

**BAĞIMLI HASAR MODELLERİNDE  
KOPULA REGRESYON YAKLAŞIMI**

**DOKTORA TEZİ**

**Emine Selin SARIDAŞ**

**İstatistik Anabilim Dalı**

**İstatistik Programı**

**Tez Danışmanı: Doç. Dr. Eylem DENİZ  
Eş Danışman: Prof. Dr. Edward. W. FREES**

**OCAK 2020**



**BAĞIMLI HASAR MODELLERİNDE  
KOPULA REGRESYON YAKLAŞIMI**

**DOKTORA TEZİ**

**Emine Selin SARIDAŞ**

**İstatistik Anabilim Dalı**

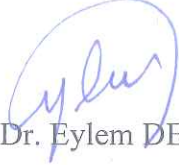
**İstatistik Programı**

**Tez Danışmanı: Doç. Dr. Eylem DENİZ  
Eş Danışman: Prof. Dr. Edward. W. FREES**

**OCAK 2020**



Emine Selin SARIDAŞ tarafından hazırlanan BAĞIMLI HASAR MODELLERİNDE KOPULA REGRESYON YAKLAŞIMI adlı bu tezin doktora tezi olarak uygun olduğunu onaylarım.

  
Doç . Dr. Eylem DENİZ HOWE  
Tez Danışmanı

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

Danışman : Doç. Dr. Eylem DENİZ HOWE

Üye : Prof. Dr. Aydın ERAR

Üye : Prof. Dr. Ali Hakan BÜYÜKLÜ

Üye : Prof. Dr. Ali KÖSE

Üye : Dr. Öğr. Üyesi Elif Özge ÖZDAMAR

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Lisansüstü Tez Yazım Kılavuzuna uygun olarak yazılmıştır.

Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım klavuzuna uygun olarak hazırladığım bu tez çalışmada;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel etik kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ücret karşılığı başka kişilere yazdırmadığımı (dikte etme dışında), uygulamalarımı yaptırmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

Emine Selin Sarıdaş

....

## ÖNSÖZ

Doktora tez sürecim boyunca desteğini esirgemeyen ve karşılaşılan güçlüklerin aşılmasında yardımcı olan danışmanım Doç. Dr. Eylem Deniz'e,

Tez çalışmamda büyük katkısı olan, Wisconsin-Madison Üniversitesi'nde bulunduğum zaman içerisinde ve sonrasında danışmanlığıyla bana yol gösteren, değerli zamanımı ayıran Prof. Dr. Edward W. Frees'e,

Tez çalışmamı, Wisconsin-Madison Üniversitesi'nde sürdürdüğüm dönemde 2214-A Yurt Dışı Doktora Sırası Araştırma Burs Programı ile destekleyen Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK)'na,

Beni destekleyen ve yanımda olduğunu hissettiren arkadaşlarıma,

Hayatımın her anında gösterdikleri sonsuz sevgi ve emek için aileme,  
teşekkürlerimi sunarım.

Emine Selin Sarıdaş

## İÇİNDEKİLER

### Sayfa

ÖNSÖZ .....	vii
İÇİNDEKİLER .....	ix
KISALTMALAR.....	xi
SEMBOLLER .....	xiii
ÇİZELGE LİSTESİ.....	xv
ŞEKİL LİSTESİ.....	xvii
ÖZET .....	xix
SUMMARY .....	xxi
<b>1. GİRİŞ .....</b>	<b>1</b>
1.1 Literatür .....	3
<b>2. HASAR MODELLERİ .....</b>	<b>7</b>
2.1 İki Parçalı Hasar Modeli.....	7
2.2 Toplamsal Hasar Modeli.....	9
2.3 Genelleştirilmiş Doğrusal Model .....	9
2.4 Hasar Sayılarının Modellenmesi .....	12
2.4.1 Poisson Genelleştirilmiş Doğrusal Modeli .....	12
2.4.2 Negatif Binom Genelleştirilmiş Doğrusal Modeli.....	13
2.4.3 Sıfır Yığılmalı Modeller .....	14
2.5 Hasar Şiddetinin Modellenmesi .....	15
2.5.1 Gamma Regresyon Modeli .....	15
2.5.2 Genelleştirilmiş Beta 2 Regresyon Modeli .....	16
<b>3. KOPULA TEORİSİ.....</b>	<b>19</b>
3.1 Bağımlılık Kavramı .....	20
3.1.1 Rank Korelasyon Katsayısı .....	20
3.1.1.1 Spearman'ın rho korelasyon katsayısı .....	21
3.1.1.2 Kendall'ın tau korelasyon katsayısı .....	21
3.2 Kopula Fonksiyonları .....	22
3.2.1 Olasılık İntegral Dönüşümü .....	22
3.2.2 Sklar Teoremi .....	23
3.2.3 Fréchet–Hoeffding Sınırları.....	25
3.2.4 Kopula Fonksiyon Aileleri .....	26
3.2.4.1 Eliptik Kopula Fonksiyonları.....	26
Normal (Gaussian) Kopula .....	26
t Kopula .....	27
3.2.4.2 Arşimedyen Kopula Fonksiyonları.....	28
Frank Kopula.....	28
Clayton Kopula.....	29



Gumbel - Hougaard Kopula .....	29
3.2.5 Kopula Tahmin Yöntemleri .....	30
3.2.5.1 En Çok Olabilirlik Yöntemi.....	30
3.2.5.2 Marjinallere İlişkin Çıkarsama Yöntemi .....	31
<b>4. TOPLAM HASAR TUTARI TAHMİNİ İÇİN KOPULA REGRESYON MODELİ.....</b>	<b>33</b>
4.1 Veri Hakkında.....	33
4.2 Üç Parçalı Toplam Hasar Modeli .....	34
4.2.1 Hasar Kategorisi için Lojistik Regresyon Modeli .....	35
4.3 Kopula Regresyon Modeli.....	38
<b>5. UYGULAMA .....</b>	<b>41</b>
5.1 Hasar Sıklığı Modellemesi Sonuçları.....	43
5.2 Hasar Şiddeti Modellemesi Sonuçları .....	45
5.3 Hasar Kategorisi Modeli Sonuçları .....	48
5.4 Kopula Regresyon Modeli Sonuçları.....	49
5.5 Model Doğrulama.....	51
<b>6. SONUÇ VE ÖNERİLER .....</b>	<b>55</b>
<b>KAYNAKLAR.....</b>	<b>57</b>
<b>EKLER .....</b>	<b>63</b>
EK A.1 .....	65
EK B.1 .....	67
EK C.1 .....	69
EK D.1 .....	71
EK E.1 .....	73
<b>ÖZGEÇMİŞ .....</b>	<b>75</b>

## KISALTMALAR

<b>GDM</b>	: Genelleştirilmiş Doğrusal Model
<b>GB2</b>	: Genelleştirilmiş Beta 2
<b>RMSE</b>	: Hata Kareler Ortalamasının Karekökü
<b>MAE</b>	: Ortalama Mutlak Hata
<b>MLE</b>	: En Çok Olabilirlik Yöntemi
<b>IFM</b>	: Marjinallere İlişkin Çıkarsama Yöntemi
<b>AIC</b>	: Akaike Bilgi Kriteri
<b>s.d.</b>	: Serbestlik derecesi



## SEMBOLLER

$n$	: Örneklem büyüklüğü
$F$	: Kümülatif dağılım fonksiyonu
$f$	: Olasılık yoğunluk fonksiyonu
$C$	: Kopula dağılım fonksiyonu
$c$	: Kopula olasılık yoğunluk fonksiyonu
$Y$	: Hasar tutarı bağımlı değişkeni
$N$	: Hasar sayısı bağımlı değişkeni
$M$	: Hasar kategorisi bağımlı değişkeni
$g_p(x)$	: Eliptik kopula üreteç fonksiyonu
$\psi(t)$	: Arşimedyan kopula üreteç fonksiyonu
$Y_j$	: j. hasar türü için ödenen hasar tutarı

## ÇİZELGE LİSTESİ

	<u>Sayfa</u>
<b>Çizelge 4.1:</b> Gözlemlenen Hasar Dağılımları (Eğitim verisi, 2012–2015) .....	36
<b>Çizelge 5.1:</b> Hasar Sayısı Frekans Tablosu .....	43
<b>Çizelge 5.2:</b> Hasar Sıklığı Modellemesi Sonuçları.....	44
<b>Çizelge 5.3:</b> Gözlenen ve Beklenen Hasar Sayıları.....	44
<b>Çizelge 5.4:</b> Gamma ve Gb2 Regresyon Modelleri AIC Değerleri.....	47
<b>Çizelge 5.5:</b> Çoklu Logit Model Sonuçları.....	49
<b>Çizelge 5.6:</b> Kopula Model Karşılaştırması.....	51
<b>Çizelge 5.7:</b> Toplam Hasar Modeli Karşılaştırma Sonuçları .....	52
<b>Çizelge 5.8:</b> Tahmin Değerlerine ilişkin İstatistikler .....	53
<b>Çizelge 5.9:</b> n= 465. gözlem için Kantil değerleri .....	53
<b>Çizelge 5.10</b> n= 465. gözlem için Kantil değerleri .....	54

## ŞEKİL LİSTESİ

	<u>Sayfa</u>
<b>Şekil 5.1</b> : Kategorik ve sürekli açıklayıcı değişkenler.....	42
<b>Şekil 5.2</b> : Poisson Gdm Sonuçları .....	45
<b>Şekil 5.3</b> : $Y_1$ Hasar değişkeni için QQ Grafikleri.....	46
<b>Şekil 5.4</b> : $Y_2$ Hasar değişkeni için QQ Grafikleri.....	46
<b>Şekil 5.5</b> : $Y_3$ Hasar değişkeni için QQ Grafikleri.....	47
<b>Şekil 5.6</b> : $Y_4$ Hasar değişkeni için QQ Grafikleri.....	47
<b>Şekil B.1</b> : $Y_1$ Hasar tutarı değişkeni için GB2 Model Sonuçları .....	67
<b>Şekil C.1</b> : $Y_2$ Hasar tutarı değişkeni için GB2 Model Sonuçları .....	69
<b>Şekil D.1</b> : $Y_3$ Hasar tutarı değişkeni için Gamma Model Sonuçları .....	71
<b>Şekil E.1</b> : $Y_4$ Hasar tutarı değişkeni için Gamma Model Sonuçları .....	73

## **BAĞIMLI HASAR MODELLERİNDE KOPULA REGRESYON YAKLAŞIMI**

### **ÖZET**

Kopula fonksiyonları çok deęişkenli ortak daęılım fonksiyonlarıdır ve deęişkenler arasındaki baęımlılık yapısının modellenmesini saęlamaktadır. Farklı türdeki baęımlılık yapılarını modelleyebilmesi ve matematiksel çeşitlilięi nedeniyle kopula fonksiyonlarının literatürdeki kullanımı yaygınlaşmaktadır.

Hasar modellerinde baęımlı deęişkenler arasındaki baęımlılık yapısı genellikle göz ardı edilmektedir. Göz ardı edilen baęımlılık yapısı toplam hasar modelinin istatistiksel tahminini etkilemektedir. Baęımlılık, hasar sıklığı ile hasar tutarı arasında ya da farklı risklerden kaynaklanan hasar tutarları arasında gözlemlenebilir. Bu çalışmada farklı hasar tutarları arasındaki baęımlılık yapısı kopula fonksiyonları ile toplam hasar modeline dahil edilmiştir. Toplam hasar modeli oluşturulurken hasar sayısı ve hasar tutarı baęımlı deęişkenlerinin yanında hasar kategorisi tanımlanarak üçüncü bir bileşen olarak modele dahil edilmiştir. Toplam hasar modeli bu üç bileşenden oluşan model yardımıyla kurulmuş ve baęımlılık yapısını içeren bir kopula regresyon modeli önerilmiştir. Böylece daha fazla bilgiyi içeren esnek bir istatistiksel model elde edilmiştir.

Bu çalışmada Türkiye’de faaliyet gösteren özel bir sigorta şirketinden alınan zorunlu trafik sigortası verisi kullanılmıştır. Dört farklı hasar tutarı deęişkeni tanımlanmış ve aralarındaki baęımlılık yapısının toplam hasar tutarının tahminini etkiledięi görülmüştür. Baęımlılık yapısının modele dahil edilmesinin toplam hasar tutarının tahmininin yanında toplam hasar daęılımının kuyruk kısımlarındaki kantil deęerlerinin tahmini de etkiledięi sonucu elde edilmiştir.

## **COPULA REGRESSION APPROACH FOR DEPENDENT LOSS MODELS**

### **SUMMARY**

Copula functions are multivariate cumulative distribution functions and provide the modeling of the dependency structure between the random variables. The applications of copula functions in the literature is becoming widespread due to its ability to model different types of dependency structures and its mathematical diversity.

The dependency structure between the response variables in the loss models is mostly ignored in the studies. The ignored dependency structure affects the statistical estimated values of the total claim model. Dependency structure can be observed between claim frequency and claim severity or between the different types of claim severities in the data. In this study, the dependency structure between different types of claim severities is included in the total claim model with copula functions. The claim category component is added as a third component to the total claim model as well as claim frequency and the claim severity components. Three component model is used for the estimation of the total claim model and copula regression model regarding the dependency structure is proposed. Thus, a flexible statistical model containing more information about the data is obtained.

In this study, traffic auto insurance data set provided by one of the Turkish insurance companies is used for the implementation of the copula regression approach. Four different claim severity variables are defined and it is seen that the dependency structure between them affects the estimation of the total claim amount. It is concluded that the considering the dependency structure in the model would affect the estimation of the quantil values in the tail parts of the total claim distribution as well as the estimation of the total claim amount.

## 1. GİRİŞ

Hasar (claim) modellemesi, aktüerya alanında poliçe priminin hesaplanması, rezerv hesaplaması ve portföy riskinin değerlendirilmesi gibi temel aktüeryal problemlerin çözümünde önemli rol oynamaktadır. Prim hesaplaması aşamasında genellikle hasar sıklığı/frekansı (frequency) ve hasar şiddeti/tutarı (severity) için değişen risk faktörlerine duyarlı modeller belirlendikten sonra fiyatlandırmanın iyi yapılabilmesi, gelecek yükümlülüklerin ve rezervlerin doğru olarak belirlenmesi amaçlanmaktadır.

Hasar modellemesinde kullanılan temel yaklaşım hasar frekansının ve hasar tutarının ayrı ayrı modellendiği iki aşamalı model yaklaşımıdır. Toplam hasar tutarını tahmin edebilmek için hasar sıklığı ve hasar tutarı yanıt değişkenleri için oluşturulacak istatistiksel modellerin doğru tahmin edilmesi önemlidir. Hasar sıklığı ve hasar tutarı rastlantı değişkenleri ile açıklayıcı değişkenler arasındaki istatistiksel ilişki alternatif regresyon modelleri aracılığıyla kurulabilmektedir. Hasar sıklığı rastlantı değişkeni kesikli yapıda ve hasar tutarı rastlantı değişkeni sürekli yapıda olduğu için en az iki ayrı regresyon modeli oluşturulmalıdır.

Hasar sıklığının modellenmesi için genellikle Poisson, Negatif Binom veya genelleştirilmiş Poisson dağılımları kullanılmaktadır. Poisson dağılımının en belirgin özelliği ortalaması ve varyansının birbirine eşit olmasıdır. Dağılımın ortalaması ve varyansının eşit olmadığı durumlarda aşırı yayılım (overdispersion) ya da az yayılım (underdispersion) durumundan söz edilebilir. Hasar sıklığı verisinde aşırı yayılım gözlemlendiğinde negatif binom dağılımı ya da quasi-Poisson dağılımı kullanılabilir. Ayrıca bu tip veride sıfır değerinde yığılma görüldüğünde sıfır yığılmalı modeller (zero-inflated models) de kullanılmaktadır.

Hasar tutarının dağılımı ise genellikle çarpık ve kalın kuyruklu olup en sık kullanılan hasar tutarı dağılımları Gamma ve Lognormal dağılımlarıdır. Ters Gaussian, Weibull dağılımları da çarpık yapı gözlemlenen hasar tutarı verisinde alternatif dağılım olabilmektedir [1].



Hasar sıklığı ve hasar tutarı dağılımları ayrı ayrı belirlenebileceği gibi kesikli ve sürekli rastlantı değişkenlerinin birlikte yer aldığı karma bir dağılım yardımıyla da modellenebilir.

Aktüeryal çalışmalarda genellikle hasar modellerinde bağımlı değişkenler arasındaki bağımlılık yapısı göz ardı edilmektedir. Göz ardı edilen bağımlılık yapısı toplam hasar modelinin istatistiksel tahminini etkilemektedir. Bağımlılık, hasar sıklığı ile hasar tutarı arasında ya da farklı risklerden kaynaklanan hasar tutarları arasında gözlemlenebilir. Bu çalışmada hasar sıklığı ve hasar tutarı arasındaki bağımlılık yapısı koşullu olasılık yaklaşımı ile modele dahil edilirken, ödenen farklı hasar tutarları arasındaki bağımlılık yapısı kopula fonksiyonları (copula functions) ile toplam hasar modeline dahil edilmiştir.

Kopula fonksiyonları matematiksel elverişliliği ve farklı türlerdeki bağımlılık yapılarını kontrol edebildiği için bağımlılık yapısının modellenmesinde tercih edilmektedir. Bu fonksiyonların temeli, birleşik dağılım fonksiyonun marjinal dağılımlar kullanılarak ifade edilmesine dayanmaktadır. Marjinal dağılımlar üzerinde bir kısıt olmaması kopula fonksiyonlarının matematiksel işlenebilirliğini arttırmaktadır.

Bu çalışmada özel bir sigorta şirketinden alınan zorunlu trafik sigortası poliçe bilgilerinden elde edilen hasar verisi için bağımlılık yapısını içeren kopula regresyon modeli önerilmektedir. Kopula fonksiyonları regresyon modelleri ile birleştirilerek çok değişkenli aktüeryal hasar modeli oluşturulmuştur.

Hasar sıklığı ve hasar tutarı ayrı olarak modellenirken hasar tutarı; hasar türü ve hasar sayısı ile koşullu olarak modellenmiştir. Hasar kategorisi olarak üçüncü bir değişken tanımlanıp hasar modeline dahil edilmiştir. Hasar modeli oluşturulurken üç ayrı temel bileşen ile çalışılmış ve önerilen modelin üçüncü bileşenini oluşturan hasar tutarı değişkeni modellenirken farklı hasar türleri arasındaki bağımlılık yapısı kopula fonksiyonları ile modele dahil edilmesi amaçlanmıştır.

Bu amaç doğrultusunda çalışmanın ikinci bölümünde hasar modelleri açıklanmış, Üçüncü bölümde rastlantı değişkenleri arasındaki bağımlılığın modellenmesinde kullanılan kopula fonksiyonları tanıtılmış ve dördüncü bölümde ise çalışmada önerilen farklı riskler arasındaki bağımlılığı modellemede kullanılan kopula regresyon modeli açıklanmıştır. Çalışmanın uygulamasının yer aldığı beşinci bölümde ise kullanılacak

olan veri seti ayrıntılı olarak açıklanmış, veri setindeki bağımlılık yapısını modelleyen kopula regresyon modeli sonuçları verilmiştir. Altıncı bölümde ise çalışmanın genel sonuçları belirtilmiş ve önerilen modele ileride yapılabilecek katkılar tartışılmıştır.

## 1.1 Literatür

Tez çalışmasının içeriğinde kopula fonksiyonları ve hasar modelleri olmak üzere iki temel başlık üzerinden ilerlendiği için bu bölümde her iki konudaki temel kaynaklara ve uygulama örneklerini içeren ilgili çalışmalara yer verilmiştir.

Literatür araştırmasına ilk olarak kopula fonksiyonları ile başlanmıştır. Kopula fonksiyonlarını olasılıksal metrik uzaylar kapsamında istatistik literatürüne tanıtan ilk kişi Abe Sklar'dır. Sklar tarafından olasılık dağılım fonksiyonları ile kopula fonksiyonları arasındaki istatistiksel ilişki teorik olarak açıklanmıştır [2].

Genest ve Mackay [3] iki değişkenli dağılım fonksiyonlarının kopula fonksiyonları yardımıyla yapılandırılabilceğini açıklamıştır. Nelson [4] ise kopula teorisini ayrıntılı olarak incelemiştir.

Genest ve Favre [5] kopula fonksiyonlarının temel teorisinin iki değişkenli bir kopula modeli oluşturulmak istendiğinde veri modellemesi aşamasında nasıl kullanılması gerektiğini açıklayan ilk çalışmalardan birini örneklemiştir.

Joe [6] çok değişkenli modeller ve bağımlılık yapılarını kopula teorisi merkezinde açıklamıştır.

Frees ve Valdez [7] hayat sigortası ürünlerinin fiyatlandırılmasını incelerken aktüeryal uygulamalardaki bağımlılık yapılarını modellemek için kopula fonksiyonlarını önermiştir.

Yang [8] çok değişkenli dağılımların bağımlık yapılarının modellenmesinde yeni kopula fonksiyonları önererek konunun gelişmesinde katkıda bulunmuştur.

Klugman [9] kopula fonksiyonlarını regresyon temelli modeller ile yapılandırarak klasik doğrusal regresyon modeline alternatif bir model oluşturmuştur.

Son yıllarda kopula fonksiyonları birden çok bağımlı değişken arasındaki ilişkileri modellemek için özellikle finans, istatistik ve aktüerya alanlarında kullanılan popüler yöntemlerden biri olmuştur.

Aktüeryal alanda kopulaların veri modellenmesindeki uygulamalarının örneklerinden bazıları Carriere (2000), Frees ve Valdez (1998), Klugman (2012), Spreeuw (2006), Luciano ve diğerleri (2008) tarafından çalışılmıştır. [7, 10–13].

Aktüerya literatürünün en geniş uygulama alanı olan sigorta sektöründe beklenen hasar tutarının ve hasar sayılarının modellenmesi önemli bir problemdir. Literatürde hasar sayısının modelleri ve hasar tutarı modellerinin ayrı ayrı incelendiği çalışmalar olduğu gibi toplam hasar tutarının birlikte modellendiği çalışmalar da bulunmaktadır.

David ve diğerleri [14] otomobil sigortası için hasar sayılarının modellenmesinde Poisson ve Negatif binom dağılımlarını kullanmıştır.

Fu ve diğerleri [15] hasar tutarını modellemek için Lognormal, Gamma ve Normal dağılım fonksiyonlarını kullanmış, yaptığı simülasyon çalışması ile Gamma dağılımının daha etkin tahmin sonuçları ürettiğini göstermiştir. Meyer [16] çalışmasında Lognormal, Gamma ve Weibull dağılımlarını kullanarak bayesyen yaklaşımla toplam hasar tutarını modellemiştir.

Hasar tutarı ve hasar şiddetinin birlikte modellendiği iki parçalı modeller ile ilgili literatürdeki çalışmaların çoğunluğu açıklayıcı değişkenlerin beklenen hasar tutarı üzerindeki etkisini incelememiştir [17, 18]. Hasar tutarı ve hasar şiddeti modellenirken açıklayıcı değişkenlerin göz önünde bulundurulduğu çalışmalar daha günceldir [19, 20].

Kramer ve diğerleri [21] çalışmasında sigorta hasar sayısı ve büyüklüklerini modellemek için bileşik kopula modeli önermiştir. Hasar sayısı ve büyüklüklerini modellemede birleşik kopula modeli öneren bir diğer çalışma Czado [22] tarafından yapılmıştır.

Hasar modelleri ile ilgili literatürde model bileşenleri arasındaki bağımlılık yapısının incelenmesi güncel bir araştırma konusudur. Son yıllarda özellikle hasar sayısı ve hasar tutarı arasındaki ilişkinin değerlendirilmesi yaygındır. Hasar sayısının toplam hasar tutarı modeli içerisinde açıklayıcı bir değişken olarak ele alındığı çalışmalar hasar sayısı ile hasar tutarı arasındaki bağımlılık varsayımını değerlendirmektedir [23–25].

Frees [19] hasar sayısı ile hasar tutarı arasındaki bağımlılık yapısını altı farklı sigorta branşı için kopula fonksiyonları ile modellemiştir. Çalışma sonucunda hasar sayısı

ile hasar tutarı arasındaki bağımlılığın farklı sigorta branşları arasındaki bağımlılık yapısına göre daha etkisiz olduğunu söylemiştir.





## 2. HASAR MODELLERİ

Hasar sayısı (claim frequency), sigortalanmış riskin gerçekleşmesi sıklığını tanımlarken, hasar tutarı/şiddeti (claim severity), sigortalı bir riskin gerçekleşmesi durumunda sigorta şirketi tarafından tanzim edilen tutarı ifade etmektedir.

Hayat dışı sigorta branşlarında düzenlenen sigorta poliçesi ile sigortalı hasarın ortaya çıkma riskine karşılık, sigortacı poliçe sahibinin maddi zararını tazmin etme sözü verilmektedir. Bu nedenle hayat dışı sigorta poliçeleri için hasar dağılımının tahmin edilmesi tüm aktüeryal hesaplamaları etkilemektedir.

Sigorta şirketleri, sigorta poliçesi risk primlerini belirleyebilmek, sermaye yeterliliği için rezerv hesaplamalarını yapabilmek için geçmiş hasar sayısı ve hasar tutarı bilgisini kullanarak istatistiksel modeller oluşturmaktadır. Hasar sayısı ve hasar tutarı iki parçalı modellenebileceği gibi (two parts model) toplamsal hasar modeli (aggregated loss model) tanımlanarak birleşik olarak da modellenebilmektedir. Model yaklaşımı seçimi araştırmacının elindeki verinin yapısına veya prim hesabında kullanılacak olan prim prensiplerine göre değişebilmektedir.

### 2.1 İki Parçalı Hasar Modeli

Hasar verisi kayıtlarından elde edilen hasar sayısı ve hasar tutarı verileri iki ayrı bağımlı değişken olarak ele alınmaktadır. İki parçalı modelde birinci bileşen hasar sayısı (sıklığı) modeli, ikinci bileşen hasar tutarı modelidir. İlk kısımda hasar ortaya çıkması olasılığı modellenirken, ikinci kısımda hasar olması koşulunda pozitif hasar tutarları modellenmektedir. Bu modelde, hasar sayısı ile hasar tutarı verisi birleştirilirken hasar ödemesi yapılmadığı durum için her iki bileşenin değeri de sıfır kabul edilmektedir [26]. İki parçalı modeller "Frekans- Şiddet Modelleri" (Frequency-Severity Models) olarak da adlandırılmaktadır.

Literatürdeki ilk çalışmalar iki-parçalı model yapısı içerisinde açıklayıcı değişkenleri modele dahil etmeden hasar frekansı ve hasar tutarını tahmin etmeye çalışmıştır. Açıklayıcı değişkenlerin modele dahil edilmesi ilk olarak ekonometrik bir çalışmada Cragg (1971) tarafından yapılmıştır [27].

Hasar kayıtlarında hasar ödemesi yapılmadığı durumlarda hasar tutarı değişkeni sıfır olarak tanımlanırken hasar sayısı da sıfır değerini almaktadır. Tanımlanan risk gerçekleştiğinde ve ödeme yapılan tutar pozitif bir değer ile tanımlandığında hasar sayısı da pozitif bir değer almaktadır. Riskin gerçekleşme olasılığı görece olarak düşük risk olarak değerlendirilirse veri yapısının içerisinde pozitif hasar ödemesi olmadığı durumlar için çok sayıda sıfır değeri gözlemlenecektir. İki parçalı model yaklaşımı, hasar verisinde yüksek miktarda gözlemlenen sıfır değerlerinin kontrolü için elverişli olmaktadır. Sadece ödenen hasar tutarları bilgisinin olduğu bir veri seti sıfır değerini alan hasar ödemelerine ilişkin poliçe bilgilerinin kullanılmaması sonucunda hasar dağılımını tahmin ederken yanlı sonuçlar ortaya çıkmasına neden olabilmektedir [28].

$r_{i,i}$  birime ait bir hasar olup olmadığını gösteren ikili (binary) değişken olarak tanımlansın.  $y_i$ ; hasar tutarı olmak üzere iki parçalı hasar modelinin en temel yapısı,

$$(hasar)_i = r_i \times y_i \quad (2.1)$$

şeklinde [28].

İki-parçalı hasar modelinde açıklayıcı değişkenlerin modele dahil edildiği yapının iki bileşeni aşağıda açıklanmıştır:

İlk Bileşen:  $r_i$  bağımlı değişken,  $\mathbf{x}_{1i}$  açıklayıcı değişken matrisi,  $\beta_1$  regresyon katsayı vektörü olmak üzere ikili (binary) regresyon modeli tanımlanabilir.  $r_i$  bağımlı değişkeni için logit ve probit regresyon modelleri kullanılabilir.

İkinci Bileşen:  $r_i=1$  koşulunda  $y_{1i}$  bağımlı değişken,  $\mathbf{x}_{2i}$  açıklayıcı değişken matrisi,  $\beta_2$  regresyon katsayı vektörü olmak üzere regresyon modeli tanımlanabilir [28].  $y_{1i}$  bağımlı değişkeni için doğrusal, lognormal ya da gamma regresyon modelleri aday olarak seçilebilmektedir.

Hasar şiddeti ve hasar frekansının birleşik olasılık fonksiyonu iki parçalı model yaklaşımı ile Eşitlik 2.2'de ifade edilmiştir.

$$f(N, y) = f(N) \times f(y|N) \quad (2.2)$$

Bu eşitlikte  $f(N, y)$  iki bileşenden oluşan birleşik olasılık fonksiyonunu göstermektedir.

Hasar frekansı olasılık fonksiyonu,  $f(N)$ ,  $N$  tane hasarın ortaya çıkma olasılığını göstermektedir. Koşullu hasar şiddeti,  $f(y|N)$ ,  $N$  bilindiğinde hasar şiddeti olasılık yoğunluğunu ifade etmektedir [11]. Koşullu olasılık yaklaşımı kullanılarak oluşturulan bu yapısal ayrıştırma ile hasar frekansı ve hasar şiddeti arasındaki bağımsızlık varsayımı esnetilmektedir. Hasar şiddeti ile hasar frekansı değişkenleri arasındaki bağımlılık yapısının modelleyen çalışmalar halen devam etmektedir [19,22].

İki-parçalı hasar modellerinin ikinci kısmında pozitif hasarlar doğrusal regresyon veya genelleştirilmiş doğrusal model ile modellenmektedir. Model bileşenlerinin olabilirlik fonksiyonları ayrı ayrı maksimize edilerek en çok olabilirlik yöntemi ile parametre tahminleri yapılabilmektedir [29].

## 2.2 Toplamsal Hasar Modeli

Hayat dışı sigorta poliçesinde bir yıl içerisinde birden fazla riskin ortaya çıktığı ve birden fazla hasar bildirim yapıldığı durumda, hasar tutarı ödemeleri kümülatif bilgisi ile çalışılmak istendiğinde toplamsal hasar modeli yaklaşımı kullanılabilir.

Her bir poliçe için  $N_i$ : hasar sayısı

$y_{ij}$ ,  $j: 1, \dots, N_i$  olmak üzere, her bir poliçe için hasar tutarını ifade etmektedir.

$S_i = y_{i1} + y_{i2} + \dots + y_{iN}$  ise  $i$ . poliçe için toplamsal hasar tutarını göstermektedir.

$N_i = 0$  için  $y_{ij}$  değeri sıfır olacaktır.  $N_i$  değişkeninin ikili (binary) seçildiği durumda Eşitlik 2.1'deki iki parçalı hasar modeli elde edilmektedir [28].

## 2.3 Genelleştirilmiş Doğrusal Model

Aktüeryal çalışmalarda hasar verisi genellikle uzun kuyruklu yapıdadır ve çarpık dağılım göstermektedir. Bu nedenle hasar verisinin modellenmesinde klasik regresyon yaklaşımı yerine genelleştirilmiş doğrusal modeller tercih edilmektedir [30].

Genelleştirilmiş doğrusal modellerde temel varsayım bağımlı değişkenin dağılımının üstel aile üyesi olmasıdır. Bağımlı değişkenin ortalamasının dönüşümü ile açıklayıcı değişkenler arasındaki ilişki incelenmektedir.

Üstel dağılım ailesi genel yapısı Eşitlik 2.3'de gösterilmiştir.



$$f(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + S(y, \phi)\right) \quad (2.3)$$

Bu eşitlikte  $y$  bağımlı değişken,  $\theta$  kanonik parametre,  $\phi$  ölçek parametresi veya yayılım parametresi olarak adlandırılmaktadır.  $S(y, \phi)$ , bağımlı değişken ve ölçek parametresinin bir fonksiyonudur. Bağımlı değişken kesikli, sürekli ya da karma yapıda olabilmektedir [28].

Üstel dağılım aile üyesi olan Normal dağılım için olasılık yoğunluk fonksiyonu eşitlik 2.4'deki formda ifade edilmektedir.  $\theta = \mu, \phi = \sigma^2, b(\theta) = \theta^2/2$  ve  $S(y, \phi) = -y^2/(2\phi) - \ln(2\pi\phi)/2$  olmak üzere,

$$f(y; \mu, \sigma^2) = \exp\left(\frac{(y\mu - \mu^2/2)}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2)\right) \quad (2.4)$$

eşitliği üstel aile yapısındadır.

GDM'de  $\eta_i = \mathbf{x}'_i \boldsymbol{\beta} = \mathbf{g}(\mu_i)$  eşitliği sistematik bileşeni tanımlamaktadır. Bu eşitlikteki  $\mathbf{g}(\cdot)$ ; bağ (link) fonksiyonudur.

$$\mu_i = \mathbf{g}^{-1}(\mathbf{x}'_i \boldsymbol{\beta}) \quad (2.5)$$

Eşitlik 2.5'de bağ fonksiyonun tersi ile ortalama fonksiyonu tanımlanmıştır.

Üstel dağılım ailesinin log-olabilirlik fonksiyonu

$$l(\theta, \phi; y) = \ln f(y; \theta, \phi) = \left\{ \frac{y\theta - b(\theta)}{\phi} + S(y, \phi) \right\} \quad (2.6)$$

ile ifade edilmektedir.

$E\left(\frac{\partial l}{\partial \theta}\right) = 0$  olmak üzere log-olabilirlik fonksiyonunun  $\theta$ 'ya göre kısmî türevi alındığında bağımlı değişkenin ortalaması bulunmaktadır.

$$\frac{\partial l}{\partial \theta} = \frac{Y - b'(\theta)}{\phi}$$

$$E\left(\frac{\partial l}{\partial \theta}\right) = \frac{E(Y) - b'(\theta)}{\phi} = 0$$

$$E(Y) = \mu = b'(\theta) \quad (2.7)$$

$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0$  olmak üzere log-olabilirlik fonksiyonunun  $\theta$ 'ya göre 2. dereceden kısmî türevi alındığında ise varyans denklemi elde edilmektedir [28, 31].

$$\frac{\partial^2 l}{\partial \theta^2} + \left(\frac{\partial l}{\partial \theta}\right)^2 = -\frac{b''(\theta)}{\phi} + \left(\frac{Y - b'(\theta)}{\phi}\right)^2$$

$$\frac{Y - b'(\theta)}{\phi} = \frac{Y - E(Y)}{\phi}$$

$$-\frac{b''(\theta)}{\phi} + \frac{\text{var}(Y)}{\phi^2} = 0$$

$$\text{Var}(Y) = b''(\theta)\phi \quad (2.8)$$

$\boldsymbol{\beta} = (\beta_1 \ \beta_2 \ \dots \ \beta_p)^T$ :  $\beta$  regresyon katsayıları vektörü olmak üzere,

$\theta_1, \dots, \theta_n$  parametreleri  $X$ 'in değerlerine bağlı iken  $b()$  ve  $g()$  fonksiyonları ile regresyon parametrelerine  $(\beta_1, \dots, \beta_p)$  de bağlıdır.

$$\sum_{j=1}^p \beta_j x_{ij} = \eta_i = g(\mu_i) = g(b'(\theta_i)) \quad (2.9)$$

$\beta$ 'nın en çok olabilirlik yöntemi ile tahmini skor fonksiyonlarının (score functions) çözümü ile bulunmaktadır.  $n$  tane gözlem için olabilirlik fonksiyonu  $\beta$  parametresinin bir fonksiyonu olarak yazıldığında,

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n l_i(\boldsymbol{\beta})$$

$\beta$  parametresinin en çok olabilirlik yöntemi ile tahminleri,

$$\frac{\partial l_i(\boldsymbol{\beta})}{\partial \beta_j} = 0, \quad j = 1, 2, \dots, p \quad (2.10)$$

eşitliği ile elde edilebilir. Zincir kuralı ile

$$\frac{\partial l_i(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \quad (2.11)$$

eşitliği yazılabilir. Kısmi türevler yerine yerleştirildiğinde

$$\frac{\partial l_i}{\partial \beta_j} = \frac{(Y_i - \mu_i) x_{ij}}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}, j = 1, 2, \dots, p \quad (2.12)$$

olabilirlik denklemi sıfıra eşitlenerek parametre tahminleri elde edilmektedir. Bu eşitliğin çözümü için iteratif yöntemlerden yararlanılmaktadır. Newton Raphson yöntemi ve Fisher Scoring algoritması sıklıkla kullanılan iteratif çözüm yöntemleridir [32,33].

## 2.4 Hasar Sayılarının Modellenmesi

Hasar sayılarının modellenmesi için genellikle Poisson, Negatif Binom veya genelleştirilmiş Poisson dağılımları kullanılmaktadır. Poisson dağılımının ortalaması ve varyansı birbirine eşit olduğundan veride aşırı yayılım görüldüğünde yetersiz kalmaktadır. Bu durumda Negatif binom modeli ya da yarı-Poisson (quasi-Poisson) modeli aşırı yayılımı modellemek için kullanılabilir [34].

Hasar sayısı verisinde sıfır değerinde yığılma görüldüğünde sıfır yığılmalı modeller (zero-inflated models) de kullanılmaktadır. Sıfır yığılmalı Poisson, sıfır yığılmalı Negatif Binom dağılımları alternatif sıfır yığılmalı dağılımlardır [35,36].

Çalışmanın uygulama bölümünde hasar sayısı bileşeni için Poisson, Negatif Binom ve yarı-Poisson modelleri çalışılmış, model tahmin sonuçları beşinci bölümde sunulmuştur.

### 2.4.1 Poisson Genelleştirilmiş Doğrusal Modeli

$(0, \infty)$  aralığında tanımlanan  $\lambda$  parametrelili Poisson dağılımının olasılık fonksiyonu eşitlik 2.13'te verilmiştir.

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!}, k = 0, 1, \dots \quad (2.13)$$

Poisson regresyon modeli oluşturulurken  $\mu$ , açıklayıcı değişkenlerin farklı gözlem değerleri için değişecektir. Dağılımın ortalaması  $\mu$  ile gözlem değerleri arasındaki ilişki,

$$\mu_i = e^{x_i\beta} \quad (2.14)$$

ile gösterilir. Burada  $x_i$ ,  $\mathbf{X}$  matrisindeki  $i$ . satırı ifade etmektedir. Bu eşitlikte logaritmik bağ fonksiyonu seçildiği için  $\mu_i$  pozitif olacaktır [37].

Eşitlik 2.5'de Poisson regresyonu için olabilirlik fonksiyonu verilmiştir.

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \left( e^{x_i\boldsymbol{\beta}} + y_i(-x_i\boldsymbol{\beta}) - \ln y_i! \right) \quad (2.15)$$

$\beta$  parametreleri olabilirlik fonksiyonunun kısmi türevi alınarak bulunmaktadır.

$$\frac{\partial}{\partial \beta} L(\boldsymbol{\beta}) \Big|_{\beta=\hat{\beta}} = \sum_{i=1}^n -x_i e^{x_i\hat{\beta}} + y_i x_i = 0 \quad (2.16)$$

Parametre türevlerini elde etmek için tanımlanan fonksiyonun kapalı formu bulunmadığı için Newton-Raphson yöntemi gibi iteratif yöntemler kullanılmaktadır [38].

#### 2.4.2 Negatif Binom Genelleştirilmiş Doğrusal Modeli

Negatif Binom dağılımı istatistik ve aktüerya literatüründe sıklıkla kullanılan bir dağılım türüdür. Hilbe, Negatif Binom dağılımının teorisi ve uygulama örneklerini ayrıntılı olarak çalışmıştır [39].

Negatif Binom dağılımı Poisson- Gamma karma dağılımından elde edilmektedir.  $\theta$  değişkeni Gamma dağılmaktadır:

$$f(\theta) = \frac{(1/\alpha)^{1/\alpha}}{\Gamma(1/\alpha)} \theta^{1/\alpha-1} \exp(-\theta/\alpha) \quad (2.17)$$

$Y_i$  rastlantı değişkeni,  $\lambda_i$  ve  $\alpha$  parametreleri ile negatif binom dağılımına sahiptir ve olasılık fonksiyonu,

$$f(y_i, \lambda_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left( \frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i} \quad (2.18)$$

ile tanımlanır.  $\alpha$  dağılım faktörü ve sabit olmak üzere  $\alpha$  sıfır değerine yaklaştıkça Negatif Binom dağılımı Poisson dağılımına yakınsamaktadır [14, 40].

Negatif Binom dağılımı için ortalama ve varyans değeri;  $E[y_i|\lambda_i, \alpha] = \lambda_i$ ,  $V[y_i|\lambda_i, \alpha] = \lambda_i(1 + \alpha\lambda_i)$  eşitlikleriyle tanımlanmaktadır [41].

Negatif Binom regresyon modeli için  $x_i$ ,  $\mathbf{X}$  matrisindeki  $i$ . satırı ifade etmek üzere koşullu ortalama fonksiyonu ve varyans fonksiyonu eşitlik 2.19 ve eşitlik 2.20'de verilmiştir.

$$E[Y_i|x_i] = \lambda_i = \exp(\beta^t x_i) \quad (2.19)$$

$$\text{Var}[Y_i|x_i] = \exp(\beta^t x_i) (1 + \alpha^2 \exp(\beta^t x_i)) \quad (2.20)$$

### 2.4.3 Sıfır Yiğilmalı Modeller

Sıfır yığilmalı modellerden sıfır yığilmalı Poisson (ZIP), sıfır yığilmalı negatif binom (ZINB) ve sıfır yığilmalı geliştirilmiş Poisson (ZIGP) modelleri, birçok alanda yaygın olarak kullanılmaktadır.

Olasılık yoğunluk fonksiyonu iki parçaya ayırarak modellenmektedir. İlk kısım verideki sıfır değeri alma olasılığını modellerken diğer kısım sıfır olmayan değerlerin ortaya çıkma olasılığını modellemektedir.

Sıfır yığilmalı Poisson regresyon modelinde, veride mevcut olan aşırı yayılımın yalnızca sıfır değerindeki yoğunluktan kaynaklandığı varsayımı yapılmaktadır. Sıfır yığilmalı negatif binom ve sıfır yığilmalı geliştirilmiş Poisson regresyon modellerinde ise aşırı yayılımın sıfır değerindeki yığılmanın yanı sıra birçok nedenden kaynaklanabileceği varsayımı mevcuttur. Dolayısıyla tüm sıfır yığilmalı regresyon modellerinin ortalaması birbirine eşit iken varyans yapıları değişmektedir [28, 37].

## 2.5 Hasar Şiddetinin Modellenmesi

Normal dağılım göstermeyen uzun kuyruklu hasar şiddeti verisi için uzun kuyruklu regresyon modelleri incelenmiştir. Gamma ve Lognormal dağılımları uzun kuyruklu çarpık hasar tutarı verisi için aday olarak seçilebilmektedir.

Dört parametrelili Genelleştirilmiş Beta (GB2) dağılımı kuyruk kısımlarında daha iyi uyum gösterdiği için alternatif bir model regresyon modeli olarak önerilmektedir [42]. Genelleştirilmiş Beta dağılımının regresyon temelli olarak çalışılması literatürde çok fazla yer almamaktadır. Gan ve Valdez (2018), hayat annüite ürünlerinde büyük portfolyolardaki riskin değerlendirmesinde GB2 dağılımı kullanılmıştır [43]. Frees ve Valdez (2008), çarpık hasar tutarı verisini modellemek için GB2 dağılımını önermiştir [44].

Uygulamada hasar tutarlarına muafiyet (deductable) uygulaması yapılarak kesildiği ya da üst limitler kullanılarak sansürlendiği durumlar olabilmektedir. Bu durumlarda hasar modellemesi yaşam modellerindeki kesilmiş ve sansürlenmiş veri yapısına benzerlik gösterebilir [45,46].

Hasar modellerinde bir diğer yaklaşım hasar şiddeti ve hasar frekansı birlikte modellenmek istendiğinde kesikli ve sürekli raslantı değişkenlerinin bir araya gelmesiyle karma bir dağılım oluşturulmasıdır. Karma bir dağılım olan Tweedie dağılımı, Gamma raslantı değişkenlerinin Poisson toplamı olarak tanımlanmaktadır. Hasar olmaması durumunda toplam hasar miktarı sıfır noktasında yığılma gösterirken bir veya daha fazla hasar oluşması durumunda toplam hasar miktarı sürekli kısımda pozitif değerlerle tanımlanmaktadır. Bu toplamda, hasar sayısını ifade eden  $N$  raslantı değişkeni  $\lambda$  ortalaması ile Poisson dağılımına sahiptir [19,47].

Bu çalışmada, hasar tutarı verisi için Gamma ve Genelleştirilmiş Beta 2 regresyonu modelleri çalışılmıştır ve model tahmin sonuçları beşinci bölümde sunulmuştur.

### 2.5.1 Gamma Regresyon Modeli

Gamma dağılımı çok tepeli bir dağılımdır. Gamma dağılımı olasılık fonksiyonu,  $Y \sim \text{Gamma}(v, \lambda)$ ,

$$f_Y(y) = \frac{\lambda}{\Gamma(v)} (\lambda y)^{v-1} e^{-\lambda y} \quad y \geq 0 \quad (2.21)$$

eşitliği ile ifade edilebilmektedir.

Gamma dağılımı beklenen değer ve varyansı,

$$\mu_i = \frac{v}{\lambda_i} \quad \sigma^2 = \frac{1}{v},$$

olmak üzere log olabilirlik fonksiyonu,

$$l(\mu, v, y) = (v - 1) \log(y) - \frac{v}{\mu} y + v \log v - v \log \mu - \log \Gamma(v) \quad (2.22)$$

olarak yazılmaktadır [22].

$x_i$ ,  $\mathbf{X}$  matrisindeki  $i$ . satırı ifade etmek üzere logaritmik bağ fonksiyonu ile oluşturulan Gamma regresyon modeli için olabilirlik fonksiyonu eşitlik 2.23'deki gibi tanımlanmaktadır ( $\mu_i = \exp \{ \mathbf{x}_i \boldsymbol{\beta} \}$ ).

$$l(y; \mu, v) = \sum_{i=1}^n \left\{ \frac{y_i / \exp(x_i \boldsymbol{\beta}) + x_i \boldsymbol{\beta}}{-v} + \frac{v+1}{v} \ln y_i - \frac{\ln v}{v} - \ln \Gamma\left(\frac{1}{v}\right) \right\} \quad (2.23)$$

## 2.5.2 Genelleştirilmiş Beta 2 Regresyon Modeli

GB2 dağılımı, üç tane şekil (shape) ve bir ölçek (scale) parametresi ile dört parametrelilik bir dağılımdır. Şekil parametreleri yardımıyla veriye özellikle kuyruk kısımlarında uyum sağlama kapasitesi fazladır.

Eşitlik 2.24'de GB2 dağılımı olasılık yoğunluk fonksiyonu verilmiştir [48, 49].

$$f(y; a, b, p, q) = \frac{a \left(\frac{y}{b}\right)^{ap-1}}{b B(p, q) \left[1 + \left(\frac{y}{b}\right)^a\right]^{p+q}}, y > 0 \quad (2.24)$$

$B(p, q)$ , Beta fonksiyonudur.  $a, p, q$  şekil parametreleri iken  $b$  ölçek parametresidir. Dağılımın tepe noktası  $a$  parametresi ile kontrol edilirken, kuyruk kısımları  $p$  ve  $q$  parametreleri ile kontrol edilmektedir.

$Y_i | \mathbf{X}_i = \mathbf{x}_i \sim GB2(a, \exp(\mathbf{x}_i \boldsymbol{\beta}), p, q)$  olmak üzere GB2 regresyon modeli için olasılık yoğunluk fonksiyonu eşitlik 2.25'deki şekilde elde edilir.

$$f[y_i; a, \exp(\mathbf{x}_i^t \boldsymbol{\beta}), p, q] = \frac{a \left[ \frac{y_i}{\exp(\mathbf{x}_i^t \boldsymbol{\beta})} \right]^{ap-1}}{\exp(\mathbf{x}_i^t \boldsymbol{\beta}) B(p, q) \left\{ 1 + \left[ \frac{y_i}{\exp(\mathbf{x}_i^t \boldsymbol{\beta})} \right]^a \right\}^{p+q}} \quad (2.25)$$

GB2 regresyon modeli için ortalama fonksiyonu

$$E(Y_i | \mathbf{X}_i = \mathbf{x}_i) = \exp(\mathbf{x}_i' \boldsymbol{\beta}) B(p + 1/a, q - 1/a) / B(p, q)$$

eşitiliği ile yazılacaktır [48].

Bu çalışmada, GB2 dağılımı ortalaması için sonlu bir sayı bulunamadığı durumda olasılık yoğunluk fonksiyonu bir limit değeri (upper limit) ile yakınsanarak ortalama değeri hesaplanmıştır. Shi tarafından [50] kitabında, uzun kuyruklu regresyon modellerinden GB2 regresyon modeli için parametre tahmini ayrıntılı olarak incelenmiştir.





### 3. KOPULA TEORİSİ

Rastgele deęişkenler arasındaki baęımlılık yapısının modellenmesi önemli bir araştırma problemidir. Baęımlılık yapısı parametre tahminlerinin istatistiksel güvenilirliğini, istatistiksel modelin anlamlılıęını etkileyebilmektedir.

İstatistiksel çalışmalarda rastgele deęişkenler ya da gözlemler arasında baęımlılık yapısının modellenmesinde kullanılan yöntemlerden biri kopula fonksiyonlarıdır. Kopula fonksiyonları rastlantı deęişkenleri arasındaki baęımlılık yapısını modellemek için kullanılan matematiksel fonksiyonlardır. Kopula yaklaşımının temeli, rastgele deęişkenlerin birleşik dağılım fonksiyonlarının marjinal dağılımların bir fonksiyonu olarak ifade edilmesine dayanmaktadır.

Kopula fonksiyonlarını aktüerya literatüründe kullanımı son yıllarda yaygınlaşmaya başlamıştır. Aktüeryal riskler arasındaki baęımlılık yapısını teorik olarak inceleyen Kaas (2005) literatüre önemli katkı sağlamıştır [51].

Kopula fonksiyonları yardımıyla rastlantı deęişkenleri arasındaki baęımlılıęın modellenmesi genel olarak iki aşamadan oluşmaktadır.  $i = 1, \dots, n$  ve  $y_1, \dots, y_d$ ; rastlantı deęişkenleri olmak üzere  $y_i = (y_{i1}, \dots, y_{id})$  için kopula fonksiyonları ile modelleme adımları;

(i)  $y_1, \dots, y_d$  her bir rastlantı deęişkeni için tek deęişkenli modelin belirlenmesi

(ii) baęımlı  $d$  deęişken için kopula fonksiyonunun belirlenmesi olarak yazılabilir [6].

İlk adımdaki tek deęişkenli modelin seçiminden sonra, aday kopula fonksiyonları belirlenerek baęımlılık yapısının modellenmesi için deęerlendirilmelidir.

Parametrik kopula ailesinin seçiminde aşağıdaki özelliklerin deęerlendirilmesi önemlidir:

(a) farklı baęımlılık yapılarını modelleyebilmesi (örneğin; koşullu baęımlılık yapısı, pozitif veya negatif baęımlılık yapısı, dönüştürülebilir (exchangeable) yapı, kuyruk baęımlılıęının yapısı (tail behavior)),

(b) matematiksel esnekliđi.

Bađımlılık yapısının modellenmesinde parametrik kopula fonksiyonlarının yanı sıra parametrik olmayan üretilmiş kopula fonksiyonlarından da bahsedilebilmektedir. Bu çalışma kapsamında parametrik yöntemler ile çalışılmıştır. Kopula fonksiyonları için parametrik olmayan tahmin yöntemleri Yang ve ark., Chen ve Huang, Rosenberg tarafından çalışılmıştır [52–54].

### 3.1 Bađımlılık Kavramı

$X$  ve  $Y$  rastlantı deđişkenleri için,  $Y$ 'nin büyük deđerleri  $X$ 'in büyük deđerleri ve  $Y$ 'nin küçük deđerleri  $X$ 'in küçük deđerleri ile ilişkili olma eğiliminde ise pozitif bađımlılıđa (positive dependency) sahip olduđu söylenmektedir. Deđerşkenlerden birinin büyük deđerleri için diđerinin küçük deđerleri ile ilişkili olma eğiliminde ise negatif bađımlılık (negative dependency) söz konusudur [55].

Rastlantı deđerşkenlerinden birinin hakkındaki bilginin, diđerini hakkında herhangi bir anlam içermediđi rastlantı deđerşkenleri bađımsızdır.

$X$  ve  $Y$  rastlantı deđerşkenleri için;

$$\Pr(X \leq x \text{ ve } Y \leq y) = \Pr(X \leq x)\Pr(Y \leq y)$$

veya

$$\Pr(X \leq X|Y) = \Pr(X \leq X)$$

ise  $X$  ve  $Y$  rastlantı deđerşkenleri bađımsızdır. Eđer rastlantı deđerşkenleri bađımsız deđerilse, bađımlı olacaktır [56].

#### 3.1.1 Rank Korelasyon Katsayısı

$X$  ve  $Y$  rastlantı deđerşkenlerinin aynı anda küçük ya da büyük deđer almaları konkordans (concordance), biri küçük (büyük) deđer alırken diđerinin büyük (küçük) deđer alma olasılıđı ise diskordans (discordance) olarak adlandırılmaktadır. Rank korelasyon katsayısı, rastlantı deđerşkenleri arasındaki konkordansı ölçmek için

kullanılmaktadır. En çok kullanılan rank korelasyon ölçütleri olan Kendall'in tau ve Spearman'ın rho korelasyon katsayılarıdır.

### 3.1.1.1 Spearman'ın rho korelasyon katsayısı

Spearman'ın rho korelasyon katsayısı doğrusal korelasyon katsayısı ile ilişkilidir. Marjinal dağılımları ile X ve Y rastlantı değişkenleri için Spearman'ın rho korelasyon katsayısı,

$$\rho_s = \rho (F_x(X), F_y(Y)) = \frac{\text{Cov}(F_x(X), F_y(Y))}{\sqrt{\text{Var}(F_x(X)) \text{Var}(F_y(Y))}}$$

eşitliği ile yazılabilir.

Sürekli rastlantı değişkenleri için Spearman'ın rho katsayısı,

$$\rho_s(X_1, X_2) = 3 \left[ \Pr \left( (X_1 - X_1^\perp) (X_2 - X_2^\perp) > 0 \right) \right] - \left[ \Pr \left( (X_1 - X_1^\perp) (X_2 - X_2^\perp) < 0 \right) \right]$$

olarak yazılabilir. Burada  $(X_1^\perp, X_2^\perp)$  bağımsız bileşenlerden oluşan  $(X_1, X_2)$  ile bağımsız olan rastgele vektördür [38].

### 3.1.1.2 Kendall'in tau korelasyon katsayısı

Bir  $(X, Y)$  sürekli rastlantı vektöründeki  $n$  gözlemden bir rastgele örneklem alındığında  $\binom{n}{2}$  tane bağımsız çift olacaktır.  $(X_i, Y_i) \stackrel{ve}{=} (X_j, Y_j)$  konkordans ya da diskordansdır. Burada Kendall'in korelasyon katsayısı,

$$\tau = \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

formülü ile bulunur. Eşitliğin sol tarafı konkordans olasılığını verirken, sağ tarafı diskordans olasılığını vermektedir. Sürekli rastgele değişkenler için Kendall'in tau katsayısı,

$$\tau = 2 \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1$$

olacaktır. Eşitlikten Kendall'in tau değerinin -1 ile +1 arasında değiştiği görülmektedir. Eğer  $(X, Y)$  rastlantı değişkenleri countermonotonic ise  $\tau$  değeri alt sınıra ulaşacak

ve -1 değerini alacak,  $(X, Y)$  rastlantı değişkenleri comotonic ise  $\tau$  değeri üst sınıra ulaşacak ve +1 değerini alacaktır [56].

Rastlantı değişkenleri arasındaki bağımlılığı ölçen Kendall'ın tau ve Spearman'ın rho katsayıları mükemmel pozitif bağımlılık durumunda +1, mükemmel negatif bağımlılık durumunda -1, değerini almaktadır [11].

Rank korelasyon ölçütleri olan Kendall'ın tau ve Spearman'ın rho katsayıları kopula fonksiyonları yardımıyla hesaplanabilmektedir.

### 3.2 Kopula Fonksiyonları

Kopulalar, olasılıksal metrik uzaylar kapsamında 1959 yılında ele alınmıştır [7]. Kopula fonksiyonu, marjinalleri  $(0, 1)$  aralığında tekdüze (uniform) dağılan çok değişkenli dağılım fonksiyonudur. Marjinal dağılımlardan bağımsız olarak çok değişkenli dağılım yapısını modellemek için kullanılmaktadır.

#### 3.2.1 Olasılık İntegral Dönüşümü

Olasılık integral dönüşümü (probability integral transform) kopula fonksiyonlarının yapılandırılmasında önemli role sahiptir. Dönüşüm rastgele değişkenler üretmede ve değişkenlerin tekdüze dönüşümlerini  $(U(0, 1))$  elde etmek için kullanılmaktadır.

Teorem 3.2.1:

$X \sim F_X$  sürekli rastlantı değişkeni olmak üzere,  $F_X(X) \sim U(0, 1)$  ise  $\bar{F}_X(X) = 1 - F_X(X) \sim U(0, 1)$  olacaktır.

Kanıt:  $0 < p < 1$  için  $F_X^{-1}(p) = \inf \{x : F_X(x) \geq p\}$ .  $F_X^{-1}$ ,  $F_X$ 'in soldan sürekli ters (inverse) fonksiyonu olmak üzere,

$$0 < u < 1, P(F_X(X) \leq u) = P(X \leq F_X^{-1}(u)) = F_X \circ F_X^{-1}(u) = u$$

olacaktır [6].

Kopula fonksiyonlarının yapılandırılmasında olasılık integral dönüşümünün özelliklerinden yararlanılmaktadır.

(i)  $U \sim U(0, 1)$  ve  $F$  tek değişkenli kümülatif dağılım fonksiyonu,  $F^{-1}$  genelleştirilmiş ters veya kantil (quantile) fonksiyonu olmak üzere,

$$X = F^{-1}(U) \sim F$$

olacaktır.

(ii)  $X \sim F$  sürekli rastlantı değişkeni için kümülatif dağılım fonksiyonu  $F(X) \sim U(0,1)$  aralığında tanımlanır.

(iii)  $j = 1,2$  için  $F_j$  sürekli dağılım fonksiyonu  $X_1 \sim F_1$  ise  $F_2^{-1}(F_1(X_1)) = F_2^{-1} \circ F_1(X_1) \sim F_2$  olacaktır.

(iv)  $j = 1,2$  için  $F_j$  sürekli dağılım fonksiyonu  $X_1 \sim F_1$  ise  $F_2^{-1}(1 - F_1(X_1)) \sim F_2$  eşitliğiyle de yazılabilir [6].

### 3.2.2 Sklar Teoremi

Sklar (1959) teoremi herhangi bir çok değişkenli dağılım fonksiyonunun marjinal dağılım fonksiyonları yardımıyla yazılabilmesini sağlamaktadır.

$U_1, \dots, U_d$  (0,1) aralığında uniform rastgele değişkenler olmak üzere,  $d$  değişkenli dağılım fonksiyonu bir kopula fonksiyonudur.

$$C(u_1, \dots, u_d) = \Pr(U_1 \leq u_1, \dots, U_d \leq u_d) \quad (3.1)$$

Kopula fonksiyonları uygulanmak istenen veride çoğunlukla tekdüze dağılım gözlemlenmemektedir. Bu durumda rastlantısal (arbitrary) marjinal dağılımlar kullanılmaktadır.

$F_1(y_1), \dots, F_p(y_d)$  rastlantısal marjinal dağılım fonksiyonları olmak üzere, çok değişkenli dağılım fonksiyonu kopula fonksiyonu yardımıyla eşitlik 3.2'de tanımlanmıştır.

$$F(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_p(y_d)) \quad (3.2)$$

Sklar Teoremi kopula teorisinde önemlidir ve çoğu uygulama için temel oluşturmaktadır. Teorem, çok değişkenli dağılım fonksiyonları ile tek değişkenli marjinal dağılımları arasındaki bağıntıda kopulaların rolünü açıklamaktadır [2].

Ayrıca eşitlik 3.2,  $F_1, \dots, F_d$  marjinal dağılım fonksiyonlarının  $C$  kopula fonksiyonundan bağımsız olarak tahmin edilmesine olanak sağlamaktadır.

Kopula fonksiyonları genellikle kümülatif dağılım fonksiyonları ile ifade edilirler. Olasılık yoğunluk fonksiyonları, kopula dağılım fonksiyonun kısmi türevleri hesaplanarak elde edilmektedir. Eğer  $C(u)$  sürekli bir kopula dağılım fonksiyonu ise olasılık yoğunluk fonksiyonu eşitlik 3.3'de verilmiştir.

$$c(\mathbf{u}) = c(u_1, \dots, u_d) = \frac{\partial^d C(\mathbf{u})}{\partial u_1 \dots \partial u_d}, \quad \mathbf{u} \in (0, 1)^d \quad (3.3)$$

$F_1, \dots, F_d$  sürekli marjinal dağılım fonksiyonları ve marjinal yoğunluk fonksiyonları  $f_j = F_j'$  ise  $F(\mathbf{y}) = C(F_1(y_1), \dots, F_d(y_d))$  kopula fonksiyonu ile tanımlanan  $F'$  nin olasılık yoğunluk fonksiyonu:

$$f(\mathbf{y}) = c(F_1(y_1), \dots, F_d(y_d)) \times \prod_{j=1}^d f_j(y_j), \quad \mathbf{y} \in \mathbb{R}^d \quad (3.4)$$

eşitliği ile tanımlanmaktadır.

Skalar teoremine göre marjinal dağılımlar süreklirse tek bir kopula fonksiyonu tanımlanmaktadır. Rastlantı değişkenlerinin kesikli olması durumunda Sklar teoremini kullanabilmek için bazı dönüşümlere ihtiyaç duyulmaktadır. Örneğin Rüschendorf, kesikli veride Sklar teoremini kullanabilmek için rastlantı değişkenlerinin dağılımsal dönüşümlerini kullanmıştır [57]. Nikoloulopoulos ise kopula fonksiyonlarının kesikli yanıt değişkenleri ile modellenmesi için kullanılan yöntemlerin performans sonuçlarını karşılaştırmıştır [58, 59].

Kopulalar bağımlılık türüne farklı karakteristikleri gerektirmektedir ve uygun kopulanın seçimi önemlidir. İdeal olan, tüm olası kopulalar arasından en iyi kopulanın kullanılmasıdır. Kopulanın seçim süreci kopulaların sınırlı bir sayısı ile kısıtlanmalıdır [12].

### **Bağımsız Kopula:**

En temel kopula fonksiyonu bağımsız kopula (independent copula) fonksiyonudur.

$$\Pi(u) = \prod_{j=1}^d u_j, \quad u \in [0, 1]^d \quad (3.5)$$

$\mathbb{P}(\mathbf{U} \leq \mathbf{u}) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d)$  olmak üzere bağımsız kopula dağılım fonksiyonlarının çarpımı olacaktır [60] :

$$\prod_{j=1}^d \mathbb{P}(U_j \leq u_j) = \prod_{j=1}^d u_j = \Pi(u). \quad (3.6)$$

### 3.2.3 Fréchet–Hoeffding Sınırları

Kopula fonksiyonları için bağımlılığın alt ve üst sınırları Fréchet–Hoeffding sınırları (Fréchet–Hoeffding bounds) ile belirtilmektedir.

Herhangi bir kopula fonksiyonu için  $u \in [0, 1]^d$  olmak üzere alt sınır  $W(u)$  ve üst sınır  $M(u)$ , eşitlik 3.7’de ve eşitlik 3.8’de tanımlanmıştır [60].

$$W(u) = \max \left\{ \sum_{j=1}^d u_j - d + 1, 0 \right\} \quad (3.7)$$

$$M(u) = \min_{1 \leq j \leq d} \{u_j\} \quad (3.8)$$

Alt sınır  $W$  sadece  $d = 2$  için bir kopula fonksiyonu iken üst sınır  $M$ ,  $d \geq 2$  koşulunda bir kopula fonksiyonunu ifade etmektedir.

İki boyutlu kopula fonksiyonları için Fréchet–Hoeffding üst sınırı mükemmel pozitif bağımlılığı ifade ederken, Fréchet–Hoeffding alt sınırı mükemmel negatif bağımlılığı ifade etmektedir.

- (i) İki boyutlu bir kopula fonksiyonu için Fréchet–Hoeffding üst sınır kopulası  $CU$  ile gösterildiğinde

$$CU(u_1, u_2) = \min(u_1, u_2), \quad (u_1, u_2) \in [0, 1] \times [0, 1] \quad (3.9)$$

eşitliği elde edilecektir.

- (ii) İki boyutlu bir kopula fonksiyonu için Fréchet–Hoeffding alt sınır kopulası  $CL$  ile gösterildiğinde ise

$$CL(u_1, u_2) = \max(0, u_1 + u_2 - 1), \quad (u_1, u_2) \in [0, 1] \times [0, 1] \quad (3.10)$$

eşitliği elde edilecektir.



Tüm iki boyutlu C kopulaları için,

$$CL(u_1, u_2) \leq C(u_1, u_2) \leq CU(u_1, u_2) \quad (3.11)$$

eşitsizliği yazılabilir [51]. Böylece negatif ya da pozitif bağımlılığın sınırları belirlenmektedir.

Kendall'ın tau değeri ne kadar büyükse, rastlantı değişkenleri arasındaki ilişki de o kadar güçlü olacaktır. En büyük değer +1'dir ve Fréchet üst sınırına karşılık gelmektedir. Çoğu Arşimedyen kopulalar için kopula parametresinin değeri arttıkça, Kendall'ın tau değeri de artmaktadır [61].

### 3.2.4 Kopula Fonksiyon Aileleri

Eliptik ve Arşimedyen kopulalar, kopula fonksiyon ailelerinin iki temel sınıfıdır. Farklı üreteç fonksiyonları (generator function) yardımıyla yapılandırılmaktadırlar.

Genellikle asimetrik bağımlılık yapısının görüldüğü finansal verilerde eliptik kopula aileleri kullanırken, simetrik bağımlılığın görüldüğü veri setlerinde arşimedyen kopula aileleri kullanılmıştır.

#### 3.2.4.1 Eliptik Kopula Fonksiyonları

Eliptik kopula fonksiyonları eliptik dağılımlardan yapılandırılmaktadır. Eliptik kopula fonksiyonlarının özellikleri kendisine ait olan eliptik dağılım ailelerinin özelliklerinden elde edilmektedir [60]. Normal kopula ve t kopula fonksiyonları eliptik kopula ailesinin başlıca üyelerindedir.

#### Normal (Gaussian) Kopula

Normal dağılım eliptik dağılımların özel bir türüdür.

$\Sigma$  diagonal elemanları 1 olan korelasyon matrisi olmak üzere, çok değişkenli Normal dağılımın olasılık yoğunluk fonksiyonu eşitlik 3.12'de verilmiştir.

$$\phi_N(z) = \frac{1}{(2\pi)^{p/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2} z' \Sigma^{-1} z\right) \quad (3.12)$$

$\Phi$  standart Normal dağılım fonksiyonu ve  $\phi$  Normal dağılım olasılık yoğunluk fonksiyonu olmak üzere, Normal kopula olasılık yoğunluk fonksiyonu 3.13'deki gibi yazılmaktadır.

$$c_N(u_1, \dots, u_p) = \phi_N(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)) \prod_{j=1}^p \frac{1}{\phi(\Phi^{-1}(u_j))} \quad (3.13)$$

$\theta$  bağımlılık parametresi olmak üzere, İki değişkenli Normal kopula fonksiyonu eşitlik 3.14'da verilmiştir.

$$\begin{aligned} C(u_1, u_2) &= \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\theta^2}} * \left( \frac{-(s^2 - 2\theta st + t^2)}{2(1-\theta^2)} \right) ds dt \end{aligned} \quad (3.14)$$

$\theta$  parametresi bağımlılık derecesini ve yönünü ifade etmektedir. Normal kopulada bağımlılık parametresi  $[-1, +1]$  aralığında sınırlandırılmıştır.  $\theta = 0$  için bağımsızlık kopulasına dönüşür.  $\theta = +1$  ve  $\theta = -1$  olduğunda sırasıyla Fréchet–Hoeffding alt sınırına (maksimum negatif bağımlılık) ve Fréchet–Hoeffding üst sınırına (maksimum pozitif bağımlılık) karşılık gelmektedir [56]. Pozitif ve negatif bağımlılığa izin verdiği için esnek bir kopula fonksiyonudur.

### t Kopula

Bağımlılık modellemesinde esnek olan bir diğer eliptik kopula t kopuladır. Farklı serbestlik dereceleri ile farklı yapıdaki bağımlılıkları modelleyebilmektedir.

Çok değişkenli t Kopula fonksiyonu, Student t dağılımından yapılandırılmaktadır. Çok değişkenli t dağılımının üreteç fonksiyonu:

$$g(t) = \frac{\Gamma(\frac{v+d}{2})}{(\pi v)^{\frac{d}{2}} \Gamma(\frac{v}{2})} \left(1 + \frac{t}{v}\right)^{-\frac{v+d}{2}}, \quad t \in [0, \infty). \quad (3.15)$$

$c_{p,v}^t$ ; t kopula olasılık yoğunluk fonksiyonu eşitlik 3.16'da verilmiştir.

$$c_{p,v}^t(u) = \frac{\Gamma(\frac{v+d}{2})}{\Gamma(\frac{v}{2}) \sqrt{\det P}} \left( \frac{\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} \right)^d \frac{\left(1 + \frac{x' P^{-1} x}{v}\right)^{-\frac{v+d}{2}}}{\prod_{j=1}^d \left(1 + \frac{x_j^2}{v}\right)^{-\frac{v+1}{2}}}, \quad u \in (0, 1)^d \quad (3.16)$$

t-kopula fonksiyonları özellikle alt ve üst kuyruk bağımlılığı gözlemlendiği durumda tercih edilmektedir [60]. Son yıllarda özellikle finans alanında hisse senedi, portföy değerlendirmesi gibi uygulamalarda kullanılmaktadır [62–64].

Farklı eliptik dağılımlar için üreteç fonksiyonları aşağıda gösterilmiştir:

	<i>Uretec</i> $g_p(x)$
Normal dağılım	$e^{-x}$
t-dağılımı (r, serbestlik derecesi)	$(1 + 2x/r)^{-(p+r)/2}$
Cauchy dağılımı	$(1 + 2x)^{-(p+1)/2}$
Logistic dağılım	$e^{-x}/(1 + e^{-x})^2$
Üstel dağılım	$\exp(-rx^s)$

### 3.2.4.2 Arşimedyan Kopula Fonksiyonları

$\psi$  üreteç fonksiyonu olmak üzere, Arşimedyan kopula fonksiyonunun genel formu eşitlik 3.17’de verilmiştir.

$$C(u) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)), \quad u \in [0, 1]^d \quad (3.17)$$

$\psi$ , sürekli, azalan ve konveks bir fonksiyondur. Arşimedyan kopulalar dönüştürülebilir (exchangeable) özelliği olan kopula fonksiyonlarıdır ( $\psi(t) = \exp(-t)$ ) [60].

Arşimedyan kopula ailesinin en sık kullanılan üyeleri Frank, Clayton (Cook-Johnson) ve Gumbel-Hougaard kopulalarıdır.

#### Frank Kopula

Frank kopula, ilk kez Frank tarafından 1979 yılında bir fonksiyonel eşitlik problemine çözüm olarak bulunmuştur [65].

İki değişkenli Frank kopula fonksiyonu,  $\psi(t)$  üreteç fonksiyonu

$$\psi(t) = -\ln\left(\frac{\exp(-\alpha t) - 1}{\exp(-\alpha) - 1}\right)$$

olmak üzere eşitlik 3.18’da verilmiştir [38].

$$C_{\alpha}(u_1, u_2) = -\frac{1}{\alpha} \ln \left( 1 + \frac{(\exp(-\alpha u_1) - 1) (\exp(-\alpha u_2) - 1)}{\exp(-\alpha) - 1} \right), \quad \alpha \neq 0 \quad (3.18)$$

$\alpha$  bağımlılık parametresini ifade etmektedir.

Frank kopulasında ise parametre değeri arttıkça, Kendall'ın tau değeri azalmaktadır. Bu yüzden Frank kopulasında bağımlılığın parametre değeriyle ters orantılı olduğu söylenebilir [61].

### Clayton Kopula

İlk olarak 1975 yılında Kimeldorf ve Sampson tarafından bulunmasına karşın, Clayton tarafından da çalışılmış ve Clayton kopula ismiyle anılmaktadır [66].

İki değişkenli Clayton kopula fonksiyonu,  $\psi(t)$  üreteç fonksiyonu

$$\psi(t) = \left( \frac{t^{-\alpha} - 1}{\alpha} \right) |$$

olmak üzere eşitlik 3.19'da verilmiştir.

$$C_{\alpha}(u_1, u_2) = \left( (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}} \right), \quad \alpha > 0 \quad (3.19)$$

### Gumbel - Hougaard Kopula

İki değişkenli Gumbel - Hougaard kopula fonksiyonu,  $\psi(t)$  üreteç fonksiyonu

$$\psi(t) = (-\ln t)^{-\alpha}$$

olmak üzere eşitlik 3.20'da verilmiştir.

$$C_{\alpha}(u_1, u_2) = \exp \left\{ - \left[ (-\log u_1)^{\alpha} + (-\log u_2)^{\alpha} \right]^{\frac{1}{\alpha}} \right\}, \quad \alpha \geq 1 \quad (3.20)$$

Tek parametrelili Arşimedyan kopula fonksiyonlarında  $\tau$  ile  $\alpha$  parametresi arasında bir ilişki gözlemlenmektedir. Başka bir deyişle katsayısı  $\tau$ ,  $\alpha$  parametresi cinsinden yazılabilmektedir. Örneğin, Clayton kopulası için,  $\alpha > 0$  olduğunda  $\tau$  değeri,

$$\tau = -4 \int_0^1 \frac{t - t^{\alpha+1}}{\alpha} dt + 1 = \frac{\alpha}{\alpha + 2}$$

olacaktır [38].

Frees, diğer arşimedyan kopula fonksiyonları için  $\tau$  ile  $\alpha$  arasındaki ilişkiyi tanımlamıştır [7].

### 3.2.5 Kopula Tahmin Yöntemleri

Kopula fonksiyonu  $C$  ve marjinal dağılım fonksiyonları,  $F_1, \dots, F_d$ , parametrik dağılımlara sahip olduğu durumda parametrik tahmin yöntemleri kullanılmaktadır.

Kopula ile bağımlılık modellemesi iki adımdan oluşmaktadır:

- $y_1, \dots, y_d$ ; bağımlı değişkenleri için marjinal modellerin tahmini
- $d$  boyutlu kopula fonksiyonunun ve bağımlılık parametrelerinin tahmini

Marjial modeller, regresyon modelleri olarak belirlendiğinde dağılım parametrelerinin yanında regresyon parametrelerinin tahmin değerleri de elde edilmelidir. Bu bölümde parametrik tahmin yöntemlerinden En Çok Olabilirlik Yöntemi (Maximum Likelihood Estimation) ve Marjinallere İlişkin Çıkarılma Yöntemi (Inference for Margins) ele alınmıştır.

#### 3.2.5.1 En Çok Olabilirlik Yöntemi

En çok olabilirlik yöntemi (MLE) ile kopula modeli için parametre tahmini yapılırken marjinal modellere ait parametreler ile bağımlılık parametreleri eşanlı tahmin edilmektedir [6].

$j \in \{1, \dots, d\}$  için  $c_{\theta}$  kopula yoğunluk fonksiyonu olmak üzere Sklar teoremine göre ortak olasılık yoğunluk fonksiyonu eşitlik 3.21'deki gibi yazılabilmektedir.

$$f(\mathbf{x}) = c_{\theta}(F_{1,y_1}(x_1), \dots, F_{d,y_d}(x_d)) \prod_{j=1}^d f_{j,y_j}(x_j) \quad (3.21)$$

Bu eşitlikte  $j \in \{1, \dots, d\}$  için  $f_{j,y_j}$  marjinal olasılık yoğunluk fonksiyonlarını ifade etmektedir. Eşitlik 3.22'de  $(\gamma_1, \dots, \gamma_d, \theta)$  parametre vektörünün tahmini için log-olabilirlik fonksiyonu verilmiştir [60].

$$l_n(\gamma_1, \dots, \gamma_d, \theta) = \sum_{i=1}^n \log c_{\theta}(F_{1,\gamma_1}(X_{i1}), \dots, F_{d,\gamma_d}(X_{id})) + \sum_{j=1}^d \sum_{i=1}^n \log f_{j,\gamma_j}(X_{ij}) \quad (3.22)$$

Eşitlik 3.22’de tanımlanan log-olabilirlik fonksiyonunda tahmin edilmek istenen parametre vektöründeki  $(\gamma_1, \dots, \gamma_d)$  parametreleri marjinal dağılım parametreleri,  $\theta$  ise kopula parametrelerini ifade etmektedir. Olabilirlik fonksiyonunu maksimum yapan parametre değerleri en çok olabilirlik tahmin edicisi olacaktır. Parametre tahmini için kullanılacak olan çok boyutlu optimizasyon yöntemlerinde parametrelerin yakınsadığı değerleri elde etmek problemin çözümünü zorlaştırmaktadır. Özellikle gözlem sayısının fazla olduğu çok boyutlu parametre vektörlerinin tahmininde Marjinallere İlişkin Çıkarsama Yöntemi alternatif olmaktadır.

### 3.2.5.2 Marjinallere İlişkin Çıkarsama Yöntemi

Marjinallere ilişkin çıkarsama yöntemi (IFM) ile parametre tahmini iki aşamadan oluşmaktadır. İlk olarak her  $j \in \{1, \dots, d\}$  için marjinal modellerin parametre vektörü  $(\gamma_{0,1}, \dots, \gamma_{0,d})$  tahmin edilmelidir.

$$\gamma_{n,j} = \operatorname{argmax}_{\gamma_j \in \Gamma_j} \sum_{i=1}^n \log f_{j,\gamma_j}(X_{ij}) \quad (3.23)$$

İkinci aşamada kopula parametre vektörü  $\theta$  tahmin edilmektedir:

$$\theta_n = \operatorname{argmax}_{\theta \in \Theta} \sum_{i=1}^n \log c_{\theta}(F_{1,\gamma_{n,1}}(X_{i1}), \dots, F_{d,\gamma_{n,d}}(X_{id})) \quad (3.24)$$

IFM’in pratikteki uygulamasında marjinal dağılımlar tahmin edildikten sonra gözlem noktalarına dönüşüm uygulanabilmektedir.  $i \in \{1, \dots, n\}$  için n birimlik bir örnekleme marjinal dağılım fonksiyonları ile dönüşüm uygulandığında eşitlik 3.25’deki sözde-gözlemler (pseudo-observations) elde edilecektir.

$$U_{i,\gamma_n} = (F_{1,\gamma_{n,1}}(X_{i1}), \dots, F_{d,\gamma_{n,d}}(X_{id})) \quad (3.25)$$

$\gamma_n = (\gamma_{n,1}, \dots, \gamma_{n,d})$  olmak üzere eşitlik 3.26’deki gibi kopulaya ait log-olabilirlik fonksiyonu maksimize edilerek kopula parametre vektörü tahmini edilmektedir [60]

$$\theta_n = \operatorname{argmax}_{\theta \in \Theta} \sum_{i=1}^n \log c_{\theta}(U_i, \gamma_n) \quad (3.26)$$

Joe [6] , marjinal dağılımların doğru belirlendiği varsayımı altında IFM tahmin edicisinin ile MLE tahmin edicisi gibi etkin bir tahmin edici olduğunu göstermiştir.

Kopula fonksiyonu ve marjinal dağılımların parametrik olmadığı ve dağılım bilgisinin olmadığı durumda parametrik olmayan tahmin yöntemleri kullanılabilir. Bu yaklaşımda verinin korelasyon ölçüleri ile kopula fonksiyonları arasındaki bağıntı kullanılarak tahmin yapılmaktadır. Spearman rho katsayısı ve Kendall tau katsayısına dayalı yöntemler sık kullanılan parametrik olmayan tahmin yöntemleridir [6, 60].



#### **4. TOPLAM HASAR TUTARI TAHMİNİ İÇİN KOPULA REGRESYON MODELİ**

Hasar modelinin bileşeni olan hasar tutarının tahmini yapılırken sigorta poliçesinin teminat altına aldığı riskler iyi değerlendirilmelidir. Sigorta şirketleri aynı sigorta poliçesi altında farklı riskleri teminat altına alabilmektedir. Toplam hasar tutarını modellerken farklı risklere karşılık gelen hasar ödemeleri birlikte modellenmek istendiğinde çok değişkenli istatistiksel yöntemlerden faydalanılabilir. Bu çalışmada Trafik sigortası hasarları modellenirken farklı hasar ödemeleri için regresyon modelleri kurulmuş, aralarındaki ilişki kopula fonksiyonu ile kurulan istatistiksel model ile ifade edilmiştir.

Kopula fonksiyonları ile regresyon modellemesi iki aşamadan oluşmaktadır. İlk aşama marjinal dağılımların belirlenmesidir. Marjinal dağılımların belirlenmesinde dağılımsal bir kısıt yoktur. Marjinal dağılımlar belirlendikten sonra kopula fonksiyonu ile birleşik modelleme aşamasına geçilir. İlk aşamada elde edilen marjinal modellerin parametreleri de kullanılarak gözlemlenen bağımlılık yapısına en uygun kopula fonksiyonu belirlenmektedir. Birleşik modelleme aşamasında marjinal model parametreleri ve kopula fonksiyonu bağımlılık parametreleri parametrik tahmin yöntemleri ile elde edilmiştir.

Bu bölümde çalışmada kullanılan üç parçalı hasar modelini açıklayabilmek için çalışma verisindeki bağımlı değişkenlerin genel yapısı hakkında ön bilgi verilmiştir.

##### **4.1 Veri Hakkında**

Veri seti; 2012-2016 yılları arasında özel bir sigorta şirketinin portföyünde bulunan Türkiye geneli Karayolları Zorunlu Trafik Sigortası kapsamındaki sigortalı bilgileri, poliçe bilgileri ve hasar tutarı bilgilerinden elde edilmiştir. Çalışma için açıklayıcı değişkenler tarafından modellenmek istenen bağımlı değişkenler;

- Maddi hasar tutarı



- Maddi muallak hasar tutarı
- Bedeni hasar tutarı
- Bedeni muallak hasar tutarı
- Hasar sayısı ve
- Hasar kategorisi deęişkenleridir.

## 4.2 Üç Parçalı Toplam Hasar Modeli

Trafik sigortası poliçe bilgilerinden elde edilen veri setinde hasar tutarı ödemeleri karşılık geldikleri teminat başlığı altında ayrılmış ve modele bağımlı deęişken olarak dahil edilmiştir. Bir hasar gerçekleştiğinde ödemesi yapılan/yapılacak olan hasar tutarı ayrımı da gözetilerek muallak hasar tutarları da toplam hasar modeli içerisinde bağımlı deęişken olarak deęerlendirilmiştir.

(1) Maddi hasar

(2) Maddi muallak hasar

(3) Bedeni hasar

(4) Bedeni muallak hasar

Trafik sigortasında sigortalı riskin ortaya çıkması durumunda yapılan ödemeler için hasar türlerinin ayrılması ve modellenmesi yaygın deęildir.

Oluşan hasar türünün ayrıntılı ve tüm kombinasyonları ile modele dahil edilebilmesi için  $M$  hasar kategorisi rastlantı deęişkeni olarak belirlenmiştir.

Çalışma için önerilen üç parçalı hasar modeli için bağımlı deęişkenler aşağıda tanımlanmıştır:

$N_i$ : Her bir poliçe için bir yıl içerisindeki hasar sayısı

$M_k$ : Hasar kategorisi,  $k=1, \dots, 15$ .

$Y_j$ : Ödenen hasar tutarı,  $j=1, \dots, 4$ .

Yukarıda tanımlanan bağımlı deęişkenlerin birleşik olasılık fonksiyonu:

$$f(N, M, Y) = f(N) \times f(M|N) \times f(Y|N, M) \quad (4.1)$$

$f(N)$ : hasar gerekleŒme olasılıđı iin yođunluk fonksiyonu

$f(M|N)$ : hasar gerekleŒtiđi bilindiđinde  $m$ . kategoriden hasarın gerekleŒmesi olasılıđı iin yođunluk fonksiyonu

$f(Y|N, M)$ :  $N$  ve  $M$  bilindiđinde gerekleŒen koŒullu hasar tutarı olasılık yođunluk fonksiyonu

Tanımlanan u paralı model uzerinde kopula fonksiyonlarına dayalı regresyon modeli uygulanmıŒtır.

Hasar sıklıđı ve hasar tr deđiŒkeninin hasar Œiddeti modeline entegre edildiđi, farklı hasar tutarları arasındaki iliŒkinin de modele dahil edildiđi u paralı model Frees [44] alıŒmasından referans alınarak geliŒtirilmiŒtir. Unc bir bileŒen olarak tanımlanan hasar kategorisi deđiŒkeninin tahmini iin Lojistik regresyon modeli kullanılmıŒtır.

#### **4.2.1 Hasar Kategorisi iin Lojistik Regresyon Modeli**

$(Y_1)$  Maddi hasar tutarı

$(Y_2)$  Maddi muallak hasar tutarı

$(Y_3)$  Bedeni hasar tutarı

$(Y_4)$  Bedeni muallak hasar tutarı

Drt farklı hasar tr iin farklı hasar tutarlarının tm kombinasyonlarının oluŒturulduđu 15 kategori tanımlanmıŒtır.

**Çizelge 4.1** : Gözlemlenen Hasar Dağılımları (Eğitim verisi, 2012–2015)

Hasar Kombinasyonu	Sayı	Oran
(Y <sub>1</sub> )	122,000	0.96830
(Y <sub>2</sub> )	0	0.00000
(Y <sub>3</sub> )	549	0.00436
(Y <sub>4</sub> )	178	0.00141
(Y <sub>1</sub> , Y <sub>2</sub> )	2,154	0.01710
(Y <sub>1</sub> , Y <sub>3</sub> )	332	0.00264
(Y <sub>1</sub> , Y <sub>4</sub> )	101	0.00080
(Y <sub>2</sub> , Y <sub>3</sub> )	0	0.00000
(Y <sub>2</sub> , Y <sub>4</sub> )	3	0.00002
(Y <sub>3</sub> , Y <sub>4</sub> )	395	0.00314
(Y <sub>1</sub> , Y <sub>2</sub> , Y <sub>3</sub> )	8	0.00006
(Y <sub>1</sub> , Y <sub>2</sub> , Y <sub>4</sub> )	7	0.00006
(Y <sub>1</sub> , Y <sub>3</sub> , Y <sub>4</sub> )	250	0.00198
(Y <sub>2</sub> , Y <sub>3</sub> , Y <sub>4</sub> )	4	0.00003
(Y <sub>1</sub> , Y <sub>2</sub> , Y <sub>3</sub> , Y <sub>4</sub> )	13	0.00010
Total	125,994	1.00000

Hasar kategorisi modellenirken çoklu lojistik regresyon modeli (multi logit model) kullanılmıştır. 4 farklı hasar türü için oluşabilecek 15 kategori tanımlanarak her bir hasar kategorisi için ortaya çıkabilecek hasar sayıları modellenmiştir.

*i.* poliçe sahibinin *j.* hasar türü için  $M_{ij}$  hasar kategorisi olarak tanımlanmıştır.

Bir poliçe sahibinin  $N_i$  hasarı olduğu bilindiğinde  $j = 1, \dots, N_i$ , her bir hasar 15 kategoriye (bir referans kategorisi olmak üzere) karşılık gelebilmektedir.

Örnekleme büyüklüğü  $B$ , ve  $i = 1, \dots, B$  olmak üzere lojistik model

$$Pr(M_{ij} = m) = \frac{e^{\beta_{0,m} + \mathbf{x}^T \boldsymbol{\beta}_m}}{\sum_{k=1}^{14} e^{\beta_{0,k} + \mathbf{x}^T \boldsymbol{\beta}_k}}$$

eşitliği ile yazılır [67].

Regresyon modelinde açıklayıcı değişkenlerin sayısı gözlem sayısına yakın ya da büyükse tahmin yönteminde cezalandırma önerilir [68, 69]. Bazı hasar kategorilerinde gözlem sayısı az olduğu için En Küçük Mutlak Daraltma ve Seçim Operatörü (Least absolute shrinkage and selection operator, Lasso) modeli ile regresyon modelinde değişken seçimi yapılmıştır.

$m_{ij}$ ,  $i = 1, \dots, B$ ,  $j = 1, \dots, N_i$ , gözlemleri için amaç cezalandırılmış olabilirlik fonksiyonunu maksimize etmektir.

$$\max_{\{\beta_{0,k}, \boldsymbol{\beta}_k | k=1, \dots, 14\}} \left[ \frac{1}{\sum_{i=1}^B N_i} \sum_{i=1}^B \sum_{j=1}^{N_i} \log Pr(M_{ij} = m_{ij}) \right] + \lambda \sum_{k=1}^{14} \|\boldsymbol{\beta}_k\|_1$$

$y_{ijk} = I(m_{ij} = k)$  olmak üzere log-olabilirlik fonksiyonu aşağıdaki şekilde de yazılabilir.

$$\ell(\{\beta_{0,k}, \boldsymbol{\beta}_k\}) = \frac{1}{\sum_{i=1}^B N_i} \sum_{i=1}^B \sum_{j=1}^{N_i} \left[ \sum_{k=1}^{14} y_{ijk} (\beta_{0,k} + \mathbf{x}_i^T \boldsymbol{\beta}_k) - \log \left( \sum_{k=1}^{14} e^{\beta_{0,k} + \mathbf{x}_i^T \boldsymbol{\beta}_k} \right) \right]$$

Maksimizasyon problemi aşağıda tanımlanmıştır.

$$\max_{\{\beta_{0,k}, \boldsymbol{\beta}_k\}} \ell(\{\beta_{0,k}, \boldsymbol{\beta}_k\}) + \lambda \sum_{k=1}^{14} \|\boldsymbol{\beta}_k\|_1$$

$\{\beta_{0,k}, \boldsymbol{\beta}_k\}$ 'nin tahmini için maksimizasyon problemi kısmi kuadratik yaklaşım kullanılarak çözülmüştür [68].

$$\ell_{Q,k}(\beta_{0,k}, \boldsymbol{\beta}_k) = -\frac{1}{2 \sum_{i=1}^B N_i} \sum_{i=1}^B \sum_{j=1}^{N_i} w_{jik} (z_{ijk} - \beta_{0,k} - \mathbf{x}_i^T \boldsymbol{\beta}_k)^2 + \tilde{C}$$

$\tilde{C}$  bir sabit olmak üzere, ve

$$z_{ijk} = \tilde{\beta}_{0,k} + \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}_k + \frac{y_{ijk} - \tilde{p}_{ij}}{\tilde{p}_{ij}(1 - \tilde{p}_{ij})}$$

$$w_{jik} = \tilde{\beta}_{0,k} + \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}_k + \frac{y_{ijk} - \tilde{p}_{ij}}{\tilde{p}_{ij}(1 - \tilde{p}_{ij})}$$

eşitlikleri ile her bir kategori için olasılık fonksiyonu

$$\tilde{p}_{ij} = \tilde{Pr}(M_{ij} = m_{ij}) = \frac{e^{\tilde{\beta}_{0,m} + \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}_m}}{\sum_{k=1}^{14} e^{\tilde{\beta}_{0,k} + \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}_k}}$$

elde edilir.

Her bir  $\lambda$  değeri için , farklı kategorilerde  $k = 1, \dots, 14$ , cezalandırılmış ağırlıklı en küçük kareler yöntemi ile çözüm elde edilmiştir.

$$\min_{\beta_{0,k}, \boldsymbol{\beta}_k} \left[ -\ell_{Q,k}(\beta_{0,k}, \boldsymbol{\beta}_k) + \lambda \|\boldsymbol{\beta}_k\|_1 \right]$$

### 4.3 Kopula Regresyon Modeli

Bölüm 4.2’de tanımlanan üç parçalı toplam hasar modelinde hasar sıklığı ve hasar kategorisi için uygun regresyon modeli belirlendikten sonra kopula fonksiyonu ile hasar tutarı modelinin belirlenmesi aşamasına geçilmektedir.

Hasar kategorisi değişkeni  $M$ ’in farklı değerleri için farklı boyutlarda kopula fonksiyonları oluşturulmalıdır.

$M=15$  için dört değişkenli kopula fonksiyonu oluşturulmuştur.

Çalışmada önerilen kopula modelinin alt modelinden biri dört değişkenli ortak dağılım fonksiyonu,

$$\begin{aligned} F(y_1, y_2, y_3, y_4) &= \Pr(y_1 \leq Y_1, y_2 \leq Y_2, y_3 \leq Y_3, y_4 \leq Y_4) \\ &= \Pr(F_1(Y_1) \leq F_1(y_1), F_2(Y_2) \leq F_2(y_2), F_3(Y_3) \leq F_3(y_3), F_4(Y_4) \leq F_4(y_4)) \\ &= H(F_1(y_1), F_2(y_2), F_3(y_3), F_4(y_4)) \end{aligned} \quad (4.2)$$

şeklinde tanımlanmıştır.

Dört değişkenli ortak olasılık yoğunluk fonksiyonu,

$$f_{y,1234} = c(F_1(y_1), \dots, F_4(y_4)) \times \prod_{j=1}^4 f_j(y_j) \quad (4.3)$$

eşitliği ile tanımlanmıştır. Eşitlik 4.3 her bir  $i$  gözlemi için revize edilmiştir:

$$f_{y,1234}(y_1, y_2, y_3, y_4) = c_4(F_{i,1}(y_1), F_{i,2}(y_2), F_{i,3}(y_3), F_{i,4}(y_4)) \prod_{j=1}^4 f_{i,j}(y_j) \quad (4.4)$$

$f_{i,j}$ ;  $i$ . gözlemin  $j$ . hasar tutarına ait olasılık yoğunluk fonksiyonudur.

Eşitlik 4.4’de oluşturulan dört değişkenli kopula olasılık fonksiyonu seçimi araştırmacının amacı ve veri yapısına uygun olarak belirli kopula fonksiyonları ile sınırlandırılmalıdır. Bu çalışmada çok değişkenli düzlemde negatif ve pozitif bağımlılığın her ikisini de modelleyebilen eliptik kopula ailelerinden Normal kopula ve t kopula, aday kopula fonksiyonları olarak belirlenmiştir.

Hasar kategorisi deęişkeni  $M$ 'nin dięer deęerleri için Eşitlik 4.2'deki kopula dağılım fonksiyonu iki ve üç deęişkenli olarak da yapılandırılmıştır.

İkiden daha fazla deęişkenli bir kopula modeli modellenmek istendiğinde eş anlı tahmin edilecek parametre sayısı ve ilişki parametrelerinin sayısının çok olması MLE yönteminin uygulanmasını zorlaştırmaktadır. Bu çalışmada da dięer bir adıyla iki aşamalı model tahmini olan IFM yöntemi kullanılmıştır. İlk aşamada marjinal modellerin regresyon parametreleri bulunmuştur. Sonrasında birleşik model için olabilirlik fonksiyonu maksimize edilirken ilk aşamada elde edilen marjinal model parametreleri kullanılmıştır.

$Y_1, Y_2, Y_3$  ve  $Y_4$  bağımlı hasar tutarlarını dahil ederek için kopula fonksiyonları ile oluşturulan olabilirlik fonksiyonu Ek A'da verilmiştir.



## 5. UYGULAMA

Tez çalışması için özel bir sigorta şirketinden alınan Türkiye geneli Karayolları Zorunlu Trafik Sigortası portföy bilgileri her bir poliçe kaydı ayrı ayrı incelenerek çalışmaya temel oluşturacak veri seti elde edilmiştir.

Çalışma için oluşturulan veri seti; araç teknik özellikleri, sigortalı demografik özellikleri, farklı hasar türlerine göre hasar kayıtları ve muallak/ödenmiş hasar tutarı bilgileri olmak üzere dört temel kategoride gruplandırılmıştır.

Veri setindeki her bir poliçeye ait olan toplamsal bilgi kullanılarak hasar sayısı kesikli bağımlı rastlantı değişkeni olarak veriye eklenmiştir.

Hasar modellerinde bir hasar verisinin müşteri bazlı ya da poliçe bazlı mı olarak takip edileceğinin belirlenmesinin sonuçlarda farklılık oluşturabileceği gözlemlenmiştir. Toplam hasar modelinin aktüeryal prim hesabını açıklaması hedeflenirken poliçe bazlı çalışmanın daha etkin olacağına karar verilmiştir.

Veri setinde iptal durumundaki poliçeler incelenip poliçelerin exposure (riske maruz değer) değerleri hesaplanmıştır.

Bu veri setinde maddi hasar tutarı, maddi muallak hasar tutarı, bedeni hasar tutarı ve bedeni muallak hasar tutarı değişkenleri bağımlı sürekli değişken olarak belirlenmiştir. Hasar sayısı ise bağımlı kesikli değişken olarak modele dahil edilirken hasar kategorisi bağımlı sınıflayıcı değişken olarak modele eklenmiştir.

Veri seti 2012 - 2016 yıllarına ait 4 136 011 adet poliçe kaydından oluşmaktadır. Veri gözlem süresi olan beş yıllık periyot içerisinde 2 284 564 sigortalının poliçelerine ait yıllık kayıtlar bulunmaktadır. Bazı müşterilerin birden fazla poliçeye sahip olduğu görülmüştür. Veri seti içerisinde ticari poliçe kayıtları çıkartılarak özel araç poliçe kayıtları üzerinden modelleme ve analizler yapılmıştır. Veri seti içerisindeki her bir poliçe kaydının birbirinden bağımsız olduğu varsayılmıştır.

Çalışmada kullanılan veri seti önerilen model sonuçlarını değerlendirmek için eğitim verisi (in sample) ve test verisi (out sample) olarak ikiye ayrılmıştır. Çalışmanın ilk



kısmı için 2012-2015 yılları arasındaki veri seti kullanılarak istatistiksel model elde edilmiş, modelin geçerliliği ise 2016 yılına ait veri seti ile değerlendirilmiştir.

Veri setinden elde edilen bağımlı değişkenler;

- Maddi hasar tutarı
- Maddi muallak hasar tutarı
- Bedeni hasar tutarı
- Bedeni muallak hasar tutarı
- Hasar sayısı ve
- Hasar kategorisi değişkenleri

iken veri setindeki açıklayıcı değişkenler araç teknik özellikleri ve sigortalı demografik özelliklerinden oluşmaktadır. Açıklayıcı değişkenler Şekil 5.1 'de gösterilmiştir.

<b>Demografik Bilgiler</b>
<i>Değişken</i>
Police sahibinin yaşı
Cinsiyet
Medeni durum (1:4): Evlü, bekar, boşanmış ve diğer
<b>Araç Özellikleri</b>
<i>Değişken</i>
Arabanın yaşı
Max hız
Hızlanma süresi
Uzunluk
Genişlik
Ağırlık
Tork
Beygir gücü kapasitesi
Motor hacmi
Kerb araç ağırlığı
Araç sınıfı: (A:L)

Şekil 5.1 : Kategorik ve sürekli açıklayıcı değişkenler

Nitel açıklayıcı değişkenler için kategoriler belirlenirken temel seviyedeki (0) gözlem sayısı seyrek olmayacak şekilde belirlenmiştir [47] .

Veri analizi ve modellemesi için R 3.5.3 istatistiksel yazılım programı kullanılmıştır. "copula", "VineCopula", "MAAS", çok değişkenli dağılım fonksiyonlarını yapılandırırken kullanılan başlıca R paketlerdir.

Bölüm 4, Eşitlik 4.1'de tanımlanan üç parçalı toplam hasar modeli adım adım tahmin edilmiş ve tüm model bileşenleri elde edildikten sonra, model doğrulama çalışması yapılmıştır.

- Çalışma veri seti içerisinde 2012- 2015 yılları poliçe kayıtları ayrılarak, eğitim verisi (in sample) olarak belirlenmiştir.
- Önerilen üç parçalı modelin ilk bileşeni hasar sıklığı modelidir. Bu nedenle ilk olarak eğitim verisi ile hasar sıklığı regresyon modeli tahminleri yapılmış ve sonuçlar Bölüm 5.1'de tartışılmıştır.
- Üç parçalı modelin ikinci bileşeni hasar kategorisi modelidir. Hasar kategorisi olasılıkları çoklu lojistik regresyon ile tahmin edilmiş ve sonuçları Bölüm 5.3'de verilmiştir.
- Modelin üçüncü bileşeni, dört farklı hasar tutarı ödemesinin Normal, t kopula ve bağımsız kopula ile birlikte modellendiği kopula regresyon modelidir. Sonuçlar Bölüm 5.4'de incelenmiştir.
- Üç model bileşeni için de nihai modeller elde edilmiştir. Bölüm 5.5'de eğitim veri seti ile tahmini yapılan üç parçalı model yapısı üzerinde test veri seti (out sample ) ile model doğrulama yapılmış ve önerilen modelin sonuçları tartışılmıştır.

### 5.1 Hasar Sıklığı Modellemesi Sonuçları

Trafik sigortası poliçe kayıtlarından elde edilen hasar kayıt bilgilerine ait frekans tablosu aşağıda gösterilmiştir:

**Çizelge 5.1** : Hasar Sayısı Frekans Tablosu

0	1	2	3	4	5
3435523	119357	6184	401	48	6

Hasar sayısı verisinde sıfır değerinde yığılma görülmektedir. Hasar modellerinde özellikle trafik sigortasında hasar frekansının sıfır değerinde yığılması sıklıkla

karşılaşılan bir durumdur. Hasar frekansının düşün olduğu poliçeler olarak değerlendirilmelerine karşın sigortalıların görece küçük hasar tutarlarını sigorta şirketine bildirmemesi de sıfır değerinde olan yığılmayı arttırmaktadır.

Poisson, Negatif Binom ve yarı- Poisson regresyon modeli AIC değerleri Çizelge 5.2’de hesaplanmıştır. ( $AIC = -2 \times L(\theta_{MLE}) + 2 \times (\text{parametre sayısı})$ )

**Çizelge 5.2 : Hasar Sıklığı Modellemesi Sonuçları**

Poisson R.M. Log bağ fonksiyonu AIC	Quasi-poisson R.M. Log bağ fonksiyonu AIC	Negatif-binom R.M. Log bağ fonksiyonu AIC
734277	826321	736008

Hasar sayısı verisinde yayılım görülmesine rağmen quasi-poisson ve negatif binom regresyon modeli için hesaplanan AIC değerleri Poisson regresyon modeli AIC sonucuna göre daha büyük bulunmuştur. Ayrıca quasi-poisson regresyon modeli tahmin sonuçlarında regresyon katsayı değerleri Poisson regresyon modeli katsayı değerleri ile çok yakın elde edilmiştir.

**Çizelge 5.3 : Gözlenen ve Beklenen Hasar Sayıları**

Hasar Sayısı	Gözlenen	Poisson	Negatif Binom
0	3435523	3421870	3410817
1	119357	134520	144728
2	6184	4737	5564
3	401	360	379
4	48	30	35
5	6	2	4

Gözlenen ve tahmin edilen hasar sayılarına bakıldığında verinin sıfır değerlerinin tahmini için Poisson modeli daha fazla uyum göstermiştir. Tek parametrelili olması ve toplam hasar modeli içerisine entegre olduğunda parametre sayısını arttırmayacağı için analizlere Poisson regresyon modeli sonuçları ile devam edilmiştir. Hasar sayısının pozitif değerler aldığı durumda ise hasar kategorisi dağılımı üç parçalı modele dahil edilmektedir. Bu durumda hasar sayısının tahmini için Negatif Binom veya Poisson dağılım parametreleri kullanıldığında kopula bağımlılık parametreleri üzerinde anlamlı farklılık olmadığı görülmüştür. Bu nedenle hasar sayısı modeli için Poisson regresyon modeli ile analizlere devam edilmiştir.

Değişken	Tahmin	Standart Hata
Sabit terim	-6.965	0.1647 *
Yaş	-0.01057	0.00024 *
Erkek	-0.05547	0.00635 *
Evli	-0.1708	0.00848 *
Boşanmış	0.1105	0.01596 *
Diğer	0.09569	0.02359 *
Exposure	0.9784	0.00502 *
Araç yaşı	0.00241	0.00079 *
Maks hız	-0,0046	0.00048 *
Hızlanma süresi	0.01363	0.00279 *
Uzunluk	-0.00012	0.00002 *
Genişlik	0.00035	0.00003 *
Yükseklik	-0.00050	0.00006 *
Tork	0.00180	0.00007 *
Beygir gücü	0.00027	0.00027
Motor hacmi	0.00065	0.00001 *
Kerb ağırlığı	0.00001	0.00033
B sınıfı	1.256	0.46140 *
C sınıfı	1.254	0.47880 *
D sınıfı	1.293	0.05177 *
E sınıfı	1.423	0.06183 *
F sınıfı	2.489	0.09232 *
G sınıfı	8.783	0.87790 *
H sınıfı	5.224	0.61010
Diğer sınıf	1.055	0.57060

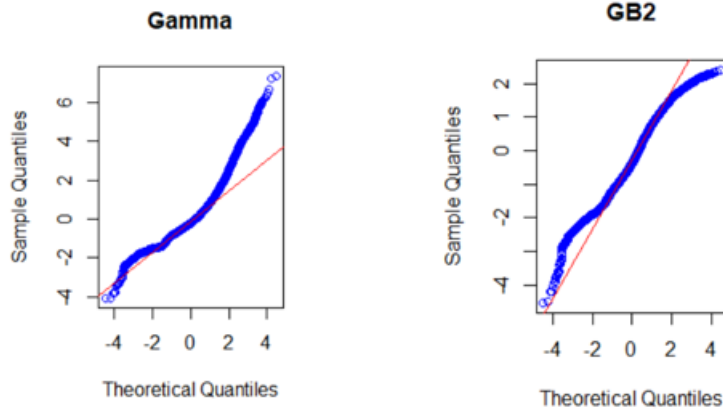
**Şekil 5.2 :** Poisson Gdm Sonuçları

Şekil 5.2'deki Poisson regresyon modeli tahmin sonuçlarına göre Beygir gücü, Kerb araç ağırlığı, H sınıfı ve Diğer sınıf değişkeni hariç diğer tüm açıklayıcı değişkenler için katsayılar anlamlı bulunmuştur (\* %95 güven düzeyinde istatistiksel olarak anlamlılık)

## 5.2 Hasar Şiddeti Modellemesi Sonuçları

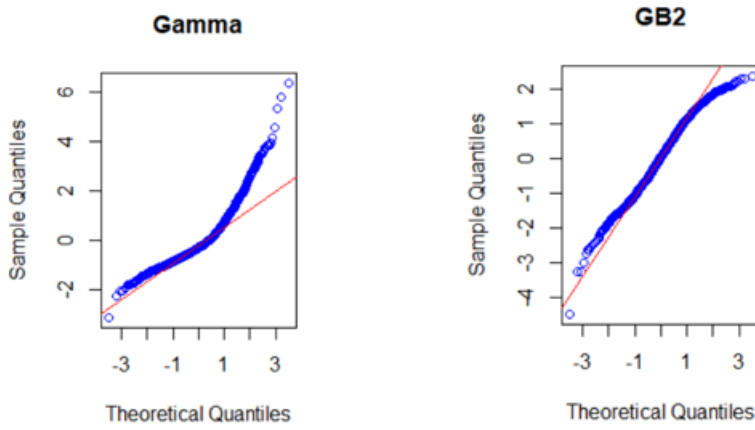
Her bir hasar tutarı için ayrı ayrı Gamma ve Genelleştirilmiş Beta 2 - GB2 dağılımları ile regresyon modellemesi yapılmıştır. Aday olarak seçilen dağılımların hasar tutarı verisine uyumlarını değerlendirebilmek için QQ (quantile- quantile) grafikleri incelenmiştir. QQ grafiği dağılımı tahmin edilmek istenen veri için iki ayrı dağılımı kantil değerleri yardımıyla karşılaştırmaktadır. Her bir hasar türü için aday olarak seçilen Gamma ve GB2 Dağılımlarına ait kantil değerleri ve ampirik dağılımın kantil

değerleri elde edilerek aşağıdaki QQ grafikleri elde edilmiştir. QQ grafiği özellikle dağılımın kuyruk kısımlarındaki uyumu değerlendirebilmek için elverişlidir [70].



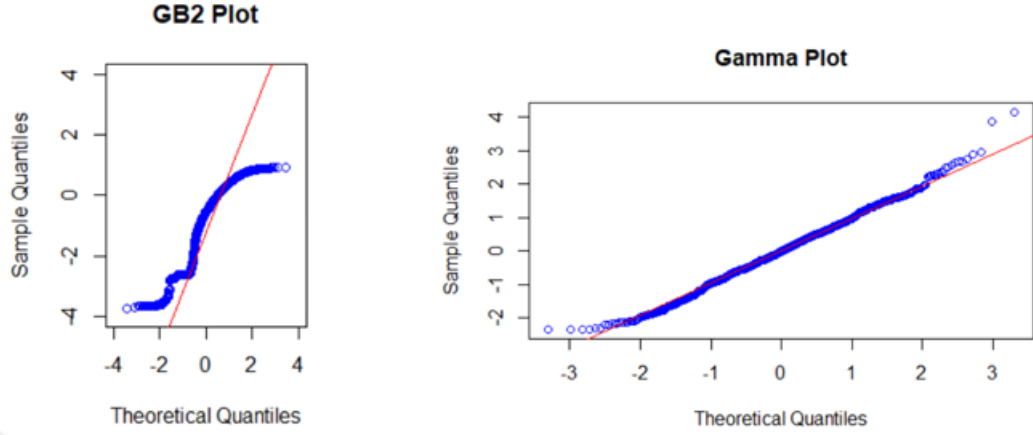
Şekil 5.3 :  $Y_1$  Hasar değişkeni için QQ Grafikleri

Maddi hasar tutarını ifade eden  $Y_1$  bağımlı değişkeni için oluşturulan Gamma ve GB2 regresyon modelleri QQ grafiklerinde GB2 dağılımının Gamma dağılımına göre veriye daha iyi uyduğu görülmüştür. Özellikle sağ kuyruk kısmında GB2 modeli Gamma modeline göre daha iyi uyum sağlamıştır.



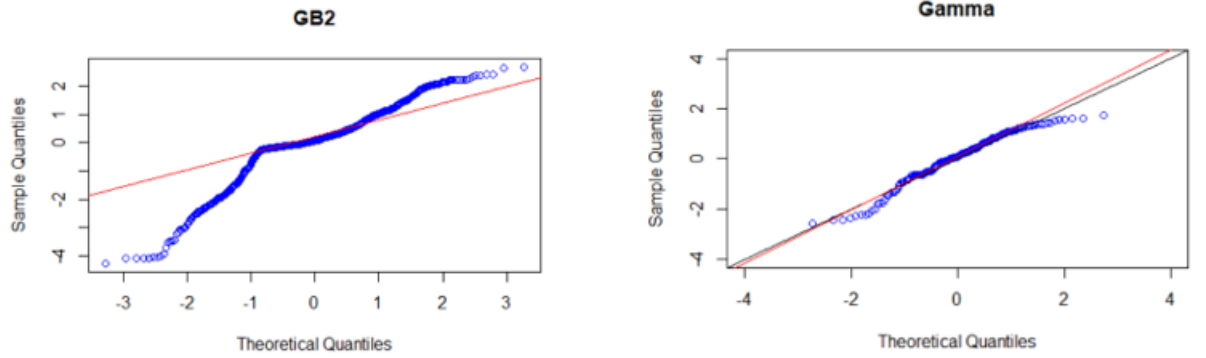
Şekil 5.4 :  $Y_2$  Hasar değişkeni için QQ Grafikleri

Maddi muallak hasar tutarını ifade eden  $Y_2$  bağımlı değişkeni için de QQ grafiklerine bakıldığında GB2 regresyon modeli daha iyi sonuç vermiştir.



Şekil 5.5 :  $Y_3$  Hasar değişkeni için QQ Grafikleri

Bedeni hasar tutarını ifade eden  $Y_3$  bağımlı değişkeni için ise Gamma regresyon modelinin GB2 regresyon modeline göre daha iyi uyduğu görülmüştür.



Şekil 5.6 :  $Y_4$  Hasar değişkeni için QQ Grafikleri

Bedeni muallak hasar tutarını ifade eden  $Y_4$  bağımlı değişkeni için de QQ grafiklerinde Gamma dağılımı daha iyi sonuç vermiştir.

Dört bağımlı hasar tutarı için de tahmin edilen Gamma ve GB2 regresyon modeli AIC değerleri Çizelge 5.4'de verilmiştir.

Çizelge 5.4 : Gamma ve Gb2 Regresyon Modelleri AIC Değerleri

	Gamma Regresyon Modeli	Gb2 Regresyon Modeli
	Log bağ fonksiyonu	Log bağ fonksiyonu
$Y_1$	22911054	1146985
$Y_2$	37691	18414
$Y_3$	17940	35968
$Y_4$	17738	23496

Dört farklı hasar tutarı regresyon modelleri için elde edilen AIC değerleri QQ grafik sonuçlarını desteklemektedir. En küçük AIC değerlerine bakıldığında Maddi hasar ve maddi muallak hasar tutarı için GB2 regresyon modeli tercih edilmiştir. Bedeni hasar tutarı ve bedeni muallak hasar tutarı için ise Gamma regresyon modeli seçilmelidir. AIC değerlerine göre elde edilen sonuçların QQ grafiklerinde öngörülen model sonuçlarını desteklediği görülmüştür.

Tüm hasar tutarı regresyon modelleri tahmin sonuçları Ek B, C, D ve E'de verilmiştir. Hasar tutarı modelleri için ayrı ayrı elde edilen regresyon parametreleri, Bölüm 5.5'te kopula modelindeki bağımlılık parametrelerinin tahmini için kullanılmıştır.

### **5.3 Hasar Kategorisi Modeli Sonuçları**

Her bir hasar kategorisi aşağıda gösterildiği şekilde 15 kategori oluşturularak kodlanmıştır.

"1000" "0100" "0010" "0001" "1100" "1010" "1001" "0110" "0101" "0011" "1110"  
"1101" "1011" "0111" "1111"

Bir kategori referans kategorisi olmak üzere tüm kategorilerin ortaya çıkma olasılıkları çoklu logit model yardımıyla modellenmiştir. Böylece her bir hasar kategorisinin ortaya çıkma olasılıkları tahmin edilmiştir.

**Çizelge 5.5 : Çoklu Logit Model Sonuçları**

Hasar Kombinasyonu	Gözlenen	Tahmin
( $Y_1$ )	0.96830	0.96025
( $Y_2$ )	0	0
( $Y_3$ )	0.00436	0.00517
( $Y_4$ )	0.00141	0.00160
( $Y_1, Y_2$ )	0.01710	0.02109
( $Y_1, Y_3$ )	0.00264	0.00349
( $Y_1, Y_4$ )	0.00080	0.00096
( $Y_2, Y_3$ )	0	0
( $Y_2, Y_4$ )	0.00002	0.00023
( $Y_3, Y_4$ )	0.00314	0.00428
( $Y_1, Y_2, Y_3$ )	0.00006	0.00017
( $Y_1, Y_2, Y_4$ )	0.00006	0.00034
( $Y_1, Y_3, Y_4$ )	0.00198	0.00200
( $Y_2, Y_3, Y_4$ )	0.00003	0.00004
( $Y_1, Y_2, Y_3, Y_4$ )	0.00010	0.00020

Gözlenen olasılık değerleri ile tahmini olasılık değerleri Çizelge 5.5’de verilmiştir.

#### **5.4 Kopula Regresyon Modeli Sonuçları**

Çalışmada farklı hasar tutarları arasındaki bağımlılığı modelleyebilmek için eliptik kopula ailesinden Normal ve t kopula fonksiyonları kullanılmıştır.

Kopula parametreleri tahmin edilirken iki aşamalı model olan Marjinallere İlişkin Çıkarsama yöntemi kullanılmıştır. Kopula ile oluşturulan birleşik model için olabirliklik fonksiyonu maksimize edilirken marjinal modellere ait regresyon parametreleri kullanılmıştır.

Maddi hasar tutarı ( $Y_1$ ), maddi muallak hasar tutarı ( $Y_2$ ), bedeni hasar tutarı ( $Y_3$ ) ve bedeni muallak hasar tutarı ( $Y_4$ ) arasındaki ilişki için farklı kopula fonksiyonları ile tahmin edilen bağımlılık parametre değerleri aşağıda gösterilmiştir.



**Normal kopula için bağımlılık parametreleri tahmin sonuçları: (standart hatalar parantez içinde belirtilmiştir)**

$$\begin{bmatrix} \sigma_{1,2} \\ \sigma_{1,3} & \sigma_{2,3} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} \end{bmatrix} = \begin{bmatrix} 0.368 (0.012) \\ 0.128 (0.038) & -0.530 (0.113) \\ 0.048 (0.036) & -0.323 (0.089) & 0.034 (0.026) \end{bmatrix}$$

**t kopula için bağımlılık parametreleri tahmin sonuçları: (standart hatalar parantez içinde belirtilmiştir)**

$$\begin{bmatrix} \sigma_{1,2} \\ \sigma_{1,3} & \sigma_{2,3} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} \end{bmatrix} = \begin{bmatrix} 0.408 (0.014) \\ 0.105 (0.052) & -0.389 (0.134) \\ 0.039 (0.036) & -0.134 (0.167) & 0.005 (0.026) \end{bmatrix}$$

Her iki model sonucunda da bağımlı değişkenler arasındaki ilişki katsayıları yakın bulunmuştur.

Maddi hasar tutarı ile maddi muallak hasar tutarı arasında pozitif yönde önemli bir bağımlılık olduğu görülmüştür. Maddi hasar tutarı ile bedeni hasar tutarı arasında daha zayıf olmakla birlikte pozitif bir ilişki olduğu söylenebilir.

Diğer taraftan maddi muallak hasar tutarı ile bedeni hasar tutarı arasında negatif yönde önemli bir bağımlılık olduğu görülmüştür. Maddi muallak hasar tutarı ile bedeni muallak hasar tutarı arasında ise daha zayıf olmakla birlikte negatif bir ilişki olduğu söylenebilir.

Muallak hasar tutarı en genel tanımıyla gerçekleşmiş ama herhangi bir sebeple ödemesi yapılmamış hasar tutarını ifade etmektedir [71]. Bu nedenle muallak hasar tutarları ile aynı riske karşılık gelen hasar tutarlarının pozitif yönde ilişkili olması beklenen bir sonuçtur.

Kopula parametrelerinin istatistiksel anlamlılıklarına bakıldığında normal kopula için  $\sigma_{1,4}$ ,  $\sigma_{3,4}$  parametreleri anlamsız bulunmuştur ( $0.048/0.036=1.33 < 1.96$ ;  $0.034/0.026=1.30 < 1.96$ ). t kopula modeli için de  $\sigma_{1,4}$ ,  $\sigma_{3,4}$  parametreleri istatistiksel olarak anlamsız bulunmuştur. Bu kategoriler için verideki gözlem sayısının az olması parametre tahminlerinin hata değerlerinin kısmen yüksek bulunmasına neden olabilmektedir.

Kopula model sonuçları Akaike bilgi kriteri kullanılarak karşılaştırma yapılmıştır.

Normal kopula, t kopula ve bağımsız kopula modelleri için AIC değerleri Çizelge 5.6'de verilmiştir. Bağımsız kopula ile oluşturulan model için hesaplanan AIC değerinin büyük olduğu görülmektedir. Normal kopula ve t kopula modellerinin AIC değerlerinin birbirine yakın olmakla birlikte t kopula (s.d. 17.182 ) için hesaplanan AIC değeri daha küçüktür.

**Çizelge 5.6 : Kopula Model Karşılaştırması**

Normal Kopula	t Kopula	Bağımsız Kopula
AIC	AIC	AIC
741.7686	709.356	6386

Önerilen üç parçalı modelde bağımlılık yapısının modele dahil edilmesi için t kopula fonksiyonu parametre tahmin sonuçları kullanılmalıdır.

### **5.5 Model Doğrulama**

Çalışmada kullanılan üç parçalı toplam hasar modeli, t kopula ile oluşturulan bağımlı üç parçalı model ve Normal kopula oluşturulan bağımlı üç parçalı model karşılaştırılmıştır.

Çalışmada önerilen üç parçalı modelde farklı hasar türleri arasındaki bağımlılık yapısının değerlendirilebilmesi için simülasyon çalışması yapılmıştır. Eğitim verisi için elde edilen tahmin modellerinin parametre vektörleri kullanılarak 50000 gözlem için bağımlı hasar verisi üretilmiştir.

Bağımlı hasar verisi üretebilmek için verideki bağımlılık yapısını modelleyen Normal kopula, t kopula ve bağımsız kopula fonksiyonlarının test verisi için elde edilen ilişki parametre değerleri kullanılmıştır.

Simülasyon- Doğrulama adımları:

1. "Tüm veri" poliçe kayıt tarihleri baz alınarak ikiye bölünmüştür.
2. Eğitim verisi analiz edilerek oluşturulan marjinal modeller ve kopula modellerinin parametre vektörleri kaydedilmiştir.
3. Test verisi içerisindeki açıklayıcı değişkenler girdi olarak kullanılmıştır.

4. Kaydedilen parametre vektörleri ve test verisinden alınan açıklayıcı değişkenlerin matrisi birleştirilmiş her bir gözlem için 10000 iterasyon yapılarak toplam hasar tutarları üretilmiştir.

5. Her bir iterasyon için hesaplanan değerler ilgili gözlem için kaydedilip hata değerlerinin hesaplanabilmesi için ortalamalar alınmıştır.

6. Her bir iterasyon sonucunda kaydedilen toplam hasar tutarları için üst kuyruk kantil (quantile) değerleri elde edilmiştir.

Aday modellerin tahminlerinin doğruluğunun test edilmesi için hata kareler ortalamasının karekökü (RMSE-Root mean square error) ve ortalama mutlak hata (MAE-Mean absolute error) ölçüleri kullanılmıştır.

Ortalama Mutlak Hata (MAE); Gözlem değeri ile tahmin değeri mutlak farkın ortalaması alınarak hesaplanmaktadır.  $N$  örneklem büyüklüğü olmak üzere MAE:

$$MAE = \frac{1}{n} \sum_i |y_i - \hat{y}_i| \quad (5.1)$$

eşitliği ile bulunmaktadır.

Hata Kareler Ortalamasının Karekökü (RMSE); Gözlem değeri ile tahmin değeri arasındaki farkın kareler ortalaması alınarak aşağıdaki biçimde hesaplanmaktadır [72]:

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2} \quad (5.2)$$

**Çizelge 5.7 :** Toplam Hasar Modeli Karşılaştırma Sonuçları

	t Kopula	Normal Kopula	Bağımsız Model
<i>MAE</i>	108.5714	108.8623	111.0656
<i>RMSE</i>	1485.522	1992.804	1596.177

Farklı hasar tutarları arasındaki bağımlılığın t kopula ile modellendiği üç parçalı toplam hasar modeli tahmin hata sonuçları Normal kopula ile oluşturulan üç parçalı toplam hasar modeli tahmin hata sonuçlarına yakın olmakla birlikte daha küçüktür.

Bağımsız kopula ile oluşturulan üç parçalı modelin hata değerleri bağımlı modellerin hata değerlerine göre daha yüksek bulunmuştur.

t kopula modeli ile üretilen toplam hasar tutarlarının tahmin değerleri ve bağımsız kopula modeli ile üretilen toplam hasar tutarlarının tahmin değerlerine ilişkin istatistikler Çizelge 5.8' da gösterilmiştir.

**Çizelge 5.8** : Tahmin Değerlerine ilişkin İstatistikler

Model	Min	1.Çeyrek	Ortalama	3.Çeyrek	Max
t kopula modeli	0	44502.82	85017.01	165531.20	218045.38
Bağımsız model	0	38576.19	77163.87	115751.54	154339.22

50000 gözlem için üretilen toplam hasar değeri dağılımının sol kuyruk ve ortalama tahmin değerlerinin yakın olduğu görülmektedir. Bağımlılık yapısı modele dahil edilmediğinde tahmin değerlerinin sağ kuyruk kısımlarında ise daha büyük değerler elde edildiği görülmektedir.

Sigorta portföyünde bulunan poliçeler için sağ üst kuyruktaki kantil değerleri portföy riskinin ölçülmesi için riske maruz değer (Var) hesaplamalarında kullanılmaktadır [73]. Bu nedenle toplam hasar tutarı için ortalama tahmin değerlerinin hesaplanmasının yanı sıra kantil değerleri ve portföyde oluşabilecek maksimum hasar değerlerinin tahmini de önemlidir.

Çalışma veri setinden rastgele seçilen hasarlı bir poliçe için üst kuyruk kantil değerleri bağımlı model ve bağımsız model varsayımı altında hesaplanarak örneklendirilmiştir.

**Çizelge 5.9** : n= 465. gözlem için Kantil değerleri

Model	% 95	% 99
t kopula	1125.321	6907.614
bağımsız model	1058.422	6659.407

Bağımlı modelde %95 ve %99 kantil değerleri için hesaplanan toplam hasar tutarının bağımlılık yapısının dahil edilmediği model sonuçlarına göre daha yüksek olduğu görülmüştür.

Bağımlı modeldeki ilişki parametrelerinin dereceleri arttığında kantil değerleri arasındaki farkın artması beklenmektedir. Bir önceki örnek için seçilen n=465. poliçe için t kopula bağımlılık parametre değerleri artırılarak kantil değerlerinin değişimi gözlenmek istenmiştir.

t kopula modelinde bağımlılık parametre değerleri artırıldığında kantil değerleri arasındaki farkın arttığı Çizelge 5.10' da görülmektedir.

**Çizelge 5.10** : n= 465. gözlem için Kantil değerleri

Model	% 95 ( $\alpha=0.05$ )	% 99 ( $\alpha=0.01$ )
t kopula	1138.36	6939.676
bağımsız model	1058.422	6659.407

Sigorta portföyünün farklı  $\alpha$  değerleri için riske maruz değeri hesaplamalarının yapılabilmesi için dağılımın kantil değerlerinin tahmini önemlidir. Ortalama hasar tutarının tahmini sigorta risk priminin hesaplanmasını sağlarken, sigorta hasar dağılımını bilmek yüzdelik veya kantil değerlerinin kontrol edilmesini sağlamaktadır. Bu nedenle model doğrulama aşamasında nokta tahmini olan ortalama değerlerinin yanı sıra dağılım bilgisinin de elde edilebilmesi için simülasyon çalışması yapılmıştır.

## 6. SONUÇ VE ÖNERİLER

Hasar modellerinde bağımlı değişkenler arasındaki göz ardı edilen bağımlılık yapısı toplam hasar modelinin istatistiksel tahminini etkilemektedir. Bu çalışmada hasar sıklığı, hasar kategorisi ve hasar tutarı için regresyon modelleri oluşturularak bağımlılık varsayımı ile toplam hasar modeline entegre edilmiştir.

Bir hasar gerçekleştiğinde oluşan hasar türleri için farklı teminat ve limitlerde hasar ödemesi yapılabilmektedir. Aynı sigorta poliçesi altında teminat altına alınan farklı hasar ödemeleri için farklı ödeme kayıtları oluşturulmaktadır. Sigorta şirketine bildirilen hasar ödemelerinin farklı başlıklar altında muhasebeleştirildiği kayıtlar beklenen hasar tutarının modellenmesi için kullanılan istatistiksel modele eklenmiştir. Bu çalışmada önerilen model ile gerçekleşebilecek hasara ilişkin daha çok bilgi modele dahil edilerek daha esnek bir model oluşturulmuştur.

Farklı hasar ödemeleri arasındaki bağımlılık yapısı kopula fonksiyonları kullanılarak modellenmiştir. Marjinal modellerden elde edilen regresyon parametreleri kopula modeli içerisine yerleştirilmiştir. Bu nedenle önerilen model kopula regresyon modeli olarak adlandırılmaktadır.

Bu tez çalışması farklı hasar türleri arasındaki bağımlılık yapısının regresyon temelli olarak modellenmek istendiğinde aday bir model önermektedir.

Uygulama verisi olarak çalışılan Trafik sigortasında gerçekleşen bir kaza sonrası yapılan ödemeler için hasar türlerinin ayrılması ve modellenmesi yaygın değildir. Çalışma verisi, Türkiye’de faaliyet gösteren ve prim üretiminde yüksek payı olan özel bir sigorta şirketinden temin edilmiştir. Uygulama sonunda önerilen model sonuçlarına göre maddi hasar tutarı ve maddi muallak hasar tutarı arasında önemli derecede pozitif bağımlılık gözlemlenmiştir. Maddi bedeni hasar tutarları ile muallak maddi hasar tutarları arasında ise önemli derecede negatif bağımlılık gözlemlenmiştir. Bu bağımlılık yapısı Normal ve t kopula fonksiyonu yardımıyla üç parçalı toplam hasar modeline entegre edilmiştir.

Bağımlılık varsayımının ihmal edilmesi beklenen hasar tutarının tahminini etkileyeceği için sigorta şirketinin risk değerlendirmesi yaparken eksik ya da fazla fiyatlandırma/ rezerv hesabı yapmasına neden olabilmektedir. Bağımlılık yapısı özellikle tahmin edilmek istenen toplam hasar tutarının dağılına ait yüzdelerle değerleri de etkileyebilmektedir. Bu nedenle sigorta portföyünün riske maruz değer hesaplamalarını direkt olarak etkileyecektir. Özellikle dağılımın kuyruk değerlerinde bağımlılık varsayımının değerlendirilmesi önemlidir.

İleriki çalışmalar için farklı riskler arasındaki bağımlılık yapısının boyutsal verilerde zaman değişkeni ile olan ilişkisi çalışılmak istenmektedir. Böylece uzun dönem ve kısa dönem bağımlılık etkisinin istatistiksel modele dahil edilmesi amaçlanmaktadır.

Ayrıca bağımlı rastgele değişkenler arasındaki ilişkiyi ifade eden kopula bağımlılık parametrelerinin değişimi ve istatistiksel tahmin modelinin vine kopula fonksiyonları için değerlendirilmesi planlanmaktadır.

## KAYNAKLAR

- [1] **GAMMA, D.U.Z.A.** (2013). Estimation of claim cost data using zero adjusted gamma and inverse Gaussian regression models, *Journal of Mathematics and Statistics*, 9(3), 186–192.
- [2] **Sklar, A.** (1973). Random variables, joint distribution functions, and copulas, *Kybernetika*, 9(6), 449–460.
- [3] **Genest, C. ve MacKay, J.** (1986). The joy of copulas: bivariate distributions with uniform marginals, *The American Statistician*, 40(4), 280–283.
- [4] **Nelsen, R.B.** (2007). *An introduction to copulas*, Springer Science & Business Media.
- [5] **Genest, C. ve Favre, A.C.** (2007). Everything you always wanted to know about copula modeling but were afraid to ask, *Journal of hydrologic engineering*, 12(4), 347–368.
- [6] **Joe, H.** (2014). *Dependence modeling with copulas*, Chapman and Hall/CRC.
- [7] **Frees, E.W. ve Valdez, E.A.** (1998). Understanding relationships using copulas, *North American actuarial journal*, 2(1), 1–25.
- [8] **Yang, X., Frees, E.W. ve Zhang, Z.** (2011). A generalized beta copula with applications in modeling multivariate long-tailed data, *Insurance: Mathematics and Economics*, 49(2), 265–284.
- [9] **Parsa, R.A. ve Klugman, S.A.** (2011). Copula regression, *Variance Advancing and Science of Risk*, 5, 45–54.
- [10] **Carriere, J.F.** (2000). Bivariate survival models for coupled lives, *Scandinavian Actuarial Journal*, 2000(1), 17–32.
- [11] **Klugman, S.A., Panjer, H.H. ve Willmot, G.E.** (2012). *Loss models: from data to decisions*, cilt715, John Wiley & Sons.
- [12] **Luciano, E., Spreeuw, J. ve Vigna, E.** (2008). Modelling stochastic mortality for dependent lives, *Insurance: Mathematics and Economics*, 43(2), 234–244.
- [13] **Spreeuw, J. ve Wang, X.** (2008). Modelling the short-term dependence between two remaining lifetimes, *Cass Business School Discussion Paper*, 2(3).
- [14] **David, M. ve Jemna, D.V.** (2015). Modeling the frequency of auto insurance claims by means of poisson and negative binomial models, *Annals of the Alexandru Ioan Cuza University-Economics*, 62(2), 151–168.



- [15] **Fu, L. ve Moncher, R.B.** (2004). Severity distributions for GLMs: Gamma or lognormal? Evidence from Monte Carlo simulations, *Casualty Actuarial Society Discussion Paper Program*, 149–230.
- [16] **Meyers, G.** (2005). On predictive modeling for claim severity, *Casualty Actuarial Society Forum*, Citeseer, s.215–253.
- [17] **Mert, M. ve Saykan, Y.** (2005). On a bonus–malus system where the claim frequency distribution is geometric and the claim severity distribution is Pareto, *Hacettepe Journal of Mathematics and Statistics*, 34, 75–81.
- [18] **Santi, D., Purnaba, I. ve Mangku, I.** (2016). Bonus-Malus System with the Claim Frequency Distribution is Geometric and the Severity Distribution is Truncated Weibull, *IOP Conference Series: Earth and Environmental Science*, cilt 31, IOP Publishing, s.012006.
- [19] **Frees, E., Lee, G. ve Yang, L.** (2016). Multivariate frequency-severity regression models in insurance, *Risks*, 4(1), 4.
- [20] **Gilenko, E.V. ve Mironova, E.A.** (2017). Modern claim frequency and claim severity models: An application to the Russian motor own damage insurance market, *Cogent Economics & Finance*, 5(1), 1311097.
- [21] **Krämer, N., Brechmann, E.C., Silvestrini, D. ve Czado, C.** (2013). Total loss estimation using copula-based regression models, *Insurance: Mathematics and Economics*, 53(3), 829–839.
- [22] **Czado, C., Kastenmeier, R., Brechmann, E.C. ve Min, A.** (2012). A mixed copula model for insurance claims and claim sizes, *Scandinavian Actuarial Journal*, 2012(4), 278–305.
- [23] **Shi, P., Feng, X. ve Ivantsova, A.** (2015). Dependent frequency–severity modeling of insurance claims, *Insurance: Mathematics and Economics*, 64, 417–428.
- [24] **Garrido, J., Genest, C. ve Schulz, J.** (2016). Generalized linear models for dependent frequency and severity of insurance claims, *Insurance: Mathematics and Economics*, 70, 205–215.
- [25] **Oh, R., Shi, P. ve Ahn, J.Y.** (2019). Implementation of frequency-severity association in bms ratemaking, *arXiv preprint arXiv:1903.05851*.
- [26] **Bowers, N., Gerber, H., Hickman, J., Jones, D. ve Nesbitt, C.** (1986). Actuarial Mathematics Society of Actuaries, *Itasca, Ill.*
- [27] **Cragg, J.G.** (1971). Some statistical models for limited dependent variables with application to the demand for durable goods, *Econometrica (pre-1986)*, 39(5), 829.
- [28] **Frees, E.W.** (2009). *Regression modeling with actuarial and financial applications*, Cambridge University Press.
- [29] **Blough, D.K., Madden, C.W. ve Hornbrook, M.C.** (1999). Modeling risk using generalized linear models, *Journal of health economics*, 18(2), 153–171.

- [30] **Ohlsson, E. ve Johansson, B.** (2010). *Non-life insurance pricing with generalized linear models*, cilt 2, Springer.
- [31] **Gill, J. ve Torres, M.** (2019). *Generalized linear models: a unified approach*, cilt134, Sage Publications, Incorporated.
- [32] **McCullagh, P.** (2018). *Generalized linear models*, Routledge.
- [33] **Karaca, O.** (2010). Genelleştirilmiş lineer modellerde model seçimi üzerine alternatif bir yaklaşım, Master's thesis, Marmara Üniversitesi, İstanbul.
- [34] **Wedderburn, R.W.** (1974). Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method, *Biometrika*, 61(3), 439–447.
- [35] **Flynn, M. ve Francis, L.A.** (2009). More flexible GLMs zero-inflated models and hybrid models, *Casualty Actuarial Soc*, 2009, 148–224.
- [36] **Yip, K.C. ve Yau, K.K.** (2005). On modeling claim frequency data in general insurance with extra zeros, *Insurance: Mathematics and Economics*, 36(2), 153–163.
- [37] **Johansson, A.**, (2014), A comparison of regression models for count data in third party automobile insurance.
- [38] **Denuit, M. ve Charpentier, A.** (2005). *Mathématiques de l'Assurance Non-Vie. Tome II: Tarification et Provisionnement*.
- [39] **Hilbe, J.M.** (2011). *Negative binomial regression*, Cambridge University Press.
- [40] **McCullagh, P. ve Nelder, J.** (1989). *Generalized linear models*, 2nd edn.(Chapman and Hall: London), *Standard book on generalized linear models*.
- [41] **Cameron, A.C. ve Trivedi, P.K.** (2013). *Regression analysis of count data*, cilt 53, Cambridge university press.
- [42] **Gan, G. ve Valdez, E.A.** (2018). Regression modeling for the valuation of large variable annuity portfolios, *North American Actuarial Journal*, 22(1), 40–54.
- [43] **Gan, G. ve Valdez, E.A.** (2018). Regression modeling for the valuation of large variable annuity portfolios, *North American Actuarial Journal*, 22(1), 40–54.
- [44] **Frees, E.W. ve Valdez, E.A.** (2008). Hierarchical insurance claims modeling, *Journal of the American Statistical Association*, 103(484), 1457–1469.
- [45] **Dickson, D.C.** (2016). *Insurance risk and ruin*, Cambridge University Press.
- [46] **Kellison, S.G. ve London, R.L.** (2011). *Risk Models and Their Estimation*, Actex Publications.
- [47] **De Jong, P., Heller, G.Z. ve diğerleri** (2008). *Generalized linear models for insurance data*, Cambridge Books.

- [48] **Anastasiade, M.C.** (2018). Estimation of a counterfactual wage distribution using survey data, *Doktora Tezi*, Université de Neuchâtel.
- [49] **Kleiber, C. ve Kotz, S.** (2003). *Statistical size distributions in economics and actuarial sciences*, cilt470, John Wiley & Sons.
- [50] **Frees, E.W., Derrig, R.A. ve Meyers, G.** (2014). *Predictive modeling applications in actuarial science*, cilt 1, Cambridge University Press.
- [51] **Kaas, R., Goovaerts, M. ve Denuit, M.,** (2005), Actuarial Theory for Dependent Risks.
- [52] **Yang, L., Frees, E.W. ve Zhang, Z.** (2019). Nonparametric estimation of copula regression models with discrete outcomes, *Journal of the American Statistical Association*, 1–25.
- [53] **Chen, S.X. ve Huang, T.M.** (2007). Nonparametric estimation of copula functions for dependence modelling, *Canadian Journal of Statistics*, 35(2), 265–282.
- [54] **Rosenberg, J.V.** (2003). Non-parametric pricing of multivariate contingent claims, *The Journal of Derivatives*, 10(3), 9–26.
- [55] **Lehmann, E.L.** (1966). Some concepts of dependence, *The Annals of Mathematical Statistics*, 1137–1153.
- [56] **Karadag, D.T.** (2008). Portfolio risk calculation and stochastic portfolio optimization by a copula based approach, Master's thesis, Bogaziçi University Istanbul.
- [57] **Rüschendorf, L.** (2009). On the distributional transform, Sklar's theorem, and the empirical copula process, *Journal of Statistical Planning and Inference*, 139(11), 3921–3927.
- [58] **Nikoloulopoulos, A.K.,** (2013). Copula-based models for multivariate discrete response data, *Copulae in Mathematical and Quantitative Finance*, Springer, s.231–249.
- [59] **Nikoloulopoulos, A.K.** (2013). On the estimation of normal copula discrete regression models using the continuous extension and simulated likelihood, *Journal of Statistical Planning and Inference*, 143(11), 1923–1937.
- [60] **Hofert, M.** (2018). *Elements of copula modeling with R*, Springer.
- [61] **Spreeuw, J.** (2006). Types of dependence and time-dependent association between two lifetimes in single parameter copula models, *Scandinavian Actuarial Journal*, 2006(5), 286–309.
- [62] **Patton, A.J.** (2006). Modelling asymmetric exchange rate dependence, *International economic review*, 47(2), 527–556.
- [63] **Burtschell, X., Gregory, J. ve Laurent, J.P.,** (2005), A comparative analysis of CDO pricing models.

- [64] **Bedoui, R., Braiek, S., Guesmi, K. ve Chevallier, J.** (2019). On the conditional dependence structure between oil, gold and USD exchange rates: Nested copula based GJR-GARCH model, *Energy Economics*, 80, 876–889.
- [65] **Frank, M.J.** (1979). On the simultaneous associativity of  $F(x, y)$  and  $x+y - F(x, y)$ , *Aequationes mathematicae*, 19(1), 194–226.
- [66] **Clayton, D.G.** (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, 65(1), 141–151.
- [67] **Hilbe, J.M.** (2009). *Logistic regression models*, Chapman and hall/CRC.
- [68] **Friedman, J., Hastie, T. ve Tibshirani, R.** (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of statistical software*, 33(1), 1.
- [69] **Hastie, T., Tibshirani, R. ve Friedman, J.** (2009). *The elements of statistical learning: data mining, inference, and prediction*, Springer Science & Business Media.
- [70] **Frees, E.** (2018). Loss Data Analytics, *arXiv preprint arXiv:1808.06718*.
- [71] **Verrall, R., Nielsen, J.P. ve Jessen, A.H.** (2010). Prediction of RBNS and IBNR claims using claim amounts and claim counts, *ASTIN Bulletin: The Journal of the IAA*, 40(2), 871–887.
- [72] **Willmott, C.J. ve Matsuura, K.** (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Climate research*, 30(1), 79–82.
- [73] **Frees, E.W., Shi, P. ve Valdez, E.A.** (2009). Actuarial applications of a hierarchical insurance claims model, *ASTIN Bulletin: The Journal of the IAA*, 39(1), 165–197.



## **EKLER**

**EK A.1** : Kopula Modeli için Log Olabilirlik Fonksiyonu

**EK B.1** :  $Y_1$  Hasar tutarı değişkeni için GB2 Model Sonuçları

**EK C.1** :  $Y_2$  Hasar tutarı değişkeni için GB2 Model Sonuçları

**EK D.1** :  $Y_3$  Hasar tutarı değişkeni için Gamma Model Sonuçları

**EK E.1** :  $Y_4$  Hasar tutarı değişkeni için Gamma Model Sonuçları





### EK A.1

$f_{Y_{ij}}(y), i = 1, \dots, k, j = 1, \dots, 4$  hasar tutarı için marjinal olasılık fonksiyonu, ve  $F_{Y_{ij}}(y)$  dağılım fonksiyonu olarak tanımlanmıştır. IFM log-olabilirlik fonksiyonu aşağıda verilmiştir.

$$\begin{aligned} \log L(\boldsymbol{\Sigma}) = & \sum_{i=1}^k \log c(F_{Y_{i3}}, F_{Y_{i4}}; \boldsymbol{\Sigma}_{3,4}) \cdot I(y_{i1} = 0, y_{i2} = 0, y_{i3} > 0, y_{i4} > 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i2}}, F_{Y_{i4}}; \boldsymbol{\Sigma}_{2,4}) \cdot I(y_{i1} = 0, y_{i2} > 0, y_{i3} = 0, y_{i4} > 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i2}}, F_{Y_{i3}}; \boldsymbol{\Sigma}_{2,3}) \cdot I(y_{i1} = 0, y_{i2} > 0, y_{i3} > 0, y_{i4} = 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i2}}, F_{Y_{i3}}, F_{Y_{i4}}; \boldsymbol{\Sigma}_{2,3,4}) \cdot I(y_{i1} = 0, y_{i2} > 0, y_{i3} > 0, y_{i4} > 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i1}}, F_{Y_{i4}}; \boldsymbol{\Sigma}_{1,4}) \cdot I(y_{i1} > 0, y_{i2} = 0, y_{i3} = 0, y_{i4} > 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i1}}, F_{Y_{i3}}; \boldsymbol{\Sigma}_{1,3}) \cdot I(y_{i1} > 0, y_{i2} = 0, y_{i3} > 0, y_{i4} = 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i1}}, F_{Y_{i3}}, F_{Y_{i4}}; \boldsymbol{\Sigma}_{1,3,4}) \cdot I(y_{i1} > 0, y_{i2} = 0, y_{i3} > 0, y_{i4} > 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i1}}, F_{Y_{i2}}; \boldsymbol{\Sigma}_{1,2}) \cdot I(y_{i1} > 0, y_{i2} > 0, y_{i3} = 0, y_{i4} = 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i1}}, F_{Y_{i2}}, F_{Y_{i4}}; \boldsymbol{\Sigma}_{1,2,4}) \cdot I(y_{i1} > 0, y_{i2} > 0, y_{i3} = 0, y_{i4} > 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i1}}, F_{Y_{i2}}, F_{Y_{i3}}; \boldsymbol{\Sigma}_{1,2,3}) \cdot I(y_{i1} > 0, y_{i2} > 0, y_{i3} > 0, y_{i4} = 0) \\ & + \sum_{i=1}^k \log c(F_{Y_{i1}}, F_{Y_{i2}}, F_{Y_{i3}}, F_{Y_{i4}}; \boldsymbol{\Sigma}) \cdot I(y_{i1} > 0, y_{i2} > 0, y_{i3} > 0, y_{i4} > 0) \end{aligned}$$

$k$  gözlem sayısı ve  $\boldsymbol{\Sigma}_{1,2}$ ,  $\boldsymbol{\Sigma}$  matrisinin birinci ve ikinci sütunlarına karşılık gelen alt matristir.





## EK B.1

Değişken	Tahmin	Standart Hata
Sabit terim	1.99612	0.01000 *
Yaş	1.01032	0.00040
Erkek	0.19167	0,00798 *
Evli	0.02744	0,00342
Boşanmış	0.11219	0,00957 *
Diğer	0.04561	0.01094 *
Exposure	0.00066	0.00003
Araç yaşı	0.03555	0.00002 *
Maks hız	0.00002	0.00862
Hızlanma süresi	0.00353	0.00004 *
Uzunluk	0.00300	0.00040 *
Genişlik	0.00003	0.00036
Yükseklik	0.00020	0.00038
Tork	0.00313	0.00027 *
Beygir gücü	0.00246	0.00043 *
Motor hacmi	0.00001	0.00001
Kerb ağırlığı	0.00008	0.00006
B sınıfı	0.00003	0.00021
C sınıfı	0.00001	0.00089
D sınıfı	0.00013	0.00005
E sınıfı	0.00008	0.00026
F sınıfı	0.00051	0.00324
G sınıfı	0.00007	0.00005
H sınıfı	0.00009	0.00139
Diğer sınıf	0.00000	0.00030

Şekil B.1 :  $Y_1$  Hasar tutarı değişkeni için GB2 Model Sonuçları



## EK C.1

Değişken	Tahmin	Standart Hata
Sabit terim	10.064	0.36500 *
Yaş	0.03413	0.17621
Erkek	0.25876	0.23504
Evli	0.03445	0.00797 *
Boşanmış	-0.05146	0.07976
Diğer	-0.11502	0.08920
Exposure	0.00848	0.00576 *
Araç yaşı	-0.01201	0.03075
Maks hız	-0.00290	0.00020 *
Hızlanma süresi	0.00098	0.00043
Uzunluk	-0.00637	0.00060 *
Genişlik	-0.00217	0.00078
Yükseklik	0.00667	0.00330
Tork	-0.00193	0.00015 *
Beygir gücü	0.00273	0.00034 *
Motor hacmi	0.52473	0.21364
Kerb ağırlığı	-0.07044	0.23125
B sınıfı	0.08552	0.26524
C sınıfı	0.10857	0.38061
D sınıfı	-0.02410	0.53641
E sınıfı	0.03506	0.06266
F sınıfı	-0.10521	0.33884
G sınıfı	0.04372	0.31388
H sınıfı	0.00378	0.46418
Diğer sınıf	-0.02690	0.10409

Şekil C.1 :  $Y_2$  Hasar tutarı değişkeni için GB2 Model Sonuçları



## EK D.1

Değişken	Tahmin	Standart Hata
Sabit terim	27.170	2.34101 *
Yaş	-0.00837	0.01181
Erkek	-0.02897	0.25580
Evli	-0.07612	0.28740
Boşanmış	-0.13270	0.01112 *
Diğer	-0.23170	0.01136
Exposure	0.01006	0.00788 *
Araç yaşı	0.12990	0.04307 *
Maks hız	0.03942	0.00029 *
Hızlanma süresi	0.01094	0.05072
Uzunluk	-0.02205	0.09471
Genişlik	0.09298	0.00115 *
Yükseklik	0.00779	0.00401
Tork	0.00044	0.00312
Beygir gücü	0.01161	0.00486
Motor hacmi	0.00673	0.00312
Kerb ağırlığı	0.05671	0.00048 *
B sınıfı	0.12560	0.04614 *
C sınıfı	0.14720	0.04788 *
D sınıfı	0.48900	0.05177 *
E sınıfı	0.08780	0.61830
F sınıfı	0.05224	0.92320
G sınıfı	1.05505	0.08779
H sınıfı	0.86501	0.07934
Diğer sınıf	0.43910	0.01755

Şekil D.1 :  $Y_3$  Hasar tutarı değişkeni için Gamma Model Sonuçları



**EK E.1**

<b>Değişken</b>	<b>Tahmin</b>	<b>Standart Hata</b>
Sabit terim	15.390	0.24602 *
Yaş	0.00738	0.47870
Erkek	0.00340	0.25753
Evli	-0.00210	0.07118
Boşanmış	-0.03811	0.03277
Diğer	0.78634	0.08871
Exposure	0.00502	0.02130
Araç yaşı	0.65020	0.04630 *
Maks hız	0.10006	0.01839 *
Hızlanma süresi	-0.06670	0.00595 *
Uzunluk	-0.01368	0.00093
Genişlik	0.00060	0.00874
Yükseklik	-0.02650	0.00126
Tork	0.00080	0.06547
Beygir gücü	0.00920	0.98310
Motor hacmi	0.10065	0.00190
Kerb ağırlığı	0.01012	0.00710 *
B sınıfı	0.00259	0.00892 *
C sınıfı	0.00001	0.00006
D sınıfı	0.00032	0.76108
E sınıfı	0.00432	0.00031 *
F sınıfı	0.08200	0.20910
G sınıfı	0.91006	0.80207
H sınıfı	0.04991	0.92460
Diğer sınıf	0.00740	0.06150

**Şekil E.1** :  $Y_4$  Hasar tutarı değişkeni için Gamma Model Sonuçları





## ÖZGEÇMİŞ

**Ad Soyad:** Emine Selin Sarıdaş

**E-Posta:** selin.saridas@msgsu.edu.tr

### ÖĞRENİM DURUMU:

- **Lisans:** Hacettepe Üniversitesi Aktüerya Bilimleri Bölümü
- **Y. Lisans:** Hacettepe Üniversitesi Aktüerya Bilimleri Bölümü

### MESLEKİ DENEYİM

- Mimar Sinan Güzel Sanatlar Üniversitesi, İstatistik Bölümü, Araştırma Görevlisi
- Marmara Üniversitesi, Aktüerya Bölümü, Araştırma Görevlisi

### TEZDEN ÜRETİLEN BİLDİRİLER

- Sarıdaş E. Selin, 2019, "Dependent Claim Severity Models Under Copula Regression Approach", Perspectives on Actuarial Risks in Talks of Young Researchers, Nisan 14-19, 2019, Sibiu, Romanya.