafına 0'a yakın bir veri eklendiğinde tahminleyici üzerinde az bir etki yaparken eklenen veri 0'dan uzaklaştıkça tahminleyici üzerindeki etkisi artmaktadır. Ancak benzer durum dağılışın sağ tarafında gerçekleşirse eklenen her veri noktası 0'dan uzaklaştıkça tahminleyici üzerindeki etkisi önce artan daha sonra nispeten sabit kalan bir yapıya sahiptir.
- Çarpıklık ($\beta_1 = 1$) ve basıklık $\beta = 4.8$ değerleri orta dereceli olduğunda, dağılışın sol tarafında benzer bir durum ortaya çıkarken, sağ tarafa eklenen veri noktası tahminleyici üzerinde daha az etki yapmaktadır.
- Çarpıklık ($\beta_1 = 1.5$) ve basıklık ($\beta_2 = 7$) değerleri diğer durumlara göre daha yüksek olduğunda dağılışın sol tarafında eklenen veri noktasının köke yaklaştıkça etkisi artarken,

dağılımın sağ tarafında eklenen veri noktası tahminleyici üzerinde diğer durumlara göre daha az etki yapmaktadır.

Her üç durum karşılaştırılırsa, dağılım çarpıklık ve basıklık değerleri artırıldığında köke yakın bölgelere eklenen veri noktası tahminleyici üzerindeki etkisini artırırken kökten uzaklaşıldığında eklenen veri noktası tahminleyici üzerindeki etkisi azalmaktadır. Köke yakın noktalara eklenen bir veri noktası dağılım yapısını bozmakta (tipini) ve kökü değiştirmektedir. Kök değiştiğinde ise PVI etki fonksiyonu dinamik yapısı nedeniyle yeniden parametre tahmini yaparak Etki ve Ağırlık fonksiyonlarını yeni parametrelere göre düzenlemektedir.

Şekil 19: PVI Dağılımına ait Etki Fonksiyonu



Etki fonksiyonu elde edildikten sonra Ağırlık fonksiyonunun tanımından;

$$w(X) = \frac{\psi(X)}{X} = \frac{M_1 R_2 + M_2 R_1 - (M_1 + M_2) X}{X(X + R_1)(X + R_2)} \quad (3.16)$$

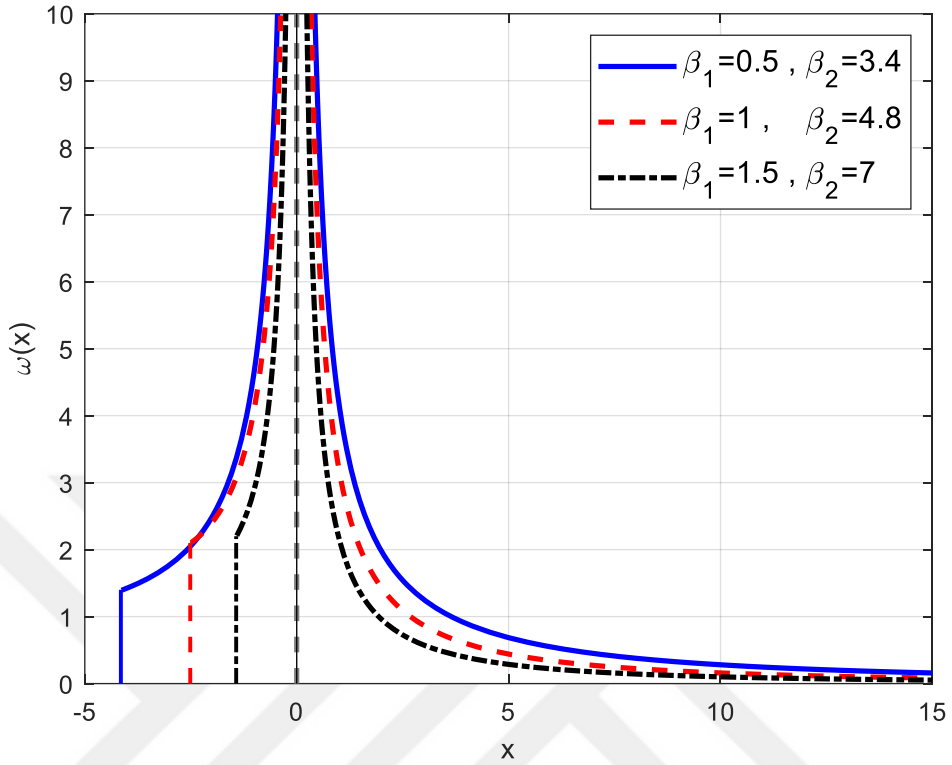
elde edilir. Ağırlık fonksiyonu incelendiğinde, regresyon parametrelerinin tahmini yapılırken kullanılan ağırlık fonksiyonu PVI dağılımının bütün parametrelerini kullanmaktadır. Burada önemli olan veri setinin mod değerinin 0' çekilmesi ve böylece verilerin çarpıklık ve basıklık değerleri ne olursa olsun en çok verinin bulunduğu noktalara en çok ağırlığın verilmesidir. Şekil

19 incelenirse Ağırlık fonksiyonunun farklı çarpıklık ve basıklık değerleri için veri noktalarına verdikleri değerler görülebilir. Bütün durumlar incelenirse;

- Çarpıklık (β_1) ve basıklık (β_2) değerleri diğer durumlara göre düşük olduğunda, dağılımın sol ve sağ tarafında bulunan verilere daha fazla ağırlık verilmiştir. Bunun nedeni ise düşük çarpıklık ve basıklık değerleri ile dağılım nispeten simetriye daha yakın olmasıdır. Özellikle basıklık değeri normal dağılımın basıklık değeri olan 3'e fazlasıyla yakın ve çarpıklık değeri ise diğer durumlara göre daha düşük seviyededir.
- Çarpıklık ve basıklık değerleri orta dereceli olursa, dağılımın merkezinde en yüksek ağırlıklar yer alırken yine sol ve sağ tarafta ağırlıklar giderek azalmaktadır. Burada önemli olan çarpıklık ve basıklık değerleri düşük olan duruma göre dağılımın merkezinde yüksek ağırlık verilen bölge daha dar ve kuyruğa doğru bakıldığında çok daha az ağırlıkların verilmesidir.
- Çarpıklık ve basıklık değerleri yüksek olduğunda ise dağılımın merkezinde ağırlıklar daha dar bir noktada en yüksek olurken, kuyruğa doğru gidildiğinde neredeyse 0' a yaklaşmaktadır.

Bütün durumlar ele alındığında, dağılımın çarpıklığı ve basıklığı (sivriliği) arttıkça mod değeri etrafında PVI Ağırlık fonksiyonu dar bir bölgede yüksek ağırlık vermektedir. Ancak Şekil 20'de kuyruğa doğru gidildiğinde çarpıklık ve basıklık değerleri dikkate alındığından PVI Ağırlık fonksiyonunun verdiği ağırlıklar hızla azalarak 0' a yaklaşmaktadır.

Şekil 20: PVI Dağılışına ait Ağırlık Fonksiyonu



Veri setinin Pearson PVI dağılışı olduğu varsayımı altında regresyon çözümlemesi yapılırken parametrelerin asimptotik davranışları önemlidir. Bu nedenle parametre vektörü $\hat{\theta}$ 'nin asimptotik varyansı elde edilmelidir. Eşitlik 1.70 ve Eşitlik 1.71 dikkate alınarak asimptotik varyans için düzeltme faktörü;

$$\frac{E[\psi(X)^2]}{(E[\psi'(X)])^2} \quad (3.17)$$

ile tanımlanmıştır. PVI dağılışının etki fonksiyonunun karesinin beklenen değeri;

$$\begin{aligned} E[\psi(X)^2] &= E\left[\left(-\frac{M_1}{X+R_1} - \frac{M_2}{X+R_2}\right)^2\right] \\ &= K^* \int_{-R_1}^{\infty} \left(-\frac{M_1}{X+R_1} - \frac{M_2}{X+R_2}\right)^2 (X+R_1)^{M_1} (X+R_2)^{M_2} dX \\ &= K^* \int_{-R_1}^{\infty} \left(\frac{M_1^2}{(X+R_1)^2} + \frac{M_2^2}{(X+R_2)^2} + \frac{2M_1M_2}{(X+R_1)(X+R_2)}\right) (X+R_1)^{M_1} (X+R_2)^{M_2} dX \end{aligned} \quad (3.18)$$

ile gösterilir. Burada $t = \frac{R_2 - R_1}{X + R_1}$ dönüşümü yapılırsa integralin sınırları $0 < t < \infty$ olur.

Dönüşümün türevi alınırsa $dx = -\left[(R_2 - R_1)/t^2\right]dt$ elde edilir. Buradan $X + R_1 = (R_2 - R_1)/t$ ve $X + R_2 = (R_2 - R_1)(1-t)/t$ alınırsa Eşitlik 3.19 içerisindeki ilk terim için;

$$\begin{aligned} & \int_0^\infty M_1^2 \left(\frac{R_2 - R_1}{t} \right)^{M_1} \left[\frac{(R_2 - R_1)(1+t)}{t} \right]^{M_2} \frac{(R_2 - R_1)}{t^2} dt \\ &= M_1^2 (R_2 - R_1)^{M_1+M_2-1} \int_0^\infty \frac{(1+t)^{M_2}}{t^{M_1+M_2}} dt \\ &= \underbrace{M_1^2 (R_2 - R_1)^{M_1+M_2-1} B(1-M_1-M_2, M_1-1)}_I \end{aligned} \quad (3.20)$$

elde edilir. Burada $\int_0^\infty t^{p-1}/(1+t)^{p+q} dt = B(p, q)$ ikinci tip beta fonksiyonudur. Eşitlik 3.19'un ikinci terimi için;

$$\begin{aligned} & \int_0^\infty M_2^2 \left(\frac{R_2 - R_1}{t} \right)^{M_1} \left[\frac{(R_2 - R_1)(1+t)}{t} \right]^{M_2-2} \frac{(R_2 - R_1)}{t^2} dt \\ &= M_2^2 (R_2 - R_1)^{M_1+M_2-1} \int_0^\infty \frac{(1+t)^{M_2-2}}{t^{M_1+M_2}} dt \\ &= \underbrace{M_2^2 (R_2 - R_1)^{M_1+M_2-1} B(1-M_1-M_2, M_1+1)}_{II} \end{aligned} \quad (3.21)$$

olur. Eşitlik 3.19 içerisindeki üçüncü terim için;

$$\begin{aligned} & \int_0^\infty 2M_1M_2 \left(\frac{R_2 - R_1}{t} \right)^{M_1-1} \left[\frac{(R_2 - R_1)(1+t)}{t} \right]^{M_2-1} \frac{R_2 - R_1}{t^2} dt \\ &= 2M_1M_2 (R_2 - R_1)^{M_1+M_2-1} \int_0^\infty \frac{(1+t)^{M_2-1}}{t^{M_1+M_2}} dt \\ &= \underbrace{2M_1M_2 (R_2 - R_1)^{M_1+M_2-1} B(1-M_1-M_2, M_1)}_{III} \end{aligned} \quad (3.22)$$

elde edilir. Bu üç ifade toplanırsa $(I + II + III)$ ve $D = (R_2 - R_1)^{M_1+M_2-1}$ olmak üzere integral sabiti eklenirse;

$$\begin{aligned}
& E\left[\psi(X)^2\right] \\
& = K^* DM_1^2 B(1-M_1-M_2, M_1-1) + \dots \\
& \quad K^* DM_2^2 B(1-M_1-M_2, M_1+1) + \dots \\
& \quad 2K^* DM_1 M_2 B(1-M_1-M_2, M_1)
\end{aligned} \tag{3.23}$$

elde edilir. PVI Etki fonksiyonunun türevinin beklenen değeri elde edilirken öncelikle türev alınır;

$$\frac{d\psi(X)}{dX} = \frac{d}{dX} \left(-\frac{M_1}{X+R_1} - \frac{M_2}{X+R_2} \right) \tag{3.24}$$

$$\psi'(X) = \frac{M_1}{(X+R_1)^2} + \frac{M_2}{(X+R_2)^2} \tag{3.25}$$

elde edilir. Bu eşitliğin beklenen değeri;

$$E[\psi'(X)] = K^* \int_{-R_1}^{\infty} \left(\frac{M_1}{(X+R_1)^2} + \frac{M_2}{(X+R_2)^2} \right) (X+R_1)^{M_1} (X+R_2)^{M_2} dX \tag{3.26}$$

ile gösterilir. Burada $t = \frac{R_2-R_1}{X+R_1}$ dönüşümü yapılırsa integralin sınırları $0 < t < \infty$ olur. Eşitlik

3.26 içerisindeki toplamlar üzerinden ayrı ayrı integral alınır birinci ifade için;

$$\begin{aligned}
& \int_{-R_1}^{\infty} M_1 (X+R_1)^{M_1-2} (X+R_2)^{M_2} dX \\
& = \int_0^{\infty} M_1 \frac{(R_2-R_1)^{M_1-2}}{t^{M_1-2}} \frac{(R_2-R_1)^{M_2}}{t^{M_2}} \frac{(1+t)^{M_2}}{t^2} \frac{(R_2-R_1)}{t^2} dt \\
& = \underbrace{DM_1 B(M_1-1, 1-M_1-M_2)}_{I^*}
\end{aligned} \tag{3.27}$$

elde edilir. Eşitlik 3.26 içerisindeki toplamın ikinci terimi için integral alınır;

$$\begin{aligned}
& \int_{-R_1}^{\infty} M_2 (X+R_1)^{M_1} (X+R_2)^{M_2-2} dX \\
& = \int_0^{\infty} M_2 \frac{(R_2-R_1)^{M_1}}{t^{M_1}} \frac{(R_2-R_1)^{M_2-2}}{t^{M_2-2}} \frac{(1+t)^{M_2-2}}{t^2} \frac{(R_2-R_1)}{t^2} dt \\
& = \underbrace{DM_2 B(1-M_1-M_2, M_1+1)}_{II^*}
\end{aligned} \tag{3.28}$$

olur. Bu iki ifade toplanır ($I^* + II^*$) ve integral sabiti eklenirse;

$$\begin{aligned}
& E[\psi'(X)] \\
& = K^*DM_1B(M_1-1, 1-M_1-M_2) + K^*DM_2B(1-M_1-M_2, M_1+1)
\end{aligned} \tag{3.29}$$

elde edilir. Böylece $\hat{\theta}$ 'in asimptotik dağılışı;

$$\hat{\theta} \sim N(\theta_0, V(\theta_0)) \tag{3.30}$$

ve burada

$$V(\hat{\theta}_0) = \sigma^2 \frac{E[\psi(X)^2]}{(E[\psi'(X)])^2} (\mathbf{X}^T \mathbf{X})^{-1} \tag{3.31}$$

olur. Uygulama bölümünde regresyon parametre tahminleri için varyans- kovaryans matrisi bulunurken Eşitlik 1.74 kullanılmıştır.

DÖRDÜNCÜ BÖLÜM

UYGULAMA

Dayanıklı regresyon uygulaması iki aşamada gerçekleştirilmiştir. Birinci aşamada Dayanıklı Regresyon literatüründe sıklıkla kullanılan Amerikan eyaletlerindeki Eğitim Harcamaları (Education Expenditure) veri seti ile Ocak 2015 – Şubat 2020 arasında Türkiye’de gerçekleşen Sanayi Üretim Endeksi, İşsizlik Oranı ve Tüfe değerleri kullanılmıştır. Uygulamanın ikinci aşamasında ise simülasyon çalışmaları yapılmıştır.

4.1 UYGULAMADA KULLANILAN GERÇEK VERİ SETLERİ

Önerilen PVI fonksiyonunun ağırlık fonksiyonuna dayanan tahminleme yöntemi için iki farklı veri seti analiz edilmiştir. Bu veri setlerinden ilki literatürde sıklıkla kullanılan Eğitim Harcamaları (Education Expenditure) veri seti ve ikincisi Türkiye’de gerçekleşen Sanayi Üretim Endeksi, İşsizlik Oranı ve Tüfe değerleri (İktisadi Veri Seti) kullanılarak kurulan regresyon modelleridir. Her bir veri seti için PVI fonksiyonuna uyum kontrol edilmiş ve elde edilen regresyon parametrelerinin standart hataları verilmiştir. Veri setleri için artıkların kutu grafiği ve histogramı EKK’dan elde edilen artıklar için çizilmiştir. Ancak Huber, Tukey ve PVI yöntemlerinden elde edilen kutu grafikleri verilmemiştir.

4.1.1. Eğitim Harcamaları (Education Expenditure) Veri Seti

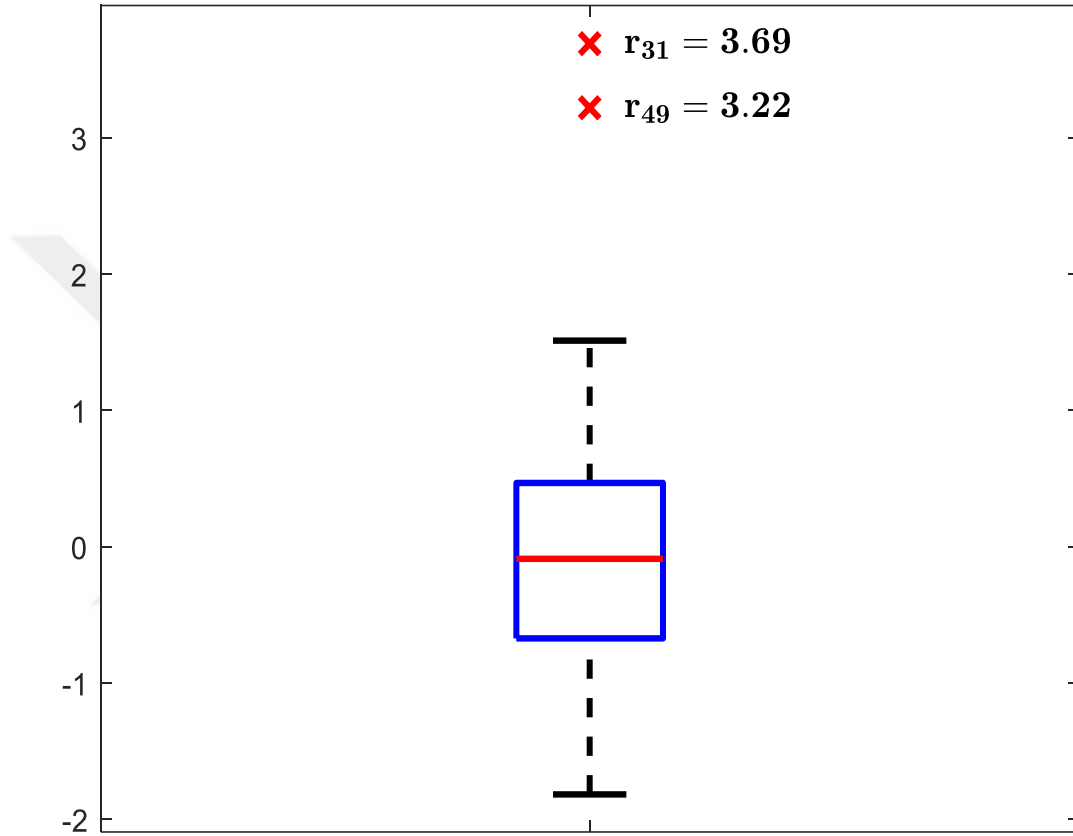
İlk defa Chatterjee ve Price (1977:108) tarafından yayınlanan veri seti Amerika’da bulunan 50 eyalette eğitim harcamalarını ele almıştır. Veri setine ait özet bilgiler;

- 1975 yılında Bir eyaletteki devlet okulunda kişi başına düşen ortalama eğitim harcaması (açıklanan değişken),
- 1970 yılında kırsal alanda ikamet eden kişi sayısı ($\times 1000$),
- 1973 yılındaki kişi başına düşen milli gelir,
- 1974’te 18 yaşın altında yaşayan birey bulunan hane sayısı ($\times 1000$).

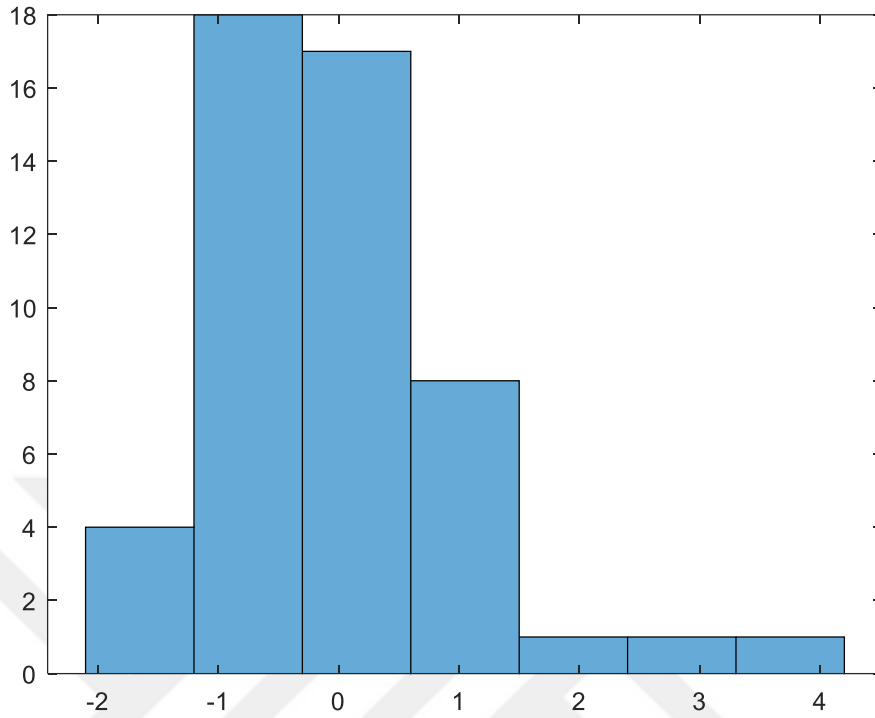
Veri seti içerisinde açıklayıcı değişkenin 31. verisinin değeri 212 iken yazım veya aktarma hatası nedeniyle 400 değeriyle değiştirilmiştir. Böylece gerçek yaşamda karşılanabilecek problem ele alınmıştır. Bu aşamadan sonra Pearson parametreleri tahmin edilmiş ve regresyon analizi

gerçekleştirilmiştir. Şekil 21 incelendiğinde 31. ve 49. verinin yüksek standardize artıklara (3.69 ve 3.22) sahip olduğu görünmektedir. Artıklara uygulanan Anderson-Darling test istatistiği sonucunda 1.046 ($p=0.000$) elde edilmiş ve verilerin dağılışının normal olmadığı görülmüştür. Ayrıca veri setinin çarpıklık değeri 1.10 ve basıklık değeri 5.30 olarak elde edilmiştir.

Şekil 21: Eğitim Harcamaları Veri Seti için Kutu Grafiği



Şekil 22: Eğitim Harcamaları Veri Seti için Histogram



Regresyon parametre tahminleri Tablo 4’te verilmiştir.

Tablo 4: Eğitim Giderleri Verisi Regresyon Analizi Sonuçları

Tahminleme Yöntemi	Sonuçlar	Değişkenler			
		Sabit Terim	Kırsal Alanda İkamet Eden Kişi Sayısı(x1000)	Kişi Başına Düşen Milli Gelir	18 Yaş Altında Yaşayan Hane Sayısı (x1000)
EKK	Tahminler	-452.203	0.001	0.063	1.346
	Std. Hata	146.891	0.061	0.014	0.375
Huber M-Tahminleyici	Tahminler	-340.860	0.044	0.053	1.067
	Std. Hata	126.629	0.053	0.012	0.323
Tukey's Bisquare M-Tahminleyici	Tahminler	-243.762	0.074	0.045	0.820
	Std. Hata	127.490	0.053	0.012	0.326
PVI	Tahminler	-307.984	0.009	0.061	0.909
	Std. Hata	72.326	0.030	0.007	0.185

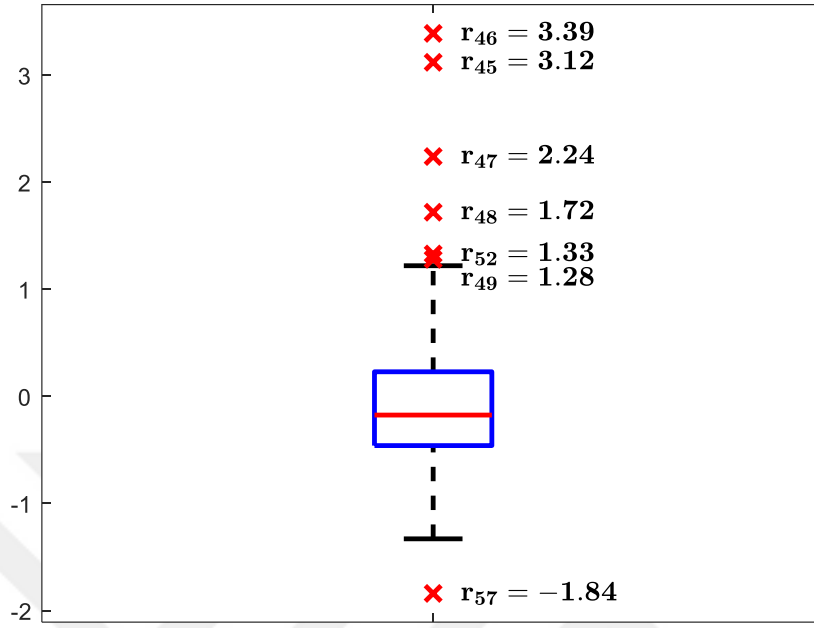
Tablo 4 incelenir ve veri seti göz önüne alınırsa, yeni geliştirilen PVI yönteminin yüksek çarpıklık ve basıklık değerlerinde bile diğer tahminleme yöntemleriyle benzer sonuçlar

vermektedir. Huber ve Tukey M- Tahminleme yöntemleri artıklar büyüdükçe çok düşük veya 0 ağırlık değerlerini vermesine rağmen PVI yöntemi yüksek standardize artık değerlerine sahip olan artıkları veri setinin bir parçası gibi analiz etmiş ve artıkların tamamına ağırlık vermiştir. Böylece regresyon tanıları yapılırken sapan gözlem gibi görünen ama veri setine ait olduğu bilinen bir gözlem ortaya çıktığında ve çarpıklık ve basıklık değerleri yüksek olduğunda eğer uyum gösteriyorsa PVI yöntemi en az diğer tahminleme yöntemleri kadar iyi sonuçlar vermektedir.

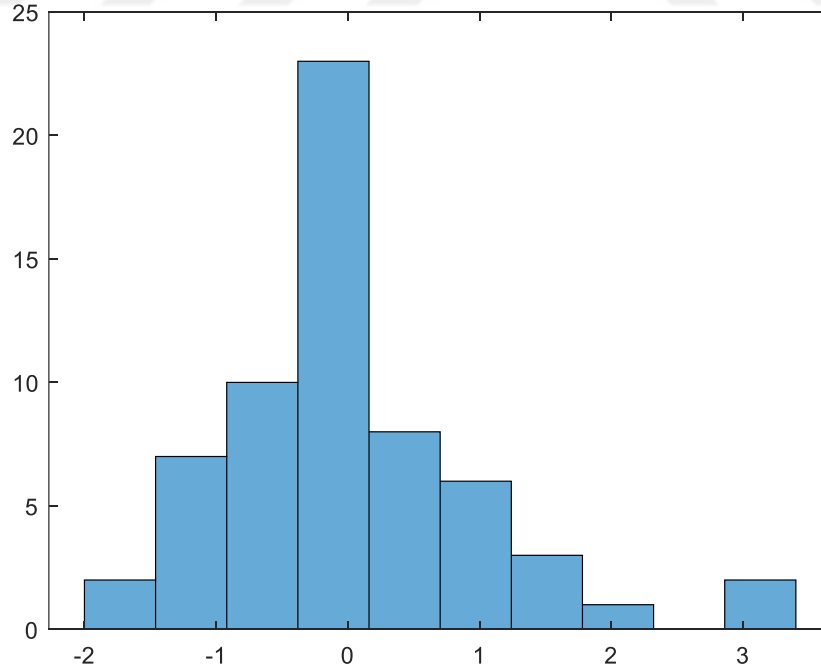
4.1.2 İktisadi Veri Seti

Ocak 2015 ile Şubat 2020 tarihleri arasında 62 aylık Sanayi Üretim Endeksi (*SÜE*, 2015=100), İşsizlik Oranı (*İO*) ve Tüfe (2003=100) değerleri ele alınmıştır. Yapılan regresyon analizinde Tüfe değerleri açıklanan değişken olarak kullanılmıştır. Veri seti üzerinde herhangi bir değişiklik yapılmamış ve değerler olduğu gibi bırakılmıştır. Bu örnekte, Türkiye’de iktisadi analizlerde sıklıkla kullanılan değişkenler ele alınarak verilerin dağılımlarını tam olarak belirlemenin önemi üzerinde durulmuştur. Şekil 23 incelenirse 45, 46, 47, 48, 49, 52 ve 57. gözlemlerin yüksek standardize artıklara sahip olduğu görülmektedir. Buna ek olarak artıkların histogramı incelenirse verilerin sağa çarpık ($\beta_1 = 1.08$) ve yüksek basıklık ($\beta_2 = 5.04$) değerlerine sahip olduğu görülmektedir. Ayrıca yapılan Anderson-Darling testi sonucunda test istatistiği 1.782 ($p=0.005$) elde edilmiş ve artıkların dağılışının normal olmadığı görülmüştür.

Şekil 23: İktisadi Veri Seti için Kutu Grafiği



Şekil 24: İktisadi Veri Seti için Histogram



Tablo 5'deki regresyon analizi sonuçları incelenirse PVI yönteminin diğer yöntemlerle çok yakın sonuçlar verdiği görülebilir. Verilerdeki çarpıklık ve basıklık değerleri yüksek olması ve

yüksek standardize artıklar olmasına rağmen katsayılar anlamlı elde edilmiştir. Tukey M- Tahminleme yöntemi artıkların neredeyse %10'una 0 ağırlığını vermiş ve Huber M- Tahminleme yöntemi ise düşük ağırlıklar vermiştir. PVI, ağırlık fonksiyonunun yapısı nedeniyle bütün artıklara ağırlık vermiş ve buna rağmen diğer yöntemlerle benzer sonuçlar elde etmiştir.

Tablo 5: İktisadi Veriler için Regresyon Analizi Sonuçları

<i>Tahminleme Yöntemi</i>	<i>Sonuçlar</i>	<i>Değişkenler</i>		
		<i>Sabit Terim</i>	<i>İşsizlik Oranı (İÖ)</i>	<i>Sanayi Üretim Endeksi (SÜE, 2015=100)</i>
EKK	<i>Tahminler</i>	-29.615	0.256	1.143
	<i>Std. Hata</i>	8.052	0.077	0.338
Huber M-Tahminleyici	<i>Tahminler</i>	-28.035	0.250	1.012
	<i>Std. Hata</i>	6.716	0.064	0.282
Tukey's Bisquare M-Tahminleyici	<i>Tahminler</i>	-17.680	0.236	0.644
	<i>Std. Hata</i>	4.722	0.045	0.128
PVI	<i>Tahminler</i>	-25.315	0.247	0.781
	<i>Std. Hata</i>	8.383	0.080	0.352

4.2. UYGULAMADA KULLANILAN SİMÜLASYON SENARYOLARI

Simülasyon çalışmasında farklı senaryolara sahip veri setleri kullanılmıştır. Hata teriminin dağılımının önceden belirlenerek ve açıklanan değişken (y) değerleri elde edilerek regresyon modeli kurulmuştur. Çalışmada regresyon parametreleri $\theta_0 = \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = 2$ olarak belirlenmiş ve açıklayıcı değişken değerleri ortalaması 0, varyansı 1 olan normal dağılıştan türetilmiştir. Böylece kurulan regresyon modeli

$$y = 2 + 2X_1 + 2X_2 + 2X_3 + 2X_4 + 2X_5 + \varepsilon, \quad X_i \sim N(0,1) \quad (4.1)$$

şeklinde dir. Hata terimine ait toplam 28 adet senaryo hazırlanmış ve regresyon modelinin çözümünden elde edilen artıkların PVI dağılımına uyumları kontrol edilmiştir. Böylece çözüm için kullanılacak Ağırlık fonksiyonunun parametreleri dinamik olarak (m_1, m_2, r_1, r_2) belirlenmiştir. Senaryolar seçilirken hem simetrik hem de asimetrik dağılışlardan yararlanılmış ve geliştirilen PVI dağılışının Ağırlık fonksiyonuna dayanan tahminleme yöntemi EKK, Huber M ve Tukey M- Tahminleme yöntemleriyle karşılaştırılmıştır. Uygulama toplam 3 aşamadan oluşmaktadır. Bu aşamalar;

Aşama 1: Regresyon artıklarının dağılışının belirlenmesi

Önceden belirlenmiş açıklayıcı değişken (X) değerleri, regresyon parametre (θ) değerleri ve hata vektörü (ε) değerleri Eşitlik 4.1’de yerine konularak açıklanan değişken (y) değerleri elde edilmiştir. Elde edilen y ve X değerleri kullanılarak regresyon parametrelerinin tahminleri elde edilmiştir.

$$\hat{\theta} = (X^T X)^{-1} X^T y.$$

Elde edilen regresyon parametre tahminleriyle regresyon artıkları bulunmuştur.

$$e = y - X\hat{\theta}$$

Artıklar bulunduktan sonra uygun bir ölçek tahmini kullanılarak standardize artıklar elde edilmiştir. Standardize artıklar elde edilirken kullanılan dönüm noktası (k) sadece Huber ve Tukey yöntemleri için kullanılmıştır. Böylece Huber ve Tukey yöntemleri standardize artıkların büyümesini ve ağırlıklarının azalmasını garantilemektedir.

$$r_i = e_i / (ks\sqrt{1-h_i})$$

Standardize artıkların ilk dört momenti hesaplanmış ve bu momentler kullanılarak Çarpıklık (β_1) ve Basıklık (β_2) değerleri hesaplanmıştır. Elde edilen β_1 ve β_2 değerleri kullanılarak diferansiyel denklemin parametreleri a, c_0, c_1, c_2 hesaplanmıştır.

$$a = \frac{-\sigma\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18}$$

$$c_0 = \frac{\sigma^2(4\beta_2 + 3\beta_1)}{10\beta_2 - 12\beta_1 - 18}$$

$$c_1 = \frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18}$$

$$c_2 = \frac{(2\beta_2 - 3\beta_1 - 6)}{10\beta_2 - 12\beta_1 - 18}$$

Diferansiyel denklemin parametreleri bulunduktan sonra bu denklemin diskriminantı parametreler kullanılarak hesaplanmıştır.

$$\Delta = c_1^2 - 4c_0c_2$$

Elde edilen diskriminant değeri ile denklemin kökleri hesaplanmıştır.

$$r_1 = \frac{-c_1 + \sqrt{\Delta}}{2c_2}$$

$$r_2 = \frac{-c_1 - \sqrt{\Delta}}{2c_2}$$

Kökler kullanılarak PVI dağılışı için üs parametreleri elde edilmiştir.

$$\frac{A_1}{c_2} = \frac{-r_1 - a}{c_2(r_2 - r_1)} = m_1$$

$$\frac{A_2}{c_2} = \frac{r_2 + a}{c_2(r_2 - r_1)} = m_2$$

Sağa çarpık PVI dağılışına ait bütün parametreler elde edildikten sonra verilerin gerçekten dağılışa uyum gösterip göstermediğinin belirlenmesi amacıyla aşağıda bulunan koşullar kontrol edilmiştir.

$$r_1 < 0,$$

$$r_2 < 0,$$

$$r_1 < r_2,$$

$$c_0 < 0$$

$$c_1 < 0,$$

$$c_2 < 0,$$

$$m_1 < 0,$$

$$m_2 > 0,$$

$$\sqrt{\delta} < |\gamma_1|,$$

$$\frac{4\delta(2+\delta)}{2+3\delta} < \beta_1$$

ve $r_1 - r_2 < 0$ koşulları sağlandıktan sonra son koşul olarak PVI için integral sabiti hesaplanmış ve integralin tanım aralığında 1 olup olmadığı kontrol edilmiştir.

$$K \int_{r_2}^{\infty} (x + r_1)^{m_1} (x + r_2)^{m_2} = \int_{r_2}^{\infty} f(x) = 1$$

Yukarıdaki koşulları sağlayan artıklara karşılık gelen açıklanan değişken y değerleri ve açıklayıcı değişken X matrisleri kaydedilmiş ve artık bu veri setlerinin dağılışının PVI dağılışına da uyum gösterdikleri belirlenmiştir. Birinci aşama geçildikten sonra elde edilen y ve X değerleri kullanılarak ikinci aşama olan PVI dağılışının parametreleri tahminlenmiştir.

Aşama 2: PVI parametrelerinin tahminlenmesi ve mod noktasının 0'a çekilmesi

Aşama 1' den elde edilen 1000 adet y ve X değerleri kullanılarak Eşitlik 4.1 yeniden EKK ile tahminlenmiş ve artıklar elde edilmiştir. Bu aşamada Pearson Diferansiyel Denkleminin parametreleri (a, c_0, c_1, c_2) ve PVI dağılışının parametreleri (m_1, m_2, r_1, r_2) Aşama 1'de olduğu gibi hesaplanmıştır. Daha sonra elde edilen mod değeri a kullanılarak artıkların modunun 0 yapılabilmesi için artıkların moddan farkları alınarak yeni artıklar elde edilmiştir.

$$r_a = r - a$$

Elde edilen yeni artıkların ilk dört momenti hesaplanmış ve bu momentler kullanılarak yeni Çarpıklık (β_1) ve Basıklık (β_2) değerleri hesaplanmıştır. Mod noktası 0' da olan dağılışın parametrelerinin hesaplanması için aşağıdaki eşitlikler kullanılmıştır.

$$c_2 = C_2$$

$$2ac_2 + c_1 = a(2c_2 + 1) = C_1$$

$$c_2a^2 + c_1a + c_0 = c_0 + a^2(1 + c_2) = C_0$$

Yeni Diferansiyel Denklem Parametreleri kullanılarak;

$$\Delta = C_1^2 - 4C_0C_2$$

$$R_1 = \frac{-C_1 + \sqrt{\Delta}}{2C_2}$$

$$R_2 = \frac{-C_1 - \sqrt{\Delta}}{2C_2}$$

$$\frac{A_1}{C_2} = \frac{-R_1}{C_2(R_2 - R_1)} = M_1$$

$$\frac{A_2}{C_2} = \frac{R_2}{C_2(R_2 - R_1)} = M_2$$

değerleri elde edilerek PVI dağılışının ağırlık fonksiyonunda kullanılacak parametreler kaydedilmiştir.

Aşama 3: EKK, Huber, Tukey ve PVI Ağırlık fonksiyonları kullanılarak regresyon parametrelerinin tahminlenmesi

Bu aşamada önceki aşamalarda elde edilen y ve X değerleri kullanılarak her bir çözüm yöntemi ile parametre tahminleri gerçekleştirilmiştir. Öncelikle EKK yöntemi ele alınırsa her bir

simülasyon için regresyon parametreleri ve standardize artıklar elde edilmiştir. Regresyon parametreleri için TMS ve OKH değerleri hesaplanmıştır. Daha sonra EKK'dan elde edilen standardize artıklar kullanılarak Huber M ve Tukey M yöntemleri için ilk ağırlık değerleri hesaplanmıştır. Bu ağırlık değerleri ile her bir simüle edilmiş veri için IRWLS yöntemi kullanılmış ve regresyon parametreleri tahmin edilmiştir. PVI yöntemi için Aşama 2'de elde edilen mod noktası 0 olan standardize artıklar kullanılarak yine IRWLS yöntemiyle regresyon parametreleri tahmin edilmiştir. Huber, Tukey ve PVI yöntemleri için TMS ve OKH değerleri elde edilmiş ve birbirleriyle karşılaştırılmıştır. IRWLS yönteminin algoritması şu şekildedir;

1. Uygun bir tahminleme yöntemi (genellikle EKK) kullanılarak parametre tahmini yapılır.

$$\hat{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

2. Tahminlenen regresyon parametreleri kullanılarak artıklar elde edilir.

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\theta}$$

3. Artıklar kullanılarak uygun bir ölçek tahmini yapılır (genellikle MAD).

$$s = \frac{MAD}{0.6745} = \frac{\text{medyan}|e_i - \text{medyan}(e_i)|}{0.6745}$$

4. Ölçek tahmini kullanılarak artıklar standardize edilir. Standardize yapılırken eğer kullanılan yöntemin dönüş sabiti (tuning constant) varsa standart sapma ile çarpılır.

$$r_i = e_i / (ks\sqrt{(1-h_i)})$$

5. Huber, Tukey veya PVI Ağırlık fonksiyonlarını kullanarak standardize artıklar için ağırlıklar hesaplanır.

$$w_{Huber}(r) = \begin{cases} 1 & , |r| < k \\ \frac{k}{|r|} & , |r| \geq k \end{cases}$$

$$w_{Tukey}(r) = \begin{cases} \left[1 - \left(\frac{r}{k}\right)^2\right]^2 & , |r| < k \\ 0 & , |r| \geq k \end{cases}$$

$$w_{PVI}(r_a) = \frac{M_1 R_2 + M_2 R_1 - (M_1 + M_2) r_a}{r_a (r_a + R_1) (r_a + R_2)}$$

6. Elde edilen ağırlıklar kullanılarak Ağırlıklı EKK yöntemi ile yeni parametreler tahminlenir.

$$\hat{\theta} = (X'WX)^{-1} X'Wy$$

7. Adım 2-7 arası aşağıdaki yakınsama kriteri sağlanana kadar devam edilir.

$$\frac{\|\hat{\theta}^i - \hat{\theta}^{i-1}\|}{\|\hat{\theta}^i\|} < 10^{-6} \text{ veya iterasyon sayısı } i > 30 \text{ olunca döngü sonlandırılır (Fox ve}$$

Weisberg, 2002:3; Maronna ve diğerleri, 2006:105; Hogg, 1979:111).

Kullanılan yöntemler karşılaştırılırken regresyon parametreleri için TMS ve OKH kriterleri hesaplanmıştır. TMS hesaplanırken;

$$\text{Toplam Mutlak Sapma} = \sum_{j=0}^5 \left| \frac{1}{1000} \sum_{i=1}^{1000} \hat{\theta}_{ij} - \theta_{ij} \right|$$

ve OKH hesaplanırken;

$$\text{Ortalama Karesel Hata} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^T (\hat{\theta}_i - \theta)$$

formülleri kullanılmıştır (Wiśniewski, 2014:951).

Önceden belirlenen simülasyon senaryoları;

- **Senaryo 1.** $\varepsilon \sim N(0,1)$, hata teriminin ortalaması 0 ve varyansı 1 olan normal dağılışa sahip olması (PXIII),
- **Senaryo 2.** $\varepsilon \sim t(1)$, hata teriminin serbestlik derecesi 1 olan t dağılışına sahip olması (PVII),
- **Senaryo 3.** $\varepsilon \sim t(5)$, hata teriminin serbestlik derecesi 5 olan t dağılışına sahip olması (PVII),
- **Senaryo 4.** $\varepsilon \sim t(10)$, hata teriminin serbestlik derecesi 10 olan t dağılışına sahip olması (PVII),
- **Senaryo 5.** $\varepsilon \sim t(20)$, hata teriminin serbestlik derecesi 20 olan t dağılışına sahip olması (PVII),
- **Senaryo 6.** $\varepsilon \sim \text{Cauchy}(0,1)$, hata teriminin konum parametresi 0 ve ölçek parametresi 1 olan Cauchy dağılışına sahip olması (PVII),
- **Senaryo 7.** $\varepsilon \sim \text{Beta}(2,3)$, hata teriminin şekil parametreleri (α, β) 2 ve 3 olan Beta dağılışına sahip olması (PI),

- **Senaryo 8.** $\varepsilon \sim \text{Beta}(2, 5)$, hata teriminin şekil parametreleri (α, β) 2 ve 5 olan Beta dağılışına sahip olması (PI),
- **Senaryo 9.** $\varepsilon \sim \text{Üstel}(1)$, hata teriminin ortalaması (β) 1 olan Üstel dağılışına sahip olması (PX),
- **Senaryo 10.** $\varepsilon \sim \text{Üstel}(2)$, hata teriminin ortalaması (β) 2 olan Üstel dağılışına sahip olması (PX),
- **Senaryo 11.** $\varepsilon \sim \text{Üstel}(10)$, hata teriminin ortalaması (β) 10 olan Üstel dağılışına sahip olması (PX),
- **Senaryo 12.** $\varepsilon \sim \text{Gama}(1, 1)$, hata teriminin şekil parametresi (α) 1 ve ölçek parametresi (β) 1 olan Gama dağılışına sahip olması (PIII),
- **Senaryo 13.** $\varepsilon \sim \text{Gama}(1, 2)$, hata teriminin şekil parametresi (α) 1 ve ölçek parametresi (β) 2 olan Gama dağılışına sahip olması (PIII),
- **Senaryo 14.** $\varepsilon \sim \text{Gama}(1, 5)$, hata teriminin şekil parametresi (α) 1 ve ölçek parametresi (β) 5 olan Gama dağılışına sahip olması (PIII),
- **Senaryo 15.** $\varepsilon \sim \text{Gama}(2, 1)$, hata teriminin şekil parametresi (α) 2 ve ölçek parametresi (β) 1 olan Gama dağılışına sahip olması (PIII),
- **Senaryo 16.** $\varepsilon \sim \text{Gama}(2, 2)$, hata teriminin şekil parametresi (α) 2 ve ölçek parametresi (β) 2 olan Gama dağılışına sahip olması (PIII),
- **Senaryo 17.** $\varepsilon \sim \text{Gama}(2, 5)$, hata teriminin şekil parametresi (α) 2 ve ölçek parametresi (β) 5 olan Gama dağılışına sahip olması (PIII),
- **Senaryo 18.** $\varepsilon \sim \chi^2(0.5)$, hata teriminin serbestlik derecesi (ν) 0.5 olan Ki-Kare dağılışına sahip olması (PIII),
- **Senaryo 19.** $\varepsilon \sim \chi^2(1)$, hata teriminin serbestlik derecesi (ν) 1 olan Ki-Kare dağılışına sahip olması (PIII),
- **Senaryo 20.** $\varepsilon \sim \chi^2(5)$, hata teriminin serbestlik derecesi (ν) 5 olan Ki-Kare dağılışına sahip olması (PIII),
- **Senaryo 21.** $\varepsilon \sim F(1, 2)$, hata teriminin serbestlik dereceleri (ν_1, ν_2) 1 ve 2 olan F dağılışına sahip olması (PVI),

- **Senaryo 22.** $\varepsilon \sim F(2,10)$, hata teriminin serbestlik dereceleri (v_1, v_2) 2 ve 10 olan F dağılışına sahip olması (PVI),
- **Senaryo 23.** $\varepsilon \sim F(5,10)$, hata teriminin serbestlik dereceleri (v_1, v_2) 5 ve 10 olan F dağılışına sahip olması (PVI),
- **Senaryo 24.** $\varepsilon \sim F(10,10)$, hata teriminin serbestlik dereceleri (v_1, v_2) 10 olan F dağılışına sahip olması (PVI),
- **Senaryo 25.** $\varepsilon \sim Weibull(1,1)$, hata teriminin şekil parametresi (α) 1 ve ölçek parametresi (β) 1 olan Weibull dağılışına sahip olması,
- **Senaryo 26.** $\varepsilon \sim Weibull(2,1)$, hata teriminin şekil parametresi (α) 2 ve ölçek parametresi (β) 1 olan Weibull dağılışına sahip olması,
- **Senaryo 27.** $\varepsilon \sim Weibull(2,2)$, hata teriminin şekil parametresi (α) 2 ve ölçek parametresi (β) 2 olan Weibull dağılışına sahip olması,
- **Senaryo 28.** $\varepsilon \sim Weibull(3,2)$, hata teriminin şekil parametresi (α) 3 ve ölçek parametresi (β) 2 olan Weibull dağılışına sahip olması olan senaryolardır.

Simülasyon çalışması sonuçları iki aşamada yorumlanmıştır. Birinci aşamada hata teriminin simetrik dağılış gösterdiği durum yorumlanırken ikinci aşamada hata teriminin asimetrik dağılış gösterdiği durumlar yorumlanmıştır.

Tablo 6'da hata teriminin simetrik dağılış gösterdiği durumlarda EKK, Huber M-Tahminleyici, Tukey M- Tahminleyici ve PVI dağılışının ağırlık fonksiyonuna dayanan M-Tahminleyici için TMS ve OKH değerleri verilmiştir. Bu değerler dikkate alınırsa hata teriminin dağılışı *Normal* olduğunda EKK yöntemi en iyi sonuçları göstermektedir. Ancak hata teriminin dağılışı *Student t* dağılışı olduğunda Huber ve Tukey M- Tahminleyiciler hem EKK'dan hem de PVI'dan daha iyi sonuçlar göstermektedirler. *Student t* dağılışının serbestlik derecesi 20 olduğunda *Normal* dağılışa yakınsadığından EKK, Huber ve Tukey yöntemleri birbirlerine yakın sonuçlar vermiştir. Hata teriminin *Cauchy* dağılışına sahip olduğu durumlarda, *Cauchy* dağılışının kalın ve uzun kuyruklara sahip olması nedeniyle örnek hacmi küçük olduğunda PVI yöntemi daha iyi sonuçlar vermektedir. Simülasyonlarda bazı dağılışlar için büyük örnek hacimlerinde PVI uyumu yapılamamıştır.

Tablo 6: Simülasyonda Kullanılan Simetrik Dağılımların TMS ve OKH Sonuçları

<i>Normal(0,1)</i>								
	EKK		Huber M		Tukey M		PVI	
	TMS	OKH	TMS	OKH	TMS	OKH	TMS	OKH
30	<u>0.0303</u>	<u>0.2589</u>	0.0808	0.2902	0.1278	0.3437	0.1651	0.3957
50	<u>0.0171</u>	<u>0.1366</u>	0.0644	0.1539	0.0933	0.1760	0.1438	0.2186
100	<u>0.0129</u>	<u>0.0648</u>	0.0445	0.0711	0.0584	0.0764	0.1177	0.1058
500	-	-	-	-	-	-	-	-
t_1								
30	15.621	230.663	0.2966	18.084	<u>0.1276</u>	<u>11.645</u>	0.0722	13.141
50	0.8723	36.184	0.2315	0.7217	<u>0.0697</u>	<u>0.5445</u>	0.0867	0.5659
100	-	-	-	-	-	-	-	-
500	-	-	-	-	-	-	-	-
t_5								
30	0.1367	0.4786	<u>0.0323</u>	<u>0.3642</u>	0.0948	0.4011	0.1306	0.4899
50	0.1096	0.2353	<u>0.0208</u>	<u>0.1949</u>	0.0568	0.2068	0.1423	0.2666
100	0.0867	0.1054	<u>0.0200</u>	<u>0.0882</u>	0.0306	0.0901	0.1091	0.1239
500	-	-	-	-	-	-	-	-
t_{10}								
30	0.0922	0.3453	<u>0.0617</u>	<u>0.3267</u>	0.1117	0.3769	0.1427	0.4490
50	0.0717	0.1842	<u>0.0330</u>	<u>0.1749</u>	0.0682	0.1928	0.1268	0.2452
100	0.0525	0.0802	<u>0.0228</u>	<u>0.0755</u>	0.0417	0.0789	0.1180	0.1130
500	0.0334	0.0153	<u>0.0189</u>	<u>0.0144</u>	0.0254	0.0146	0.0934	0.0266
t_{20}								
30	<u>0.0564</u>	<u>0.2831</u>	0.0614	0.2992	0.1119	0.3510	0.1347	0.4124
50	<u>0.0398</u>	<u>0.1554</u>	0.0619	0.1583	0.0954	0.1748	0.1380	0.2303
100	<u>0.0282</u>	<u>0.0703</u>	0.0288	0.0711	0.0461	0.0751	0.1110	0.1088
500	<u>0.0159</u>	<u>0.0136</u>	0.0167	0.0137	0.0217	0.0139	0.0875	0.0256
<i>Cauchy(0,1)</i>								
30	0.3869	0.1570	0.3753	0.1513	0.3676	0.1484	<u>0.3649</u>	<u>0.1465</u>
50	0.3875	0.1521	0.3800	0.1475	0.3754	0.1451	<u>0.3693</u>	<u>0.1404</u>
100	-	-	-	-	-	-	-	-
500	-	-	-	-	-	-	-	-

Hata teriminin asimetrik dağılım gösterdiği her durumda (Tablo 7) EKK, Huber M ve Tukey M- Tahminleyiciler çarpıklık ve yüksek basıklığı dikkate almadıkları için PVI'ya göre daha kötü sonuçlar vermektedirler. PVI ağırlık fonksiyonu dağılımın parametrelerine göre esnek bir yapı oluşturmaktadır. Bu nedenle hem çarpık hem de yüksek basıklığa sahip verilerde en iyi TMS ve OKH değerlerine sahip olmaktadır. Ayrıca simülasyonda kullanılan asimetrik

dağılıkların kuyruk yapısı incelenirse (Bkz: Ek 6) hepsinin PVI'dan daha ince kuyruklara sahip oldukları görülmektedir.

Beta (2,3) ve *Beta* (2,5) dağılımlarında büyük örnek hacminde PVI dağılışı elde edilememektedir. *Beta* dağılışı (0,1) aralığında değerler aldığı için örnek hacmi büyüdükçe PVI dağılışının ortaya çıkması zorlaşmaktadır. Bu nedenle örnek hacmi küçük olduğunda Tukey ve PVI yöntemlerinin sonuçları arasındaki fark azalmaktadır. Üstel dağılışı ele alındığında, dağılışın J- biçimli olması nedeniyle PVI yöntemi diğer yöntemlerden oldukça iyi TMS ve OKH değerleri verirken PVI'ya en yakın değerleri Tukey yöntemi vermektedir. Ayrıca Üstel dağılışın ortalaması arttığında kuyruklar hem kalınlaşmakta hem de uzamaktadır (Bkz: Ek 6).

Hata teriminin *Gama* dağılışı olduğu durumlarda ise dağılışın ölçek parametresi arttığı zaman kuyrukları kalınlaşmakta ve uzun olmaktadır. Bu nedenle PVI yüksek TMS ve OKH değerlerine sahip olsa bile diğer yöntemlerden daha iyi sonuçlar vermektedir.

Hata teriminin dağılışının χ^2 dağılışı olduğu durumlarda PVI yöntemi her serbestlik derecesinde en iyi TMS ve OKH değerlerine sahip olmaktadır. *F* ve *Weibull* dağılışı dikkate alınırsa, diğer tahminleyicilere göre PVI en iyi sonuçları vermektedir. *F* dağılışının serbestlik dereceleri düşük olduğunda kuyruğun çok hızlı azalmasından dolayı PVI dağılışına uyum yapılamamıştır. *Weibull* dağılışında ise şekil ve ölçek parametreleri azaldığında dağılış kalın kuyruklu hale gelirken şekil ve ölçek parametreleri arttığında ince kuyruklara sahip olmaktadır (Bkz. Ek 6). Bu koşullar dikkate alındığında bile en iyi çözümü PVI yöntemi vermektedir.

Elde edilen simülasyon sonuçlarına göre PVI yöntemi, eğer veriler çarpık ve yüksek basıklık değerine sahipse ağırlık fonksiyonunun bu parametreleri dikkate alması nedeniyle en iyi TMS ve OKH değerlerine sahip olmaktadır. Ayrıca PVI dağılışının kalın ve uzun kuyruklu olması verilerde sapan gibi görülebilen gözlemlere bile ağırlık vererek parametre tahmini yapmasını sağlamaktadır. Ancak hataların dağılışı simetrik olduğunda PVI yöntemi diğer yöntemlerden geride kalmaktadır.

Tablo 7: Simülasyonda Kullanılan Asimetrik Dağılımların TMS ve OKH Sonuçları

<i>Beta(2,3)</i>								
	EKK		Huber M		Tukey M		PVI	
	TMS	OKH	TMS	OKH	TMS	OKH	TMS	OKH
30	0.3877	0.1572	0.3768	0.1518	0.3709	0.1489	<u>0.3696</u>	<u>0.1476</u>
50	0.3873	0.1533	0.3793	0.1486	0.3745	0.1461	<u>0.3666</u>	<u>0.1417</u>
100	-	-	-	-	-	-	-	-
500	-	-	-	-	-	-	-	-
<i>Beta(2,5)</i>								
30	0.2768	0.0812	0.2646	0.0750	0.2581	0.0723	<u>0.2542</u>	<u>0.0719</u>
50	0.2775	0.0794	0.2658	0.0734	0.2609	0.0711	<u>0.2513</u>	<u>0.0674</u>
100	0.2762	0.0776	0.2654	0.0720	0.2623	0.0704	<u>0.2493</u>	<u>0.0643</u>
500	-	-	-	-	-	-	-	-
<i>Üstel(1)</i>								
30	10.178	12.658	0.8284	0.8349	0.7569	0.7206	<u>0.6891</u>	<u>0.6547</u>
50	10.124	11.453	0.8290	0.7662	0.7588	0.6577	<u>0.6682</u>	<u>0.5368</u>
100	10.139	10.663	0.8412	0.7300	0.7853	0.6420	<u>0.6485</u>	<u>0.4534</u>
500	10.067	10.105	0.8418	0.7115	0.7961	0.6382	<u>0.6209</u>	<u>0.3910</u>
<i>Üstel(2)</i>								
30	20.356	50.631	16.567	33.395	15.138	28.822	<u>13.779</u>	<u>26.184</u>
50	20.248	45.813	16.579	30.648	15.176	26.306	<u>13.365</u>	<u>21.472</u>
100	20.279	42.652	16.823	29.198	15.707	25.678	<u>12.970</u>	<u>18.137</u>
500	20.133	40.421	16.835	28.458	15.922	25.529	<u>12.418</u>	<u>15.639</u>
<i>Üstel(10)</i>								
30	101.781	1.265.773	82.837	834.860	75.692	720.546	<u>68.899</u>	<u>654.653</u>
50	101.240	1.145.326	82.895	766.194	75.879	657.656	<u>66.823</u>	<u>536.811</u>
100	101.394	1.066.307	84.116	729.949	78.534	641.947	<u>64.850</u>	<u>453.435</u>
500	100.666	1.010.528	84.177	711.460	79.612	638.232	<u>62.092</u>	<u>390.964</u>
<i>Gama(1,1)</i>								
30	10.133	12.765	0.8252	0.8327	0.7501	0.7077	<u>0.6987</u>	<u>0.6668</u>
50	10.047	11.351	0.8415	0.7712	0.7787	0.6624	<u>0.6821</u>	<u>0.5434</u>
100	10.069	10.666	0.8338	0.7312	0.7770	0.6404	<u>0.6407</u>	<u>0.4492</u>
500	10.047	10.117	0.8435	0.7114	0.7991	0.6385	<u>0.6197</u>	<u>0.3887</u>

Tablo 7 (Devam): Simülasyonda Kullanılan Asimetrik Dağılımların TMS ve OKH Sonuçları

<i>Gama(1,2)</i>								
	EKK		Huber M		Tukey M		PVI	
	TMS	OKH	TMS	OKH	TMS	OKH	TMS	OKH
30	20.327	50.465	16.761	33.600	15.272	28.808	<u>14.036</u>	<u>26.913</u>
50	20.297	45.436	16.676	30.566	15.349	26.368	<u>13.211</u>	<u>21.222</u>
100	20.149	41.781	16.648	28.711	15.523	25.246	<u>12.844</u>	<u>17.945</u>
500	20.068	40.414	16.837	28.390	15.944	25.447	<u>12.435</u>	<u>15.592</u>
<i>Gama(1,5)</i>								
30	50.819	315.408	41.904	210.001	38.179	180.047	<u>35.089</u>	<u>168.198</u>
50	50.742	283.978	41.691	191.035	38.373	164.799	<u>33.031</u>	<u>132.652</u>
100	50.372	261.131	41.620	179.442	38.809	157.784	<u>32.109</u>	<u>112.153</u>
500	50.170	252.585	42.093	177.435	39.860	159.041	<u>31.087</u>	<u>97.447</u>
<i>Gama(2,1)</i>								
30	20.204	45.536	17.983	36.199	17.140	33.240	<u>16.354</u>	<u>31.576</u>
50	20.195	43.043	18.272	35.018	17.600	32.523	<u>16.153</u>	<u>28.518</u>
100	20.157	41.289	18.307	34.139	17.775	32.323	<u>15.819</u>	<u>26.369</u>
500	20.088	40.229	18.382	33.655	18.067	32.532	<u>15.757</u>	<u>24.764</u>
<i>Gama(2,2)</i>								
30	40.409	182.145	35.965	144.797	34.280	132.960	<u>32.708</u>	<u>126.308</u>
50	40.390	172.170	36.544	140.072	35.200	130.091	<u>32.306</u>	<u>114.070</u>
100	40.315	165.155	36.613	136.557	35.551	129.292	<u>31.638</u>	<u>105.477</u>
500	40.177	160.915	36.763	134.621	36.134	130.128	<u>31.513</u>	<u>99.056</u>
<i>Gama(2,5)</i>								
30	101.022	1138	89.912	904.979	85.701	831.000	<u>81.770</u>	<u>789.426</u>
50	100.975	1076	91.359	875.446	88.001	813.066	<u>80.767</u>	<u>712.947</u>
100	100.787	1032	91.534	853.479	88.876	808.076	<u>79.096</u>	<u>659.232</u>
500	100.442	1005	91.909	841.378	90.334	813.301	<u>78.784</u>	<u>619.100</u>
$\chi^2_{0.5}$								
30	0.5240	0.4829	0.2564	0.1175	0.1701	0.0659	<u>0.1382</u>	<u>0.0529</u>
50	0.5164	0.3790	0.2304	0.0793	0.1350	0.0342	<u>0.1065</u>	<u>0.0228</u>
100	0.5196	0.3056	0.2144	0.0537	0.1165	0.0180	<u>0.0865</u>	<u>0.0113</u>
500	-	-	-	-	-	-	-	-

Tablo 7 (Devam): Simülasyonda Kullanılan Asimetrik Dağılımların TMS ve OKH Sonuçları

χ_1^2								
	EKK		Huber M		Tukey M		PVI	
	TMS	OKH	TMS	OKH	TMS	OKH	TMS	OKH
30	10.416	16.044	0.7200	0.7256	0.6027	0.5310	<u>0.5276</u>	<u>0.4956</u>
50	10.230	12.796	0.6916	0.5761	0.5603	0.4031	<u>0.4618</u>	<u>0.3061</u>
100	10.223	11.186	0.6817	0.4995	0.5540	0.3442	<u>0.4249</u>	<u>0.2209</u>
500	10.044	10.205	0.6648	0.4464	0.5433	0.2987	<u>0.3742</u>	<u>0.1474</u>
χ_5^2								
30	50.487	274.566	46.236	227.205	44.150	209.838	<u>43.171</u>	<u>203.966</u>
50	50.890	265.389	46.795	223.527	45.095	209.263	<u>42.211</u>	<u>187.777</u>
100	50.460	255.428	46.692	218.271	45.643	208.684	<u>41.509</u>	<u>174.865</u>
500	50.192	250.864	46.733	217.276	46.163	211.924	<u>41.184</u>	<u>169.547</u>
$F(1,2)$								
30	39.548	584.343	15.028	38.513	0.9399	17.042	<u>0.8689</u>	<u>16.584</u>
50	31.350	165.382	13.005	23.005	0.8569	0.9925	<u>0.7198</u>	<u>0.8707</u>
100	27.882	94.235	11.789	15.475	0.7506	0.6534	<u>0.5915</u>	<u>0.4391</u>
500	-	-	-	-	-	-	-	-
$F(2,10)$								
30	13.022	21.606	0.9670	11.523	0.8499	0.9246	<u>0.7715</u>	<u>0.8434</u>
50	12.630	18.317	0.9635	10.395	0.8466	0.8245	<u>0.7286</u>	<u>0.6531</u>
100	12.355	16.233	0.9512	0.9471	0.8509	0.7636	<u>0.6935</u>	<u>0.5346</u>
500	12.143	14.840	0.9429	0.8946	0.8541	0.7340	<u>0.6426</u>	<u>0.4249</u>
$F(5,10)$								
30	13.372	20.590	11.000	13.247	10.019	11.227	<u>0.9545</u>	<u>10.795</u>
50	12.767	17.393	10.798	12.246	10.074	10.714	<u>0.9088</u>	<u>0.9122</u>
100	12.478	16.039	10.699	11.670	10.103	10.410	<u>0.8882</u>	<u>0.8129</u>
500	12.278	15.109	10.634	11.287	10.170	10.316	<u>0.8511</u>	<u>0.7267</u>
$F(10,10)$								
30	12.908	18.987	11.268	13.736	10.614	12.138	<u>10.215</u>	<u>11.770</u>
50	12.662	16.900	11.113	12.744	10.521	11.409	<u>0.9710</u>	<u>10.071</u>
100	12.545	15.769	11.150	12.339	10.673	11.334	<u>0.9655</u>	<u>0.9347</u>
500	12.338	15.199	11.027	12.119	10.674	11.353	<u>0.9308</u>	<u>0.8639</u>

Tablo 7 (Devam): Simülasyonda Kullanılan Asimetrik Dağılımların TMS ve OKH Sonuçları

<i>Weibull(1,1)</i>								
	EKK		Huber M		Tukey M		PVI	
	TMS	OKH	TMS	OKH	TMS	OKH	TMS	OKH
30	10.187	12.668	0.8270	0.8381	0.7502	0.7172	<u>0.6917</u>	<u>0.6737</u>
50	10.232	11.371	0.8441	0.7636	0.7765	0.6565	<u>0.6834</u>	<u>0.5333</u>
100	10.082	10.613	0.8385	0.7282	0.7850	0.6413	<u>0.6432</u>	<u>0.4505</u>
500	10.045	10.113	0.8438	0.7113	0.7995	0.6381	<u>0.6250</u>	<u>0.3925</u>
<i>Weibull(2,1)</i>								
30	20.374	50.672	16.539	33.522	15.004	28.686	<u>13.834</u>	<u>26.946</u>
50	20.465	45.484	16.882	30.545	15.530	26.259	<u>13.668</u>	<u>21.333</u>
100	20.164	42.451	16.769	29.129	15.699	25.651	<u>12.864</u>	<u>18.020</u>
500	20.089	40.450	16.875	28.451	15.990	25.525	<u>12.500</u>	<u>15.701</u>
<i>Weibull(2,2)</i>								
30	17.646	32.608	16.791	29.843	16.346	28.650	<u>16.053</u>	<u>28.109</u>
50	17.723	31.984	16.951	29.370	16.646	28.379	<u>15.990</u>	<u>26.575</u>
100	17.708	31.551	17.018	29.212	16.829	28.590	<u>15.901</u>	<u>25.913</u>
500	17.726	31.369	17.115	29.273	17.061	29.098	<u>15.993</u>	<u>25.656</u>
<i>Weibull(3,2)</i>								
30	26.468	73.368	25.186	67.147	24.519	64.463	<u>24.082</u>	<u>63.250</u>
50	26.584	71.964	25.426	66.083	24.969	63.853	<u>23.986</u>	<u>59.798</u>
100	26.562	70.990	25.527	65.727	25.243	64.328	<u>23.852</u>	<u>58.305</u>
500	26.588	70.581	25.673	65.865	25.592	65.470	<u>23.989</u>	<u>57.726</u>

SONUÇ

Dayanıklı (Robust) Regresyon Analizinde birçok tahminleme yöntemi yer almaktadır. Bu tahminleme yöntemlerinin başında M- Tahminleyiciler gelmektedir. Ancak geleneksel M- Tahminleyiciler veri setinde bulunan aykırı davranışları (kalın- uzun kuyruk, çarpıklık ve basıklık) dikkate almamaktadırlar.

Bu tez çalışmasında Pearson Diferansiyel Denklemi ile Etki Fonksiyonu arasındaki ilişkiden yola çıkılarak verilerin dağılışının PVI dağılışı gösterdiği durumlarda PVI dağılışının ağırlık fonksiyonuna dayanan M- Tahminleyicinin performansı incelenmiştir. Hata teriminin simetrik olduğu *Normal* ve *Student t* dağılışlarında PVI yönteminin diğer yöntemlerden daha kötü performansa sahip olduğu gözlenmektedir. Ancak *Cauchy* dağılışı simetrik olsa bile kalın ve uzun kuyruklara sahip olması nedeniyle küçük örnek hacminde PVI yöntemi diğer yöntemlerden daha iyi sonuçlar vermektedir.

Hata teriminin dağılışı asimetrik ($\beta_1 > 0$) ve yüksek basıklığa ($\beta_2 > 3$) sahip olduğu durumlarda PVI dağılışının ağırlık fonksiyonu bu parametreleri içerdiğinden diğer M- Tahminleyicilere göre daha iyi bir performansa sahip olmaktadır. Ayrıca PVI dağılışının kalın ve uzun kuyruklu (bkz. Ek 6) olması sebebiyle kendisine yakın veya daha ince kuyruklara sahip asimetrik dağılışlarda iyi performans gösterdiği gözlemlenmiştir. Geliştirilen yöntemin performansı iki farklı gerçek veri seti ile 9 farklı dağılışa sahip 28 senaryo üzerinden simülasyon çalışması yapılarak incelenmiştir. Simülasyonlarda kullanılan asimetrik dağılışlar PVI'ya göre daha ince kuyruk yapısına sahiptirler.

Gelecek çalışmalarda, PI ve PII dağılışının ağırlık fonksiyonları üzerinde durulması planlanmaktadır. Bu dağılışların ağırlık fonksiyonları U- biçimli olduğundan Huber ve Tukey M- tahminleme yöntemlerinde olduğu gibi bir dönüş sabiti (tuning constant) ile şans değişkeninin değeri arttığında azalan bir ağırlık fonksiyonuna dönüşüm yapılmaya çalışılabilir. Özellikle PII dağılışı simetrik yapısı nedeniyle hata teriminin dağılışının simetrik olduğu durumlar için iyi bir çözüm elde edebileceği düşünülmektedir. Bunun yanında PI dağılışının geniş bir yelpazede ortaya çıkabilmesi ve asimetrik yapısı sebebiyle bir M- tahminleme yöntemi olarak kullanılabileceği düşünülmektedir. Ayrıca Pearson Diferansiyel Denklemi içerisinde bulunan mod (a) değeri için dayanıklı bir tahminleme yönteminin geliştirilmesi üzerine çalışılabilir.

KAYNAKÇA

Abramowitz, M. ve Stegun, I. A. (1965). Handbook of mathematical functions with formulas, graphs, and mathematical table. In *US Department of Commerce*. National Bureau of Standards Applied Mathematics series 55.

Agee, W. S. ve Turner, R. H. (1978). *Application of Robust Statistical Methods to Data Reduction* (No. ACD-TR-65). National Range Operations Directorate White Sands Missile Range N Mex Analysis And Computation Div.

Allende, H. Frery, A. C. Galbiati, J. ve Pizarro, L. (2006). M-estimators with asymmetric influence functions: the distribution case. *Journal of Statistical Computation and Simulation*, 76(11), 941-956.

Andersen, R. (2008). *Modern Methods for Robust Regression* (No. 152). Sage.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. ve Tukey, J. W. (1972). Robust Estimates of Location: Survey and Advances.

Andrews, D. F. (1974). *A Robust Method for Multiple Linear Regression*. *Technometrics*, 16(4), 523-531.

Beaton, A. E. ve Tukey, J. W. (1974). The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic data. *Technometrics*, 16(2), 147-185.

Balkema, G. ve Embrechts, P. (2018). Linear regression for heavy tails. *Risks*, 6(3), 93.

Barnett, V., ve Lewis, T. (1984). Outliers in statistical data. osd.

Bell, R. M. (1980). *An adaptive choice of the scale parameter for M-estimators* (No. TR-3). Stanford Univ Ca Dept of Statistics.

Bickel, P. J., Holm, S., Rosén, B., Spjøtvoll, E., Lauritzen, S., Johansen, S. ve Barndorff-Nielsen, O. (1976). Another look at robustness: a review of reviews and some new developments [with discussion and reply]. *Scandinavian Journal of Statistics*, 145-168.

Bosse, M., Agamennoni, G. ve Gilitschenski, I. (2016). Robust estimation and applications in robotics. *Foundations and Trends® in Robotics*, 4(4), 225-269.

Box, G. E. (1953). Non-normality and tests on variances. *Biometrika*, 40(3/4), 318-335.

Brascamp, J. W., Berg, A. V. ve Ee, R. (2004). Shared neural circuitry for switching between perceptual states and ocular motor states?. *Journal of Vision*, 4(8), 255-255.

Bryson, M. C. (1974). Heavy-Tailed Distributions: Properties and Tests. *Technometrics*, 16(1), 61-68.

Cantoni, E. ve Ronchetti, E. (2006). A robust Approach for Skewed and Heavy-Tailed Outcomes in the Analysis of Health Care Expenditures. *Journal of Health Economics*, 25(2), 198-213.

Lehmann, E. L., ve Casella, G. (2006). *Theory of Point Estimation*. Springer Science & Business Media.

Chatterjee, S. ve Price, B. (1977). Selection of Variables in a Regression Equation. *Regression Analysis by Example*, 201-203.

Chatterjee, S. ve Mächler, M. (1997). Robust Regression: A Weighted Least Squares Approach. *Communications in Statistics-Theory and Methods*, 26(6), 1381-1394.

Chen, C. (2013). Distributed Iteratively Reweighted Least Squares and Applications. *Statistics and Its Interface*, 6(4), 585-593.

Craig, C. C. (1936). A new Exposition and Chart for the Pearson System of Frequency Curves. *The Annals of Mathematical Statistics*, 7(1), 16-28.

Crooks, G. E. (2010). Field Guide to Continuous Probability Distributions. URL: <http://threeplusone.com/fieldguide>. (Accessed: 29.01. 2019).

Croux, C. ve Reusseeuw, P. J. (1992). Time-efficient Algorithms for two Highly Robust Estimators of Scale. In *Computational Statistics* (pp. 411-428). Physica, Heidelberg.

Davies, L., & Gather, U. (1993). The Identification of Multiple Outliers. *Journal of the American Statistical Association*, 88(423), 782-792.

Donoho, D. L. ve Huber, P. J. (1983). The Notion of Breakdown Point. *A Festschrift for Erich L. Lehmann*, 157-184.

Draper, N. R. ve Smith, H. (2014). Applied Regression Analysis (Vol. 326). John Wiley & Sons.

Dzhun', I. V. (2011). Method for Diagnostics of Mathematical Models in Theoretical Astronomy and Astrometry. *Kinematics and Physics of Celestial Bodies*, 27, 260-264.

Eicker, F. (1963). Asymptotic Normality and Consistency of the least Squares Estimators for Families of Linear Regressions. *The Annals of Mathematical Statistics*, 447-456.

Elderton, W. P. (1906). Frequency Curves and Correlation Cambridge University. New York.

Elderton, W. P. (1953). Frequency Curves and Correlation Cambridge University. New York.

Fiori, A. M. ve Zenga, M. (2009). Karl Pearson and the Origin of Kurtosis. *International Statistical Review*, 77(1), 40-50.

Foss, S., Korshunov, D., ve Zachary, S. (2011). *An Introduction to Heavy-Tailed and Subexponential Distributions* (Vol. 6, pp. 0090-6778). New York: Springer.

Fox, J. ve Weisberg, S. (2002). Robust Regression. *An R and S-Plus Companion to Applied Regression*, 91.

Godambe, V. P. (1960). An Optimum Property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4), 1208-1211.

Gross, A. M. (1977). Confidence Intervals for Bisquare Regression Estimates. *Journal of the American Statistical Association*, 72(358), 341-354.

Hald, A. (2008). A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935. Springer Science & Business Media.

Hampel F.R., Ronchetti E.M., Rousseuw P.J. ve Stahel W.A. (1986). *Robust Statistics. The Approach Based on Influence Functions*. Wiley, New York.

Hampel, F. R. (1968). Contribution to the Theory of Robust Estimation. *Ph. D. Thesis, University of California, Berkeley*.

Hampel, F. R. (1974). The Influence Curve and its role in Robust Estimation. *Journal of the American Statistical Association*, 69(346), 383-393.

Harvey, A. C. (1977). A Comparison of Preliminary Estimators for Robust Regression. *Journal of the American Statistical Association*, 72(360a), 910-913.

Hinich, M. J. ve Talwar, P. P. (1975). A Simple Method for Robust Regression. *Journal of the American Statistical Association*, 70(349), 113-119.

Hodges Jr, J. L. (1967). Efficiency in Normal Samples and Tolerance of Extreme Values for Some Estimates of Location. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 163-186).

Hogg, R. V. (1979). Statistical robustness: One view of its use in applications today. *The American Statistician*, 33(3), 108-115.

Hogg, R. V., McKean, J., ve Craig, A. T. (2005). *Introduction to mathematical statistics*. Pearson Education.

Holland, P. W. ve Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*, 6(9), 813-827.

Huber, P. J. (1964). Robust version of a location parameter. *Annals of Mathematical Statistics*, 36, 1753-1758.

Huber, P. J. (1967, June). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 1, pp. 221-233).

Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics*, 1(5), 799-821.

Huber, P. J., ve Ronchetti, E. M. (1981). Robust Statistics, ser. Wiley Series in Probability and Mathematical Statistics. New York, NY, USA, Wiley-IEEE, 52, 54.

Huber, P. J. (2011). *Robust Statistics* (pp. 1248-1251). Springer Berlin Heidelberg.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H., ve Lee, T. C. (1988). *Introduction to the Theory and Practice of Econometrics* (No. 330.015195 I61 1988). J. Wiley.

Kauermann, G., ve Carroll, R. J. (2001). A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *Journal of the American Statistical Association*, 96(456), 1387-1396.

Khalil, U., Ali, A., Khan, D. M., Khan, S. A. Ve Qadir, F. (2016). Efficient Uk'S re-Descending M-Estimator For Robust Regression. *Pakistan Journal Of Statistics*, 32(2).

Kilany, N. M. (2016). Weighted Lomax distribution. *SpringerPlus*, 5(1), 1862.

Kotz, S., Balakrishnan, N. ve Johnson, N. L. (1972). *Continuous Multivariate Distributions, Volume 1: Models and Applications* (Vol. 1). John Wiley & Sons.

Krasker, W. S., Kuh, E., ve Welsch, R. E. (1983). Estimation for Dirty Data and Flawed Models. *Handbook of Econometrics*, 1, 651-698.

Lenth, R. V. (1977). Robust splines. *Communications in Statistics-Theory and Methods*, 6(9), 847-854.

Maronna, R. A., Martin, D., ve Yohai, R. S. (2006). Wiley Series in Probability and Statistics. *Robust Statistics: Theory and Methods*, 404-414.

Mielke Jr, P. W. ve Johnson, E. S. (1974). Some Generalized Beta Distributions of the Second Kind Having Desirable Application Features in Hydrology and Meteorology. *Water Resources Research*, 10(2), 223-226.

Mises, R. V. (1947). On the Asymptotic Distribution of Differentiable Statistical Functions. *The Annals of Mathematical Statistics*, 18(3), 309-348.

Moghaddam, M. D., Liu, J. ve Serota, R. A. (2019). Implied and Realized Volatility: A Study of Distributions and the Distribution of Difference. *arXiv preprint arXiv:1906.02306*.

Mohebbi, M., Nourijelyani, K. ve Zeraati, H. (2007). A simulation Study on Robust Alternatives of Least Squares Regression. *Journal of Applied Sciences*, 7(22), 3469-3476.

Muthukrishnan, R. ve Radha, M. (2010). M-estimators in Regression Models. *Journal of Mathematics Research*, 2(4), 23.

Nagahara, Y. (1999). The PDF and CF of Pearson Type IV Distributions and the ML Estimation of the Parameters. *Statistics & Probability Letters*, 43(3), 251-264.

Nagahara, Y. (2008). A Method of Calculating the Downside Risk by Multivariate Nonnormal Distributions. *Asia-Pacific Financial Markets*, 15(3-4), 175-184.

Pearson, K. (1895). X. Contributions to the Mathematical Theory of Evolution.—II. Skew variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London*, (A.), (186), 343-414.

Pearson, K. (1901). XI. Mathematical Contributions to the Theory of Evolution.—X. Supplement to a Memoir on Skew Variation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 197(287-299), 443-459.

Pearson, K. (1914). Tables for Statisticians and Biometricians.

Pearson, E. (1963). Some Problems Arising in Approximating to Probability Distributions, Using Moments. *Biometrika*, 50(1/2), 95-112. doi:10.2307/2333751

Rasheed, B. A., Adnan, R., Saffari, S. E. ve dano Pati, K. (2014). Robust Weighted Least Squares Estimation of Regression Parameter in the Presence of Outliers and Heteroscedastic Errors. *Jurnal Teknologi*, 71(1).

Ryan, T. P. (2008). *Modern Regression Methods* (Vol. 655). John Wiley & Sons.

Salazar, J. S. I. (2000). Detection Schemes for Synthetic-Aperture Radar Imagery Based on a Beta Prime Statistical Model.

Schrader, R. ve Hettmansperger, T. (1980). Robust Analysis of Variance Based Upon a Likelihood Ratio Criterion. *Biometrika*, 67(1), 93-101.

Siegel, A. F. (1982). Robust Regression Using Repeated Medians. *Biometrika*, 69(1), 242-244.

Staudte, R. G. ve Sheather, S. J. (2011). *Robust Estimation and Testing* (Vol. 918). John Wiley & Sons.

Stefanski, L. A. ve Boos, D. D. (2002). The Calculus of M-Estimation. *The American Statistician*, 56(1), 29-38.

Street, J. O., Carroll, R. J. ve Ruppert, D. (1988). A note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares. *The American Statistician*, 42(2), 152-154.

Stuart, C. (2011). Robust Regression. *Department of Mathematical Sciences, Durham University*, 169.

Susanti, Y. ve Pratiwi, H. (2014). M estimation, S estimation, and MM Estimation in Robust Regression. *International Journal of Pure and Applied Mathematics*, 91(3), 349-360.

Sumarni, C., Sadik, K., Notodiputro, K. A. ve Sartono, B. (2017, March). Robustness of Location Estimators Under t-Distributions: a Literature Review. In *IOP Conference Series: Earth and Environmental Science* (Vol. 58, No. 1, p. 012015). IOP Publishing.

Şehirlioğlu, A. K., Dündar, S. (2014). *Pearson Dağılım Ailesi*. İzmir: Ege Üniversitesi Basımevi.

Tukey, J. W. (1960). A Survey of Sampling from Contaminated Distributions. *Contributions to Probability and Statistics*, 448-485.

Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, 33(1), 1-67.

Tulupyyev, A., Suvorova, A., Sousa, J. ve Zelterman, D. (2013). Beta Prime Regression with Application to Risky Behavior Frequency Screening. *Statistics in Medicine*, 32(23), 4044-4056.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct test for Heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817-838.

Wilcox, R. R. (2011). *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press.

Wiśniewski, Z. (2014). M-estimation with Probabilistic Models of Geodetic Observations. *Journal of Geodesy*, 88(10), 941-957.

Wood, F. S. ve Gorman, J. W. (1971). *Fitting Equations to Data: Computer Analysis of Multifactor Data For Scientists and Engineers*. Wiley-Interscience.

Wu, L. L. (1985). Robust M-estimation of Location and Regression. *Sociological Methodology*, 15, 316-388.

Ye, Y., Oluyede, B. O. ve Pararai, M. (2012). Weighted Generalized beta Distribution of the Second kind and Related Distributions. *Journal of Statistical and Econometric Methods*, 1(1), 13-31.

Yulita, T., Notodiputro, K. A. ve Sadik, K. (2018). M-Estimation Use Bisquare, Hampel, Huber, and Welsch Weight Functions in Robust Regression.

Zeigler, R. K. ve Ferris, Y. (1973). Half-life of Plutonium-241. *Journal of Inorganic and Nuclear Chemistry*, 35(10), 3417-3418.

Zuo, Y. (2001). Some Quantitative Relationships Between Two Types of Finite Sample Breakdown Point. *Statistics & Probability Letters*, 51(4), 369-375.

EKLER



Ek 1: Değişmezlik (Invariant) Özelliği

Eğer bir tahminleyici düzgün bir şekilde dönüştürülmüş ise buna Değişmezlik denir (Rousseeuw ve Leroy, 1987: 116). Bir regresyon tahminleyicisi için değişmezlik şu 2 özelliği sağlamalıdır;

1. Konum değişmezliği

Bir tahminleyici T ancak ve ancak şu özelliği sağlıyorsa değişmezdir;

$$T\left(\left\{\left(\mathbf{x}_i^T, y_i + x_i^T v\right); i = 1, \dots, n\right\}\right) = T\left(\left\{\left(\mathbf{x}_i^T, y_i\right); i = 1, \dots, n\right\}\right) + v$$

Konum değişmezliği vardır denir.

2. Ölçek değişmezliği

Eğer açıklanan değişken Y herhangi bir sabit c değeri ile çarpılırsa (ölçüm birimi değişirse), tahminleyici aynı c değeri ile çarpılır.

$$T\left(\left\{\left(\mathbf{x}_i^T, cy_i\right); i = 1, \dots, n\right\}\right) = cT\left(\left\{\left(\mathbf{x}_i^T, y_i\right); i = 1, \dots, n\right\}\right)$$

Böylece eğer $y \rightarrow cy$ olursa $\hat{\theta} \rightarrow c\hat{\theta}$ olur. Eğer bir tahminleyici ölçek değişmezliğine sahip değilse artıklar parametre tahmini yapılmadan önce mutlaka standardize edilmelidir (Stuart, 2011:10).

Ek 2: Medyan Mutlak Sapma (MAD)

Bir tahminleyici her zaman ölçek değişmezliğine sahip olamayabilir. Böyle durumlarda örneklem standart sapması s eğer örnekte sapan gözlem veya sapan olduğundan şüphelenilen gözlem varsa böyle gözlemlerden etkilenecek dayanıklı bir tahmin olmayacaktır (Hogg, 1979:109). Bu nedenle sapan gözlemlere daha dayanıklı ve daha az etkilenen örneklem meydanına dayanan bir ölçek tahminleyicisi Medyan Mutlak Sapma (MAD) kullanılabilir (Andersen, 2008:51; Draper and Smith 2014: 572; Street et. Al, 1988:153; Hogg, 1979:109). MAD;

$$s = \frac{MAD}{0.6745} = \frac{\text{medyan}|x_i - \text{medyan}(x)|}{0.6745}$$

ile gösterilir. Paydada bulunan 0.6745 bölüni n büyük olduğunda ve verilerin dağılışı normal dağılışa yakınsadığında $s \approx \sigma$ olur.

$n \rightarrow \infty$ ve $x_i \sim N(0, \sigma^2)$ olduğunda;

$$P(|x_i| < MAD) \approx 0.5$$

$$P\left(\left|\frac{x_i - 0}{\sigma}\right| < \frac{MAD}{\sigma}\right) \approx 0.5$$

$$P\left(|Z| < \frac{MAD}{\sigma}\right) \approx 0.5$$

$$\frac{MAD}{\sigma} \approx \Phi^{-1}(0.75)$$

$$\frac{MAD}{\Phi^{-1}(0.75)} \approx \sigma$$

$$\frac{MAD}{0.6745} \approx \sigma$$

Burada $Z \sim N(0,1)$ ve Φ standart normal dağılışın kümülatif fonksiyonunu göstermektedir (Stuart, 2011:14).

Ek 3: Düzgünlük (Regularity) Koşulları

Düzgünlük koşulları, genellikle $o(y)$ 'nin türevlenebilirliği ve türev ile integral işaretinin yer değiştirebilir olması ile ilgilenmektedir. Genel olarak düzgünlük koşulları, bağımsız ve aynı $f(x|\theta)$ dağılışıma sahip $X_1, \dots, X_n$ şans örneği ele alınırsa gerçek değerli bir θ parametresi için;

- a. Parametre uzayı Ω açık bir aralığa sahiptir (sonlu olması gerekmez),
- b. $A = \{x : f(x|\theta) > 0\}$ koşulunu sağlayan herhangi bir küme θ 'dan bağımsızdır.
- c. Her bir $x \in A$ için, $f(x|\theta)$, θ 'ya göre iki kere türevlenebilir olmalıdır ve ikinci türev θ için sürekli olmalıdır.
- d. $\int f(x|\theta) dx$ integrali integral işareti altında iki kere türevlenebilir olmalıdır.
- e. Fisher Bilgisi $I(\theta)$ sonlu olmalıdır ($0 < I(\theta) < \infty$).
- f. Verilen herhangi bir $\theta_0 \in \Omega$ için, θ_0 'a bağlı bir fonksiyon $M(x)$ ve pozitif bir sayı c 'nin var olduğu kabul edilirse;

$$\left| \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right| \leq M(x), \quad \forall x \in A, \quad \theta_0 - c < \theta < \theta_0 + c$$

ve

$$E_{\theta_0} [M(X)] < \infty$$

koşulları sağlanmalıdır.

g. $\left| \frac{\partial \log f(X|\theta)}{\partial \theta} \right| = 0$

h. $E \left[-\frac{\partial^2 \log f(X|\theta)}{\partial \theta^2} \right] = E \left[\frac{\partial \log f(X|\theta)}{\partial \theta} \right]^2 = I(\theta)$ ile gösterilir (Lehmann ve

Casella, 2006; Hogg ve diğerleri, 2005).

Ek 4: Slutsky's Teoremi

Slutsky's teoreminden bahsedilmek için öncelikle olasılıksal yakınsama (convergence in probability) ve dağılımsal yakınsama (convergence in distribution) ele alınacaktır.

- Her bir $\varepsilon > 0$ için $X_1, \dots, X_n$ şans değişkenleri dizisi, bir X şans değişkenine aşağıdaki şartı sağlıyorsa olasılıksal olarak yakınsıyor denir.

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$$

veya

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1.$$

- $X_1, \dots, X_n$ şans değişkenleri dizisi, bir X şans değişkenine aşağıdaki şartı sağlıyorsa dağılımsal olarak yakınsıyor denir.

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

Burada $F_X(x)$, sürekli bir dağılım fonksiyonudur.

Olasılıksal ve dağılımsal yakınsama bilgileri kullanılarak, eğer X_n , X 'e dağılımsal olarak yakınsıyor ($X_n \xrightarrow{d} X$) ve Y_n bir a sabitine olasılıksal olarak yakınsıyorsa ($Y_n \xrightarrow{p} a$) Slutsky's teoremi;

- $Y_n X_n \xrightarrow{d} aX$
- $X_n + Y_n \xrightarrow{d} X + a$
- $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a}$

olarak gösterilir (Casella ve Berger, 2002).

Ek 5: Sandwich Varyans Tahminleyicisi

Bölüm 1’de ele alınan regresyon modeli kullanılırsa, eğer hatalar eş varyanslı ve dağılışı ortalaması 0 varyanslı σ^2 olan normal dağılış ise ($\varepsilon_i \sim N(0, \sigma^2)$) regresyon parametrelerinin varyans- kovaryans matrisi $V(\hat{\theta}) = s^2 (X^T X)^{-1}$ ile gösterilir. Öncelikle, z^T , $1 \times p$ boyutlu ortonormal bir vektör ($z^T z = 1$) olmak üzere bir doğrusal eşitlik $z^T \theta$ ile ilgilenilirse $V(z^T \theta) = s^2 z^T (X^T X)^{-1} z$ olur ve burada $s^2 = \sum_{i=1}^n e_i^2 / (n - p)$ olarak gösterilebilir. Ancak Ekonometrik yöntemler ve Dayanıklı yöntemler dikkate alınırse regresyon parametreleri için varyans tahmini yapmak, hipotez testleri ve güven aralıkları elde edebilmek için önemli olmaktadır. Eğer hatalar değişen varyanslı ise $z^T \theta$ için varyans- kovaryans matrisi;

$$V_{\text{sand}}(z^T \theta) = z^T (X^T X)^{-1} \left(\sum_i x_i^T x_i e_i^2 \right) (X^T X)^{-1} z = \sum_{i=1}^n a_i^2 e_i^2$$

ile gösterilir. Burada $a_i = z^T (X^T X)^{-1} x_i$ ’dir. Bu tahminleyiciye Sandwich Varyans Tahminleyicisi (Sandwich Variance Estimator), Dayanıklı (Robust) Varyans Tahminleyicisi veya Değişen Varyansa Karşı Tutarlı Kovaryans Tahminleyicisi denir (Huber, 1967; Eicker, 1963; Kauerman ve Carrol, 2001; White, 1980).

Ek 6: Bir Olasılık Dağılışının Kuyruk Özellikleri

Reel sayılar uzayında tanımlı f olasılık yoğunluk fonksiyonuna sahip herhangi bir dağılım fonksiyonu F için kuyruk fonksiyonu $\bar{F}(x)$;

$$\bar{F}(x) = F(x, \infty), \quad x \in \mathbb{R}$$

şeklinde tanımlanabilir. Kuyruk fonksiyonu ve oyf dikkate alınarak bir dağılışın Kalın Kuyruklu, Şışman Kuyruklu veya Uzun Kuyruklu olup olmadığı aşağıdaki üç koşul dikkate alınarak belirlenebilir.

Koşul 1. Kalın Kuyruklu (Heavy- Tailed) Dağılışlar

Bir oyf'nin "kalın kuyruklu dağılış" olarak tanımlanabilmesi için bütün üstel momentlerinin sonsuz olması gerekmektedir. Bir F dağılışının kalın kuyruklu olması ancak ve ancak;

$$\int_{\mathbb{R}} e^{\lambda x} F(dx) = \infty, \quad \forall \lambda > 0$$

koşulunun sağlanması yeterlidir. Eşitlik f için yazılırsa;

$$\limsup_{x \rightarrow \infty} f(x) e^{\lambda x} = \infty, \quad \forall \lambda > 0$$

elde edilir. Eğer dağılış ince (light) kuyruklu ise;

$$\int_{\mathbb{R}} e^{\lambda x} F(dx) < \infty, \quad \lambda > 0$$

şartını sağlaması gerekir. Burada $\lambda > 0$ şartı " λ 'nın herhangi bir değeri için" anlamında kullanılmıştır.

Koşul 2. Uzun Kuyruklu (Long Tailed) Dağılışlar

Uzun kuyruklu dağılışlar aynı zamanda kalın kuyruklu dağılışların bir alt sınıfıdır. Bir oyf'nin uzun kuyruklu olması ancak ve ancak;

$$\lim_{x \rightarrow \infty} \frac{f(x + \lambda)}{f(x)} = 1, \quad \forall \lambda > 0.$$

koşulu gerçekleştiğinde olur. Eşitlik kuyruk fonksiyonu cinsinden yazılırsa;

$$\lim_{x \rightarrow \infty} \bar{F}(x + \lambda) \sim \bar{F}(x), \quad \forall \lambda > 0$$

olur. Eğer f dağılış uzun kuyruklu ise herhangi bir sabit $\lambda > 0$ için $g(x) := f(\log x)$ alınır;

$$\lim_{x \rightarrow \infty} \frac{g(\lambda x)}{g(x)} \rightarrow 1$$

şartı sağlandığında f dağılış sonsuza doğru yavaşça azalan bir kuyruğa sahiptir denir.

Koşul 3. Şişman Kuyruklu (Fat Tailed) Dağılımlar

Bir oyf'nin şişman kuyruklu olarak adlandırılması için aşağıda belirtilen koşulu sağlaması gerekir.

$$\lim_{x \rightarrow \infty} P(X > x) \sim x^{-\lambda}, \quad \forall \lambda > 0.$$

Burada $P(X > x) = \bar{F}(x)$ olduğu görülmektedir. Araştırmacılar aynı zamanda grafiksel olarak dağılımlarının kontrolünün yapılmasını önermişlerdir (Bryson, 1974:62; Foss ve diğerleri, 2011: 18).

Bu tez çalışmasında kullanılan PVI dağılımının kuyruk özellikleri aşağıda gösterilmiştir. PVI dağılımının oyf'si;

$$f(x) = K(x + r_1)^{m_1} (x + r_2)^{m_2}, \quad r_1 < r_2 < x < \infty.$$

Burada $m_1 > 0, m_2 < 0, m_1 < |m_2|$ 'dir. Buna göre Koşul 1 dikkate alınır ve integral sabiti atılırsa;

$$\begin{aligned} \lim_{x \rightarrow \infty} e^{\lambda x} f(x) &= \lim_{x \rightarrow \infty} e^{\lambda x} (x + r_1)^{m_1} (x + r_2)^{m_2} \\ &= \lim_{x \rightarrow \infty} e^{\lambda x} \frac{(x + r_1)^{m_1}}{(x + r_2)^{m_2}} \\ &= \infty \end{aligned}$$

elde edilir. Buna göre PVI dağılımı kalın kuyruklu bir dağılımdır. Dağılım kalın kuyruklu olduğu belirlendikten sonra kuyruklarının uzunluğu belirlemek amacıyla herhangi bir $\lambda > 0$ için;

$$f(x + \lambda) = (x + r_1 + \lambda)^{m_1} (x + r_2 + \lambda)^{m_2}$$

elde edilir. Koşul 2 dikkate alınır;

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{f(x + \lambda)}{f(x)} &= \frac{(x + r_1 + \lambda)^{m_1} (x + r_2 + \lambda)^{m_2}}{(x + r_1)^{m_1} (x + r_2)^{m_2}} \\ &= \lim_{x \rightarrow \infty} \left(\frac{x + r_1 + \lambda}{x + r_1} \right)^{m_1} \left(\frac{x + r_2 + \lambda}{x + r_2} \right)^{m_2} \\ &= \lim_{x \rightarrow \infty} \left(1 + \frac{\lambda}{x + r_1} \right)^{m_1} \left(1 + \frac{\lambda}{x + r_2} \right)^{m_2} \\ &= 1 \end{aligned}$$

bulunur. Buna göre PVI dağılımı uzun kuyruklu bir dağılımdır. Eğer PVI uzun kuyruklu ise;

$$\begin{aligned}
\lim_{x \rightarrow \infty} \frac{g(\lambda x)}{g(x)} &= \frac{(\log x + r_1 + \log \lambda)^{m_1} (\log x + r_2 + \log \lambda)^{m_2}}{(\log x + r_1)^{m_1} (\log x + r_2)^{m_2}} \\
&= \lim_{x \rightarrow \infty} \left(\frac{\log x + r_1 + \log \lambda}{\log x + r_1} \right)^{m_1} \left(\frac{\log x + r_2 + \log \lambda}{\log x + r_2} \right)^{m_2} \\
&= \lim_{x \rightarrow \infty} \left(1 + \frac{\log \lambda}{\log x + r_1} \right)^{m_1} \left(1 + \frac{\log \lambda}{\log x + r_2} \right)^{m_2} \\
&= 1
\end{aligned}$$

elde edilir. Buna göre PVI dağılışı hem uzun kuyruklu hem de sonsuza doğru yavaşça azalan kuyruklara sahiptir. Bu aşamadan sonra PVI dağılışının kuyruklarının şişman olup olmadığının belirlenebilmesi için kuyruk fonksiyonu;

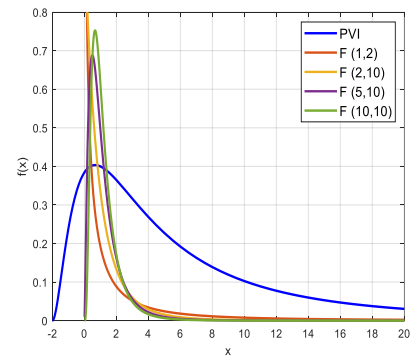
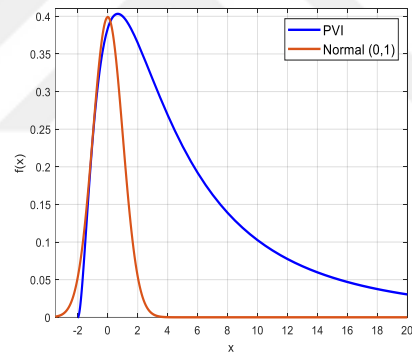
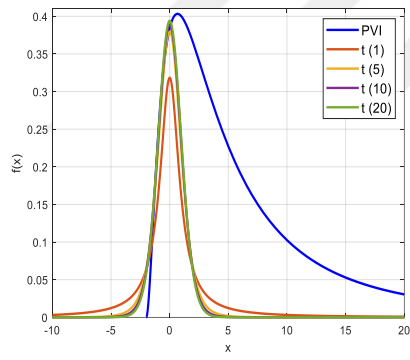
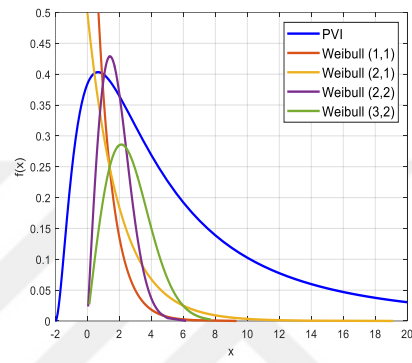
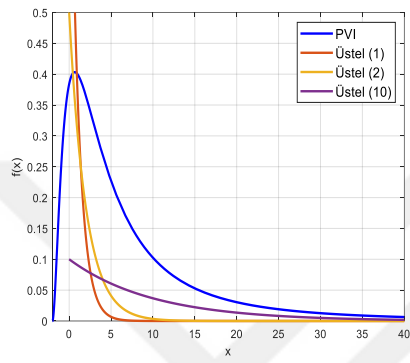
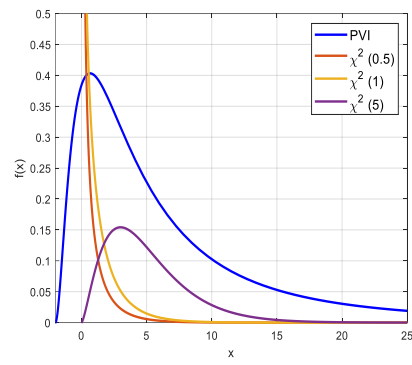
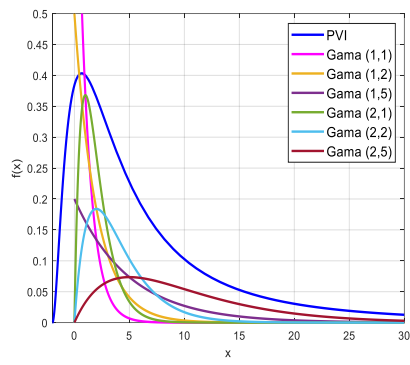
$$\begin{aligned}
P(X > x) &= 1 - F(x) \\
&= \bar{F}(x) \\
&= 1 - \frac{x^{m_1+1} {}_2F_1 \left[\begin{matrix} m_1+1 & -m_2 \\ m_1+2 \end{matrix} ; -x \right]}{(m_1+1) B(-m_1-m_2-1; m_2+1)}
\end{aligned}$$

elde edilir. Koşul 3 dikkate alınır ve kuyruk fonksiyonunda bulunan sabitler atılırsa;

$$\begin{aligned}
&\lim_{x \rightarrow \infty} \bar{F}(x) \\
&= \lim_{x \rightarrow \infty} (1 - x^{m_1+1}) \\
&= -\infty
\end{aligned}$$

olur. Böylece PVI dağılışının şişman kuyruklara sahip olduğu söylenemez.

Bütün koşulları dikkate aldığımızda PVI dağılışı kalın, yavaşça azalan bir kuyruğa sahip ancak kuyruklarının şişman olmadığı görülmektedir. PVI dağılışının simülasyonda kullanılan asimetrik dağılışlar ile grafiksel karşılaştırmaları aşağıda verilmiştir.



Ek 7: M- Tahminleyicilerin Asimtotik Normalligi

$\theta = T(F)$ için $\psi(x; \hat{\theta})$ 'e, θ etrafında Taylor açılımı uygulanırsa;

$$\begin{aligned}\sum_{i=1}^n \psi(x; \hat{\theta}) &= \sum_{i=1}^n \psi(x; \theta) + (\hat{\theta} - \theta) \sum_{i=1}^n \psi'(x; \theta) + \dots = 0 \\ \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n \psi(x; \theta) \right\} + \sqrt{n} (\hat{\theta} - \theta) \left\{ \frac{1}{n} \sum_{i=1}^n \psi'(x; \theta) \right\} + O_p(1/\sqrt{n}) &= 0\end{aligned}$$

elde edilir. Merkezi Limit Teoreminden (dağılımsal yakınsama);

$$\sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n \psi(x; \theta) \right\} \xrightarrow{d} \mathbf{Z} \sim N\left(0, E_F \left[\psi(X; \theta)^2 \right]\right)$$

elde edilir. Buradan Büyük Sayıların Zayıf Kanunu (olasılıksal yakınsama) kullanılarak;

$$\frac{1}{n} \sum_{i=1}^n \psi'(x; \theta) \xrightarrow{p} E_F \left[\psi'(X; \theta) \right]$$

olur. Böylece Slutsky's teoreminden;

$$\sqrt{n}(\theta - \hat{\theta}) \xrightarrow{d} -\mathbf{Z} / E_F \left[\psi'(X; \theta) \right] \sim N(0, \sigma^2)$$

ve burada;

$$\sigma^2 = E_F \left[\psi(x; \theta)^2 \right] / E_F \left[\psi'(x; \theta) \right]^2 = E_F \left[IF(X; T(F))^2 \right]$$

olur.

Ek 8: Etki Fonksiyonunun Beklenen Değeri

$\int f(x; \theta) dx = 1$ olmak üzere düzgünlük koşulları (bkz. EK 3) altında bu eşitlik;

$$\int \frac{\delta}{\delta \theta} f(x; \theta) = 0$$

olarak yazılabilir. Bu eşitliğe denk bir eşitlik;

$$\int \frac{\delta}{\delta \theta} (-\log f(x; \theta)) f(x; \theta) dx = 0$$

yazılırsa burada $-\log f(x; \theta) = \rho(x; \theta)$ olur ve $\psi(x; \theta) = \frac{\delta}{\delta \theta} \rho(x; \theta)$ olduğundan;

$$\int \psi(x; \theta) f(x; \theta) dx = 0$$

$$E(\psi(x; \theta)) = 0$$

elde edilir (Godambe, 1960: 1210).