

155957

**REGRESYON MODELİNDE
KUVVET DÖNÜŞÜMLERİNİN
MONTE-CARLO DEĞERLENDİRMESİ**

Hüseyin GÜLER

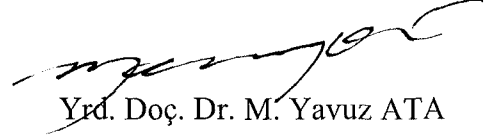
**YÜKSEK LİSANS TEZİ
(İSTATİSTİK)**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

Temmuz 2004

ANKARA

Hüseyin GÜLER tarafından hazırlanan REGRESYON MODELİNDE KUVVET DÖNÜŞÜMLERİNİN MONTE-CARLO DEĞERLENDİRMESİ adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.


Yrd. Doç. Dr. M. Yavuz ATA
Tez Yöneticisi

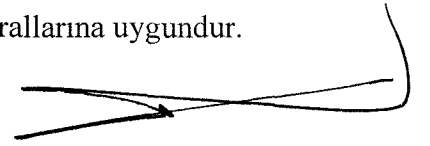
Bu çalışma, jürimiz tarafından İstatistik Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Müslim EKİCİ 

Üye : Prof. Dr. Ayden Ünsal 

Üye : Yrd. Doç. Dr. Mustafa Y. ATA 

Bu tez, Gazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.



İÇİNDEKİLER

	Sayfa
ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ÇİZELGELERİN LİSTESİ	iv
ŞEKİLLERİN LİSTESİ	v
EKLERİN LİSTESİ	vi
SİMGELER VE KISALTMALAR	vii
1. GİRİŞ	1
2. KUVVET DÖNÜŞÜMLERİ	4
2.1. Görelî Çarpıklık	5
2.2. Tukey Dönüşümü	8
2.3. Box-Cox Dönüşümü	9
2.4. Manly Dönüşümü	11
2.5. John-Draper Dönüşümü	12
2.6. Bickel-Doksum Dönüşümü	13
2.7. Yeo-Johnson Dönüşümü	14
3. BAĞIMLI DEĞİŞKENE KUVVET DÖNÜŞÜMÜ UYGULANMIŞ REGRESYON MODELİ	16
3.1. En Çok Olabilirlik Yöntemi	18
3.1.1. Ağ taraması tekniği	20

4. MONTE-CARLO SANAL DENEY	23
4.1. Modeller	23
4.1.1. Normallikten ayrılma	23
4.1.2. Türdeş olmama	27
4.2. Normallik ve Türdeşlik Testleri	28
4.2.1. Lilliefors Normallik Testi	28
4.2.2. Bartlett Türdeşlik Testi	29
4.3. Monte-Carlo Sanal Deney Algoritması	30
4.4. Monte-Carlo Deney Sonuçları.....	38
4.4.1. Normallikten ayrılma	38
4.4.2. Türdeşlikten ayrılma	44
5. SONUÇ VE ÖNERİLER.....	47
KAYNAKLAR	50
EKLER	53
ÖZGEÇMİŞ	64

**REGRESYON MODELİNDE
KUVVET DÖNÜŞÜMLERİNİN
MONTE-CARLO DEĞERLENDİRMESİ
(Yüksek Lisans Tezi)**

Hüseyin GÜLER

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

Temmuz 2004

ÖZET

Klasik regresyon modelinde, hata terimlerinin birbirinden bağımsız, aynı ve normal dağıldığı varsayılır. Bu çalışmada hata terimlerinin aynı ve normal dağıldığı varsayımlarının sağlanmadığı durumlar modellenmiş ve böyle durumlarda uygulanan kuvvet dönüşümlerinin bu varsayımları sağlama başarımları Monte-Carlo sanal bir deneyle karşılaştırılmıştır. Sonuçlar, Tukey dönüşümünün normallikten ve Manly dönüşümünün de türdeşlikten ayrılma durumunda diğerlerine göre daha iyi olduklarını göstermiştir.

Bilim Kodu : 212
Anahtar Kelimeler : Regresyon; kuvvet dönüşümleri; Monte-Carlo tahmin
Sayfa Adedi : 64
Tez Yöneticisi : Yrd. Doç. Dr. Mustafa Yavuz ATA

**MONTE-CARLO EVALUATION OF
POWER TRANSFORMATIONS
IN REGRESSION MODEL
(M. Sc. Thesis)**

Hüseyin GÜLER

**GAZİ UNIVERSITY
INSTITUTE OF SCIENCE AND TECHNOLOGY
July 2004**

ABSTRACT

In the classical regression model, it is assumed that the error terms are independently, identically and normally distributed. In this study, the cases where the assumptions of normality and homoskedasticity are not valid were modelled and the performances of the power transformations were compared by a Monte-Carlo virtual experiment. The results showed that Tukey's transformation and Manly's transformation performed better than others, respectively in case of the departure from normality and from homoskedasticity.

Science Code : 212
Key Words :Regression; power transformation; Monte-Carlo estimation
Page Number : 64
Adviser : Ass. Prof. Dr. Mustafa Yavuz ATA

TEŞEKKÜR

Bu çalışmanın ortaya çıkmasında büyük emeği geçen ve bana her zaman güvenen tez danışmanım Yrd. Doç. Dr. Mustafa Yavuz ATA'ya teşekkürü bir borç bilirim. Ayrıca istatistik bilimini sevmeme ve öğrenmeme neden olan Gazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik bölümündeki tüm hocalarıma teşekkür ederim.

Bana sürekli destek olan ve çalışma boyunca çıkan aksaklıklara rağmen beni teselli edip moralimi düzelter, çalışmam için beni teşvik eden müstakbel eşim Dr. Ebru ÖZGÜR'e sonsuz teşekkürler.

Eğitim hayatım boyunca en iyi noktaya gelmem için desteklerini hiçbir zaman esirgemeyen ve fedakarlık yapmaktan hiç tereddüt etmeyen babam Ferhat ve annem Gülsen GÜLER'e ne kadar teşekkür etsem azdır. Yine her zaman yanımda olan ve akademik hayata adım atmam konusunda bana örnek olan ağabeyim üst teğmen Musa Hakan GÜLER'e ve her zaman beni neşelendirip güldüren biricik kardeşim Dilek GÜLER'e çok teşekkürler.

Ayrıca Gazi Üniversitesi Ekonometri bölümündeki tüm hocalarım ve çalışma arkadaşlarıma teşekkürü bir borç bilirim. Uzun süren gece ve gündüzler boyunca bilgisayarlarını simülasyon çalışmam için kullanmama izin veren Arş. Gör. Yeliz YALÇIN, Arş. Gör. Furkan EMİRMAHMUTOĞLU ve Arş. Gör. Nükhet DOĞAN'a çok teşekkür ederim. Çalışma arkadaşlarım Arş. Gör. Hasan TÜRE ve Arş. Gör. Gaye ÇATALBAŞ'a sürekli destekleri için, oda arkadaşlarım Arş. Gör. Yeliz YALÇIN ve Arş. Gör. Nükhet DOĞAN'a da bana destek olup kahrımı çektikleri için çok teşekkür ederim. Yine simülasyon aşamasında Gazi Üniversitesi, İktisadi ve İdari Bilimler Fakültesi bilgisayar laboratuvarlarını kullanmama izin veren dekan yardımcısı ve hocam Yrd. Doç. Dr. Nezir KÖSE'ye teşekkürü bir borç bilirim. Ayrıca çalışmalarım boyunca bana anlayış gösteren bölüm başkanım ve hocam Prof. Dr. Aydın ÜNSAL'a teşekkür ederim.

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 4.1. α ve k değerleri için çarpık dağılımlı hata teriminin çarpıklık katsayıları.....	25
Çizelge 4.2. Çok tepe değerli hata teriminin simülasyonunda kullanılan f ve σ değerleri.....	27
Çizelge 4.3. Adım 1, 2, 3 ve 5 sonunda elde edilen değerler.....	36
Çizelge 4.4. Çarpık dağılımlı hata terimleri için Monte-Carlo tahminleri.....	39
Çizelge 4.5. Çok tepe değerli hata terimleri için Monte-Carlo tahminleri.....	42
Çizelge 4.6. Türdeş olmayan hata terimleri için Monte-Carlo tahminleri.....	45

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Y için içbükey (Y^2) ve dışbükey ($\sqrt{Y}$) bir dönüşüm örneği.....	6
Şekil 2.2. (0;2) aralığındaki tekdüze dağılıma Y^2 dönüşümü uygulandığında elde edilen dağılım.....	7
Şekil 2.3. (0;2) aralığındaki tekdüze dağılıma $\sqrt{Y}$ dönüşümü uygulandığında elde edilen dağılım.....	8
Şekil 2.4. $\lambda = 2$ ve $\lambda = 0,5$ için Box-Cox dönüşümü.....	10
Şekil 2.5. $\lambda = 2$ ve $\lambda = -2$ için Manly dönüşümü.....	11
Şekil 2.6. $\lambda = 2$ ve $\lambda = -2$ için John-Draper dönüşümü.....	12
Şekil 2.7. $\lambda = 2$ ve $\lambda = 0,5$ için Bickel-Doksum dönüşümü.....	13
Şekil 2.8. $\lambda = 2$ ve $\lambda = 0,1$ için Yeo-Johnson dönüşümü.....	14
Şekil 4.1. Monte-Carlo sanal deneyin akış diyagramı.....	31

EKLERİN LİSTESİ

Ek	Sayfa
Ek-1 Regresyon Modelinde Dönüşümlerin Jacobienleri.....	54
Ek-2 Gamma Dağılımı İçin Çarpıklık Katsayısı.....	55
Ek-3 [4.3a] Rassal Değişkeninin Beklenen Değeri.....	57
Ek-4 TURBO C Kodu.....	58



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklama
$OI(X)$	X'in gerçekleşme olasılığı
λ	Dönüşüm parametresi
ω	İşaret fonksiyonu
$E(.)$, μ	Beklenen değer
$V(.)$, σ^2	Değişke (Varyans)
$X \sim N(a,b)$	X, a ortalama ve b değişke ile normal dağılır
$X \sim \text{Gamma}(a,b)$	X, a ve b parametreleri ile gamma dağılır
$J(.)$	Jacobien
η	Ağ taraması tekniğinde adım uzunluğu
I	Birim matris
μ_3	Ortalama etrafındaki 3. moment
α_3	Çarpıklık katsayısı

Kısaltmalar	Açıklama
TUK	Tukey dönüşümü
BC	Box-Cox dönüşümü
MAN	Manly dönüşümü
JD	John-Draper dönüşümü
BD	Bickel-Doksum dönüşümü
YJ	Yeo-Johnson dönüşümü

1. GİRİŞ

Regresyon analizi ve varyans analizi gibi istatistiksel yöntemlerin çoğunda, bağımlı değişkenin, açıklayıcı değişken veya değişkenlere olan bağımlılığı, incelenen olayı betimleyen bir model oluşturularak analiz edilmektedir. Bu modellerin hepsinde hata terimi ve bu terimin dağılımına dair çeşitli varsayımlar mevcuttur. Özel olarak doğrusal regresyon modelinde, parametrelerin tahmini için; hata terimlerinin birbirleri ve açıklayıcı değişkenlerle ilişkisiz¹, 0 ortalamalı bir dağılıma sahip oldukları varsayılır. Ayrıca etkin parametre tahminleri elde etmek için hata terimleri üzerinde sabit değişke (varyans) varsayımı, istatistiksel çıkarsama yapmak amacıyla da hata terimlerinin dağılımları üzerinde normal dağılım varsayımı yapılır. Ancak bu varsayımların tamamıyla geçerli olduğu durumlarla uygulamada çok nadir karşılaşılır. Örneğin zaman serisi verilerinde hata terimlerinde otokorelasyon; kesit verilerinde ise değişen değişke sorununun bulunması, bir istisnadan çok bir kuraldır. Benzer şekilde normal dağılıma ilişkin varsayımın sağlanmadığı durumlarla da oldukça sık karşılaşılır.

Bu varsayımların sağlanmaması durumunda izlenebilecek yöntemleri Graybill (1) şu şekilde özetlemiştir:

- (i) Varsayımdan sapmaları önemsemeyip, analize tüm varsayımlar sağlanıyormuş gibi devam etmek.
- (ii) Sağlanmayan varsayımın yerine geçecek doğru bir varsayım geliştirmek ve bu yeni varsayımı sağlayan bir yöntem uygulamak.
- (iii) Orijinal modelin gerekli özelliklerini taşıyan ve geçerliliği sağlanan varsayımlara dayalı yeni bir model tasarlamak. Bunun için verilere uygun bir dönüşüm uygulanabilir ya da kuşkulu uç değerlerin etkisi giderilmeye çalışılabilir.
- (iv) Varsayımların sağlanmadığı durumlarda bile geçerliliğini sürdüren dağılımdan bağımsız yöntemleri uygulamak.

¹ Hata terimlerinin birbirlerinden ilişkisiz olmaları, en çok olabilirlik yöntemi için gerekli olup, en küçük kareler yönteminde bu varsayıma ihtiyaç yoktur.

Sakia’da (2) belirtildiği gibi çoğu araştırmacı uygulamadaki sorunlarını bu yöntemleri kullanarak çözmeye çalışmıştır. Bir çok araştırmacı da (iii) numaralı yöntem üzerinde çalışmış ve bu sayede varsayımları sağlayan ve kendi verilerine uygun modeller kullanmıştır. Özellikle değişkenleri dönüştürerek modelin varsayımlarını sağlamaya çalışmak oldukça sık başvurulmuş bir yöntemdir. Bu bağlamda literatürde önerilen önemli dönüşümler; Tukey (3), Box-Cox (4), Manly (5), John-Draper (6), Bickel-Doksum (7) ve Yeo-Johnson (8) olarak verilebilir. Ancak bu dönüşümler, tek başlarına, varsayımlardan sapmaların tümünü giderememektedirler. Box-Cox dönüşümü ve genel olarak kuvvet dönüşümleri, “normallikten” ve “türdeşlikten (sabit değişkeden)” ayrılma sorunlarının giderilmesinde etkili olup (4, 9), hata terimlerinin bağımlı olduğu durumlarda etkili olamamaktadırlar. Spitzer’in (10), Box-Cox dönüşümünün küçük örnek çaplarındaki davranışını inceleyen çalışması dışında, bu dönüşümlerin varsayımlardan sapmaları ne ölçüde giderdiklerine ve hangi durumda hangi dönüşümün daha üstün olduğuna dair ayrıntılı bir inceleme bulunmamaktadır. İşte bu çalışmada, adı geçen dönüşümlerin, regresyon modelinin varsayımlarının sağlanmaması durumundaki başarımları incelenecek ve hangi durumda hangi dönüşümün daha iyi sonuç verdiği bulunmaya çalışılacaktır. Normal dağılım varsayımının sağlanmadığı durumda, dönüşüm sonucu normalliğin sağlanma oranı her bir dönüşüm için elde edilecektir. Ayrıca türdeşliğin sağlanmadığı durumlarda, dönüşümlerin türdeşliği sağlama oranları yine her bir dönüşüm için elde edilecek ve bu sayede dönüşümleri karşılaştırabilmek mümkün olacaktır. Bu işlemler, normallik ve türdeşliğin sağlanmadığı durumlarda farklı örnek çapları için tekrarlanacak, böylece örnek çapının dönüşümlerin başarımları üzerindeki etkisi de kıyaslanacaktır.

Bu amaçla çalışmanın ikinci bölümü olan “Kuvvet Dönüşümleri” bölümünde, öncelikle normal dağılımla ilişkili olan “görelî çarpıklık” tanımı verilecek, ardından Tukey, Box-Cox, Manly, John-Draper, Bickel-Doksum ve Yeo-Johnson dönüşümleri incelenecektir. Öncelikle bu dönüşümlerin yapılarından bahsedilecek, daha sonra çarpıklık durumundaki davranışları incelenecek ve teorik anlamda birbirlerine olan üstünlükleri tartışılacaktır.

Üçüncü bölümde, regresyon modelinde kuvvet dönüşümlerinin nasıl kullanıldığı açıklanacak, dönüşüm uygulanmış model parametrelerinin tahmini için literatürde öngörülen yöntemlere değinilecektir. Ardından çalışmanın amacına en uygun yöntem olduğu düşünülen ve dördüncü bölümündeki Monte-Carlo uygulamada kullanılacak olan “en çok olabilirlik yöntemi” ile “ağ taraması tekniği” ele alınacaktır.

Çalışmanın dördüncü bölümünde “Monte-Carlo Deneyi” verilecektir. Dönüşümlerin karşılaştırıldığı bu bölümde, ilk olarak deneyde kullanılacak modeller tanıtılacaktır. Bunlar sırasıyla,

- normallikten ayrılma ve
- türdeşlikten ayrılma

durumlarının gerçekleştiği modellerdir. Ardından her bir dönüşüm için varsayımlardan sapmaların ne ölçüde giderildiği ayrı ayrı incelenecek ve en iyi sonucu veren dönüşüm belirlenmeye çalışılacaktır.

“Sonuç ve Öneriler” kısmında ise “Monte-Carlo Deneyi” ile bulunan sonuçlar özetlenecek, araştırmacılara varsayımlardan sapmalardaki farklı durumlar için hangi dönüşümü kullanmanın daha uygun olduğu önerilecektir. Ayrıca daha ileriki çalışmalara ilişkin çeşitli öneriler de bu bölümde yer alacaktır.

2. KUVVET DÖNÜŞÜMLERİ

Her istatistiksel tahmin tekniği yığına ilişkin çeşitli varsayımlar içerir ve bu varsayımların geçerliliği altında söz konusu teknik uygulanıp sonuçlar yorumlanır. Örneğin yığın ortalamasının belli bir değere eşit olup olmadığını sınırlarken, ölçülen değişkenin yığın değerlerinin genellikle normal dağıldığı varsayılır. Araştırmacı bu varsayımın geçerliği altında deneyi tasarlar, verileri elde eder, sınamayı gerçekleştirir ve seçtiği anlamlılık düzeyinde sonuçlarını yorumlar. Uygulanan tekniğin, ancak varsayımların geçerli olduğu durumda, anlamlı ve tutarlı sonuçlar vereceği açıktır.

Bunun farkında olan araştırmacılar kullandıkları tekniğin varsayımlarının geçerli olup olmadığını istatistiksel yöntemlerle sınırlar ve ardından tekniğin varsayımları geçerli ise bu tekniği uygulamaya devam ederler. Aksi halde araştırmacı, daha az varsayım gerektiren ya da daha farklı varsayımları olan bir teknik uygulayabilir. Ancak bazen araştırmacının amacına uygun alternatif bir teknik bulması o kadar kolay olmayabilir. Özellikle incelenen konu karmaşık olduğunda daha az veya daha farklı varsayımları olan bir teknik bulmak oldukça güçtür. Örneğin bazı çok değişkenli analizlerde değişkenlerin çok değişkenli normal dağılıma uydukları varsayılır. Eğer eldeki veri çok değişkenli normal dağılım varsayımını sağlamıyorsa yeni bir çok değişkenli analiz tekniği bulmak zordur. Bu durumda araştırmacı verilerini tercih ettiği tekniğin varsayımlarını sağlayacak şekilde dönüştürebilir.

Verilere dönüşüm uygulama yaklaşımı oldukça sık başvuru alan yöntemlerden birisidir. Eğer modelin varsayımları sağlanmıyorsa, doğrusal olmayan dönüşümler uygulanarak varsayımların geçerli olduğu bir model kurmak mümkündür. Aslında verileri dönüştürmek, sadece varsayımları sağlayacak şekilde bir dönüşüm uygulamaktan ziyade, analizlerin geçerli olmasını sağlayan bir yeniden ölçeklemedir (4).

Literatürde dönüşümlerle ilgili pek çok çalışma bulunmaktadır. Bartlett'in (11) çalışmasında dönüşümlerle ilgili geniş bir açıklama yer almakta, dönüşümlerin ne

amaçla kullanılabileceği belirtilmektedir. Bartlett'e göre dönüşümler özellikle sabit bir hata değişkesi elde etmek için kullanılır. Dönüşümler, özellikle incelenen değişkenin değişkesi, binom ve Poisson dağılımlarında olduğu gibi, ortalamasının bir fonksiyonu olduğunda türdeş hata terimleri elde etmede kullanılmaktadır.

Doğrusal regresyon modelinde de hata terimleri ile ilgili varsayımlar mevcuttur. Bunlar hata terimlerinin 0 ortalamalı, türdeş ve birbirlerinden bağımsız olmalarıdır. Ayrıca parametrelerle ilgili hipotez testleri yapmak ve güven aralıkları oluşturmak için hata terimlerinin normal dağıldıkları da varsayılır. Giriş bölümünde belirtildiği gibi kuvvet dönüşümleri, normallikten ve türdeşlikten ayrılma durumlarında etkili olmakta ve bu dönüşümler sayesinde varsayımları sağlayan bir model elde edilmektedir.

Bu bölümde varsayımları sağlayan bir regresyon modeli elde etmede kullanılabilecek olan dönüşümlerden bahsedilecektir. *Görelî çarpıklık*, normallikten ayrılma durumlarından birisi olan simetrik olmama durumunda dönüşümlerin davranışlarını açıklamaya yardımcı olacak bir kavramdır. Bu nedenle, önce görelî çarpıklık kavramının açıklanması gerekir.

2.1. Görelî Çarpıklık

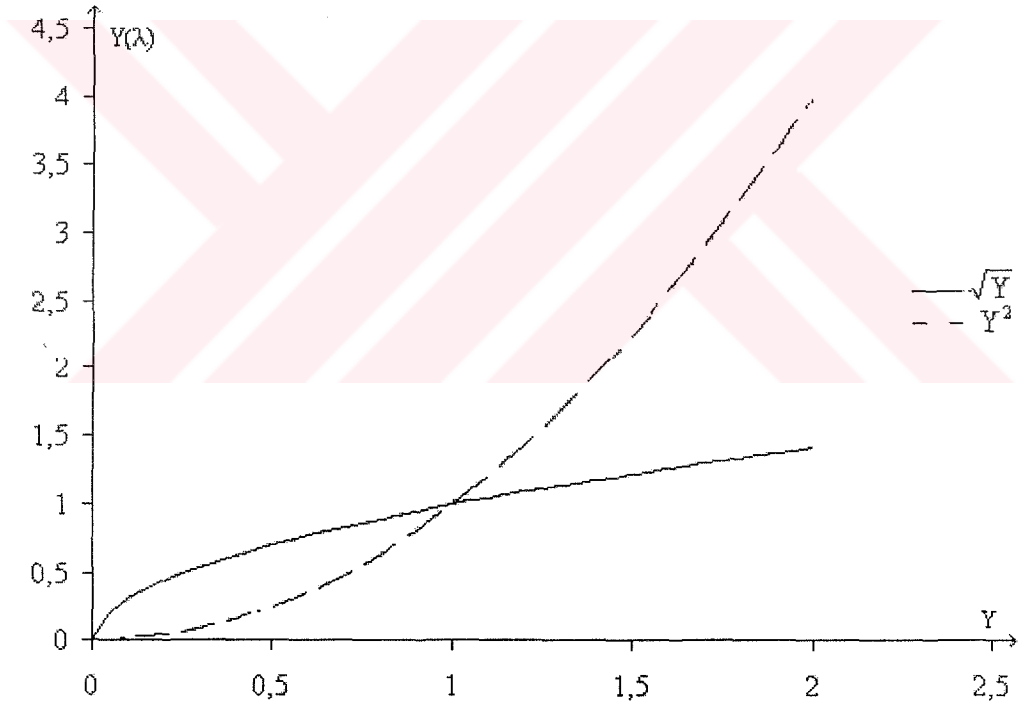
Çarpık bir dağılımın simetrik bir dağılıma dönüştürülmesinde uygulanacak dönüşümün seçiminde, van Zwet'in (12) görelî çarpıklık kavramından yararlanılabilir (8).

Y ; sürekli dağılım fonksiyonu F ve bu fonksiyonun tersi F^{-1} olan bir rassal değişken olsun. I_F ise $Ol(Y \in I_F) = 1$ olasılığını sağlayacak en dar aralık olsun. Bu durumda, eğer $\forall y \in I_F$ için $F(y_0 + y) + F(y_0 - y) = 1$ ise, F dağılımı y_0 etrafında simetriktir. Z ise sürekli dağılım fonksiyonu G ve bu fonksiyonun tersi G^{-1} olan başka bir rassal değişken olsun. $Y(\lambda) = G^{-1}\{F(y)\}$ olarak tanımlansın. Bu durumda $Y(\lambda)$ 'nin dağılım fonksiyonu, Z 'nin dağılım fonksiyonu olan G olur. Buna göre, $G^{-1}F$ 'nin içbükey

(dışbükey) olması için gerek ve yeter koşul G 'nin, Y 'nin azalmayan içbükey (dışbükey) dönüşümü olan $Y(\lambda)$ 'nin dağılım fonksiyonu olmasıdır. Bu tanımların ışığında, görelî çarpıklık aşağıdaki gibi tanımlanabilir:

Tanım: $G^{-1}\{F(.)\}$ azalmayan içbükey (dışbükey) bir fonksiyon ise, dağılım fonksiyonu G , dağılım fonksiyonu F 'den daha çok sağa (sola) çarpıktır.

Bu durum Şekil 2.1'deki örnekle gösterilebilir. Burada negatif olmayan Y değişkeninin içbükey dönüşümü olarak, Y^2 ; dışbükey dönüşümü olarak, $\sqrt{Y}$ kullanılmıştır.

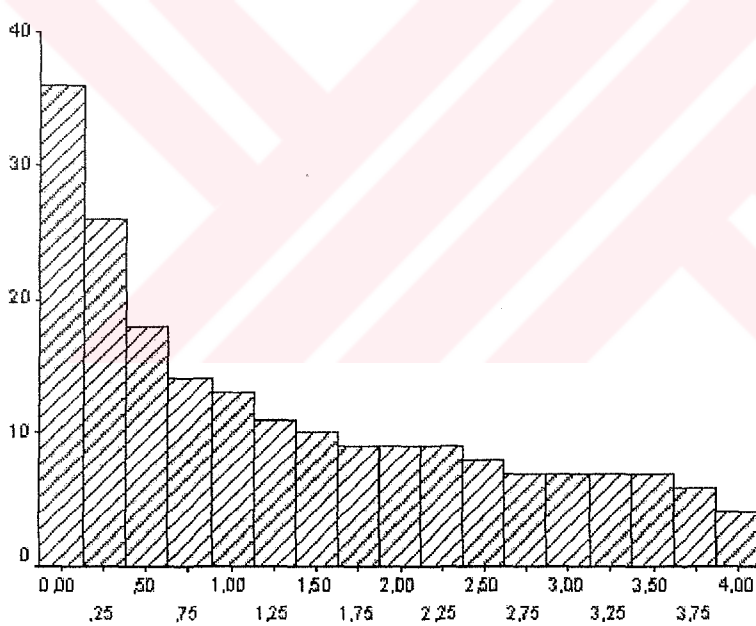


Şekil 2.1. Y için içbükey (Y^2) ve dışbükey ($\sqrt{Y}$) bir dönüşüm örneği.

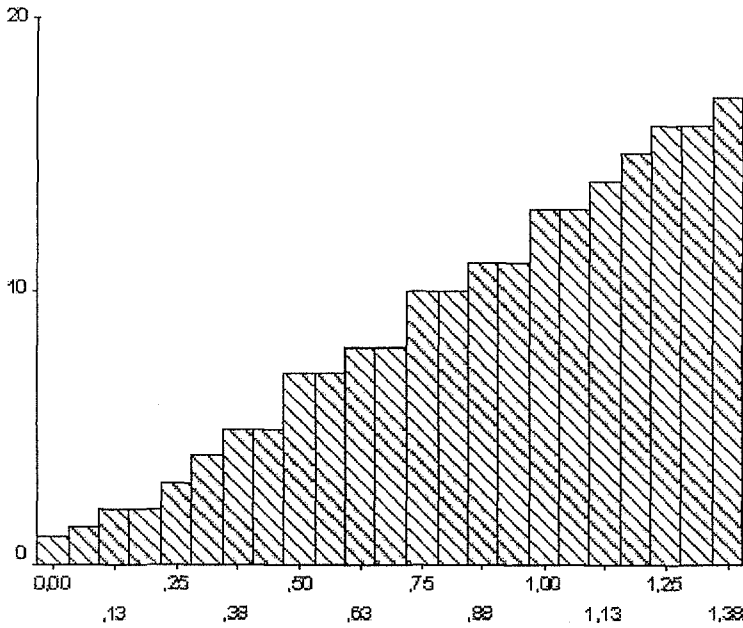
van Zwet'in görelî çarpıklık tanımı, Şekil 2.1 aracılığı ile daha kolay anlaşılabilir. Şekilden görüleceği üzere, içbükey bir dönüşümün; orijinal değişkenin büyük değerleri üzerindeki etkisi, küçük değerlere olan etkisine oranla daha çoktur. Bir başka deyişle içbükey dönüşüm, değişkenin büyük değerlerini küçük değerlerine göre daha büyük oranda büyütür. Bu durumun bir sonucu olarak,

içbükey bir dönüşümle dönüştürülmüş bir değişkenin dağılımının sağ kuyruğunun yayılması ve sol kuyruğunun daralması, dolayısıyla dönüştürülmüş değişkenin dağılımının, orijinal değişkene göre daha çok sağa çarpık olması beklenir. Dışbükey dönüşüm için bunun tam tersi söylenebilir. Dışbükey dönüşüm sonucunda, dönüştürülmüş değişkenin dağılımının sağ kuyruğu daralacak ve sol kuyruğu uzayacaktır. Böylece dönüştürülmüş değişken, orijinal değişkene göre daha çok sola çarpık olacaktır.

Bu durum, $(0;2)$ aralığındaki tekdüze bir dağılıma, Şekil 2.1’de verilen dönüşümlerin uygulanması sonucu elde edilen ve Şekil 2.2 ile Şekil 2.3’te verilen histogramlarda görülmektedir.



Şekil 2.2. $(0;2)$ aralığındaki tekdüze dağılıma Y^2 dönüşümü uygulandığında elde edilen dağılım.



Şekil 2.3. (0;2) aralığındaki tekdüze dağılıma $\sqrt{Y}$ dönüşümü uygulandığında elde edilen dağılım.

Bu bilgilerin ışığı altında, eğer eldeki değişkenin dağılımı sola çarpık ise bu değişkene içbükey bir dönüşüm uygulandığında, elde edilen değişkenin dağılımının daha az sola çarpık, dolayısıyla daha simetrik olması beklenir. Benzer şekilde değişkenin dağılımı sağa çarpıksa, bu değişkene dışbükey bir dönüşüm uygulanarak sağa çarpıklık giderilebilir yorumu yapılabilir.

İzleyen altbölümlerde literatürde yer alan bazı dönüşümler ve bunların, başta çarpıklık durumundaki davranışları olmak üzere, çeşitli özellikleri incelenecektir.

2.2. Tukey Dönüşümü

Önerilen kuvvet dönüşümlerden ilki Tukey (3) dönüşümüdür. Bu dönüşüm $Y_i > 0$ için;

$$Y_i(\lambda) = \begin{cases} Y_i^\lambda & , \quad \lambda \neq 0 \\ \log_e Y_i & , \quad \lambda = 0 \end{cases} \quad [2.1]$$

şeklinde tanımlanmaktadır. Bu dönüşümün önemli eksikliği $\lambda = 0$ noktasında sürekli olmamasıdır. Temeli Tukey dönüşümüne dayanan Box-Cox dönüşümü ise bu durumda da geçerlidir.

2.3. Box-Cox Dönüşümü

Box ve Cox dönüşümü, Tukey dönüşümü

$$Y_i(\lambda) = \begin{cases} (Y_i^\lambda - 1)/\lambda & , \quad \lambda \neq 0 \\ \log_e Y_i & , \quad \lambda = 0 \end{cases} \quad [2.2]$$

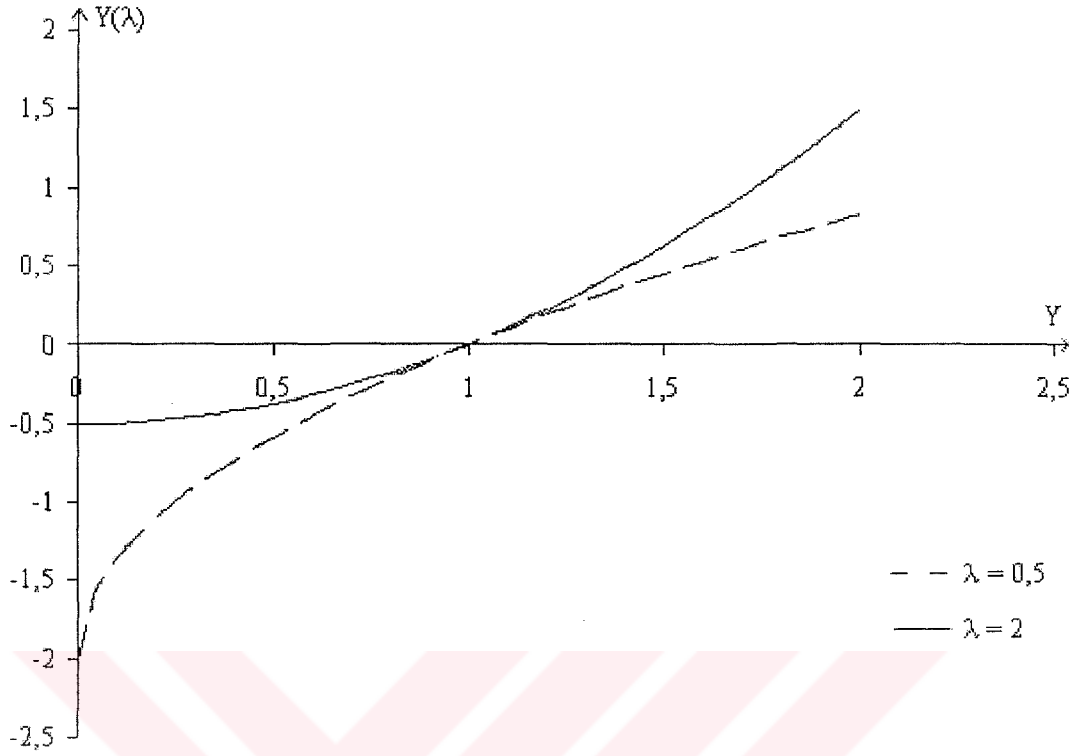
biçiminde yeniden düzenlenerek elde edilebilir (4).

Bu biçimi ile Box-Cox dönüşümü, $\lambda = 0$ noktasında sürekli olmasına karşın, $Y_i > 0$ için geçerlidir. Bu kısıdın olmadığı biçimi ise,

$$Y_i(\lambda) = \begin{cases} [(Y_i + \lambda_2)^{\lambda_1} - 1]/\lambda_1 & , \quad \lambda_1 \neq 0 \\ \log_e (Y_i + \lambda_2) & , \quad \lambda_1 = 0 \end{cases} \quad [2.3]$$

olarak tanımlanır.

Burada λ_1 dönüşüm parametresidir. λ_2 ise tüm i değerleri için $Y_i > -\lambda_2$ olacak şekilde seçilir ve böylece tüm Y değerleri pozitif olur. [2.2] ve [2.3]'te verilen Box-Cox dönüşümleri, dönüşüm parametresi 1'den büyük olduğunda içbükey, 1'den küçük olduğunda dışbükey bir yapıya sahiptir. Bu durum Şekil 2.4'te gösterilmektedir.



Şekil 2.4. $\lambda = 2$ ve $\lambda = 0,5$ için Box-Cox dönüşümü.

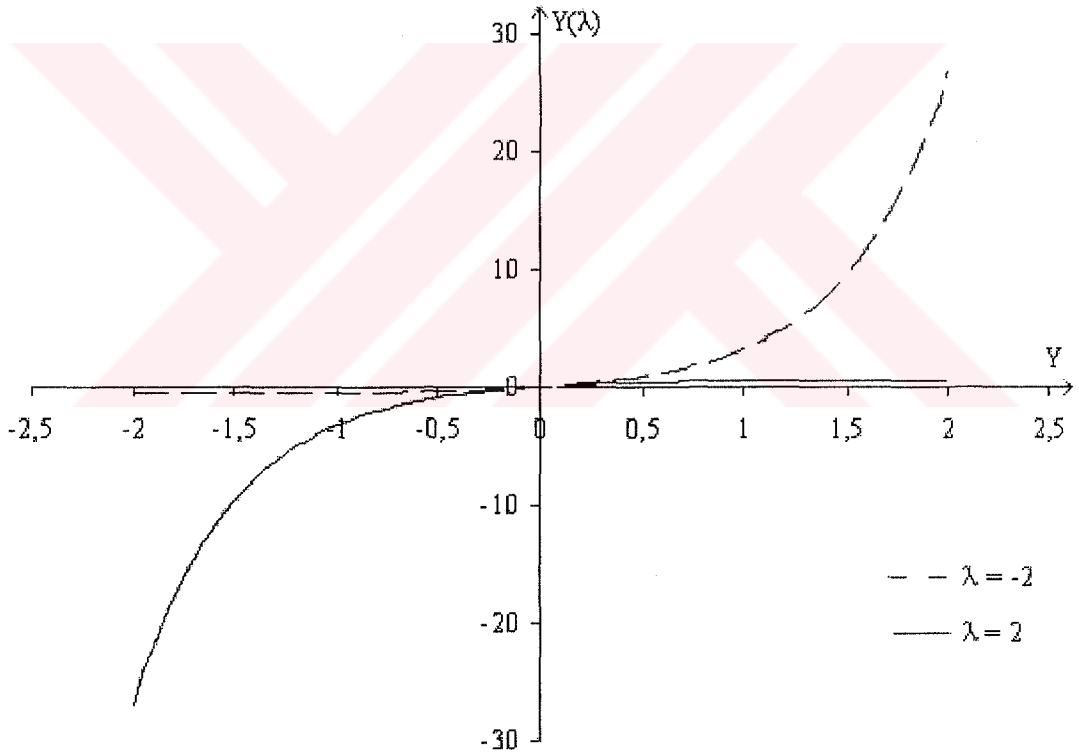
Her ne kadar bu dönüşüm negatif değerli değişkenlerde de kullanılabilse de, dönüşüm sonrası elde edilen değişken, $\lambda_1 = 0$ olmadığı sürece $(-\infty, +\infty)$ aralığında değerler alamaz. Daha açık bir ifadeyle, $\lambda_1 > 0$ için $Y(\lambda) > -1/\lambda_1$ ve $\lambda_1 < 0$ için de $Y(\lambda) < -1/\lambda_1$ olur. Normal dağılmış bir değişken $(-\infty, +\infty)$ aralığında değişirken, dönüştürülmüş değişken değerleri sınırlı bir aralıkta değiştiğinden, dönüştürülmüş dağılım ancak “yaklaşık normal” sayılabilir ve bu nedenle varsayımların tam olarak sağlandığı söylenemez. Bu durumda kullanılabilecek alternatif yöntemlerden birisi (13)’de verilmiştir. Ancak yine de (2)’de de belirtildiği gibi, Box-Cox dönüşümü varsayımları geçerli kılmak için yaygın olarak kullanılan yöntemlerden birisidir.

2.4. Manly Dönüşümü

Negatif değerler alabilen değişkenleri dönüştürmek amacıyla bir diğer dönüşüm ise Manly (5) tarafından önerilmiştir. Manly'nin önerdiği dönüşüm;

$$Y_i(\lambda) = \begin{cases} (e^{\lambda Y_i} - 1)/\lambda & , \quad \lambda \neq 0 \\ Y_i & , \quad \lambda = 0 \end{cases} \quad [2.4]$$

şeklinde olup, negatif değerler de alabilen ve tek tepe değeri olan çarpık dağılımlı bir değişkeni, simetrik ve normale yakın bir dağılıma sahip yeni bir değişkene dönüştürdüğü kabul edilir (2). Manly'nin dönüşümü, dönüşüm parametresi pozitif iken dışbükey, negatif iken içbükeydir. Bu durum Şekil 2.5'te gösterilmektedir.



Şekil 2.5. $\lambda = 2$ ve $\lambda = -2$ için Manly dönüşümü.

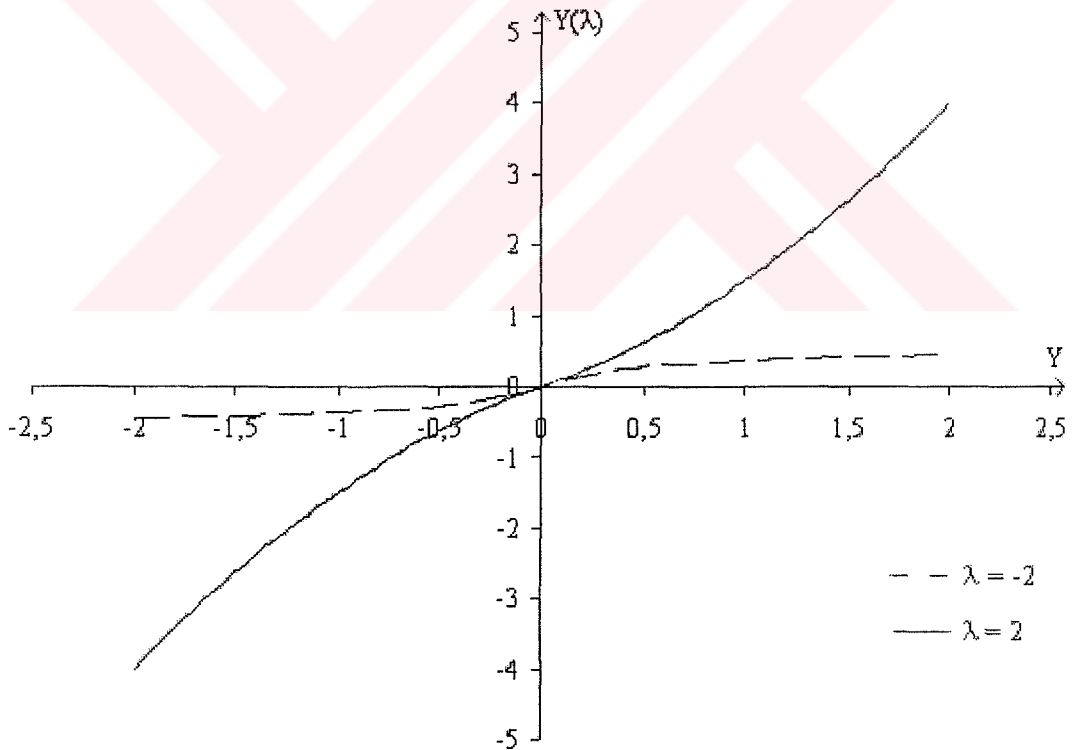
2.5. John-Draper Dönüşümü

$$\omega_i = \begin{cases} 1 & , & Y_i > 0 \\ -1 & , & Y_i < 0 \end{cases}$$

şeklinde tanımlanan işaret fonksiyonu olsun. John ve Draper (6), bir miktar simetrikliğe sahip değişkenleri normalleştirdiği düşünülen,

$$Y_i(\lambda) = \begin{cases} \omega_i \left[(|Y_i| + 1)^\lambda - 1 \right] / \lambda & , & \lambda \neq 0 \\ \omega_i \left[\log_e (|Y_i| + 1) \right] & , & \lambda = 0 \end{cases} \quad [2.5]$$

dönüşümünü önermişlerdir. Fakat bu dönüşüm, orijinal değişkenin simetrikliği doğrultusunda başarılıdır ve simetriklik azaldıkça etkisini yitirir. John-Draper dönüşümüne ilişkin grafik Şekil 2.6'da verilmektedir.



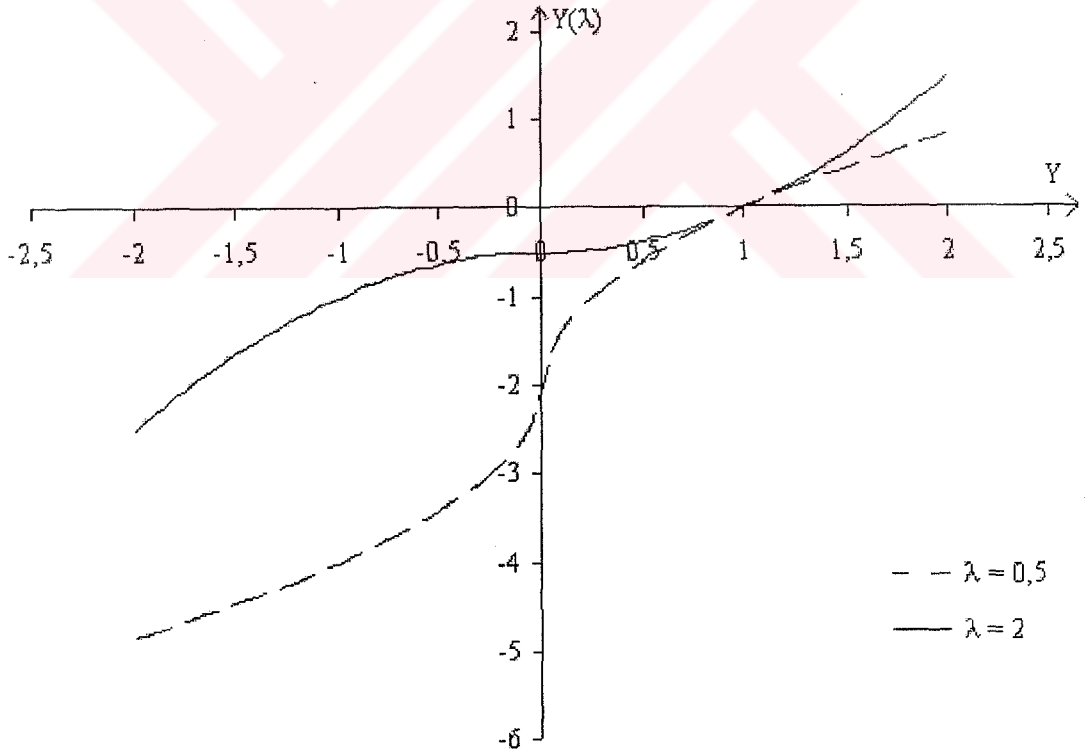
Şekil 2.6. $\lambda = 2$ ve $\lambda = -2$ için John-Draper dönüşümü.

2.6. Bickel-Doksum Dönüşümü

Bickel ve Doksum'un (7) önerdiği dönüşüm ise dönüştürülmüş değişkenin değişim aralığı üzerinde bir sınırlama yapmaz ve [2.6]'da verilen kalıba uyar.

$$Y_i(\lambda) = (\omega_i |Y_i|^\lambda - 1) / \lambda, \quad \lambda > 0 \quad [2.6]$$

Bu dönüşüm dağılımdaki basıklığı düzeltmekle beraber, simetrik olmayan dağılımlara uygulandığında, dönüştürülmüş değişken çift tepe değere sahip olduğundan, uygun bir çözüm yöntemi olmaz. Bunun nedeni, orijinal veri işaret değiştirirken dönüşümün, içbükeyden dışbükeye (ya da tam tersi) geçmesidir. Bu durum Şekil 2.7'de gösterilmektedir.



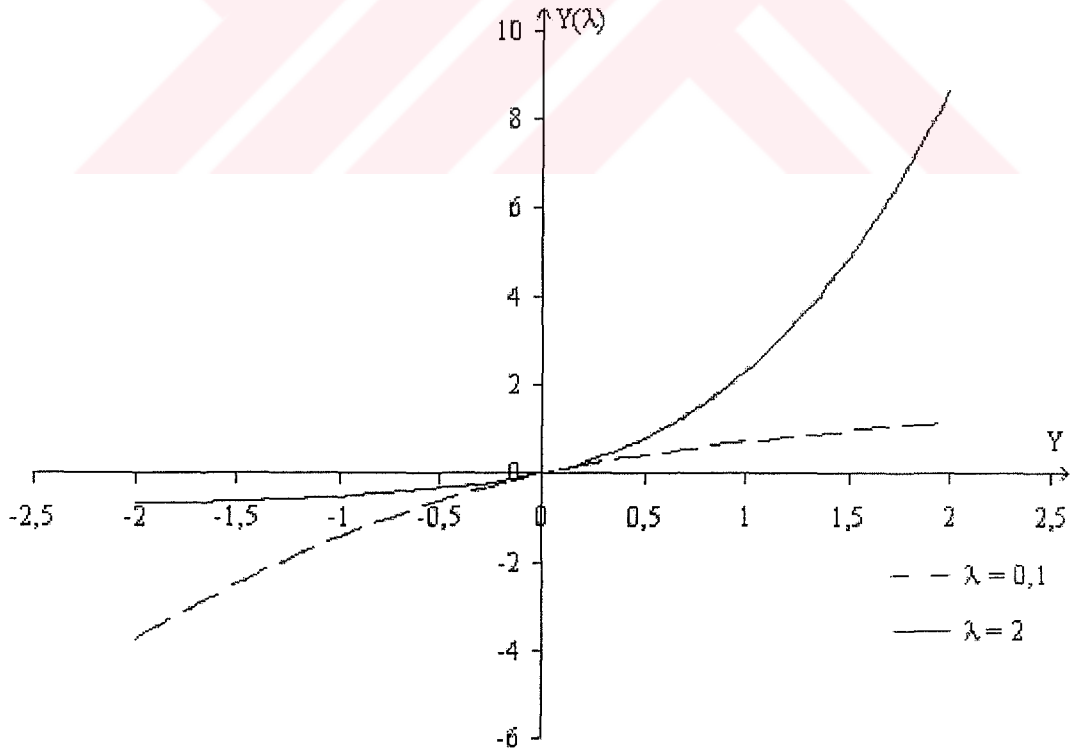
Şekil 2.7. $\lambda = 2$ ve $\lambda = 0,5$ için Bickel-Doksum dönüşümü.

2.7. Yeo-Johnson Dönüşümü

Yeo ve Johnson'ın (8) önerdiği dönüşüm ise, başlangıçta pozitif ve negatif değerler için iki farklı parametreye sahipken, çeşitli kısıtlamalar konulduğunda [2.7]'de verilen tek parametrelili biçim elde edilir.

$$Y_i(\lambda) = \begin{cases} [(Y_i + 1)^\lambda - 1]/\lambda & , & Y \geq 0, \lambda \neq 0 \\ \log_e(Y_i + 1) & , & Y \geq 0, \lambda = 0 \\ -[(-Y_i + 1)^{2-\lambda} - 1]/(2-\lambda) & , & Y < 0, \lambda \neq 2 \\ -\log_e(-Y_i + 1) & , & Y < 0, \lambda = 2 \end{cases} \quad [2.7]$$

Bu dönüşümün önemli özellikleri, hem pozitif hem negatif değerler için tanımlı olması, dönüştürülmüş değişkenin $(-\infty, +\infty)$ aralığında değerler alması ve özellikle çarpık dağılıma sahip değişkenlerde iyi sonuç vermesidir. İçbükey ve dışbükeyliği Box-Cox dönüşümünününe benzeyen [2.7]'deki dönüşüm Şekil 2.8'de gösterilmiştir.



Şekil 2.8. $\lambda = 2$ ve $\lambda = 0,1$ için Yeo-Johnson dönüşümü.

Özet olarak Tukey, Box-Cox ve John-Draper dönüşümlerinde, dönüştürülmüş değişkenin değişim aralığı kısıtlıdır, bu yüzden dönüştürülmüş değişken için sadece yaklaşık normallik varsayımı öne sürülebilir. Bickel-Doksum dönüşümü ise çarpıklık durumunda, dönüştürülmüş değişkenin çift tepe değerli olmasından dolayı, çözümden çok bir sorundur. Bu durumda incelenen dönüşümler içinde, çarpık dağılımlı ve negatif değerler alabilen değişkenler için uygun dönüşümlerin Manly ve Yeo-Johnson dönüşümleri olduğu söylenebilir.

Öte yandan kuvvet dönüşümleri ve özellikle Box-Cox dönüşümü türdeşliği sağlamada sıkça kullanılır (14). Regresyon modelinde bağımlı değişkenin dağılımı dikkate alınarak değişkenler dönüştürüldüğünde, türdeş hata terimleri elde edilebilir.

Bir sonraki bölümde bu dönüşümlerin uygulandığı durumlar için parametre tahmin yöntemleri incelenecektir.

3. BAĞIMLI DEĞİŞKENE KUVVET DÖNÜŞÜMÜ UYGULANMIŞ REGRESYON MODELİ

Regresyon modelinin hem bağımlı hem de açıklayıcı değişkenlerinin dönüştürülmesi düşünülebilir. Ancak bağımlı değişkenin ve hata terimlerinin dağılımının biçimi, açıklayıcı değişken(ler)den etkilenmemektedir (15). Bu çalışmada, hata terimlerine ilişkin varsayımlar inceleneceğinden, yalnızca bağımlı değişkene dönüşüm uygulanmış modeli incelemek yeterlidir.

Bağımlı değişkenin gözlem değerlerinden oluşan n elemanlı yöney (vektör) $\mathbf{Y} = (Y_1 \ Y_2 \ \dots \ Y_n)'$; $k-1$ tane açıklayıcı değişkenin n tane gözlem değerini içeren dizey (matris),

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,k-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2,k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,k-1} \end{bmatrix};$$

kuvvet dönüşümü parametresi λ ; dönüştürülmüş bağımlı değişken yöneyi $\mathbf{Y}(\lambda) = (Y_1(\lambda) \ Y_2(\lambda) \ \dots \ Y_n(\lambda))'$; dönüştürülmüş modelin regresyon katsayılar yöneyi $\boldsymbol{\beta} = (\beta_1 \ \beta_2 \ \dots \ \beta_k)'$ ve hata terimleri yöneyi $\boldsymbol{\varepsilon} = (\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n)'$ olmak üzere, bağımlı değişkenin dönüştürülmüş değerlerini içeren regresyon modeli,

$$\mathbf{Y}(\lambda) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad [3.1]$$

şeklinde yazılabilir. Bu modelde, uygun dönüşüm parametresi seçildiğinde, $E(\varepsilon_i \varepsilon_j) = 0, (i \neq j)$ ve $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ olduğu varsayılır (4, 16, 17, 18).

Literatürde bu varsayımları göz önüne alan birkaç tahmin yöntemi önerilmiştir. Bu yöntemlerden en önemlileri, *en çok olabilirlik yöntemi*, *Bayesgil tahmin yöntemi* ve *Kullback-Leibler bilgi kriteri yöntemi* (2).

En çok olabilirlik ve Bayesgil tahmin yöntemleri, Box ve Cox tarafından önerilmiş (4), daha sonra üzerlerinde çeşitli uyarlamalar yapılmıştır (2). Bu yöntemlerin, genel

olarak birbirlerine denk oldukları (4)'te gösterilmiştir. Ancak açıklayıcı değişken sayısı fazla olduğunda, yöntemlerin denklikleri bozulmaya başlar. Box ve Cox, bu durumda Bayesgil yöntemin kullanılmasını önermektedirler. Fakat Bayesgil yöntem, teorik yapısının karmaşıklığı ve hesaplama zorluğundan dolayı uygulamada pek tercih edilmemektedir. Buna karşın, en çok olabilirlik yöntemi, Box ve Cox'ın çalışmasına sonradan yapılan katkılarla, yaygın kullanım olanağı bulmuştur.

Bir diğer tahmin yöntemi ise, Kullback-Leibler bilgi kriterini kullanan Hernandez ve Johnson'ın önerdiği yöntemdir (19). Kullback-Leibler bilgi kriteri, dönüştürülmüş değişkenin dağılımı ile, hedeflenen normal dağılım arasındaki sapmayı ölçen bir kriterdir. Buradan yola çıkan Hernandez ve Johnson, Box-Cox dönüşümü için, dönüştürülmüş hata terimlerinin dağılımı ile hedef normal dağılım arasındaki sapmayı minimum yapacak λ 'yı seçmeyi önermişlerdir. Hernandez ve Johnson, Box ve Cox'ın en çok olabilirlik ve Bayes tahmin edicilerinin, asimptotik olarak Kullback-Leibler tahmin edicisine yakınsadığını göstermişlerdir. Ancak önerdikleri yöntem orijinal değişkenin dağılımını bilmeyi gerektirir ve uygulamada genellikle bu dağılım bilinmez. Bu durum, Kullback-Leibler tahmin yönteminin uygulanmasını güçleştirmektedir.

Yukarıdaki açıklamaların ışığında, anlaşılması ve uygulaması kolay, uygulamada yaygın olarak kullanılan ve çoğu yeni paket programda mevcut olan en çok olabilirlik yöntemi, kuvvet dönüşümü içeren regresyon modelinin parametre tahminlerini elde etmede uygun bir araç olarak görülmektedir. Bu yüzden bu çalışmada parametre tahminleri en çok olabilirlik yöntemi kullanılarak elde edilecektir.

3.1. En Çok Olabilirlik Yöntemi

[3.1] numaralı regresyon modelinde, hata terimlerinin 0 ortalama ve sabit değişke ile normal dağıldıkları varsayıldığından, i _inci hata terimi için olasılık yoğunluk fonksiyonu;

$$f_{\varepsilon_i}(\varepsilon_i; \sigma_\varepsilon^2) = \begin{cases} \frac{1}{\sqrt{\sigma_\varepsilon^2 2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2} \varepsilon_i^2}, & -\infty < \varepsilon_i < +\infty \quad \sigma_\varepsilon > 0 \\ 0, & \text{dd} \end{cases} \quad [3.2]$$

şeklinde. Öte yandan, hata terimleri birbirlerinden bağımsız olduklarından, ortak olasılık yoğunluk fonksiyonu için

$$f_\varepsilon(\varepsilon; \sigma_\varepsilon^2) = \frac{1}{(\sigma_\varepsilon^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma_\varepsilon^2} \varepsilon' \varepsilon} \quad [3.3]$$

eşitliği yazılabilir. Regresyon modelinde, açıklayıcı değişken değerlerinin sabit olduğu durumda, bağımlı değişkenin dağılımının biçimi, hata terimlerinin dağılımının biçimi ile aynıdır (15). Bu yüzden, [3.1] numaralı regresyon modeli için, uygun dönüşüm parametresi seçildiğinde, dönüştürülmüş bağımlı değişkenin dağılımı normal olur. Açıklayıcı değişken değerleri sabit olarak alındığında, dönüştürülmüş bağımlı değişkenin ortalaması için $E(\mathbf{Y}(\lambda) | \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$ ve değişkesi için $V(\mathbf{Y}(\lambda) | \mathbf{X}) = \sigma_\varepsilon^2$ eşitlikleri geçerli olur.

Tüm bu ifadeler dikkate alındığında, dönüştürülmüş bağımlı değişken için ortak olasılık yoğunluk fonksiyonu;

$$f_{\mathbf{Y}(\lambda)}(\mathbf{Y}(\lambda); \mathbf{X}, \lambda, \boldsymbol{\beta}, \sigma_\varepsilon^2) = \frac{1}{(\sigma_\varepsilon^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y}(\lambda) - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y}(\lambda) - \mathbf{X}\boldsymbol{\beta})} \quad [3.4]$$

şeklinde yazılabilir. Bu fonksiyon yardımıyla olabilirlik fonksiyonu türetilmekle birlikte, bağımlı değişkenin dönüştürülmüş değerleri değil, orijinal değerleri gözlenmektedir. Bu yüzden [3.4] ifadesi kullanılarak, orijinal değişkenler için ortak olasılık yoğunluk fonksiyonu ve bu fonksiyon yardımıyla olabilirlik fonksiyonu türetilir. Kuvvet dönüşümünün Jacobien'ini $J(\lambda; \mathbf{Y})$ ile gösterirsek,

$J(\lambda; \mathbf{Y}) = \left| \frac{\partial \mathbf{Y}(\lambda)}{\partial \mathbf{Y}} \right| = \prod_{i=1}^n \frac{\partial Y_i(\lambda)}{\partial Y_i}$ yazılabilir. Bu çalışmada geçen dönüşümlerin Jacobienleri EK-1’de verilmiştir. Bu durumda, bağımlı değişkenin orijinal değerleri cinsinden olabilirlik fonksiyonu, [3.4]’te verilen fonksiyonun, dönüşümün Jacobien’i ile çarpılmasıyla;

$$l(\lambda, \beta, \sigma_\varepsilon^2 | \mathbf{Y}, \mathbf{X}) = \frac{1}{(\sigma_\varepsilon^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y}(\lambda) - \mathbf{X}\beta)' (\mathbf{Y}(\lambda) - \mathbf{X}\beta)} |J(\lambda; \mathbf{Y})| \quad [3.5]$$

olarak elde edilir (4, 20). Olabilirlik fonksiyonu monoton bir fonksiyon olduğundan, bu fonksiyonu maksimum yapan parametre değerleri ile fonksiyonun logaritmasını maksimum yapan parametre değerleri birbirlerine eşittir. Bu yüzden, [3.5] ifadesinin e tabanında logaritması alınarak, daha basit olan [3.6] numaralı log-olabilirlik fonksiyonu elde edilir:

$$L(\lambda, \beta, \sigma_\varepsilon^2 | \mathbf{Y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y}(\lambda) - \mathbf{X}\beta)' (\mathbf{Y}(\lambda) - \mathbf{X}\beta) + \ln |J(\lambda; \mathbf{Y})| \quad [3.6]$$

[3.6] gibi doğrusal olmayan fonksiyonları basitleştirmek için, Lawton ve Sylvestre (21), doğrusal parametrelerin fonksiyondan elenmesini önermiş, Spitzer (22) ise Box-Cox dönüşümü içeren regresyon denkleminde bu yöntemi kullanmıştır. Bu yöntemde, doğrusal olmayan parametre değerlerinin bilindiği varsayılır ve doğrusal parametreler bu değerlere bağlı olarak tahmin edilir. Dolayısıyla [3.6] fonksiyonu için doğrusal olmayan λ parametresinin bilindiği varsayılır ve bu fonksiyonun doğrusal parametreler olan β ve σ_ε^2 ’a göre türevleri alınırsa,

$$\begin{aligned} \frac{\partial L(\beta, \sigma_\varepsilon^2 | \mathbf{Y}, \mathbf{X}, \lambda)}{\partial \beta'} &= -\frac{1}{2\hat{\sigma}_\varepsilon^2} (-2\mathbf{X}'\mathbf{Y}(\lambda) + 2\mathbf{X}'\mathbf{X}\hat{\beta}) \\ \frac{\partial L(\beta, \sigma_\varepsilon^2 | \mathbf{Y}, \mathbf{X}, \lambda)}{\partial \sigma_\varepsilon^2} &= -\frac{n}{2} \frac{1}{\hat{\sigma}_\varepsilon^2} + \frac{1}{2(\hat{\sigma}_\varepsilon^2)^2} (\mathbf{Y}(\lambda) - \mathbf{X}\hat{\beta})' (\mathbf{Y}(\lambda) - \mathbf{X}\hat{\beta}) \end{aligned} \quad [3.7]$$

denklem sistemi elde edilir. Log-olabilirlik fonksiyonunun maksimumu bulunmak istendiğinden, [3.7]’de verilen eşitlikler sıfıra eşitlenir ve

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}(\lambda) \quad [3.8]$$

$$\hat{\sigma}_e^2 = \frac{1}{n} (\mathbf{Y}(\lambda) - \mathbf{X}\hat{\beta})' (\mathbf{Y}(\lambda) - \mathbf{X}\hat{\beta}) \quad [3.9]$$

elde edilir. [3.9]'da verilen hata değişkesi tahmini, [3.6]'da yerine yazılırsa aşağıdaki fonksiyon elde edilir:

$$L(\beta, \sigma_e^2 | \mathbf{Y}, \mathbf{X}, \lambda) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \hat{\sigma}_e^2 - \frac{n}{2} + \ln |J(\lambda; \mathbf{Y})| \quad [3.10]$$

[3.10] ile ifade edilen fonksiyon *profil olabilirlik fonksiyonu* adını alır ve bu fonksiyon, log-olabilirlik fonksiyonundan daha basit bir fonksiyondur. Profil olabilirlik fonksiyonunun daha basit bir yapıya sahip olması parametre tahminlerinin bulunmasını kolaylaştırmaktadır. Eğer [3.10]'da verilen fonksiyonu maksimum yapan β ve λ değerleri bulunabilirse, tahmin problemi çözülmüş olur. Diğer taraftan [3.10] fonksiyonu sadece bir tek doğrusal olmayan parametre içermektedir. Bu yüzden [3.10]'un maksimumu bulunurken *ağ taraması tekniği* kullanılabilir.

3.1.1. Ağ taraması tekniği

Maksimumu aranan fonksiyon bir tek doğrusal olmayan parametre içerdiğinde ağ taraması tekniği hem hesaplama hem de maksimumu bulma açısından kullanışlı bir tekniktir. Bu teknik maksimum noktanın içinde bulunduğu kabul edilen bir aralıkta belli sayıda nokta seçip, bu noktalarda fonksiyonu incelemeye dayanır. Kim'de (23) verilen ağ taraması tekniği, aşağıdaki gibi tanımlanabilir.

Fonksiyonun maksimumunun arandığı aralık $A = [a, b]$ olsun. $x_0 = a$, $x_n = b$ ve adım uzunluğu η ile ifade edilsin. Bu durumda fonksiyonun aldığı değerlerin inceleneceği noktalar $x_0, x_1, \dots, x_n$ olarak tanımlanabilir. Her bir nokta bir öncekinden η kadar uzaklıkta olacak şekilde seçildiğinde $x_k = x_{k-1} + \eta$ ($1 \leq k \leq n$) yazılabilir. Burada tanım gereği $n = (x_n - x_0)/\eta$ eşitliği geçerlidir. η parametresi ne kadar küçük olursa, fonksiyonun o kadar çok noktadaki değeri inceleneceğinden, bu

parametre tahminin hassaslığını belirleyecektir. Bu yüzden adım uzunluğunun küçük seçilmesi uygun olacaktır.

Tarama aralığı ve adım uzunluğu belirlendikten sonra maksimumu aranan $f(x)$ fonksiyonunun $x_0, x_1, \dots, x_n$ noktalarındaki değerleri yani $f(x_0), f(x_1), \dots, f(x_n)$ hesaplanır. Ardından maksimum $f(x_i)$ 'yi sağlayan i^* değeri bulunur ve fonksiyonun maksimumunun $f(x_{i^*})$ olduğuna karar verilir.

Bu yaklaşım, [3.10]'da verilen fonksiyonun maksimumunu bulmada kullanılabilir. [3.8] numaralı eşitlikte verilen $\hat{\beta}$, [3.9]'da yerine yazılır ve buradan elde edilen $\hat{\sigma}_\varepsilon^2$ eşitliği [3.10]'da yerine konulup gerekli işlemler yapıldığında,

$$L(Y, X, \lambda) = -\frac{n}{2} \ln(2\pi) + \frac{n}{2} \ln n - \frac{n}{2} + \ln |J(\lambda; Y)| - \frac{n}{2} \ln \left(Y'(\lambda) [I - X(X'X)^{-1} X'] Y(\lambda) \right) \quad [3.11]$$

fonksiyonu elde edilir. Burada I , $n \times n$ boyutlu birim matristir. n değeri tahmin sürecinde sabit olacağından [3.11] fonksiyonundaki ilk 3 terim ihmal edilirse [3.12]'de verilen fonksiyon elde edilir.

$$\varphi(\lambda) = \ln |J(\lambda; Y)| - \frac{n}{2} \ln \left(Y'(\lambda) [I - X(X'X)^{-1} X'] Y(\lambda) \right) \quad [3.12]$$

[3.12] fonksiyonu sadece λ 'ya bağlı olduğundan bu fonksiyonun maksimumu ağ taraması tekniği kullanılarak bulunabilir. Bu amaçla kullanılabilecek bir algoritma aşağıda verilmektedir:

Başlangıç Adımı:

$\eta > 0$ olacak şekilde η değeri seçilir. Maksimumu barındırdığı düşünülen aralık $A = [a, b]$ olmak üzere $\lambda_0 = a$, $\lambda_n = b$ ve $\lambda_k = \lambda_{k-1} + \eta$ ($1 \leq k \leq n$) olarak atanır. Ardından tarama adımına geçilir.

Tarama Adımı:

Başlangıç adımında belirlenen $\lambda_0, \lambda_1, \dots, \lambda_n$ değerleri [3.12]'de yerine yazılarak $\varphi(\lambda_0), \varphi(\lambda_1), \dots, \varphi(\lambda_n)$ değerleri hesaplanır. Ardından maksimum $\varphi(\lambda_i)$ 'yi sağlayan i^* değeri bulunur ve [3.12] fonksiyonunun maksimumunun $\varphi(\lambda_{i^*})$ olduğuna karar verilir.

Algoritma sonlandırıldığında λ 'nın tahmin değeri λ_{i^*} olarak belirlenir. Bu $\hat{\lambda}$ değeri kullanılarak $\hat{\beta}$ ve $\hat{\sigma}_\varepsilon^2$ tahminleri [3.8] ve [3.9] formülleri aracılığıyla elde edilir. Böylece bağımlı değişkeni dönüştürülmüş regresyon modelinin parametreleri tahmin edilmiş olur.

Bir sonraki bölümde, bu bölümde incelenen en çok olabilirlik yöntemi ve ağ taraması tekniği kullanılarak, ikinci bölümde incelenen dönüşümlerin normallikten ve türdeşlikten ayrılma durumlarındaki davranışları Monte-Carlo simülasyon tekniği kullanılarak analiz edilecektir.

4. MONTE-CARLO SANAL DENEY

Bu bölümde dönüştürülmüş modele ilişkin hata terimlerinin normalliğine ve türdeş olmalarına ilişkin varsayımların geçerlilik kazanıp kazanmadığı incelenecektir. Bu amaçla öncelikle deney noktalarının tanımlandığı modeller verilecek, ardından varsayımların sınanmasında kullanılan istatistiksel testler üzerinde durulacaktır. Daha sonra Monte-Carlo deneyinde kullanılacak algoritma ve yakınsama kıstası verilecektir. Son olarak her bir model için Monte-Carlo deney sonuçlarına göre dönüşümlerin başarımları karşılaştırılacaktır.

4.1. Modeller

Sanal deney verilerinin üretileceği modeller, hata terimlerinin

- normal dağılıma sahip olmadıkları, “normallikten ayrılma” ve
- değişkelerinin sabit olmadığı, “türdeş olmama”

durumlarını kapsayacak şekilde seçilecektir.

4.1.1. Normallikten ayrılma

Regresyon analizindeki amaçlardan biri, parametre tahminlerini kullanarak, parametrelerin aldıkları değerlere ilişkin hipotez testleri yapmak veya güven aralıkları oluşturmaktır. Bu amacı gerçekleştirmek için parametre tahminlerinin dağılımı hakkında bir bilgiye ihtiyaç duyulur. Doğrusal regresyon modelinde hata terimlerinin normal dağıldıkları varsayıldığında, parametre tahminlerinin dağılımının da normal olacağı bilinmektedir (24).

Dolayısıyla hata terimlerinin normallikten ayrıldığı durumlarda, parametrelerle ilgili hipotezlerin testinde ve güven aralıklarının oluşturulmasında normal dağılım kullanılamaz. Bunun yerine uygun bir dönüşüm yapılır ve hata terimlerinin normal dağılmaları sağlanabilir. Ardından hipotez testleri yapılabilir ve güven aralıkları oluşturulabilir.

Hata terimlerinin normallikten ayrıldığı durumlarda dönüşümlerin davranışlarını incelemek ve uygun dönüşümü önermek amacıyla bu çalışmada

- sağa veya sola çarpık ya da
- çok tepe değerli dağılımlar

kullanılacaktır.

4.1.1.1 Çarpık dağılımlı hata terimleri

Dönüşümlerin başarısını etkilemede orijinal değişkenlerin çarpıklığının önemli olduğu düşünüldüğünden, bu çalışmada farklı çarpıklık durumlarında dönüşümlerin davranışları incelenecektir. Bu amaçla çeşitli çarpıklık katsayıları için sağa ve sola çarpık hata terimleri üretilecek ve bu hata terimlerine bağlı olarak dönüşümlerin başarıları karşılaştırılacaktır.

Farklı çarpıklık katsayılarına sahip hata terimleri elde etmek için Gamma dağılımı kullanılabilir. α ve β parametreleri ile Gamma dağılımına sahip bir X rassal değişkeni için olasılık yoğunluk fonksiyonu;

$$f_X(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0, \alpha, \beta > 0 \\ 0, & \text{dd} \end{cases} \quad [4.1]$$

şeklinde tanımlanabilir. [4.1] ile ifade edilen olasılık yoğunluk fonksiyonuna sahip $X \sim \text{Gamma}(\alpha, \beta)$ rassal değişkeni için $E(X) = \alpha\beta$, $V(X) = \alpha\beta^2$ yazılabilir (25).

Ayrıca μ_3 ortalama etrafındaki 3. moment ve $\alpha_3 = \frac{\mu_3}{\sigma^3}$ çarpıklık katsayısı olmak üzere, $\alpha_3 = \frac{2}{\sqrt{\alpha}}$ olduğu gösterilebilir (Bkz. EK-2).

Bu özellik kullanılarak, α 'nın farklı değerleri için, farklı çarpıklık katsayılarına sahip hata terimleri üretilebilir. Bu sayede farklı çarpıklık katsayıları için dönüşümler karşılaştırılabilir ve uygun olan dönüşüm önerilebilir.

Ancak Gamma dağılımının yukarıda verilen çarpıklık katsayısı daima pozitif olacağından, bu yolla ancak sağa çarpık hata terimleri üretilebilir. Sola çarpık hata

terimleri üretmek için uygulanabilecek bir yöntem, sağa çarpık olarak üretilen Gamma değişkeninin değerlerini -1 ile çarpma işlemidir. Bu çarpma işlemi sonucunda elde edilen yeni rassal değişkenin çarpıklık katsayısı mutlak değerce orijinal değişkeninki ile aynı olacak ancak çarpıklığın yönü sola doğru olacaktır. Bu sayede sola çarpık hata terimleri üretilmiş olacaktır.

Bu yöntemde karşılaşılabilecek bir diğer sorun ise üretilen terimlerin ortalamalarının 0 olmamasıdır. Hata terimlerinin 0 ortalamalı olmaları gerektiğinden, üretilen rassal değişkenlerden teorik ortalamaları çıkartılarak, 0 ortalamalı hata terimleri üretilir.

Ayrıca dönüşümler karşılaştırılırken sadece çarpıklık katsayısının değişmesi, hata terimlerinin değişkesi gibi diğer deney parametrelerinin sabit tutulması doğru olacaktır. Ancak α parametresinin aldığı değerler değiştiğinde Gamma dağılımlı hata terimlerinin değışkeleri de doğrudan değışecektir. Bu amaçla her bir α parametresi için farklı β parametresi değeri seçilerek hata terimlerinin değışkelerinin sabit olması sağlanmıştır. Deney boyunca hata terimlerinin değışkeleri 25 değerde tutulmuştur.

Tüm bu unsurlar göz önüne alındığında, hata terimlerinin üretilmesinde [4.2] ile verilen modelin kullanılması uygun görülmüştür.

$$\varepsilon_i \sim (-1)^k \text{Gamma}\left(\alpha, \frac{5}{\sqrt{\alpha}}\right) - 5\sqrt{\alpha}(-1)^k \quad [4.2]$$

Seçilen α ve k değeri ile çarpıklık katsayıları Çizelge 4.1’de özetlenmiştir.

Çizelge 4.1. α ve k değeri için çarpık dağılımlı hata teriminin çarpıklık katsayıları.

k	1					2				
α	0,16	0,25	0,44	1	4	4	1	0,44	0,25	0,16
α_3	-5	-4	-3	-2	-1	1	2	3	4	5

4.1.1.2. Çok tepe değerli hata terimleri

Hata terimlerinin normal olmamasına neden olabilecek bir diğer durum ise, bu terimlerin birden fazla tepe değerlerinin olmasıdır. İşte bağımlı değişkenin birden fazla tepe değerine sahip olduğu durumlarda, hata terimleri de birden fazla tepe değerine sahip olabilecek, bu ise normal dağılım varsayımının geçerli olmasını engelleyecektir. Bu durumda yine modelde kuvvet dönüşümleri uygulanabilir ve normal dağılım varsayımının geçerliliği sağlanabilir. Bu bölümde, 2 tepe değerli hata terimlerinin simülasyonunda kullanılacak bir model tanımlanacak ve Monte-Carlo deney düzeni verilecektir.

ε_1 ve ε_2 aynı şekilli fakat farklı ortalamalı normal dağılımlı rassal değişkenler olsun. Bu durumda $\varepsilon_1 \sim N(\mu_1, \sigma^2)$ ve $\varepsilon_2 \sim N(\mu_2, \sigma^2)$ yazılabilir. I ise 0 ve 1 değerlerini alan kesikli tekdüze değişken olarak tanımlansın. Ayrıca I , ε_1 ve ε_2 ’den bağımsız olsun. Bu tanımlar ile 2 tepe değerli hata terimini üretmede kullanılabilecek simülasyon modeli

$$\varepsilon_i = I_i \varepsilon_{1i} + (1 - I_i) \varepsilon_{2i} \quad [4.3a]$$

şeklinde verilir. Bu model ile üretilen hata terimlerinin ortalaması $(\mu_1 + \mu_2)/2$ olacaktır (Bkz. EK-3). Dolayısıyla 0 ortalamalı hata terimleri elde etmek için, [4.3a]’da verilen hata terimlerinden teorik ortalamaları çıkartılmalıdır. Sonuç olarak 2 tepe değerli hata terimleri, [4.3b] modeli ile üretilebilir.

$$\varepsilon_i = I_i \varepsilon_{1i} + (1 - I_i) \varepsilon_{2i} - (\mu_1 + \mu_2)/2 \quad [4.3b]$$

Özel olarak $\mu_1 = \mu_2$ için ε_1 ve ε_2 rassal değişkenleri aynı dağılımlı olacaklarından, [4.3b] modeli ile üretilen hata terimleri tek tepe değerli olacaklardır. Dolayısıyla $f = |\mu_1 - \mu_2|$ farkı arttıkça, ε_1 ve ε_2 değişkenlerinin tepe değerleri birbirinden uzaklaşacak ve çift tepe değerlilik problemi daha da belirgin hale gelecektir. Ayrıca f sabitken, σ^2 ’nin artması da ε_1 ve ε_2 ’nin aldığı değerleri birbirlerine yaklaştıracığından 2 tepe değerinin etkisini azaltıcı bir etken olabilir. Bu yüzden 2 tepe

değerli hata terimlerinin simülasyonunda farklı f ve σ^2 değerleri kullanılmıştır. Kuvvet dönüşümlerinin 2 tepe değerli hata terimlerindeki başarımlarını değerlendirmede Çizelge 4.2’de verilen deney düzeni tercih edilmiştir.

Çizelge 4.2. Çok tepe değerli hata teriminin simülasyonunda kullanılan f ve σ değerleri.

f	0				5				10				15			
σ	0,5	1	5	10	0,5	1	5	10	0,5	1	5	10	0,5	1	5	10

4.1.2. Türdeş olmama

Hata terimlerinin değişkelerinin birbirlerinden farklı olmaları, bu terimlerin türdeş olmadığı anlamına gelir. Eğer hata terimlerinin açıklayıcı değişken(ler)e göre koşullu değişkeleri sabit değilse en küçük kareler tahmin edicileri etkinlik özelliklerini yitirirler. Ayrıca en küçük kareler ile elden edilen değişke tahmin edicileri tutarsız olurlar (26). Bu durumda klasik t ve F testleri geçerliliğini yitirirler (27).

Özellikle yatay kesit verilerinde, modelin açıklama gücü gözlemler arasında farklılık gösterebileceğinden, türdeş olmayan hata terimleriyle karşılaşılması oldukça olasıdır (27). Bu durumda da dönüşümler kullanılarak hata değişkelerinin türdeş olmaları sağlanabilir ve bunun sonucunda sapmasız değişke tahmin edicileri elde edilmiş olur.

Bağımlı değişkenin, dolayısıyla hata teriminin değişkesine ilişkin çeşitli fonksiyonel formlar önerilebilir. [4.4]’teki eşitlik hata terimlerinin koşullu değişkesinin her bir birey için giderek arttığını ifade etmektedir.

$$V(\varepsilon_i | \mathbf{X}_i) = [(n - \delta) + i(\delta - 1)] / (n - 1) \quad [4.4]$$

Burada $\mathbf{X}_i$, i _inci birey için $k-1$ tane açıklayıcı değişkenin gözlem değerlerinden oluşan yöney ve δ değeri bilinen bir sabittir.

[4.4]'de önerilen eşitlik, hata teriminin koşullu değişkesinin, n , i ve δ 'nın bir fonksiyonu olduğunu ifade etmektedir. [4.4] fonksiyonu, gözlem sayısı (n) ve $\delta > 1$ değeri sabit olduğunda hata terimlerinin koşullu değişkesinin giderek arttığını ifade etmektedir. [4.4]'te verilen koşullu değişkenin değeri, $i = 1$ için 1'e; $i = n$ için δ 'ya eşit olmaktadır. Dolayısıyla $\delta = 1$ için hata terimleri türdeş; $\delta > 1$ için türdeş olmayacaklardır. δ 'nın aldığı değer büyüdükçe, türdeş olmama sorununun derecesi de artacaktır.

Monte-Carlo deneyinde hata terimleri 0 ortalamalı ve [4.4] eşitliğinde verilen değişkeli normal dağılım kullanılarak türetilmişlerdir. [4.4]'de kullanılan δ değerleri ise 1, 2, 2,5 ve 3 olarak seçilmiştir.

4.2. Normallik ve Türdeşlik Testleri

Bu çalışmada normal dağılımın testinde Lilliefors, türdeşliğin sınanmasında ise Bartlett testi kullanılacaktır. Test istatistikleri ve testlerin çeşitli özellikleri ilerleyen kısımlarda verilmektedir.

4.2.1. Lilliefors Normallik Testi

Birimlerin geldikleri yığının dağılımının normal dağılıma uyup uymadığını test etmede oldukça sık başvuru alan testlerden birisi Kolmogorov-Smirnov testidir. Bu test parametrik bir test olmayıp, iddia edilen dağılım fonksiyonu ile gözlenen birikimli dağılım fonksiyonu arasındaki farka dayalı bir testtir. Ancak bu testte yığının parametrelerinin (normal dağılımın testinde ortalama ve değişke) aldığı değerlerin bilinmesi gerekir. Eğer bu parametreler bilinmiyorsa Kolmogorov-Smirnov testinin yerine Lilliefors testi kullanılabilir. Lilliefors testi Kolmogorov-Smirnov testine benzer bir test olup, benzer bir test istatistiği ile sınama işlemi gerçekleştirilir. Aradaki tek fark orijinal gözlem değerleri yerine standartlaştırılmış gözlem değerlerinin kullanılmasıdır (28).

[3.1] dönüştürülmüş regresyon denkleminde, hata terimlerinin normal dağılımları amaçlanmaktadır. Dönüşümlerin başarılarını karşılaştırırken, dönüştürülmüş modelin hata terimlerinin normal dağılıp dağılmadıkları sınanabilir. Burada kullanılacak hipotezler,

$$H_0: \varepsilon_i \text{ 'lerin geldiği dağılım normaldir}$$

$$H_1: \varepsilon_i \text{ 'lerin geldiği dağılım normal değildir}$$

şeklindedir. ε_i 'lerin değişkesinin tahmininde,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n \varepsilon_i^2$$

kullanılabilir. Ardından hata terimleri

$$\zeta_i = \varepsilon_i / s$$

ile standartlaştırılır. $F_0(\cdot)$ standart normal dağılım için birikimli dağılım fonksiyonu, $S(\cdot)$ ise örneğin birikimli dağılım fonksiyonu olmak üzere, test istatistiği [4.5]'deki gibi tanımlanabilir.

$$L = \text{eküs}_{\zeta} [F_0(\zeta) - S(\zeta)] \quad [4.5]$$

Burada “eküs” en küçük üst sınır anlamına gelmektedir. Eğer hesaplanan L değeri anlamlı ise, hata terimlerinin normal dağıldıklarını iddia eden H_0 hipotezi reddedilir (29). [4.5] ile hesaplanan L değerinin anlamlı olup olmadığı, Conover’de (29) verilen tablo değerleri kullanılarak tespit edilebilir.

4.2.2. Bartlett Türdeşlik Testi

Literatürde türdeşliğin sınanması için pek çok test önerilmiştir (30, 31, 32). Bu testler içinde en çok bilinen ve kabul görmüş testlerden birisi Bartlett (33) testidir. Bartlett’in önerdiği test, regresyon modeline aşağıdaki gibi uyarlanabilir.

[3.1]'de verilen dönüştürülmüş regresyon modelinde, i _inci gözlem birimi için j _inci deneydeki hata terimi ε_{ij} ile gösterilsin. Bu durumda hata terimlerinin türdeşliğini sınamada kullanılacak hipotezler,

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$$

$$H_1 : \text{En az biri farklıdır}$$

şeklindedir. Burada σ_i^2 , i _inci gözlem birimi için hata teriminin değişkesidir. m_i ise i _inci gözlem birimi için, Bartlett testinde kullanılacak sanal gözlem sayısını belirtsin. Monte-Carlo deneyinde her bir gözlem birimi için sanal gözlem sayısı eşit olacağından $m_1 = m_2 = \dots = m_n = m$ alınabilir. Bu durumda Bartlett'in önerdiği test istatistiği,

$$B = -\frac{m-1}{c} \sum_{i=1}^n \ln(s_i^2/s^2) \quad [4.6]$$

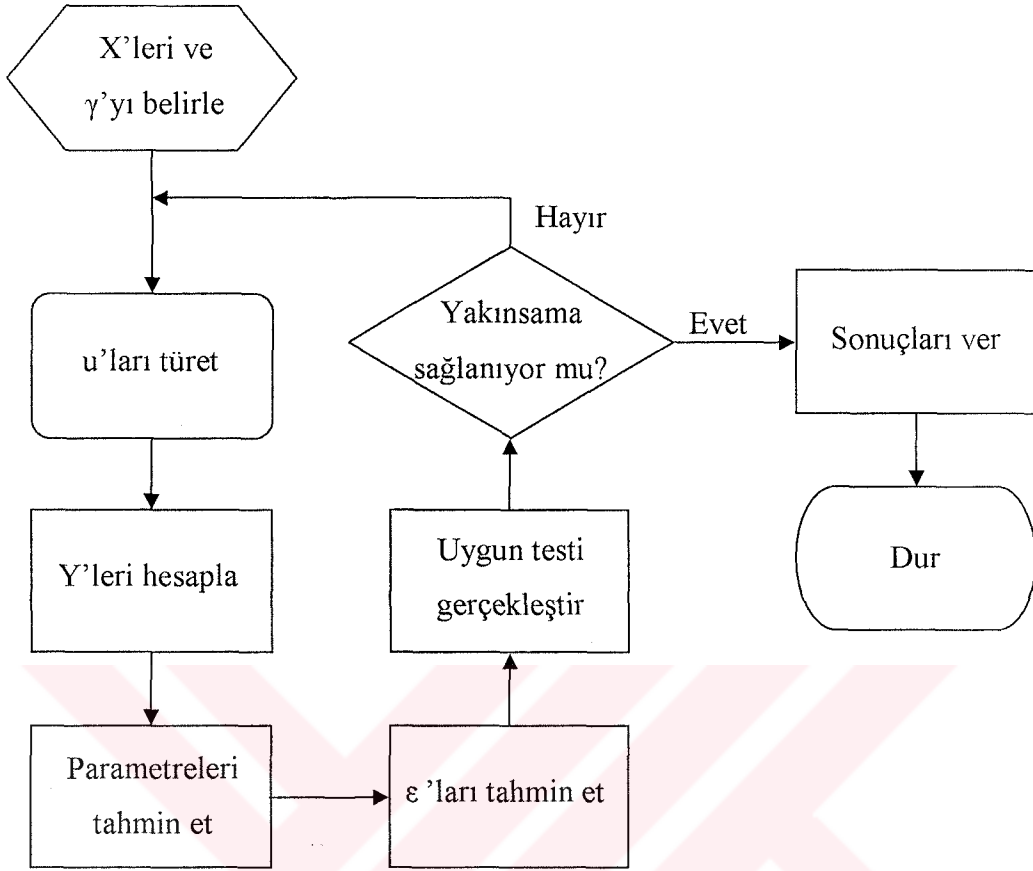
şeklindedir. Burada,

$$s_i^2 = \frac{1}{m-1} \sum_{j=1}^m (\varepsilon_{ij} - \bar{\varepsilon}_i)^2, \quad \bar{\varepsilon}_i = \frac{1}{m} \sum_{j=1}^m \varepsilon_{ij}, \quad c = \frac{3nm - 2n + 1}{3n(m-1)}, \quad s^2 = \frac{1}{n} \sum_{i=1}^n s_i^2$$

ile ifade edilmektedir. H_0 hipotezi geçerli olduğunda, [4.6] ile tanımlanan B istatistiği asimptotik olarak χ_{n-1}^2 dağılımına sahip olacaktır (33, 34). Bu test istatistiğinin, normal dağılım varsayımının sağlandığı durumlarda iyi sonuçlar verdiği yapılan bir Monte-Carlo çalışması ile gösterilmiştir (30). Özellikle m 'nin değeri arttıkça test istatistiğinin gücü de artmaktadır. Bu yüzden bu çalışmada m değeri, testin gücü için tatmin edici bir değer olan 25 olarak belirlenmiştir.

4.3. Monte-Carlo Sanal Deney Algoritması

Deneyde kullanılan algoritma 8 adımdan oluşmaktadır. Bu adımlar sırasıyla açıklayıcı değişken değerleri ve model parametrelerinin atanması; hata terimlerinin türetilmesi; bağımlı değişken değerlerinin hesaplanması; dönüştürülmüş modelin parametrelerinin tahmini; dönüştürülmüş modelin artık terimlerinin hesaplanması; dönüşümün başarılı olup olmadığını tespit etmek üzere uygun testin gerçekleştirilmesi; Monte-Carlo test istatistiğinin yakınsayıp yakınsamadığının belirlenmesi; yakınsama gerçekleşmişse durulması, aksi halde yeniden hata terimlerinin türetilerek işlemlere devam edilmesi olarak belirlenmiştir. Şekil 4.1'de akış diyagramı ile özetlenen bu adımların açıklaması aşağıda sırayla verilmiştir.



Şekil 4.1. Monte-Carlo sanal deneyin akış diyagramı.

Adım 1: X'lerin ve γ 'nın belirlenmesi

Monte-Carlo deneyinin ilk aşaması olan bu adımda açıklayıcı değişken ve parametre değerleri tespit edilmektedir. Açıklayıcı değişken sayısı hata terimlerinin dağılımını etkilemeyeceğinden, tüm modellerde 2 açıklayıcı değişken kullanılması tercih edilmiştir. Ayrıca değişkenlere ilişkin parametrelerin değerlerinin de hata teriminin dağılımı üzerinde herhangi bir etkisi olmadığından, bu parametre değerleri de deney boyunca sabit değerler alacak şekilde belirlenmiştir.

Doğrusal regresyon modelinde açıklayıcı değişken(ler)in sabit değerleri için bağımlı değişkenin alacağı değerler üzerinde durulduğundan öncelikle açıklayıcı değişken değerleri türetilmiş ve tüm deney boyunca bu değerler sabit tutulmuştur. Bu değerler türetilirken X_1 ve X_2 değişkenlerinin birbirlerinden bağımsız olmaları gerektiği göz önüne alınmıştır.

Tüm simülasyon işlemleri boyunca (0,1) aralığında rassal sayıların üretilmesinde (35)'de verilen algoritma kullanılmıştır.

X_1 ve X_2 değişkenlerinin gözlem değerlerinin türetilmesinde, başlangıç değerleri farklı 2 rassal sayı dizisi olan r_1 ve r_2 dizileri kullanılmıştır. Açıklayıcı değişken değerleri deney başında belirlenip her bir regresyonda aynı kalacağından, bu değişkenlerin dağılımları deney sonucunu etkilemeyecektir. Bu yüzden bu değişkenlerin dağılımlarının tekdüze seçilmesinde bir sakınca görülmemiştir. X_1 değerleri [10,30]; X_2 değerleri ise [40,80] aralığında kesikli tekdüze dağılım kullanılarak türetilmiştir. Kesikli tekdüze dağılımdan simülasyon yapılırken Law ve Kelton'da (36) verilen algoritma kullanılarak, ilk sütunu 1'lerden, ikinci sütunu X_1 , üçüncü sütunu X_2 değerlerinden oluşan açıklayıcı değişken dizeyi elde edilmiştir.

Ardından $Y = X\gamma + u$ doğrusal regresyon modelinin parametre yöneyi olan γ 'nın belirlenmesine sıra gelmiştir. Doğrusal regresyon modelinin parametreleri $\gamma' = (5 \ 2 \ 3)$ olarak atanmıştır. Bu aşamadan sonra Monte-Carlo algoritmasının 2. adımına geçilebilir.

Adım 2: u'ların türetilmesi

Bu aşamada modele ilişkin varsayımların sağlanmadığı durumları simüle etmek için hata terimlerinin türetilmesine geçilmiştir. Hata terimleri türetilirken [4.2], [4.3b] ve [4.4] denklemlerinde belirtilen modeller kullanılmıştır. [4.2] modelinde gamma dağılımlı hata terimleri türetilirken Law ve Kelton'da (36) verilen algoritma kullanılmıştır. [4.3b] modelinde [0,1] kesikli tekdüze dağılımına sahip olan I değişkeninin değerleri türetilirken bağımsız değişkenlerin türetilmesinde verilen algoritmaya başvurulmuştur. Yine [4.3b] ve [4.4] modellerindeki normal dağılımlı hata terimleri türetilirken Bratley, Fox ve Schrage'in (37) belirttiği Box-Muller algoritmasının kullanılması uygun görülmüştür.

Adım 3: Y'lerin hesaplanması

1. adımda belirlenen $\mathbf{X}$ ve γ ile 2. adımda türetilen ε yöneyi kullanılarak $\mathbf{Y} = \mathbf{X}\gamma + \mathbf{u}$ modelindeki $\mathbf{Y}$ değerleri hesaplanabilir. Bu sayede varsayımların sağlanmadığı “1 doğrusal regresyon veri seti” elde edilmiş olur.

Adım 4: Parametrelerin tahmin edilmesi

3. adım sonunda bağımlı değişken yöneyi $\mathbf{Y}$ ve açıklayıcı değişken dizeyi $\mathbf{X}$ elde edilmiş olur. Uygulamada da araştırmacının elinde sadece bu veriler vardır ve bu veriler yardımıyla uygun modelin parametreleri tahmin edilmeye çalışılır.

Bu adımda, $\mathbf{Y} = \mathbf{X}\gamma + \mathbf{u}$ modelinin dönüştürülmesiyle elde edilen [3.1] modelinin parametreleri, 3. bölümde bahsedilen ağ taraması tekniği ile tahmin edilmiştir. Ağ taraması için gerekli olan η değeri 0,001 olarak belirlenmiştir. Başlangıç aralıkları ise dönüşüm parametrelili negatif değerler alabilen dönüşümler için $(-4,4)$; sadece pozitif değerler alabilenler içinse $(0,5)$ olarak tespit edilmiştir. Ağ taraması tekniği ile dönüşüm parametresi λ ve doğrusal regresyon modeli parametresi β tahmin edilmiştir.

Adım 5: ε 'ların tahmin edilmesi

[3.1]'in parametreleri tahmin edildikten sonra, modeldeki hata terimleri yöneyi $\hat{\varepsilon} = \mathbf{Y}(\lambda) - \mathbf{X}\hat{\beta}$ ile tahmin edilebilir. Bu yöney dönüştürülmüş modelin hata terimlerini içerdiğinden, $\hat{\varepsilon}$ 'nın model varsayımlarını sağlaması beklenir.

Adım 6: Uygun testin gerçekleştirilmesi

Dönüştürülmüş modelin hata terimleri tahmin edildikten sonra, bu terimlerin model varsayımlarını sağlayıp sağlamadığı uygun bir test kullanılarak sınanabilir. Hata terimleri [4.2] ve [4.3b] düzenekleri ile türetildiğinde Lilliefors normallik testi, [4.4] ile türetildiğinde Bartlett türdeşlik testi kullanılarak varsayımların sağlanıp sağlanmadığı sınanabilir. Lilliefors normallik testi her bir $\hat{\varepsilon}$ için gerçekleştirilmekle birlikte, Bartlett türdeşlik testini gerçekleştirmek için $m = 25$ adet $\hat{\varepsilon}$ yöneyine ihtiyaç duyulmaktadır. Bu yüzden normallik testi yapılırken her bir regresyon sonunda bir,

türdeşlik testi yapılırken ise her 25 regresyon sonunda bir sinama işlemi gerçekleştirilmiştir.

Her iki sinama sonunda da yokluk hipotezi kabul edildiğinde 1, red edildiğinde 0 değerini alan θ_i , i_inci denemeye ilişkin sanal gözleme karşılık gelmek üzere,

Monte-Carlo istatistik m_inci deneme sonunda $\bar{\theta}_m = \frac{1}{m} \sum_{i=1}^m \theta_i$ biçiminde elde edilir.

Adım 7: Yakınsama sağlanıyor mu?

Monte-Carlo deneyin 2-6. adımları belli bir yakınsama kıstası sağlanana kadar sürdürülmelidir. Eğer Monte-Carlo istatistiği belli bir değere yakınsıyorsa, daha fazla sanal gözlem yapmaya gerek yoktur. θ 'nın yakınsanıp yakınsanmadığını tespit etmek için (38)'deki yakınsama kıstası kullanılabilir.

Bu kıstasa göre, ε_{MC} Monte-Carlo tahmin aralığı yarı genişliği ve

$$\bar{\theta}_{enb} = \begin{cases} \bar{\theta}_m, \bar{\theta}_m > \bar{\theta}_{enb} \\ \bar{\theta}_{enb}, \bar{\theta}_{enk} \leq \bar{\theta}_m \leq \bar{\theta}_{enb} \end{cases}, \quad \bar{\theta}_{enk} = \bar{\theta}_{enb} - 2\varepsilon_{MC}$$

$$\bar{\theta}_{enk} = \begin{cases} \bar{\theta}_m, \bar{\theta}_m < \bar{\theta}_{enk} \\ \bar{\theta}_{enk}, \bar{\theta}_{enk} \leq \bar{\theta}_m \leq \bar{\theta}_{enb} \end{cases}, \quad \bar{\theta}_{enb} = \bar{\theta}_{enk} + 2\varepsilon_{MC}$$

olsun. m denemenin en son ζ denemesinde, Monte-Carlo istatistik $\bar{\theta}_{enb} - \bar{\theta}_{enk} = 2\varepsilon_{MC}$ genişliğindeki Monte-Carlo tahmin aralığı dışına çıkmadı ise θ parametresi yakınsanmış olur ve Adım 8'e geçilir. Aksi durumda Adım 2'ye dönülür. Bu deneyde yakınsama kıstasının parametreleri $\varepsilon_{MC} = 0,001$ ve $\zeta = 200$ olarak alınmıştır.

Adım 8: Sonuçları ver ve dur

Bu adımda Monte-Carlo istatistiğinin yakınsanan değeri

$$\hat{\theta}_{MC} = (\bar{\theta}_{enb} + \bar{\theta}_{enk})/2$$

biçiminde elde edilir.

Yukarıda verilen algoritmanın işleyişini göstermek ve Monte-Carlo sanal deneyi gerçekleştirilecek yazılımın sınanmasında yararlanmak üzere aşağıda sayısal bir örnek verilmiştir. Örnek çapı 20 ve hata terimlerinin üretileceği model [4.2] olarak belirlenmiştir. Çarpıklık katsayısı $\alpha_3 = -2$ olmak üzere Tukey dönüşümü için Monte-Carlo algoritmasının 1 kez işleyişi ve her bir adımın açıklaması aşağıda belirtildiği gibidir:

Adım 1:

X_1 ve X_2 değişkenlerinin simülasyonunda kullanılacak olan r_1 ve r_2 rassal sayı dizilerinin başlangıç değerleri sırasıyla 0,123456 ve 0,635437 olarak atanmıştır. Ardından algorithmada belirtilen dağılımlar kullanılarak X_1 ve X_2 değişkenleri için 20 adet gözlem değeri türetilmiştir. Bu değerler Çizelge 4.3'de gösterilmiştir. Ardından parametre yöneyi $\gamma' = (5 \ 2 \ 3)$ olarak atayıp 2. adıma geçilmiştir.

Adım 2:

Hata terimleri türetilirken [4.2] modeli ve bu model için gerekli gamma dağılımı kullanılmıştır. [4.2] modelinde $\alpha = 1$ ve $k = 1$ alınarak $\alpha_3 = -2$ olan 20 adet hata terimi türetilmiş ve bu terimler Çizelge 4.3'de gösterilmiştir.

Adım 3:

Bu adımda Çizelge 4.3'de verilen açıklayıcı değişken gözlem değerleri ve hata terimleri ile Adım 1'de belirtilen γ yöneyi kullanılarak Y değerleri, $Y = X\gamma + u$ eşitliği ile hesaplanmıştır. Hesaplanan bu değerler Çizelge 4.3'de verilmiştir. Bu sayede 1 doğrusal regresyon veri seti elde edilmiştir.

Çizelge 4.3. Adım 1, 2, 3 ve 5 sonunda elde edilen değerler.

i	Adım 1				Adım 2	Adım 3	Adım 5
	X_{0i}	X_{1i}	X_{2i}	γ_i	u_i	Y_i	$\hat{\varepsilon}_i$
1	1	19	77	5	2,610	276,609	0,000024
2	1	16	70	2	-2,129	244,871	0,000014
3	1	12	59	3	-1,053	204,947	0,000016
4	1	27	56		2,408	229,408	0,000037
5	1	26	60		4,260	241,260	0,000051
6	1	18	61		3,426	227,426	0,000055
7	1	30	58		-4,200	234,800	-0,000018
8	1	22	54		-1,809	209,191	0,000001
9	1	22	77		4,424	284,424	0,000025
10	1	12	79		-10,206	255,794	-0,000048
11	1	24	44		1,308	186,308	-0,000005
12	1	24	73		-3,633	268,367	-0,000022
13	1	10	46		4,340	167,340	-0,000004
14	1	27	69		4,748	270,748	0,000036
15	1	18	41		5,703	169,703	0,000002
16	1	29	69		-2,849	267,151	-0,000021
17	1	25	77		-4,843	281,157	-0,000046
18	1	20	74		5,135	272,135	0,000047
19	1	25	43		4,268	188,268	0,000020
20	1	23	53		-11,909	198,091	-0,000088

Adım 4:

4. adımda [3.1] modelinin parametreleri ağ taraması tekniği ile tahmin edilmiştir. λ parametresinin başlangıç değeri -4 seçilmiş ve bu değer her seferinde $\eta = 0,001$ kadar arttırılmıştır. Her bir λ değeri için [3.12] fonksiyonunun değeri hesaplanmış ve bu işlem λ 'nın değeri 4'e eşit olana değin sürdürülmüştür. Deney, Tukey dönüşümü için gerçekleştirildiğinden, $Y(\lambda)$ ve $J(\lambda; Y)$ değerleri Tukey dönüşümüne

göre elde edilmiştir. Bu işlemler tamamlandığında [3.12] fonksiyonunun değeri 8000 farklı λ için hesaplanmış olur. Bunlar içinde en büyük $\varphi(\lambda)$ değeri $-10,441$ olup, fonksiyonun maksimumunu sağlayan değer, $\hat{\lambda} = 0,002$ 'dir. $\hat{\lambda} = 0,002$ için regresyon modeli parametrelerinin tahmini, [3.8] numaralı eşitlik kullanılarak $\hat{\beta} = (1,009406 \quad 0,000017 \quad 0,000023)$ şeklinde elde edilmiştir.

Adım 5:

5. adımda dönüştürülmüş modelin hata terimleri, $\hat{\varepsilon} = Y(\lambda) - X\hat{\beta}$ ile tahmin edilmiş ve Çizelge 4.3'te sunulmuştur.

Adım 6:

Bu adımda, Çizelge 4.3'te verilen hata terimlerine Lilliefors normallik testi uygulanmış ve L istatistiğinin değeri 0,0436 olarak hesaplanmıştır. Örnek çapı 20 olduğunda L istatistiğine ilişkin %5 anlamlılık düzeyindeki kritik değer 0,190'dır. Bu durumda " $H_0: \varepsilon_i$ 'lerin geldiği dağılım normaldir" şeklindeki yokluk hipotezi %5 anlamlılık düzeyinde reddedilemez. Bu durumda $\theta_1 = 1$ olarak atanır.

Adım 7:

$\theta_1 = 1$ olduğundan, $\bar{\theta}_1 = 1/1 = 1$ olarak atanır. $\varepsilon_{MC} = 0,001$ olduğundan $\bar{\theta}_{enb} = 1,001$ ve $\bar{\theta}_{enk} = 0,999$ atamaları yapılır. Henüz yakınsama sağlanmadığından 2. adıma geçilir.

Yukarıda bahsedilen adımlar yakınsama sağlanana kadar tekrarlanır. Her bir denemede yeni bir regresyon veri seti elde edilir, dönüştürülmüş model parametreleri tahmin edilir ve bu parametreler aracılığı ile dönüştürülmüş modelin hata yöneyi tahmin edilir. Ardından bu terimlere normallik testi uygulanır ve yakınsama sağlanana kadar bu işlemler yinelenir.

Yakınsama sağlandığında bir dönüşüm için istatistiğin yakınsadığı değer tespit edilmiş olur. Daha sonra aynı deney düzeninde diğer dönüşümler için aynı işlemler

tekrarlanır. Bu sayede dönüşümlerin her birinin aynı deney içinde normal dağılımlı ya da türdeş hata terimleri elde etme oranları tespit edilmiş olur.

4.4. Monte-Carlo Deney Sonuçları

Şekil 4.1’de verilen algoritmanın, Turbo C programlama dilindeki kodu EK-4’te verilmiştir. Örnek çapı 20, 60 ve 90 olmak üzere 3 farklı şekilde seçilmiş, [4.2], [4.3b] ve [4.4] modellerinin her birinde tüm dönüşümler için $\varepsilon_{MC} = 0,001$ ’lik hata bandı içinde yakınsanan θ değerleri tespit edilmiştir. Deney sonuçları normallikten ayrılma ve türdeşlikten ayrılma başlıkları altında verilecektir.

4.4.1. Normallikten ayrılma

Bu başlık altında, [4.2] modeli ile verilen çarpık dağılımlı hata terimleri ve [4.3b] modeli ile türetilen çift tepe değerli hata terimlerine sahip regresyon modelinde, dönüşümlerin normal dağılım varsayımını sağlamadaki başarı oranları verilmektedir.

4.4.1.1 Çarpık dağılımlı hata terimleri

Çarpık dağılımlı hata terimlerini üretmede [4.2] modeli kullanılmıştır. Ardından her bir dönüşüm, hata terimleri çarpık dağılan regresyon modeline ayrı ayrı uygulanmıştır. Bu sayede [3.1] modelinin artıkları elde edilmiş ve bu terimlere Lilliefors normallik testi uygulanmıştır.

i _inci denemede hata terimlerinin normal dağıldığını iddia eden H_0 hipotezi %5 anlamlılık düzeyinde reddedilemiyorsa $\theta_i = 1$ aksi halde $\theta_i = 0$ olarak atanmıştır.

Ardından $\hat{\theta}_{MC}$ istatistiğinin yakınsadığı θ değeri tespit edilmiştir. Böylece her bir dönüşüm için normal dağılım hipotezinin reddedilememe oranları elde edilmiştir. Bu değer, dönüşümden önce normal dağılmayan hata terimlerinin dönüşümden sonra normal dağılma oranlarını vermektedir. Bu oran ne kadar büyükse, ilgili dönüşümün

o kadar başarılı olduğu söylenebilir. Her bir dönüşüm için Monte-Carlo tahmin sonuçları Çizelge 4.4'te verilmektedir.

Çizelge 4.4. Çarpık dağılımlı hata terimleri için Monte-Carlo tahminleri.

n	α_3 Dönüşüm	-5	-4	-3	-2	-1	1	2	3	4	5
20	TUK	0,962	0,964	0,986	0,990	1,000	1,000	0,986	0,969	0,940	0,905
	BC	0,927	0,950	0,977	0,986	1,000	0,992	0,964	0,918	0,875	0,826
	MAN	0,931	0,955	0,979	0,983	1,000	0,992	0,955	0,930	0,901	0,872
	JD	0,927	0,950	0,977	0,982	1,000	0,992	0,964	0,919	0,875	0,829
	BD	0,927	0,949	0,977	0,982	1,000	0,992	0,964	0,916	0,877	0,829
	YJ	0,927	0,950	0,977	0,982	1,000	0,992	0,964	0,919	0,875	0,829
60	TUK	0,644	0,716	0,829	0,955	0,979	0,978	0,959	0,771	0,656	0,604
	BC	0,481	0,508	0,613	0,833	0,983	0,948	0,766	0,553	0,461	0,433
	MAN	0,483	0,512	0,608	0,842	0,978	0,947	0,784	0,559	0,474	0,451
	JD	0,476	0,509	0,611	0,833	0,975	0,949	0,770	0,552	0,461	0,430
	BD	0,474	0,510	0,609	0,833	0,983	0,947	0,775	0,551	0,459	0,426
	YJ	0,476	0,509	0,611	0,833	0,975	0,949	0,770	0,552	0,461	0,430
90	TUK	0,841	0,631	0,702	0,862	0,914	0,968	0,965	0,830	0,713	0,649
	BC	0,461	0,441	0,476	0,640	0,953	0,897	0,631	0,493	0,466	0,436
	MAN	0,456	0,433	0,481	0,649	0,949	0,894	0,643	0,496	0,470	0,463
	JD	0,461	0,439	0,476	0,640	0,951	0,899	0,631	0,493	0,466	0,435
	BD	0,460	0,440	0,476	0,640	0,954	0,899	0,633	0,493	0,466	0,427
	YJ	0,461	0,439	0,476	0,640	0,951	0,899	0,631	0,493	0,466	0,435

Çizelge 4.4 incelendiğinde, orijinal hata terimlerinin çarpıklık katsayısı olan α_3 'ün değeri 0'dan uzaklaştıkça dönüşümlerin normal dağılımı sağlama oranlarının düştüğü gözlenmektedir. Regresyon denklemindeki gözlem sayısı 20 ve çarpıklık katsayısı -1 iken tüm dönüşümler için $\hat{\theta}_{MC} = 1$ olduğu görülmektedir. Benzer şekilde gözlem sayısı 20 ve çarpıklık katsayısı 1 iken Tukey dönüşümü için Monte-Carlo tahmin

değerinin 1, diğer dönüşümler içinse 0,992 olduğu göze çarpmaktadır. Öte yandan, çarpıklık arttıkça dönüşümlerin normal dağılımı sağlama oranları giderek azalmaktadır. Dolayısıyla orijinal hata terimlerinin çarpıklık katsayısı ne kadar küçükse, dönüşümlerin de o oranda başarılı oldukları gözlenmektedir. Ayrıca gözlem sayısı artsa dahi bu özelliğin devam ettiği dikkat çekmektedir.

Çizelge 4.4'deki diğer bir özellik ise çarpıklığın yönünün, dönüşümlerin başarısını çok fazla etkilemediğidir. Örnek çapı 20 ve 60 iken, dönüşümler sağa çarpık hata terimlerini daha yüksek oranda normalleştirirken, örnek çapı 90 olduğunda çarpıklığın yönünün, hata terimlerini normalleştirme oranları üzerinde sabit bir etkisi olmadığı gözlenmektedir. Örnek çapı düşük olduğunda, çarpıklığın yönünün, normalliği sağlama oranları üzerinde çok büyük farklılıklara yol açmadığı dikkat çekmektedir. Örneğin örnek çapı 20, çarpıklık katsayısı -2 ve 2 iken Box-Cox dönüşümünün normalliği sağlama oranları sırasıyla 0,986 ve 0,964 olup bu iki oran birbirine oldukça yakındır. Diğer çarpıklık katsayısı ve dönüşümlerde de benzer bir durum söz konusudur.

Örnek çapının, dönüşümlerin normal dağılımı sağlama oranları üzerindeki etkisi incelendiğinde, örnek çapı arttıkça, dönüşümlerin normal dağılımı sağlama oranlarının düştüğü dikkat çekmektedir. Çizelge 4.4 incelendiğinde, hemen hemen tüm dönüşümler için çarpıklık katsayısı sabitken, örnek çapının artırıldığı durumlarda, normal dağılımı sağlama oranlarının düştüğü görülmektedir. Bu durum, örnek çapı arttıkça dönüşümlerin başarısının azalmasından kaynaklanabileceği gibi, Lilliefors normallik testinden de kaynaklanabilir. Conover'in (29) belirttiği gibi, hemen hemen tüm uyum iyiliği testleri, n oldukça büyük olduğunda, H_0 hipotezinin reddine karar vermektedirler. Bunun nedeni, eldeki verilerin iddia edilen dağılıma tamı tamına uymamasıdır. İşte bu yüzden örnek çapı arttıkça yokluk hipotezinin reddedilememe olasılığı da azalmaktadır. Bu yüzden n değeri arttıkça dönüşümlerin başarımlarının düşmesi dikkatle yorumlanmalıdır. Ancak tüm dönüşümlere aynı test uygulandığından ve örnek çapı sabitken testin gücü değişmeyeceğinden, aynı test ile dönüşümlerin karşılaştırılmasında bir sakınca olmadığı söylenebilir.

Dönüşümler, birbirleri ile kıyaslandığında Tukey dönüşümünün diğer dönüşümlerden daha iyi sonuç verdiği göze çarpmaktadır. Örnek çapı 20 iken, seçilen tüm çarpıklık katsayıları için Tukey dönüşümünün normal dağılımı sağlama oranı diğer dönüşümlerinkilere göre daha yüksektir. Örnek çapı 60 ve 90 olduğunda, çarpıklık katsayısının -1 olduğu durum haricinde yine seçilen tüm çarpıklık katsayılarında Tukey dönüşümünün diğerlerine oranla daha başarılı olduğu görülmektedir. Başarılı olan bir diğer dönüşüm ise Manly dönüşümüdür. Örnek çapı 20 iken, çarpıklık katsayısının ± 2 olduğu durumlar haricinde, Tukey dönüşümüne en yakın normalleştirme oranlarını Manly dönüşümü vermektedir. Örnek çapı 60'a çıktığında α_3 'ün $-5, -4, -2, -1, 2, 3, 4$ ve 5 ; 90'a çıktığında ise α_3 'ün $-3, -2, 2, 3, 4$ ve 5 değerleri için Tukey dönüşümünden sonra en iyi normalleştirme oranlarını yine Manly dönüşümü vermektedir. Diğer dönüşümlerin normalleştirme oranları Manly dönüşümününkine yakın görülmektedir. Ancak hem Manly hem de diğer dönüşümlerin normalleştirme oranları Tukey dönüşümününkine oranla düşüktür.

4.4.1.2. Çok tepe değerli hata terimleri

2 tepe değerli hata terimleri türetilirken [4.3b] modeli kullanılmıştır. Hata terimleri 2 tepe değerli olan regresyon modelinde tüm dönüşümler uygulanmış ve [3.1] modelinin parametreleri tahmin edilmiştir. Ardından dönüştürülmüş modelin artıkları hesaplanmış ve bu terimlere Lilliefors normallik testi uygulanmıştır.

Daha sonra çarpık dağılımlı hata terimleri durumu incelenirken uygulanan yöntem izlenmiştir. Bu modelde de i _inci denemede hata terimlerinin normal dağıldığını iddia eden H_0 hipotezi %5 anlamlılık düzeyinde reddedilemiyorsa $\theta_i = 1$ aksi halde $\theta_i = 0$ olarak belirlenmiş ve $\hat{\theta}_{MC}$ istatistiğinin yakınsadığı θ değeri tespit edilmiştir. Bu sayede dönüştürülmüş modelin hata terimleri için normal dağılım hipotezinin reddedilememe oranları elde edilmiştir. Daha önce olduğu gibi bu oran ne kadar büyükse, ilgili dönüşümün o kadar başarılı olduğu söylenebilir. Her bir dönüşüm için Monte-Carlo tahmin sonuçları Çizelge 4.5'te verilmektedir.

Çizelge 4.5. Çok tepe değerli hata terimleri için Monte-Carlo tahminleri.

n	f	0				5				10				15			
		0,5	1	5	10	0,5	1	5	10	0,5	1	5	10	0,5	1	5	10
	Dönüşüm																
20	TUK	1,000	1,000	1,000	0,997	1,000	1,000	1,000	0,997	1,000	1,000	1,000	1,000	0,990	0,996	1,000	0,997
	BC	0,998	0,997	0,995	0,997	0,976	0,986	0,998	0,998	0,949	0,962	0,993	1,000	0,942	0,945	0,977	1,000
	MAN	0,994	0,994	0,994	0,993	0,978	0,988	0,999	1,000	0,954	0,962	0,994	1,000	0,951	0,950	0,975	1,000
	JD	0,997	0,997	0,995	0,997	0,975	0,987	0,998	1,000	0,948	0,962	0,994	1,000	0,942	0,943	0,979	1,000
	BD	0,996	0,997	0,995	0,997	0,975	0,986	0,998	1,000	0,949	0,960	0,993	1,000	0,942	0,943	0,977	1,000
	YJ	0,997	0,997	0,995	0,997	0,975	0,987	0,998	1,000	0,948	0,962	0,994	1,000	0,942	0,943	0,979	1,000
60	TUK	0,787	0,896	0,987	0,994	0,988	0,993	1,000	0,997	0,970	0,980	1,000	1,000	0,875	0,900	0,961	0,985
	BC	1,000	1,000	1,000	1,000	0,725	0,876	1,000	1,000	0,617	0,651	0,922	0,986	0,587	0,610	0,729	0,890
	MAN	1,000	1,000	0,998	1,000	0,724	0,882	1,000	1,000	0,613	0,644	0,918	1,000	0,593	0,608	0,745	0,890
	JD	1,000	1,000	1,000	0,998	0,731	0,875	1,000	1,000	0,615	0,651	0,920	0,986	0,588	0,609	0,738	0,891
	BD	1,000	1,000	1,000	1,000	0,730	0,874	1,000	1,000	0,614	0,653	0,919	0,987	0,587	0,610	0,739	0,892
	YJ	1,000	1,000	1,000	0,998	0,731	0,875	1,000	1,000	0,615	0,651	0,920	0,986	0,588	0,609	0,738	0,891
90	TUK	0,933	0,874	0,973	0,993	1,000	0,995	0,997	0,996	0,837	0,880	0,971	0,989	0,690	0,730	0,869	0,933
	BC	0,999	0,998	0,996	0,996	0,585	0,739	1,000	1,000	0,516	0,534	0,811	0,971	0,504	0,511	0,590	0,763
	MAN	1,000	1,000	1,000	1,000	0,589	0,743	1,000	1,000	0,512	0,536	0,819	0,977	0,504	0,513	0,588	0,758
	JD	0,998	0,999	0,996	0,999	0,584	0,740	1,000	1,000	0,516	0,533	0,811	0,971	0,501	0,511	0,589	0,764
	BD	0,998	0,999	0,996	0,996	0,588	0,740	1,000	1,000	0,516	0,534	0,811	0,970	0,504	0,511	0,589	0,763
	YJ	0,998	0,999	0,996	0,999	0,584	0,740	1,000	1,000	0,516	0,533	0,811	0,971	0,501	0,511	0,589	0,764

Çizelge 4.5 incelendiğinde $n = 20$ iken genel olarak Tukey dönüşümünün diğer dönüşümlerden daha fazla normalliği sağladığı görülmektedir. $n = 20$ için sadece $f = 5$, $\sigma = 10$ ve $f = 15$, $\sigma = 10$ değerlerini aldığıında Tukey dönüşümü diğer dönüşümlerin gerisine düşmektedir. Ancak bu durumlarda da Tukey dönüşümünün başarımı diğer dönüşümlerinkinden çok fazla uzak değildir. Bu iki durumda diğer dönüşümlerin normalliği sağlama oranları 1 iken, Tukey dönüşümününki 0,997'dir.

$n = 20$ için Tukey dönüşümünün ardından normalliği sağlama oranı en yüksek olan dönüşüm genelde Manly dönüşümüdür. Ancak bu dönüşümün diğer dönüşümlere belirgin bir üstünlük sağladığı söylenemez. Manly dönüşümünün normalliği sağlama oranı ile diğer dönüşümlerin normalliği sağlama oranları arasındaki fark en fazla $f = 15$ ve $\sigma = 0,5$ iken gerçekleşen 0,009'dur.

n 'nin değeri arttıkça $f \neq 0$ için Tukey dönüşümünün diğer dönüşümlere olan üstünlüğü devam etmektedir. Üstelik bu durumda Tukey dönüşümü ile diğer dönüşümlerin normalliği sağlama oranları arasındaki fark giderek artmaktadır. Özellikle σ 'nın değeri küçüldükçe, Tukey dönüşümü diğer dönüşümlerden daha başarılı görülmektedir. $f = 5$, $n = 60$ ve $\sigma = 0,5$ için Tukey dönüşümü ile bu dönüşüme en yakın dönüşümün normalliği sağlama oranı arasındaki fark 0,257 iken $f = 5$, $n = 90$ ve $\sigma = 0,5$ için aynı fark 0,411 olmaktadır. Dolayısıyla n 'nin değeri arttıkça Tukey dönüşümünün diğer dönüşümlere oranla daha başarılı olduğu söylenebilir. Ancak $f = 0$ değerini aldığıında Tukey dönüşümünün normalliği sağlama oranları diğer dönüşümlere oranla daha düşük görülmektedir. $f = 0$ için hata terimleri tek tepe değerli olacaklarından, $f = 0$ için normal dağılım varsayımı çığnenmez. Dolayısıyla zaten normal dağılan hata terimlerine Tukey dönüşümü uygulamanın uygun olmadığı söylenebilir. Diğer dönüşümler $f = 0$ için oldukça başarılı görülürken, Tukey dönüşümünün başarımının düşmesi ilgi çekici bir durumdur.

n 'nin değeri arttıkça Tukey dönüşümü dışındaki dönüşümlerin normalliği sağlama oranları arasında fazla bir fark gözlenmemektedir. Regresyon modelinin gözlem

sayısı arttıkça diğer dönüşümlerin birbirlerine yakın başarımlar gösterdikleri, birinin diğerine baskın olmadığı söylenebilir.

Çizelge 4.5 incelendiğinde, her bir örnek çapında, $f = 0$ için σ 'nın aldığı değerin dönüşümlerin normalliği sağlama oranlarını pek fazla etkilemediği gözlenmektedir. Sadece Tukey dönüşümü için $n = 60$ iken σ 'nın değeri arttıkça dönüşümün normalliği sağlama oranı artmaktadır. Yine aynı dönüşüm için $n = 90$ iken genel olarak aynı yapı gözlenmektedir. Ancak diğer tüm dönüşümler, tepe değerleri arasındaki uzaklık 0 iken, σ 'nın aldığı değerden fazla etkilenmemektedir.

Ancak f 'nin değeri 0'dan farklı iken σ 'nın etkisi de artmaktadır. Farklı n değerleri ve tüm dönüşümler için, σ 'nın değeri arttıkça dönüşümlerin normalliği sağlama oranları hemen hemen her zaman artma eğilimindedir. Bu artış, σ 'nın değeri büyüdükçe 2 farklı tepe değerine sahip hata terimlerinin aldıkları değerlerin birbirlerine yaklaşması ile açıklanabilir.

Çizelge 4.5'den çıkarılabilecek bir diğer sonuç ise, örnek çapı arttıkça, dönüşümlerin normal dağılımı sağlama oranlarının genelde azalmasıdır. f ve σ 'nın değeri sabitken, n 'nin değeri arttırıldığında, genel olarak dönüşümlerin normal dağılımı sağlama oranları düşmekte ya da hemen hemen aynı kalmaktadır.

4.4.2. Türdeşlikten ayrılma

Hata terimlerinin türdeşlikten ayrıldığı durumu incelemek için [4.4] modeli kullanılmıştır. Ardından kullanılan bu model için [3.1] modelinin parametreleri tahmin edilmiş ve modelin artıkları hesaplanmıştır. Bu işlem 25 kere tekrarlanarak 1 deneme için gerekli veriler elde edilmiştir.

Daha sonra Bartlett testi kullanılarak dönüştürülmüş modelin hata terimlerinin türdeş olup olmadığı sınanmıştır. i _inci denemede hata terimlerinin türdeş olduğunu iddia eden H_0 hipotezi %5 anlamlılık düzeyinde reddedilemiyorsa $\theta_i = 1$ aksi halde $\theta_i = 0$

olarak belirlenmiş ve $\hat{\theta}_{MC}$ istatistiğinin yakınsadığı θ değeri tespit edilmiştir. Bu sayede dönüştürülmüş modelin hata terimleri için türdeşlik hipotezinin reddedilememe oranları elde edilmiştir. Bu oran ne kadar büyükse, ilgili dönüşümün o kadar başarılı olduğu söylenebilir.

[4.4] modeli için elde edilen Monte-Carlo tahminleri Çizelge 4.6'da verilmiştir.

Çizelge 4.6. Türdeş olmayan hata terimleri için Monte-Carlo tahminleri.

n	Dönüşüm	δ			
		1	2	2,5	3
20	TUK	0,000	0,000	0,000	0,000
	BC	0,447	0,269	0,088	0,020
	MAN	0,954	0,754	0,519	0,110
	JD	0,440	0,268	0,086	0,020
	BD	0,446	0,269	0,089	0,020
	YJ	0,440	0,268	0,086	0,020
60	TUK	0,000	0,000	0,000	0,000
	BC	0,765	0,623	0,044	0,000
	MAN	0,977	0,908	0,183	0,102
	JD	0,758	0,616	0,043	0,000
	BD	0,766	0,621	0,048	0,000
	YJ	0,758	0,616	0,043	0,000
90	TUK	0,000	0,000	0,000	0,000
	BC	0,920	0,831	0,008	0,000
	MAN	0,998	0,979	0,024	0,000
	JD	0,921	0,829	0,007	0,000
	BD	0,922	0,832	0,006	0,000
	YJ	0,921	0,829	0,007	0,000

Çizelge 4.6 incelendiğinde tüm örnek çaplarında Manly dönüşümünün diğer dönüşümlere oranla daha iyi sonuçlar verdiği görülmektedir. Özellikle n'nin küçük

değerleri için Manly dönüşümünün, diğer dönüşümlere oranla çok daha başarılı olduğu söylenebilir. Göze çarpan bir diğer sonuç ise Tukey dönüşümünün her bir δ ve n değeri için türdeşliği sağlayamadığıdır. Bu sonuç, Tukey dönüşümünün türdeşliği sağlamada uygun bir dönüşüm olmadığını göstermektedir.

Çizelge 4.6'dan çıkan bir diğer sonuç ise Tukey ve Manly dönüşümü haricindeki dönüşümlerin birbirlerine oldukça yakın sonuçlar verdiğidir. Dolayısıyla bu dönüşümler arasında bir diğerine üstün olan bir dönüşüm olmadığı söylenebilir.

Dikkati çeken bir diğer sonuç ise δ 'nin değeri arttıkça dönüşümlerin türdeşliği sağlama oranlarının giderek azaldığıdır. Özellikle δ 'nin değeri 3 olduğunda dönüşümlerin başarımları oldukça düşük değerlere ulaşmaktadır. Bu durumda değişkelerin değişimi arttıkça, dönüşümlerin başarımlarının azaldığı söylenebilir.

Çizelge 4.6'dan elde edilebilecek bir diğer sonuç ise, n 'nin durumudur. δ 'nin 2,5 ve 3 değeri haricinde, örnek çapı arttıkça tüm dönüşümler daha iyi sonuçlar vermektedirler. Ancak δ 'nin 2,5 ve 3 değerleri için örnek çapı arttıkça dönüşümlerin başarımları da düşmektedir. Dolayısıyla türdeşlikten sapmanın düşük olduğu durumlarda, örnek çapının artmasının, dönüşümlerin türdeşliği sağlama oranlarını genelde arttırdığı söylenebilir. Buna karşın türdeşlikten sapma arttıkça, örnek çapını arttırmak dönüşümlerin türdeşliği sağlama oranlarını düşürmektedir.

Bartlett testinde yokluk hipotezi, dönüştürülmüş modelin tüm hata terimlerinin değişkelerinin eşit olduğu şeklindedir. Dolayısıyla dönüştürülmüş modelin değişkelerinden biri bile diğerlerinden farklı olursa H_0 hipotezinin reddedilmesi beklenir. [4.4] modeli ile türetilen hata terimlerinin tümünün değişkeleri farklı olduğundan, δ değeri arttıkça dönüşümlerin başarımlarının çok çabuk düşmesi normal olarak karşılanabilir. Özellikle örnek çapı arttıkça, birbirinden farklı değişkelerin de sayısı artacak, dolayısıyla dönüştürülmüş modelde H_0 hipotezinin sağlanması oldukça güçleşecektir. δ ve n 'nin değeri birlikte arttığında, dönüşümlerin türdeşliği sağlama oranlarının düşmesi bu şekilde açıklanabilir.

5. SONUÇ VE ÖNERİLER

Regresyon modelinin varsayımları sağlanmadığında, değişkenleri dönüştürerek modelin varsayımlarını sağlamaya çalışmak oldukça sık başvurulan bir yöntemdir. Bu bağlamda literatürde önerilen önemli dönüşümler; Tukey (3), Box-Cox (4), Manly (5), John-Draper (6), Bickel-Doksum (7) ve Yeo-Johnson (8) olarak verilebilir. Box-Cox dönüşümü ve genel olarak kuvvet dönüşümleri, “normallikten” ve “sabit değişkeden (türdeşlik)” ayrılma sorunlarının giderilmesinde etkili olup (4, 9), hata terimlerinin bağımlı olduğu durumlarda etkili olamamaktadırlar. Spitzer’in (10), Box-Cox dönüşümünün küçük örnek çaplarındaki davranışını inceleyen çalışması dışında, bu dönüşümlerin varsayımlardan sapmaları ne ölçüde giderdiklerine ve hangi durumda hangi dönüşümün daha üstün olduğuna dair ayrıntılı bir inceleme bulunmamaktadır. İşte bu çalışmada, adı geçen dönüşümlerin, regresyon modelinin varsayımlarının sağlanmaması durumundaki başarımları incelenmiş ve hangi durumda hangi dönüşümün daha iyi sonuç verdiği bulunmaya çalışılmıştır. Normal dağılım varsayımının sağlanmadığı durumda, dönüşüm sonucu normal dağılımın sağlanma oranları her bir dönüşüm için elde edilmiştir. Ayrıca türdeşliğin sağlanmadığı durumlarda, dönüşümlerin türdeşliği sağlama oranları yine her bir dönüşüm için elde edilmiştir. Bu işlemler farklı örnek çapları için tekrarlanmış, böylece farklı örnek çapları için dönüşümler kıyaslanmıştır.

Öncelikle orijinal modelin bağımlı değişkeni normallikten ayrıldığında, dönüştürülmüş modellerde normalliğin sağlanma oranları incelenmiştir. Bu amaçla orijinal modelde hata terimlerinin çarpık dağılımlı ve çok tepe değerli oldukları durumlar göz önüne alınmıştır.

Hata terimlerinin dağılımı çarpık olarak türetildiğinde, orijinal modelin hata terimlerinin çarpıklık katsayısı azaldıkça, dönüşümlerin normal dağılımı sağlama oranlarının daha yüksek oldukları bulunmuştur. Ayrıca çarpıklığın yönünün dönüşümlerin başarımlarını etkileyip etkilemedikleri incelenmiş fakat çarpıklığın yönünün, dönüşümlerin normal dağılımı sağlama oranları üzerinde önemli bir etkisi olmadığı gözlenmiştir.

İncelenen bir başka nokta ise örnek çapının etkisidir. Örnek çapı arttıkça genel olarak dönüşümlerin başarımları azalmaktadır. Ancak bu durum, Conover'in (29) belirttiği gibi Lilliefors testinden de kaynaklanabilir.

Çarpıklık durumunda dönüşümlerin başarımları birbirleriyle karşılaştırıldığında, Tukey dönüşümünün hemen hemen tüm durumlarda daha iyi sonuçlar verdiği gözlenmektedir. Bu dönüşümün ardından Manly dönüşümü gelmekle birlikte, Manly dönüşümü ile diğer dönüşümler arasında, normalliği sağlama açısından çok büyük farklar gözlenmemektedir. Bickel-Doksum dönüşümünün teorik olarak çarpık dağılımlı değişkenlerde sorunlara yol açtığı iddia edilmesine rağmen, yapılan bu çalışmada, Bickel-Doksum dönüşümünün Tukey haricindeki dönüşümlere yakın başarımlar verdiği görülmüştür.

Daha sonra normalliğin sağlanmadığı bir diğer durum olan çok tepe değerlilik incelenmiştir. Özel olarak hata terimlerinin 2 tepe değerli olduğu durum incelenmiş ve yine bu durumda da dönüşümlerin normalliği sağlamaya etkisi olup olmadığı araştırılmıştır.

Öncelikle iki tepe değeri arasındaki fark ve hata terimlerinin değişkelerinin etkisi incelenmiştir. İki tepe değeri arasındaki fark 0 iken, hata terimlerinin değişkesinin dönüşümlerin başarımları üzerinde pek fazla etkisi olmadığı, ayrıca Tukey dönüşümü dışındaki dönüşümlerin normalleştirme oranlarının, 1'e oldukça yakın oldukları görülmektedir. Tepe değerleri arasındaki fark 0'dan farklı olduğunda ise hata terimlerinin değişkesi arttıkça dönüşümlerin başarımları da yükselmektedir.

Hata terimleri 2 tepe değerli olduğunda da, örnek çapı arttıkça genel olarak dönüşümlerin başarımları azalmaktadır. Bu durum, daha önce belirtildiği gibi Lilliefors testinden kaynaklanabilir.

Hata terimleri 2 tepe değerli iken dönüşümler birbirleriyle karşılaştırıldığında, genel olarak Tukey dönüşümünün diğer dönüşümlere oranla daha başarılı olduğu görülmektedir. Yine bu modelde de Tukey dönüşümünün ardından Manly

dönüşümünün başarılı olduğu ve diğer dönüşümlerin Manly dönüşümüne yakın sonuçlar verdiği söylenebilir.

Son olarak türdeşlikten ayrılma durumu incelenmiştir. Burada türdeşlikten ayrılma sorununun derecesi arttıkça, dönüşümlerin türdeşliği sağlama oranlarının düştüğü gözlenmiştir. Türdeşlikten ayrılma sorununun derecesi düşükken örnek çapı arttıkça dönüşümlerin daha iyi; sorunun derecesi yüksekken örnek çapı arttıkça dönüşümlerin daha kötü sonuçlar verdikleri gözlenmiştir.

Türdeşlikten ayrılma durumunda dönüşümler birbirleriyle karşılaştırıldığında Manly dönüşümünün diğer dönüşümlere oranla oldukça iyi sonuçlar verdiği gözlenmektedir. Burada dikkati çeken bir diğer nokta ise Tukey dönüşümünün tüm deney noktalarında türdeşliği sağlama oranının 0 olmasıdır. Ayrıca Tukey ve Manly dönüşümü dışındaki tüm dönüşümler birbirlerine oldukça yakın sonuçlar vermektedirler.

Sonuç olarak regresyon modelinde normallikten ayrılma durumunda en başarılı dönüşümün Tukey, türdeşlikten ayrılma durumunda ise Manly dönüşümü olduğu söylenebilir. Araştırmacı regresyon modelindeki sorunu belirledikten sonra, bu sonuçlar doğrultusunda uygun bir dönüşüm seçerek sorunu çözebilir.

Daha ileriki çalışmalarda, oldukça uzun süren Monte-Carlo deneylerini daha kısa sürede sonuçlandırmak için [3.1] modelinin parametreleri tahmin edilirken ağ taraması tekniği yerine Newton-Rapson yöntemi kullanılabilir. Ayrıca yine Monte-Carlo deneyindeki deneme sayılarını düşürmek için değişke küçültme tekniklerinin kullanılması uygun olabilir.

KAYNAKLAR

1. Graybill, F.A., "The Theory and Applications of the Linear Model", *Duxbury Press*, London, 213 (1976).
2. Sakia, R.M., "The Box-Cox transformation technique: A review", *The Statistician*, 41: 169-178 (1992).
3. Tukey, J.W., "The comparative anatomy of transformations", *Annals of Mathematical Statistics*, 28: 602-632 (1957).
4. Box, G.E.P., Cox, D.R., "An analysis of transformations", *Journal of the Royal Statistical Society, Series B*, 26: 211-252 (1964).
5. Manly, B.F., "Exponential data transformations", *The Statistician*, 25: 37-42 (1976).
6. John, J.A., Draper, N.R., "An alternative family of transformations", *Applied Statistics*, 29: 190-197 (1980).
7. Bickel, P.J., Doksum, K.A., "An analysis of transformations revisited", *Journal of the American Statistical Association*, 76: 296-311 (1981).
8. Yeo, I.K., Johnson, R.A., "A new family of power transformations to improve normality or symmetry", *Biometrika*, 87(4): 954-959 (2000).
9. McCullagh, P., Nelder, J.A., "Generalized Linear Models, 2nd ed.", *Chapman and Hall*, London, 105 (1989).
10. Spitzer, J.J., "A Monte-Carlo investigation of the Box-Cox transformation in small samples", *Journal of the American Statistical Association*, 73: 488-495 (1978).
11. Bartlett, M.S., "The use of transformations", *Biometrics*, 3: 39-52 (1947).
12. van Zwet, W.R., "Convex transformations of random variables", *Mathematisch Centrum*, Amsterdam (1964).
13. Poirier, D.J., "The use of the Box-Cox transformation in limited dependent variable models", *Journal of the American Statistical Association*, 73: 284-287 (1978).
14. Atkinson, A.C., Cox, D.R., "Transformations", *Encyclopedia of Statistical Sciences*, 9, Kotz, S., Johnson, N.L., *John Wiley & Sons*, New York, 312-318 (1981).

15. Seber, G.A.F., "Linear Regression Analysis", *John Wiley & Sons*, New York, 50 (1977).
16. Zarembka, P., "Functional form in the demand for money", *Journal of the American Statistical Association*, 63: 502-511 (1968).
17. Schlesselman, J., "Power families: A note on the Box and Cox transformation", *Journal of the Royal Statistical Society, Series B*, 33: 307-311 (1971).
18. Cook, R.D., Weisberg, S., "Transforming a response variable for linearity", *Biometrika*, 81(4): 731-737 (1994).
19. Hernandez, F., Johnson, R.A., "The large sample behavior of transformations to normality", *Journal of the American Statistical Association*, 75: 855-861 (1980).
20. Spitzer, J.J., "A primer on Box-Cox estimation", *The Review of Economics and Statistics*, 64: 307-313 (1982).
21. Lawton, W.H., Sylvestre, E.A., "Elimination of linear parameters in nonlinear regression", *Technometrics*, 13(3): 461-467 (1971).
22. Spitzer, J.J., "A fast and efficient algorithm for the estimation of parameters in models with the Box-and-Cox transformation", *Journal of the American Statistical Association*, 77: 760-766 (1982).
23. Kim, J., "Iterated grid search algorithm on unimodal criteria", Doktora Tezi, *Virginia Polytechnic Institute and State University*, Virginia, 13-14 (1997).
24. Searle, S.R., "Linear Models", *John Wiley & Sons*, New York, 99 (1971).
25. Mendenhall, W., Wackerly, D.D. and Scheaffer, R.L., "Mathematical Statistics with Applications", *Duxbury Press*, California, 165-166 (1990).
26. Patterson, K., "An Introduction to Applied Econometrics: A Time Series Approach", *Palgrave Publications*, Wiltshire, 161 (2000).
27. Greene, W.H., "Econometric Analysis, 2nd ed.", *Prentice-Hall International, Inc.*, New Jersey, 358-360 (1993).
28. Kotz, S., Johnson, N.L., "Lilliefors Test", Encyclopedia of Statistical Sciences, 4, Kotz, S., Johnson, N.L., *John Wiley & Sons*, New York, 650 (1981).
29. Conover, W.J., "Practical Nonparametric Statistics, 2nd ed.", *John Wiley & Sons*, New York, 357-361, 367 (1980).

30. Layard, M.W.J., "Robust Large-sample tests for homogeneity of variances", *Journal of the American Statistical Association, Theory and Methods Section*, 68: 195-198 (1973).
31. Szroeter, J., "A class of parametric test for heteroscedasticity in linear econometric models", *Econometrica*, 46(6): 1311-1327 (1978).
32. Scheffé, H., "The Analysis of Variance", *John Wiley & Sons*, New York, 83-87 (1959).
33. Bartlett, M.S., "Properties of sufficiency and statistical tests", *Proceedings of the Royal Society A*, 160: 262-282 (1937).
34. Glaser, R.E., "Bartlett's test of homogeneity of variances", Encyclopedia of Statistical Sciences, 1, Kotz, S., Johnson, N.L., *John Wiley & Sons*, New York, 189-191 (1981).
35. Fishman, G.S., "Monte-Carlo: Concepts, Algorithms, and Applications", *Springer-Verlag*, New York, 577 (1996).
36. Law, A.M., Kelton, W.D., "Simulation Modeling & Analysis, 2nd ed.", *McGraw-Hill, Inc.*, New York, 497, 487-490 (1991).
37. Bratley, P., Fox, B.L., Schrage, L.E., "A Guide to Simulation", *Springer Verlag*, New York, 150-153 (1983).
38. Ata, Mustafa Y., "Sanal Deneylerde Monte-Carlo Tahmin İçin Deneysel Bir Yakınsama Kısıtası", *Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Teknik Rapor No: 2002-01T*, Ankara (2002).



EKLER

EK-1 REGRESYON MODELİNDE DÖNÜŞÜMLERİN JACOBIENLERİ

“n” gözlemden oluşan regresyon modelinde dönüşümlerin Jacobieni

$$J(\lambda; \mathbf{Y}) = \left| \frac{\partial \mathbf{Y}(\lambda)}{\partial \mathbf{Y}} \right| = \prod_{i=1}^n \frac{\partial Y_i(\lambda)}{\partial Y_i} \quad [1]$$

şeklinde gösterilebilir. Her bir dönüşüm için [1] numaralı eşitlik türetilirse aşağıdaki Jacobienler elde edilir:

Dönüşüm	Jacobien
Tukey [2.1]	$J(\lambda; \mathbf{Y}) = \left(\prod_{i=1}^n Y_i \right)^{\lambda-1}$
Box-Cox [2.2]	$J(\lambda; \mathbf{Y}) = \left(\prod_{i=1}^n Y_i \right)^{\lambda-1}$
Box-Cox [2.3]	$J(\lambda; \mathbf{Y}) = \left(\prod_{i=1}^n (Y_i + \lambda_2) \right)^{\lambda_1-1}$
Manly [2.4]	$J(\lambda; \mathbf{Y}) = e^{\lambda \sum_{i=1}^n Y_i}$
John-Draper [2.5]	$J(\lambda; \mathbf{Y}) = \left(\prod_{i=1}^n (Y_i + 1) \right)^{\lambda-1}$
Bickel-Doksum [2.6]	$J(\lambda; \mathbf{Y}) = \left(\prod_{i=1}^n Y_i \right)^{\lambda-1}$
Yeo-Johnson [2.7]	$J(\lambda; \mathbf{Y}) = \left(\prod_{i=1}^n (Y_i + 1)^{\sigma_i} \right)^{\lambda-1}$

EK-2 GAMMA DAĞILIMI İÇİN ÇARPIKLIK KATSAYISI

X rassal değişkeni α ve β parametreli Gamma dağılımına sahipse,

$$f(X; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0, \quad \alpha, \beta > 0$$

$$= 0, \quad \text{dd}$$
[1]

yazılabilir. Olasılık yoğunluk fonksiyonu [1] ile ifade edilen X rassal değişkeninin moment çıkaran fonksiyonu [2] numaralı eşitlikteki gibidir:

$$M_X(t) = E(e^{tx}) = (1 - \beta t)^{-\alpha}$$
[2]

[2] numaralı eşitliğin t'ye göre 1. türevi alınırsa;

$$M'_X(t) = -\alpha(1 - \beta t)^{-\alpha-1}(-\beta) = \alpha\beta(1 - \beta t)^{-\alpha-1}$$
[3]

elde edilir. [3] numaralı eşitlikte t yerine 0 yazılarak, X değişkeninin ortalaması,

$$E(X) = \alpha\beta$$
[4]

bulunabilir. [2] numaralı ifadenin t'ye göre 2. türevi alınırsa,

$$M''_X(t) = -\alpha\beta(\alpha + 1)(1 - \beta t)^{-\alpha-2}(-\beta) = \alpha\beta^2(\alpha + 1)(1 - \beta t)^{-\alpha-2}$$
[5]

bulunur. [5] numaralı eşitlikte t yerine 0 yazıldığında,

$$E(X^2) = \alpha^2\beta^2 + \alpha\beta^2$$
[6]

elde edilir. Böylece X değişkeninin varyansı,

$$V(X) = E(X^2) - [E(X)]^2 = \alpha\beta^2$$
[7]

olur.

Öte yandan [2] numaralı ifadenin t'ye göre 3. türevi alınırsa,

$$M'''_X(t) = \alpha\beta^3(\alpha + 1)(\alpha + 2)(1 - \beta t)^{-\alpha-3}$$
[8]

bulunur. [8] numaralı eşitlikte t yerine 0 yazılırsa,

$$E(X^3) = \alpha^3\beta^3 + 3\alpha^2\beta^3 + 2\alpha\beta^3$$
[9]

olur. Ortalama etrafındaki 3. moment,

$$\mu_3 = E(X - \mu)^3 = E(X^3) - [E(X)]^3 - 3E(X)V(X)$$
[10]

şeklinde ifade edilebilir. [4], [7] ve [9] numaralı eşitlikler, [10] numaralı eşitlikte yerine yazılırsa,

$$\mu_3 = \alpha^3 \beta^3 + 3\alpha^2 \beta^3 + 2\alpha(\beta^3) - \alpha^3 \beta^3 - 3\alpha^2 \beta^3 = 2\alpha(\beta^3) \quad [11]$$

olarak elde edilir.

Diğer taraftan çarpıklık katsayısı, $\alpha_3 = \frac{\mu_3}{\sigma^3}$,

$$\alpha_3 = \frac{2\alpha(\beta)^3}{\alpha^{3/2}\beta^3} = \frac{2}{\sqrt{\alpha}}$$

olarak bulunur.



EK-3 [4.3a] RASSAL DEĞİŞKENİNİN BEKLENEN DEĞERİ

[4.3a] ile ifade edilen hata terimi,

$$\varepsilon_i = I_i \varepsilon_{1i} + (1 - I_i) \varepsilon_{2i} \quad [1]$$

şeklindedir. Burada I_i , $[0,1]$ aralığında değişen kesikli tekdüze değişken, $\varepsilon_{1i} \sim N(\mu_1, \sigma^2)$ ve $\varepsilon_{2i} \sim N(\mu_2, \sigma^2)$ 'dir. [1] numaralı ifadenin beklenen değeri alınır,

$$E(\varepsilon_i) = E\{I_i \varepsilon_{1i} + (1 - I_i) \varepsilon_{2i}\} = E(I_i \varepsilon_{1i}) + E[(1 - I_i) \varepsilon_{2i}] \quad [2]$$

elde edilir. Öte yandan I_i , ε_{1i} ve ε_{2i} 'den bağımsız olduğundan,

$$\begin{aligned} E(\varepsilon_i) &= E(I_i)E(\varepsilon_{1i}) + E(1 - I_i)E(\varepsilon_{2i}) \\ &= \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 = (\mu_1 + \mu_2)/2 \end{aligned} \quad [3]$$

elde edilir.

EK-4 TURBO C KODU

```

#include<stdio.h>
#include<stdlib.h>
#include<conio.h>
#include<math.h>
#include<string.h>

/*Alt program tanımlamaları*/
void MODEL();
void GIRDI();
void MONTECARLO();
void YAZ();
int ISARET(float sayi);
float TUK(float sayi,float LAMBDA1,float LAMBDA2);
float BC1(float sayi,float LAMBDA1);
float BC2(float sayi,float LAMBDA1,float LAMBDA2);
float MAN(float sayi,float LAMBDA1);
float JAD(float sayi,float LAMBDA1);
float BAD(float sayi,float LAMBDA1);
float YAJ(float sayi,float LAMBDA1);
float GAMMA(float ALFA,float BETA,float rsayil,float rsayi2);
float NORMAL(float ORTALAMA,float VARYANS,float standart);
float STNORMAL1(float rsayil,float rsayi2);
float STNORMAL2(float rsayil,float rsayi2);
float TD(float A,float B,float rsayil);
int KTD(int A,int B,float rsayil);
float RAS(double sayi);
float DONUSTUR(int J,float DL1,float DL2);
void HATMAT();
float HATKARTOP(float ZI[]);
float logJACOBIEN(float j11,float j12);
void AGTAR();
float LILLIEFORS();
float BARTLETT(int m);

/*Global değişken tanımlamaları*/
float
ESAY,PI,Y[99],HAT[99][99],Z[99],X[99][9],E[99],BETA[9],XUXI[9][9];
float LAMBDA1,LAMBDA2,R1,R2,ZTAB[381],ZPROB[381],MCHATA,MCTAHMIN;
float istatistik,WTOP,WORT,WENB,WENK;
int ORCAP,BSAY,DONUSUM;
long N,I,YGS,SGS;
float far XUXIXU[9][120],far XXUXIXU[120][120],far XUXYeni[9][9],far
XUX[9][9];

/*Ana program*/
main()
{
    int j,l;

    FILE *SONUC;
    SONUC=fopen("SONUC.TXT","w");
    fclose(SONUC);

    for(DONUSUM=1;DONUSUM<=7;DONUSUM++)
    {
        if(DONUSUM==2) DONUSUM=3;
    }

```

```

clrscr();
GIRDI();

/*X değerlerini simüle eden kısım*/
for(j=1;j<=ORCAP;j++)
    for(l=1;l<=BSAY;l++)
    {
        if(l==1)
            X[j][l]=1;
        else
            if(l==2)
            {
                R1=RAS(R1);
                X[j][l]=KTD(10,30,R1);
            }
            if(l==3)
            {
                R2=RAS(R2);
                X[j][l]=KTD(40,80,R2);
            }
    }

HATMAT();

for(I=1;I<=N;I++)
{
    FILE *bosalt;
    bosalt=fopen("BART.TXT","w");
    fclose(bosalt);

    for(l=1;l<=25;l++)
        MODEL();
    istatistik=BARTLETT(25);
    MONTECARLO();
    YAZ();
    if(SGS==YGS) goto YENIDONUSUM;
}
YENIDONUSUM:
FILE *SONUC;
SONUC=fopen("SONUC.TXT","a");
fprintf(SONUC,"Secilen
donusum:%i\nMCTAHMIN:%f\n\n",DONUSUM,MCTAHMIN);
fclose(SONUC);

}

}

/*Alt fonksiyonlar burada başlıyor*/

/*Deneyde kullanılan MODEL*/
void MODEL()
{
    float BETA[4]={0,5,2,3}; /*Dizilerin ilk elemanı 0. eleman
olduğundan
aslında 3 elemanlı olması gereken dizi 4
elemanlı tanımlanıyor*/

```

```

int j,k;

float eort=0;

for(j=1;j<=ORCAP;j++)
{
    R1=RAS(R1); R2=RAS(R2);
    E[j]=NORMAL(0,((20-1)+j*(1-1))/19,STNORMAL1(R1,R2));
    eort=eort+E[j];
}
eort=eort/ORCAP;

for(j=1;j<=ORCAP;j++)
{
    E[j]=E[j]-eort;
    Y[j]=E[j];
    for(k=1;k<=BSAY;k++)
        Y[j]=Y[j]+X[j][k]*BETA[k];
}

AGTAR();

/*Bu program parçası Bartlett testi alt programı
   için gerekli olan artıkları bir dosyaya yazar.*/
FILE *BART;
BART=fopen("BART.TXT","a");
for(j=1;j<=ORCAP;j++)
    fprintf(BART,"%10f\n",E[j]);
fclose(BART);
}

/*GIRDI alt programı*/
void GIRDI()
{
    float Tohum1,Tohum2;
    int j;

    FILE *ZDEGER;
    ZDEGER=fopen("ZTAB.DAT","r");
    for(j=1;j<=381;j++)
        fscanf(ZDEGER,"%f %f",&ZTAB[j],&ZPROB[j]);
    fclose(ZDEGER);

    N=10000; YGS=200; SGS=0; ORCAP=20; BSAY=3; MCHATA=0.001;
    #define PI 3.14159265358979
    #define ESAY 2.7182818284591
    LAMBDA1=1; LAMBDA2=0;
    WENB=0; WENK=1; WTOP=0;
    Tohum1=0.123456; Tohum2=0.635437;
    R1=RAS(Tohum1);
    R2=RAS(Tohum2);
}

/*MONTECARLO tahminini yapan alt program*/
void MONTECARLO()
{

```

```

int W;
if(istatistik<30.1435) W=1; else W=0;
WTOP=WTOP+W;
WORT=WTOP/I;
SGS=SGS+1;
if(WORT>WENB || WORT<WENK)
{
    if(WORT>WENB)
    {
        WENB=WORT;
        WENK=WENB-2*MCHATA;
        MCTAHMIN=(WENB+WENK)/2;

        SGS=0;
    }
    if(WORT<WENK)
    {
        WENK=WORT;
        WENB=WENK+2*MCHATA;
        MCTAHMIN=(WENB+WENK)/2;
        SGS=0;
    }
}

/*Sonuçları ekrana yazan program*/
void YAZ()
{
    clrscr();
    printf("Örnek çapı:%4i",ORCAP);
    printf("\nSeçilen dönüşüm:%2i",DONUSUM);
    printf("\nDeney=%8i",I);
    printf("\nHata bandının dışına çıkılmayan gözlem sayısı=%8i",SGS);
    printf("\nMC ortalama istatistiği=%3.8f",WORT);
    printf("\nMC ort. ist. en büyük değeri=%3.8f",WENB);
    printf("\nMC ort. ist. en küçük değeri=%3.8f",WENK);
    printf("\nMCTahmin değeri=%3.8f",MCTAHMIN);
    printf("\nMCHata değeri=%3.8f",MCHATA);
    printf("\nLambda değeri=%3.8f",LAMBDA1);
    printf("\nİstatistik değeri=%5.8f",istatistik);
}

/*İşaret fonksiyonu*/
int ISARET(float sayi)
{
    int sonuc;
    if(sayi>0){sonuc=1;}
    if(sayi<0){sonuc=-1;}
    if(sayi==0){sonuc=0;}
    return sonuc;
}

/*Kuvvet dönüşümlerini yapan fonksiyonlar
Tukey          : TUK
Box&Cox1       : BC1
Box&Cox2       : BC2
Manly          : MAN

```

```

John&Draper   : JAD
Bickel&Doksum: BAD
Yeo&Johnson  : YAJ */

/*Tukey dönüşümünü yapan fonksiyon*/
float TUK(float sayi,float LAMBDA1,float LAMBDA2)
{
    float sonuc;
    if(LAMBDA1!=0) sonuc=pow(sayi+LAMBDA2,LAMBDA1);
    else sonuc=log(sayi+LAMBDA2);
    return sonuc;
}

/*Box&Cox1 dönüşümünü yapan fonksiyon*/
float BC1(float sayi,float LAMBDA1)
{
    float sonuc;
    if(LAMBDA1!=0) sonuc=(pow(sayi,LAMBDA1)-1)/LAMBDA1;
    else sonuc=log(sayi);
    return sonuc;
}

/*Box&Cox2 dönüşümünü yapan fonksiyon*/
float BC2(float sayi,float LAMBDA1,float LAMBDA2)
{
    float sonuc;
    if(LAMBDA1!=0) sonuc=(pow(sayi+LAMBDA2,LAMBDA1)-1)/LAMBDA1;
    else sonuc=log(sayi+LAMBDA2);
    return sonuc;
}

/*Manly dönüşümünü yapan fonksiyon*/
float MAN(float sayi,float LAMBDA1)
{
    float sonuc;
    if(LAMBDA1!=0) sonuc=(exp(LAMBDA1*sayi)-1)/LAMBDA1;
    else sonuc=sayi;
    return sonuc;
}

/*John&Draper dönüşümünü yapan fonksiyon*/
float JAD(float sayi,float LAMBDA1)
{
    float sonuc;
    if(LAMBDA1!=0) sonuc=ISARET(sayi)*(pow(fabs(sayi)+1,LAMBDA1)-1)/LAMBDA1;
    else sonuc=ISARET(sayi)*log(fabs(sayi)+1);
    return sonuc;
}

/*Bickel&Doksum dönüşümünü yapan fonksiyon*/
float BAD(float sayi,float LAMBDA1)
{
    float sonuc;
    sonuc=((ISARET(sayi)*pow(fabs(sayi),LAMBDA1))-1)/LAMBDA1;
    return sonuc;
}

```

```
/*Yeo&Johnson dönüşümünü yapan fonksiyon*/
float YAJ(float sayi,float LAMBDA1)
{
    float sonuc;
    if(sayi>=0)
    {
        if(LAMBDA1!=0) sonuc=(pow(sayi+1,LAMBDA1)-1)/LAMBDA1;
        else sonuc=log(sayi+1);
    }
    else
    {
        if(LAMBDA1!=2) sonuc=(pow(-sayi+1,2-LAMBDA1)-1)/(2-LAMBDA1);
        else sonuc=-log(-sayi+1);
    }
    return sonuc;
}
```



ÖZGEÇMİŞ

Hüseyin GÜLER, 1980 yılında Ankara’da doğmuştur. Lise öğrenimini Ankara Kurtuluş Lisesinde tamamladıktan sonra 1997 yılında Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik bölümünde lisans öğrenimine başlamıştır. 2001 yılında aynı bölümü birincilikle bitirerek mezun olan Hüseyin GÜLER, 2001 Eylül ayında Gazi Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik anabilim dalında yüksek lisansa başlamıştır. 2002 yılında Gazi Üniversitesi, Ekonometri bölümünde araştırma görevlisi olarak iş hayatına başlayan Hüseyin GÜLER, halen aynı bölümde hizmet vermektedir.

