

**BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK  
BİLEŞİK MODELLEMESİ**

**PARAMETRIC JOINT MODELLING OF  
LONGITUDINAL AND SURVIVAL DATA**

**ELİF DİL**

**PROF. DR. DURU KARASOY**

**Tez Danışmanı**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı için Öngördüğü

**YÜKSEK LİSANS TEZİ**

olarak hazırlanmıştır.

2019



ELİF DİL'in hazırladığı "Boylamsal ve Yaşam Verilerinin Parametrik Bileşik Modellemesi" adlı bu çalışma aşağıdaki juri tarafından İSTATİSTİK ANABİLİM DALI'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Prof. Dr. Yasemin YAVUZ  
Başkan

Prof. Dr. Duru KARASOY  
Danışman

Doç. Dr. Nihal Ata TUTKUN  
Üye

Doç. Dr. Nursel KOYUNCU  
Üye

Dr. Öğr. Üyesi Uğur KARABEY  
Üye

Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından YÜKSEK LİSANS TEZİ  
...../...../..... tarihinde onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU  
Fen Bilimleri Enstitüsü Müdürü



## ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına 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 ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

30/01/2019



Elif DİL



## YAYINLAMA ve FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açılma iznini Hacettepe üniversitesine verdiğim bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım ben de kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezimin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*" kapsamında tezimin aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi/ H. Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.
- Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren .... ay ertelenmiştir.
- Tezimin ile ilgili gizlilik kararı verilmiştir.

30/01/2019

Elif DİL





## ÖZET

# BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİ

**Elif DİL**

**Yüksek Lisans, İstatistik Bölümü  
Tez Danışmanı: Prof. Dr. Duru KARASOY**

**Ocak, 2019, 64 Sayfa**

Boylamsal veriler, aynı birimlerden belirli zaman aralıklarında elde edilen tekrarlı ölçümlerden oluşurken, yaşam verileri takip edilen herhangi bir olayın gerçekleşmesine kadar geçen süreden oluşmaktadır. Bu iki tip verinin ayrı analizleri için literatürde farklı yöntemler bulunmaktadır. Ancak birlikte toplanan bu iki veri, aralarında ilişki olduğunda yansız ve etkin sonuçlar elde etmek için birlikte analiz edilmelidir. Bileşik model, bu iki verinin paylaşılmış parametre modeli ile birleştirilmesiyle elde edilmekte ve boylamsal alt model ve yaşam çözümlemesi alt modellerinden oluşmaktadır.

Literatürde sıkça kullanılan standart bileşik model yapısı, boylamsal verilerin doğrusal karma etkili model ve yaşam verilerinin Cox regresyon modelinin paylaşılmış parametre modeliyle birleştirilmesiyle elde edilmektedir. Ancak Cox regresyon modelinin uygulanabilmesi için orantılı tehlikeler varsayımının sağlanması gerekmektedir. Varsayımın sağlanmadığı durumlarda ve yaşam verilerinin bilinen bir dağılıma sahip olduğu durumlarda parametrik regresyon yöntemleri kullanılmalıdır. Bileşik modelleme de orantılı tehlikeler varsayımının sağlanmadığı durumlarda da yaşam çözümlemesi alt modeli parametrik yaşam çözümlemesi modelleri ile yapılmalıdır.

Çalışmada standart bileşik model, iki sürecin ayrı analizleri ve Üstel, Weibull, Log-lojistik, Log-normal ve Gamma parametrik alt modelleri ile elde edilen bileşik modeller literatürde yer alan Primer Biliyer Siroz verilerine uygulanmıştır. İlk olarak orantılı

tehlikeler varsayımı test edilmiş ve varsayımın sağlanmadığı görülmüştür. Varsayım sağlanmadığından dolayı parametrik bileşik modeller incelenmiş ve Weibull parametrik alt model ile doğrusal karma etkili modelin bileşik modellemesi en iyi model olarak belirlenmiştir. Standart bileşik model ile Weibull parametrik bileşik model sonuçları karşılaştırıldığında, istatistiksel açıdan önemli farklılıklar bulunmuştur. Boylamsal ve yaşam verilerinin ayrı analizi için doğrusal karma etkili model ve Weibull parametrik model sonuçları incelenmiş ve Weibull parametrik bileşik model ile kıyaslanmıştır. Buna göre, Weibull parametrik model parametrelerinin tehlike oranlarının Weibull parametrik bileşik modelden yüksek olduğu tespit edilmiştir. Weibull parametrik modelde boylamsal gözlem zamana bağlı açıklayıcı değişken olarak alınmış ve yaşam süresine etkileri araştırılmıştır. Analiz sonucunda Weibull parametrik modelin, boylamsal gözlemin yaşam süresine etkisini, Weibull parametrik bileşik modelden daha küçük verdiği gözlemlenmiştir.

**Anahtar Kelimeler:** Yaşam çözümlemesi, Boylamsal veri, Bileşik model, Parametrik bileşik model

## **ABSTRACT**

# **PARAMETRIC JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA**

**Elif DİL**

**Master of Science, Department of Statistics**

**Supervisor: Prof. Dr. Duru KARASOY**

**January, 2019, 64 pages**

Longitudinal data consist of repeated measurements obtained from the same units at certain time intervals, while survival data consists of time until the occurrence of any event under consideration. There are different methods in the literature for separate analysis of longitudinal and survival data. Nevertheless, these two data, particularly collected together in clinical studies, should be analyzed together to obtain unbiased and effective results when there is a relationship between each other. The joint model is obtained by combining these two data with the shared parameter model, and consists of longitudinal sub models and survival sub models.

The standard joint model structure frequently used in the literature is obtained by combining the linear mixed effect model for longitudinal data and shared parameter models of Cox regression model for survival data. However, in order to apply Cox regression model the proportional hazard assumption must be satisfied. Parametric regression methods should be used in cases where the assumption is not provided, and when the survival data has a known distribution. In cases where the assumption of proportional hazard is not provided in joint modelling, the survival analysis sub model should be made with parametric survival analysis models.

In this study, standard joint model, separate analysis of longitudinal and survival data and joint model obtained with Exponential, Weibull, Log-logistic, Log-normal and Gamma parametric sub models have been applied to data set of Primary Biliary

Cirrhosis in the literature. Firstly, the assumption of proportional hazard has been checked and found that the assumption is not provided. Because the assumption is not satisfied, parametric joint models have been examined and the joint modeling of the linear mixed effect model with parametric sub model is determined as the best model. When the standard joint model and Weibull parametric joint model results have been compared, statistically significant differences have been found. For separate analysis of longitudinal and survival data, the results of the linear mixed effect model and the Weibull parametric model have also been investigated and compared with the results of Weibull parametric joint model. Accordingly, the parameters of Weibull parametric model are determined to have higher hazard ratios than the parameters of Weibull parametric joint model. In addition, while Weibull parametric model is established, longitudinal observation have been considered as the independent variable but dependent on time and its effect on survival time has been investigated. At the end of the analysis, it has been observed that the effect of Weibull parametric model on the survival times of longitudinal observation is smaller than the Weibull parametric joint model.

**Keywords:** Survival analysis, Longitudinal data, Joint model, Parametric joint model

## TEŐEKKÜR

Tezimin oluŐmasında en büyük desteęi veren, alıŐmalarımın her aŐamasında deęerli bilgileri ve yardımlarıyla bana yol gÖsteren, sabırla ve büyük bir ilgiyle her zaman beni teŐvik eden ve bana yol gÖsteren danıŐmanım Sayın Duru KARASOY'a; tez alıŐmam boyunca desteklerini esirgemeyen ve tecrübesi ile katkıda bulunan Sayın Do. Dr. Nihan Ata TUTKUN'a; yüksek lisans Öęrenim süresince destekleri ve katkılarıyla her zaman beni motive eden ve destekleyen Sayın Do. Dr. Nursel KOYUNCU'ya; hayatımın her aŐamasında maddi manevi desteklerini esirgemedikleri, cesaret verdikleri, bana güvendikleri ve hep yanımda oldukları için sevgili babam Zekeriya YILDIRIM, annem Sevim YILDIRIM ve abim Emirhan YILDIRIM'a; tez alıŐmam boyunca desteklerini esirgemeyen dayım GüngÖr ÖZYEDİERLER ve yengem Semra ÖZYEDİERLER'e ve yüksek lisans Öęrenim süresince ve akademik hayatımda bana yön gÖsteren, sabırla destekleyen ve yardımlarını esirgemeyen sevgili eŐim Emre DİL'e ve tüm aileme sonsuz teŐekkürlerimi sunarım.

# İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	iii
TEŞEKKÜR.....	v
İÇİNDEKİLER .....	vi
ÇİZELGELER .....	viii
ŞEKİLLER.....	ix
KISALTMALAR .....	x
1. GİRİŞ.....	1
2. BOYLAMSAL VERİ .....	7
2.1. Genelleştirilmiş Tahmin Denklemleri.....	7
2.2. Doğrusal Karma Etkili Model .....	8
2.2.1. Doğrusal Karma Etkili Modellerde Parametre Tahmini.....	9
3. YAŞAM ÇÖZÜMLEMESİ .....	11
3.1. Cox Regresyon Modeli.....	13
3.2. Parametrik Modeller .....	15
3.2.1. Üstel Dağılım .....	15
3.2.2. Weibull Dağılımı .....	15
3.2.3. Log-lojistik Dağılım .....	16
3.2.4. Log-normal Dağılım .....	16
3.2.5. Gamma Dağılımı .....	17
4. BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİ.....	18
4.1. Yaşam Çözümlemesi Alt Modeli.....	19
4.2. Parametrik Yaşam Çözümlemesi Alt Modelleri.....	20
4.2.1. Weibull Yaşam Çözümlemesi Alt Modeli.....	21
4.2.2. Üstel Yaşam Çözümlemesi Alt Modeli .....	22

4.2.3. Log-lojistik Yaşam Çözümlemesi Alt Modeli .....	22
4.2.4. Log-normal Yaşam Çözümlemesi Alt Modeli .....	23
4.3. Boylamsal Alt Model .....	24
4.4. Bileşik Model Parametre Tahmini .....	25
4.4.1. İki Aşamalı Yaklaşım .....	25
4.4.2. Bileşik Olabilirlik Yöntemi.....	26
4.4.2.1. EM Algoritması .....	28
4.4.2.2. İntegrasyon Yaklaşımı.....	31
4.4. Bileşik Model Seçim Kriterleri.....	34
5. UYGULAMA .....	36
5.1. Boylamsal ve Yaşam Verilerinin Bileşik Çözümlemesi .....	40
5.2. Boylamsal ve Yaşam Verilerinin Ayrı Çözümlemesi.....	50
6. SONUÇLAR .....	52
KAYNAKLAR .....	54
EKLER .....	64
EK 1 - Tezden Türetilmiş Bildiriler .....	64

## ÇİZELGELER

<b>Çizelge 5.1.</b> Kullanılan açıklayıcı değişkenler ve düzeyleri. ....	38
<b>Çizelge 5.2.</b> Değişkenlerin orantılı tehlikeler varsayımı için Schoenfeld artıkları ile yaşam süresinin rankı arasındaki ilişki sonuçları. ....	42
<b>Çizelge 5.3.</b> Standart bileşik model sonuçları. ....	43
<b>Çizelge 5.4.</b> Üstel parametrik bileşik model sonuçları.....	44
<b>Çizelge 5.5.</b> Weibull parametrik bileşik model sonuçları. ....	45
<b>Çizelge 5.6.</b> Log-normal parametrik bileşik model sonuçları. ....	46
<b>Çizelge 5.7.</b> Log-lojistik parametrik bileşik model sonuçları. ....	47
<b>Çizelge 5.8.</b> Gamma parametrik bileşik model sonuçları. ....	48
<b>Çizelge 5.9.</b> Parametrik bileşik modeller için AIC ve BIC değerleri. ....	49
<b>Çizelge 5.10.</b> Doğrusal karma etkili model sonuçları.....	50
<b>Çizelge 5.11.</b> Weibull parametrik model sonuçları.....	51



## ŞEKİLLER

Şekil 5.1. Serum bilirubin değerlerinin histogram grafiği.....	38
Şekil 5.2. İlaç değişkeninin Kaplan-Meier ve boylamsal ölçüm grafikleri.....	39
Şekil 5.3. Log(serBilir) değerlerinin zamansal grafikleri .....	40
Şekil 5.4. Değişkenlerin orantılı tehlikeler varsayımı için grafikler .....	41

## KISALTMALAR

DKEM	Doğrusal Karma Etkili Model
GTD	Genelleştirilmiş Tahmin Denklemleri
EÇO	En Çok Olabilirlik
EÇOT	En Çok Olabilirlik Tahminleri
EKK	En Küçük Kareler
REÇOT	Kısıtlanmış En Çok Olabilirlik Tahminleri
KO	Kısmi Olabilirlik
EM	Beklenti Maksimizasyonu
LRT	Olabilirlik Oran Testi
AIC	Akaike Bilgi Kriteri
BIC	Bayesci Bilgi Kriteri
PBS	Primer Biliyer Siroz

# 1. GİRİŞ

Yaşam verileri, takip edilen herhangi bir olayın gerçekleşmesine kadar geçen süreden oluşurken, boylamsal veriler aynı birimlere ait özelliklerin zaman içerisinde tekrarlı olarak ölçülmesinden elde edilmektedir.

Boylamsal ve yaşam süresi verilerini ayrı ayrı analiz eden birçok popüler yöntem bulunmaktadır (Guo ve Carlin, 2004). Yaşam süresi verilerini analiz etmek için yaygın olarak kullanılan yöntemler; yarı parametrik bir yöntem olan Cox regresyon modeli ve parametrik regresyon yöntemleridir. Cox regresyon modeli kullanılmadan önce orantılı tehlikeler varsayımının sağlanıp sağlanmadığı kontrol edilmelidir. Varsayımın sağlanmadığı durumlarda geliştirilmiş Cox regresyon, tabakalandırılmış Cox regresyon veya parametrik regresyon yöntemleri kullanılmalıdır.

Boylamsal verilerin çözümlemesinde ise popüler olarak kullanılan yöntemler; Doğrusal karma etkili model (DKEM - Linear Mixed Effect Models) ve geliştirilmiş tahmin denklemleridir (GTD - Generalized Estimating Equations) (Mondal, 2017).

Boylamsal ve yaşam süresi verilerinin ilişkili olduğu durumlarda ayrı analiz yapmak uygun olmayabilir (Guo ve Carlin, 2004). Bu nedenle, bazı araştırmacılar alternatif olarak bu iki tip verinin bileşik modellemesini önermişlerdir.

Bileşik modelleme genellikle, tekrarlı ölçümlerden oluşan boylamsal verileri ve yaşam süresini içeren yaşam verilerini birleştirmek için kullanılmaktadır (Wulfsohn ve Tsiatis, 1997). Bileşik modellemenin temel amacı, bağımsız değişkenlerin boylamsal ve yaşam verileri üzerine etkilerini incelemektir. Bu modelleme yaklaşımı biyomedikal araştırmalarda sıklıkla kullanılmaktadır. Özellikle klinik çalışmalarda, boylamsal gözlemler yaşam süresi verileriyle birlikte elde edilmektedir. Örneğin; literatürde bileşik modelleme uygulaması olarak sıklıkla kullanılan HIV çalışmasında, hastalığın göstergesi olarak kullanılan CD4 hücre sayıları zaman boyunca tekrarlı olarak ölçülmüş ve AIDS hastalığı olana kadar geçen süre veya her birim için ölüm gözlemlenene kadar geçen süre incelenmiştir (Wulfsohn ve Tsiatis, 1997; Guo ve Carlin, 2004; Wu ve ark., 2012). Kanser çalışmaları da bileşik model literatüründe sıklıkla kullanılmıştır (Pauler ve Finkelstein, 2002; Lima ve Taylor, 2009). Prostat kanseri çalışmalarında, prostata ait

özel değerlerin zaman içerisinde sürekli ölçülmesiyle elde edilen ölçümler ile hastalığın yeniden nüksetmesi durumu birlikte modellenmiştir (Pauler ve Finkelstein, 2002).

Tek değişkenli yaşam ve boylamsal veri için temel bileşik model üzerine ilk çalışmalar Self ve Pawitan (1992), DeGruttola ve Tu (1994), Tsiatis, DeGruttola ve Wulfsohn (1995), Faucett ve Thomas (1996) ve Wulfsohn ve Tsiatis (1997) tarafından yapılmıştır. Self ve Pawitan (1992), bileşik modelin parametre tahminlerini elde etmek için iki aşamalı yaklaşımı önermişlerdir. Birinci aşamada rastgele etkileri tahmin etmek için en küçük kareler yöntemini kullanmışlar ve ikinci aşamada bu tahminleri Cox regresyon modelinin kısmi olabilirlik fonksiyonunda kullanarak, boylamsal ölçümlerin tahminlerini elde etmişlerdir. DeGruttola ve Tu (1994) yaşam süresini parametrik olarak modelleyen basit olabilirlik çıkarımına dayanan bileşik modellemeyi incelemişlerdir. Tsiatis, DeGruttola ve Wulfsohn (1995), Self ve Pawitan'ın (1992) önerdiği iki aşamalı yaklaşımdan farklı olan iki aşamalı yaklaşımı önermişlerdir. Boylamsal ölçümler için normal dağılmış hatalara sahip bir model varsaymışlar ve EM algoritmasıyla modeli tahmin etmişlerdir. Daha sonra bu tahminleri, yaşam verilerinin parametrelerini tahmin etmek için orantılı tehlikeler modelleri ve Cox regresyon modelinde kullanmışlardır. Faucett ve Thomas (1996) sadece gözlemlenmiş verilerden oluşan tüm modelin bilinmeyen parametrelerinin bileşik sonsal dağılımını elde etmek için Markov zincirleri, Gibbs algoritmasını kullanmışlardır. Wulfsohn ve Tsiatis (1997) Cox regresyon modeli ve boylamsal süreç için doğrusal karma etkili modelden oluşan bileşik model için tam olabilirlik yaklaşımını önermişlerdir.

Henderson, Diggle ve Dobson (2000) boylamsal ölçümler ve yaşam süreleri için sıralı korelasyon ve rastgele etkiler içeren iki durağan Gauss sürecini benimseyen genel bir model oluşturmuşlardır. Sıralı korelasyon süreci, eğilimin (trend) zamanla değişmesine olanak tanır ve süreçteki düzensiz eğilimle ilgili olarak gelişen biyolojik dalgalanmalardan kaynaklandığı düşünülebilecek birimler arası korelasyon yapısına izin verir (Henderson, Diggle ve Dobson, 2000; Wang ve Taylor, 2001).

Tsiatis ve Davidian (2001) rastgele etkilerin üzerinde dağılım varsayımı gerektirmeyen bileşik model parametrelerinin tahmin edilmesi için basit bir model önermişlerdir. Koşullu skor yaklaşımı olarak bilinen bu yaklaşımda, rastgele etkiler sorunlu parametreler olarak kabul edilir ve uygun bir yeterli istatistik üzerinde belirlenerek

bileşik model için yarı parametrik tahminler elde edilir. Bu modelin parametrik modellerden daha az kullanışlı olduğu ifade edilmektedir (San, 2013).

Song, Davidian ve Tsiatis (2002) bileşik modelin tahmininde kullanılan rastgele etkilerin normallik varsayımının ihlal edildiği durumlarda parametre tahminleri için olabilirlik temelli bir yaklaşım önermişlerdir. Yaklaşımı EM algoritması kullanarak HIV klinik verisi üzerinde ve simülasyonlar ile test etmişlerdir. Dobson ve Henderson (2003) bileşik model varsayımlarını test edebilmek için koşullu artıkları kullanmışlardır.

Tseng, Hsieh ve Wang (2005) hızlandırılmış başarısızlık süresi ve ölçüm hataları içeren boylamsal veriler varlığında, bileşik modelleme olabilirlik fonksiyonunun maksimize edilmesi üzerine çalışmışlardır. Modeldeki bilinmeyen parametrelerin tahminini elde etmek için Monte Carlo EM algoritmasını kullanmışlar ve bu prosedürün performansını simülasyonlar ve gerçek veri üzerinde incelemişlerdir. Ayrıca, modelin yanlış tanımlanmasına karşılık sağlamlık elde etmek için, standart hataların tahminlerinin sıkıştırılmış (sandwich) tahmin ediciler ile bulunmasını önermişlerdir.

Hsieh, Tseng ve Wang (2006) bileşik modelin olabilirlik tahminlerinden elde edilen standart hataların güvenilir tahminlerini elde etmek için bootstrap yaklaşımını önermişlerdir. Rizopoulos, Verbeke ve Molenberghs (2008) bileşik modellemede rastgele etkilerin dağılımının yanlış tanımlanmasındaki etkisini araştırmışlardır. Araştırma sonucu, belirli tahmin ediciler için birim başına boylamsal ölçümlerin artmasıyla, rastgele etkilerin dağılımının yanlış tanımlanmasının çıkarım prosedürü üzerinde daha az etkiye sahip olacağını göstermişlerdir.

Birçok klinik çalışmada, hastalar takip süresi içinde birden fazla olay yaşayabilir ve bu tip olaylar yarışan riskler olarak adlandırılmaktadır. Elashoff, Li ve Li (2008), Williamson ve ark. (2008) ve Hu, Li ve Li (2009) yarışan riskler ve çoklu başarısızlık durumunda bileşik modelleme üzerine çalışmalar yapmışlardır. Li ve ark. (2009) yarışan riskler ve boylamsal sıralı ölçümlerden oluşan bileşik modellemeyi önermişlerdir. Boylamsal sıralı ölçümler için kısmi orantılı odds modellerini gizli rastgele değişkenler ile yaşam sürelerine bağlamışlardır. Kısmi orantılı odds modeli, sıralı gözlemler için kullanılan popüler orantılı odds modelinin genişletilmiş versiyonudur. Oluşturdukları bileşik modelin parametre tahminlerini elde etmek için EM algoritmasını kullanmışlardır. Rizopoulos (2010) boylamsal ve yaşam verilerinin bileşik model tahminlerini elde etmek için R paket programında “JM” paketini

geliştirmiştir. Rizopoulos, Verbeke ve Molenberghs (2010) model varsayımlarını test etmek için çoklu etki artıklarını önermişlerdir. Albert ve Shih (2010) çoklu boylamsal gözlemler ve kesikli yaşam süresi verilerinin bileşik modellemesi için yeni bir yaklaşım önermişlerdir. Bu yaklaşım, olay süresinde koşullanan ikili doğrusal karma model ile boylamsal gözlemlerin tüm ikili kombinasyonlarını modelleyerek, olay süresindeki çoklu boylamsal gözlemlerin koşullu dağılımlarını tahmin etmeyi amaçlamaktadır. Sweeting ve Thompson (2011) iki aşamalı yaklaşım ile paylaşılmış rastgele parametre modelini karşılaştırmışlardır. Çalışma sonucunda, iki aşamalı yaklaşımın boylamsal ve yaşam süresi verileri arasındaki ilişkiyi yanlış tahmin ettiğini bulmuşlardır.

Su ve Wang (2012) soldan kesilmiş (left truncation) yaşam verileri ve boylamsal gözlemlerin bileşik modellemesini incelemişlerdir. Literatürde yer alan standart bileşik modelleme olabirlik yaklaşımlarının, durdurulmuş gözlemlerin olduğu yaşam verileri ve boylamsal verilerin bileşik modelleme tahminlerinde yansız ve etkin sonuçlar verse de soldan kesilmiş verilerin olması durumunda aynı performansı göstermediğini açıklamışlar ve bu sorunu ortadan kaldırmak için yeni bir olabirlik yaklaşımı önermişlerdir. Önerdikleri yaklaşımın yansız ve etkin tahminler verdiğini, AIDS kohort çalışması ve simülasyonlar ile göstermişlerdir. Lima ve ark. (2012) parametre tahminleri için bileşik modellemede gizli sınıf (latent class) durumunu incelemiş ve bileşik gizli sınıf modelini önermişlerdir. Bileşik gizli sınıf modeli, boylamsal bağımsız değişkenler ve olay riski arasındaki ilişkiyi tamamen yakaladığını varsayarak elde edilir. Önerdikleri modeli, paylaşılmış rastgele etkili modelle karşılaştırmışlar ve modelin performansını değerlendirmişlerdir. Crowther ve ark. (2012) bileşik modelin yaşam çözümlemesi kısmı için, kısıtlanmış kübik splineleri kullanarak logaritmik temel tehlike modelleri için esnek parametrik yaklaşımı tanımlamışlardır.

Efendi ve ark. (2013) sürekli ve ikili boylamsal gözlemler ile yaşam süresi verilerini birleştiren bir model önermişlerdir. Liu ve Li (2014) çok değişkenli boylamsal gözlemler ile yaşam verilerinin bileşik model tahmininde Bayesci bileşik model tahmin edicilerini kullanmışlardır. Tang ve Tang (2015) çok değişkenli yaşam süresi ve boylamsal verileri için yarı parametrik çok değişkenli çarpık-normal (skew-normal) bileşik modeli önermişlerdir.

Hickey ve ark. (2016) bileşik modelleme ile literatürde geliştirilen yöntemleri, çok değişkenli boylamsal gözlemler içeren bileşik modelleme yöntemlerini, model

varsayımlarını, birleştirme ve parametre tahmin yöntemlerini ve bileşik modellemenin yazılım araçlarını özetleyen bir çalışma yapmışlardır. Rizopoulos (2016) bileşik model parametre tahmini için Monte Carlo algoritmasını kullanarak bayesci tahminleri elde eden, R programında bulunan “JMBayes” paketini geliştirmiştir.

Hickey ve ark. (2018) birden fazla boylamsal gözlemleri içeren bileşik modelin parametre tahminlerini elde etmek için, R paket programında “JoineRML” paket programını geliştirmişlerdir.

Bileşik modelleme üzerine Türkiye’de sınırlı sayıda çalışma bulunmaktadır. Ülgen ve Asar (2017) bileşik modelleme için eğitsel ve derleme niteliğindeki çalışmalarında, boylamsal alt model için doğrusal karma etkili model ve yaşam verileri için genişletilmiş Cox regresyon modelini PAQUID verisine uygulamışlar ve bileşik model tahminlerini, doğrusal karma etkili model ve genelleştirilmiş Cox regresyon modelinin ayrı analizleri ile kıyaslamışlardır. Turgal (2016) bileşik modelin yaşam çözümlemesi kısmı için Cox regresyon modeli ve boylamsal kısım için doğrusal karma etkili modeli kullanmıştır. Turgal, Erdoğan ve Köse (2017) simülasyon çalışmaları ile zamana bağlı değişkenlerin çeşitli örneklem büyüklüklerinin ve varyans-kovaryans yapılarının performanslarını incelemişlerdir.

Çalışmanın ikinci bölümünde boylamsal veri ve bu verilerin analizi hakkında genel bilgiler verilmiştir. Boylamsal veri analizinde parametrik regresyon yöntemi olarak sıklıkla kullanılan GTD ve DKEM incelenmiştir. Bileşik modellemenin boylamsal alt modelinde kullanılan doğrusal karma etkili modelin parametre tahmin yöntemlerine değinilmiştir.

Çalışmanın üçüncü bölümünde yaşam çözümlemesi detaylı şekilde incelenmiştir. Yarı parametrik regresyon yöntemi olan Cox regresyon yöntemi ve yaşam çözümlemesinde kullanılan parametrik yaşam çözümlemesi yöntemleri tanıtılmıştır.

Çalışmanın dördüncü bölümünde ise boylamsal ve yaşam çözümlemesi sürecini paylaşılmış parametre modelleri ile bağlayan bileşik model ayrıntılı olarak incelenmiştir. İlk olarak boylamsal alt model için doğrusal karma etkili model ve yaşam çözümlemesi için Cox regresyon modelinin kullanıldığı standart bileşik model yapısı ele alınmıştır. Daha sonra orantılı tehlikeler varsayımının sağlanmadığı ve yaşam

süresinin belirli bir parametrik dağılım gösterdiği durumlar için parametrik alt modeller ile bileşik modelleme yöntemleri tanıtılmış ve bileşik model parametre tahmin yöntemleri ayrıntılı olarak incelenmiştir.

Beşinci kısımda ise literatürde yer alan Primer Biliyer Siroz (PBS) verisi üzerinde standart bileşik model, parametrik bileşik modeller ve iki sürecin ayrı ayrı analizleri yapılarak, sonuçlar yorumlanmıştır.

Bu çalışmanın amacı, boylamsal ve yaşam verilerinin ilişkili olduğu durumlarda kullanılan bileşik modellemeyi incelemek ve literatürde sıklıkla kullanılan standart bileşik modellemenin orantılı tehlikeler varsayımını sağlamadığı durumlarda alternatif olarak kullanılacak parametrik bileşik model yapılarını tanıtmaktır. Bu amaçla veri seti üzerinde standart bileşik modelleme, iki sürecin ayrı analizi ve farklı parametrik bileşik modeller incelenip, yorumlar yapılmıştır.



## 2. BOYLAMSAL VERİ

Panel veri olarak da bilinen boylamsal veriler, aynı birimlere ait özelliklerin belirli zaman aralıklarında tekrarlı olarak ölçülmesi ile elde edilmektedir (Sousa, 2011). Birimler; haneler, kuruluşlar ve bireylerden oluşabilir. Boylamsal veriler farklı zaman noktalarında tekrarlı ölçümlerden elde edildiğinden dolayı, zaman serileri ve klasik regresyon analizinden ziyade özel istatistiksel yöntemlerle analiz edilmelidir (Fitzmaurice, Laird ve Ware, 2011).

Boylamsal çalışmalarda kayıp veri oldukça yaygındır (Ibrahim ve Molenberghs, 2009). Örneğin; çalışmanın tüm birimleri belirli bir zamanda çalışmaya katılmayabilir veya bazı birimler çalışma sona ermeden önce çalışmayı bırakabilirler. Tüm birimler aynı zaman noktasında aynı tekrarlı ölçümlere sahip olamayacağından dolayı veri dengesiz (unbalanced) olarak elde edilebilir. Bu nedenle dengesiz ve kayıp veri varlığında boylamsal verileri analiz edecek teknikler kullanılmalıdır (Fitzmaurice, Laird ve Ware, 2011).

Boylamsal veri analizinde amaç, bağımlı değişkende zaman içerisindeki değişimi ve bu değişime neden olan etkenleri ortaya çıkartmaktır. Bu veri setinin analizinde, parametrik regresyon, parametrik olmayan regresyon ve yarı parametrik regresyon gibi çeşitli regresyon yöntemleri kullanılmaktadır. En sık kullanılan parametrik yöntemler; GTD ve DKEM'dir (Kalkan, 2014).

### 2.1. Genelleştirilmiş Tahmin Denklemleri

Liang ve Zeger (1986), boylamsal verilerin analizi için DKEM'e alternatif olarak GTD'yi önermişlerdir (Ballinger, 2004). Klinik ve biyomedikal çalışmalarda yaygın olarak kullanılmaktadır (Wang, 2014). GTD sürekli, sıralı, ikili ve sayım verisi yapısına sahip verilerin analizinde kullanılmaktadır (Ghisletta ve Spini, 2004).

GTD ve DKEM dengeli ve dengesiz boylamsal verilerin analizi için kullanılabilir (Kalkan, 2014). Ancak, bu iki yöntem model tahminlerinde farklılık göstermektedir. DKEM, aynı birimlerin gözlemleri arasında korelasyonu belirlemek için rastgele etkileri kullanarak bireysel bir yaklaşım sergilerken, GTD

parametrelerin tahminleri için kitle temelli yarı olabilirlik yöntemlerini kullanmaktadır (Wedderburn, 1974; Crowder, 1995).

Tez çalışmasında, bileşik modelin boylamsal kısmı için DKEM modeli kullanılacağından dolayı, DKEM modeli ayrıntılı olarak incelenmiştir.

## 2.2. Doğrusal Karma Etkili Model

Laird ve Ware (1982), birimler arası ve birimler içindeki değişimi içeren boylamsal veriler için doğrusal bir model önermişlerdir. Eksik veri ve denekler arasındaki dengesizlik durumunda model tahminleri yansız ve tutarlı sonuçlar verdiği için en çok tercih edilen yöntemlerden biridir (Hedeker ve Gibbons, 2006; Doğanay, 2007).

Doğrusal karma veya rastgele etkili modelde bağımlı değişken, birimden birime rastgele etkiyle değişen regresyon katsayılarına sahip bağımsız değişkenlerin doğrusal bir formu olarak kabul edilir. Birimler arasındaki bu farklılaşma ölçülmeyen faktörlerden dolayı ortaya çıkmaktadır (Diggle ve ark., 2002). Bağımlı değişken rastgele ve sabit etkiler kullanılarak modellenmektedir (Bates ve ark., 2015).

Değişkenlerin düzeyleri önceden belirlenmişse bu değişken düzeyleri sabit etkiler olarak adlandırılır. Eğer değişkenin düzeyleri daha geniş bir kitleden rastgele olarak seçilmişse bu etkiler rastgele olarak adlandırılmaktadır. Karma etkili model hem sabit hem rastgele etkileri içinde barındırmaktadır (Doğanay, 2007). DKEM’de, sabit etkiler tüm birimler için ortak kabul edilir ve kitle etkisini gösterirken, rastgele etkiler birimden birime değişir ve birimlerin kitle değerlerinden sapmalarını göstermektedir (Vittinghoff ve ark., 2005; Mondal, 2017).

Genel olarak DKEM,

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i \quad (2.1)$$

şeklindedir (Harville, 1977; Laird ve Ware, 1982; Verbeke ve Molenberghs, 2000).

Burada;

$y_i = (y_{i1}, y_{i2}, \dots, y_{im_i})'$ , i. birim için tekrarlı ölçümlerden oluşan  $m_i \times 1$  boyutlu vektör,

$\beta$ :  $p \times 1$  boyutlu sabit etkiler vektörü,

$X_i$ :  $m_i \times 1$  boyutlu sabit etkilere karşılık gelen tasarım matrisi,

$b_i$ :  $q \times 1$  boyutlu rastgele etkiler vektörü,

$Z_i$ : rastgele etkiler  $b_i$ 'ye karşılık gelen  $m_i \times q$  boyutlu tasarım matrisi,

$\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{im_i})'$   $m_i \times 1$  boyutlu örneklem veya ölçüm hataları vektörüdür.

Rastgele etkilerin  $b_i \sim N(0, G)$  şeklinde çok değişkenli normal dağılıma sahip olduğu varsayılmaktadır. Burada,  $G$  rastgele etkiler için kovaryans matrisidir. Örneklem hatası  $\varepsilon_i$ 'nin ise rastgele etkilerden bağımsız ( $Cov(b_i, \varepsilon_i) = 0$ ) ve  $\varepsilon_i \sim N(0, R_i)$  şeklinde çok değişkenli normal dağılıma sahip olduğu varsayılmaktadır.  $R_i$  örneklem hatası için kovaryans matrisini göstermektedir (Fitzmaurice, Laird ve Ware, 2011).

### 2.2.1. Doğrusal Karma Etkili Modellerde Parametre Tahmini

DKEM'de parametre tahmini için en çok olabilirlik (EÇO) yöntemi kullanılır (Rizopoulos, 2012a). En çok olabilirlik tahminleri (EÇOT) olabilirlik fonksiyonunun maksimize edilmesi ile elde edilmektedir (Fitzmaurice, Laird ve Ware, 2011). Bir modelde gözlemlerin bağımsız olduğu varsayılırsa, olabilirlik fonksiyonu her birimin yoğunluk fonksiyonu ile basitçe hesaplanabilir (Asimow ve Maxwell, 2015). Ancak, karma modellerdeki gibi tekrarlı ölçümler olduğunda gözlemler bağımsız olmayacağından dolayı ölçümler için bileşik olabilirlik fonksiyonu kullanılmalıdır (Fitzmaurice, Laird ve Ware, 2011).

Eşitlik 2.1'de tekrarlı ölçümler  $y_i$ 'nin bağımsız olduğu varsayılırsa, olabilirlik fonksiyonu  $(\ell(\theta))$  tekrarlı ölçümlerin çok değişkenli normal olasılık yoğunluk fonksiyonlarının toplamı şeklinde yazılabilir.  $\theta$ 'nın EÇOT'sini hesaplamak için aşağıda belirtilen log-olabilirlik fonksiyonunun maksimize edilmesi gerekmektedir (Fitzmaurice, Laird ve Ware, 2011):

$$l(\theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log |V_i| - \frac{1}{2} \left\{ \sum_{i=1}^n (y_i - X_i \beta)' V_i^{-1} (y_i - X_i \beta) \right\} \quad (2.2)$$

Burada  $N = \left( \sum_{i=1}^n m_i \right)$  toplam gözlem sayısını,  $m_i$  i. birim için tüm gözlem sayısını ve

$|V_i|$ , varyans-kovaryans matrisi  $V_i = Z_i G Z_i' + \sigma^2 I_{n_i}$ 'nin determinantını göstermektedir (Rizopoulos, 2012a).

Skor fonksiyonunu (log-olabilirlik fonksiyonunun (Eşitlik 2.2)  $\beta$ 'ya göre türevi) sıfıra eşitleyerek ve denklemini çözerek  $\beta$ 'nin EÇOT elde edilebilir. İlk iki terim  $\beta$ 'yı içermediğinden dolayı log-olabilirlik fonksiyonunu maksimize etmek için bu terimler ihmal edilebilir. Ayrıca üçüncü terim negatif işaretli olduğundan dolayı, log-olabilirlik fonksiyonunun  $\beta$ 'ya göre maksimizasyonu basitçe elde edilebilir ve aşağıdaki gibidir (Fitzmaurice, Laird ve Ware, 2011):

$$\sum_{i=1}^n (y_i - X_i \beta)' V_i^{-1} (y_i - X_i \beta). \quad (2.3)$$

Eşitlik 2.3 minimize edilerek,  $\beta$ 'nin en küçük kareler (EKK) tahmin edicisi

$$\hat{\beta} = \left\{ \sum_{i=1}^n (X_i' V_i^{-1} X_i) \right\}^{-1} \sum_{i=1}^n (X_i' V_i^{-1} y_i) \quad (2.4)$$

biçiminde elde edilir. Bu formül  $V_i$ 'nin bilindiği varsayımı altında uygulanır ancak genellikle  $V_i$  bilinmez ve veriden tahmin edilerek elde edilir (Fitzmaurice, Laird ve Ware, 2011).  $\beta$ 'nin verilen bir değeri için  $\ell(\theta_b, \sigma^2)$  fonksiyonunun maksimize edilmesiyle  $V_i$ 'nin EÇOT'si elde edilebilir (Rizopoulos, 2012a).  $V_i$ 'nin EÇOT'si alındıktan sonra, Eşitlik 2.4'te  $V_i$ 'nin tahmin edicisi yerine  $\hat{V}_i$  yazılarak  $\beta$ 'nin EÇOT'si aşağıdaki gibi elde edilebilir:

$$\hat{\beta} = \left\{ \sum_{i=1}^n (X_i' \hat{V}_i^{-1} X_i) \right\}^{-1} \sum_{i=1}^n (X_i' \hat{V}_i^{-1} y_i) \quad (2.5)$$

$V_i$ 'nin tahmini için kısıtlanmış en çok olabilirlik tahmin edicileri (REÇOT) de kullanılabilir (Harville, 1977; Rizopoulos, 2012a).  $V_i$ 'nin REÇOT tahmininde, olabilirlik  $\beta$ 'yı içermez ve  $V_i$  cinsinden tanımlanmıştır (Fitzmaurice, Laird ve Ware, 2011; Rizopoulos, 2012a).

### 3. YAŞAM ÇÖZÜMLEMESİ

Yaşam çözümlemesi, takip edilen olayın gerçekleşmesine kadar geçen sürenin çözümlemesini içermektedir (Chernick ve Friis, 2003). Süre; yaşam süresi, başarısızlık süresi veya olay süresini ifade etmektedir (Rizopoulos, 2012a). Örneğin, hastaneden çıkana kadar geçen süre, HIV hastalarında ölüm veya AIDS tanısı konulana kadar geçen süre olabilir.

Genellikle olaylarda birimlerin bir kısmı için olay gözlemlenebilir. Diğer olaylar için olayın gerçekleşme zamanı mevcut takip zamanından daha büyüktür ve olay gözlemlenmeyebilir. Bu tip olaylarla yaşam çözümlemesinde sık sık karşılaşılmakta ve bu durumda süre durdurulmuş süre olarak adlandırılmaktadır (Sullivan, 2012). Durdurma temel olarak üç nedenden dolayı görülmektedir (Kleinbaum ve Klein, 2012):

- Çalışmanın sonuna kadar ilgilenilen olayın gözlemlenmemesi,
- Birimlerin, çalışma devam ederken çalışmadan çekilmesi,
- Birimlerin ilgilenilen olayın dışında başka bir olaydan dolayı çalışmadan ayrılması.

Durdurma bilgi içeren (informative) veya bilgi içermeyen (non-informative) şeklinde de ortaya çıkabilir. Bir kişi test sonuçlarının kötüleşmesinden dolayı çalışmadan çıkarsa durdurma bilgi içeren (rastgele olmayan) olarak kabul edilir ve başarısızlık oranı hala çalışmada olanlardan farklı olabilir. Eğer bir kişinin çalışmadan ayrılması test sonuçlarından bağımsızsa durdurma bilgi içermeyen olarak tanımlanır (Rizopoulos, 2012a).

Standart yaşam çözümlemesinde durdurmanın bilgi içermeyen olduğu varsayılır (Allison, 2010). Bu çözümlemede genellikle yaşam sürelerinin dağılımları yaşam fonksiyonu (survival function) ve tehlike fonksiyonu (hazard function) ile gösterilmektedir (Bewick, Cheek ve Ball, 2004).

#### **Yaşam Fonksiyonu**

$T$ ,  $F(t) = P(T \leq t)$  birikimli dağılım fonksiyonu ve  $f(t)$  olasılık yoğunluk fonksiyonu ile yaşam süresi için sürekli rastgele değişkeni gösterebilir. Yaşam fonksiyonu  $S(t)$ , olayın  $t$  zamanından sonra meydana gelme olasılığı olarak tanımlanır ve

$$S(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(x) dx \quad (3.1)$$

biçiminde ifade edilir (Pintilie, 2006). Yaşam fonksiyonunun özellikleri;

- Artmayan bir fonksiyondur
- $t = 0$  iken;  $S(t) = S(0) = 1$
- $t = \infty$  iken;  $S(t) = S(\infty) = 0$

şeklinde (Kleinbaum ve Klein, 2012).  $t$  zamanına karşı  $S(t)$  yaşam olasılıklarının grafiği çizilerek yaşam eğrisi elde edilebilir (Bewick, Cheek ve Ball, 2004).

### Tehlike Fonksiyonu

Tehlike fonksiyonu,  $t$  zamanında yaşayan bir birimin anlık olay hızını göstermektedir ve

$$\begin{aligned} h(t) &= \lim_{\delta t \rightarrow 0} \left\{ \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} \right\}, t > 0 \\ &= \lim_{\delta t \rightarrow 0} \left\{ \frac{P(t \leq T \leq t + \delta t)}{\delta t P(T > t)} \right\} \\ &= \lim_{\delta t \rightarrow 0} \left\{ \frac{F(t + \delta t) - F(t)}{\delta t P(T > t)} \right\} \\ &= \frac{f(t)}{S(t)} \end{aligned} \quad (3.2)$$

biçiminde tanımlanır (Pintilie, 2006). Burada,  $f(t)$  yaşam süresi için  $T$  rastgele değişkeninin olasılık yoğunluk fonksiyonunu göstermektedir.  $H(t)$  birikimli tehlike fonksiyonu,  $t$  zamanına kadar oluşan birikimli tehlike olarak tanımlanabilir (Rizopoulos, 2012a):

$$H(t) = \int_0^t h(u) du \quad (3.3)$$

$S(t)$ ,  $h(t)$  veya  $H(t)$  fonksiyonlarının biri biliniirse, diğer iki fonksiyon,

$$h(t) = -\frac{\delta}{\delta t} \log(S(t))$$

$$H(t) = -\log(S(t))$$

$$S(t) = \exp(-H(t)) = \exp\left\{-\int_0^t h(u) du\right\} \quad (3.4)$$

şeklinde tanımlanır (Pintilie, 2006).

Yaşam çözümlemesi, temel olarak çalışmanın başlangıç noktasından itibaren herhangi bir zamanda meydana gelen başarısızlığın tehlikesi veya riski ile ilgilenir. Bu nedenle çözümlemelerde çoğunlukla tehlike fonksiyonu modellenmektedir. Bu modellemenin iki ana hedefi vardır: 1) Hangi bağımsız (açıklayıcı) değişkenlerin tehlike fonksiyonu üzerinde etkisi olduğunu incelemek, 2) Çalışmadaki birimlerin tehlike fonksiyonlarını incelemektir (Collett, 2003).

Yaşam süresi ve bir veya daha fazla bağımsız değişken (yaş, cinsiyet, ırk gibi) arasındaki ilişkiyi incelemek için yarı parametrik ve parametrik yaşam çözümlemesi modelleri kullanılmaktadır (Fox, 2008).

### 3.1. Cox Regresyon Modeli

Cox (1972) tarafından önerilen Cox regresyon modeli (orantılı tehlikeler modeli), yaşam çözümlemesinde popüler olarak kullanılan yarı parametrik regresyon modelidir (Fox ve Weisberg, 2011). Bu model yaşam süresi ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi incelemektedir (Singh ve Mukhopadhyay, 2011).  $i$ . birim için Cox regresyon modeli;

$$h_i(t | x_i) = h_0(t) \exp(\beta' x_i) \quad (3.5)$$

biçimindedir (Rizopoulos, 2012a). Burada,  $h_0(t)$  bağımsız değişken içermeyen birimin temel tehlike fonksiyonu,  $x_i' = (x_{i1}, \dots, x_{ip})$   $p$  bağımsız değişken vektörünü ve  $\beta$  modeldeki bağımsız değişkenlerin katsayılar vektörünü göstermektedir (Ata, Karasoy ve Sözer, 2008).

Cox regresyon modelinde, temel tehlike fonksiyonu belirlenmemiş olduğundan dolayı yarı parametrik bir modeldir (Fox ve Weisberg, 2011). Model tahmini için temel tehlike fonksiyonun şekli hakkında bir varsayım bulunmamaktadır (Ahmed, Vos ve Holbert, 2007). Cox regresyon modeli, tehlike oranının zaman boyunca sabit olduğunu varsaymaktadır (Bewick, Cheek ve Ball, 2004). Bu nedenle, Cox regresyon modeli uygulanmadan önce “orantılılık” varsayımının kontrol edilmesi gerekmektedir (Persson, 2002).

Cox regresyon modelinin regresyon katsayılarının tahmininde kısmi olabilirlik (Partial likelihood - KO) tahmin edicileri kullanılmaktadır (Cox, 1972; Lewis, 2017).  $r$  tane ayrıklı başarısızlık (olay) sürelerine ve  $n - r$  tanesi durdurulmuş yaşam sürelerine sahip  $n$

birime ait bir veri seti olsun. Her başarısızlık zamanında sadece bir tane olay olduğunu varsayalım. KO yöntemi için sıralanmış başarısızlık sürelerine ihtiyaç duyulmaktadır (Collett, 2003):

$$t_{(1)} < t_{(2)} < \dots < t_{(r)} \quad (3.6)$$

Burada  $t_{(j)}$  j. sıralı başarısızlık süresini göstermektedir.

Cox (1972), Eşitlik (3.5)'de verilen Cox regresyon modeli için kısmi olabilirlik fonksiyonu aşağıdaki gibi formüle etmiştir:

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' x_l)} \quad (3.7)$$

Burada,

$x_{(j)}$ ; j. sıralı başarısızlık süresi  $t_{(j)}$  'de başarısız olan bağımsız değişken vektörü,

$R(t_{(j)})$  (risk kümesi);  $t_{(j)}$  'de riskte olan birimlerin kümesidir.

Olabilirlik fonksiyonunun tahmini, başarısızlık süresine sahip birimler üzerindedir ancak risk altında bulunan birimler olabilirlik fonksiyonun paydasında katkıda bulunurlar (Collett, 2003).

n birim için gözlenen yaşam süresi  $t_1, t_2, \dots, t_n$  ve  $\delta_i$  gösterge değişkeni olsun.  $i = 1, 2, \dots, n$  olmak üzere i. yaşam süresi  $t_i$ , durdurulmuş ise  $\delta_i = 0$ , diğer durumlarda  $\delta_i = 1$  değerini alır (Collett, 2003). Eşitlik (3.7)'deki olabilirlik fonksiyonu,

$$L(\beta) = \prod_{i=1}^n \left\{ \frac{\exp(\beta' x_i)}{\sum_{l \in R(t_i)} \exp(\beta' x_l)} \right\}^{\delta_i} \quad (3.8)$$

biçiminde yazılabilir. Burada  $R(t_i)$   $t_i$  zamanındaki risk kümesini göstermektedir ve bu eşitliğe karşılık gelen log-olabilirlik fonksiyonu aşağıdaki gibi ifade edilmektedir:

$$\log L(\beta) = \sum_{i=1}^n \delta_i \left\{ \beta' x_i - \log \sum_{l \in R(t_i)} \exp(\beta' x_l) \right\} \quad (3.9)$$

Newton-Raphson gibi bazı sayısal maksimizasyon yöntemleri kullanılarak log-olabilirlik fonksiyonundaki  $\beta$  parametrelerinin EÇOT'leri elde edilebilir (Collett, 2003).



### 3.2. Parametrik Modeller

Yaşam süresinin belli bir dağılıma sahip olduğu durumlarda parametrik yöntemlerden yararlanılmalıdır (Collett, 2003; Klein ve Moeschberger, 1997; Hosmer, Lemeshow ve May, 2008). Yaygın olarak kullanılan dağılımlar: Üstel, Weibull, Log-lojistik, Log-normal ve Gamma dağılımıdır.

#### 3.2.1. Üstel Dağılım

Üstel model, yaşam çözümlemesinde kullanılan en basit parametrik yöntemdir ve tehlikenin zaman boyunca sabit olma varsayımına dayanmaktadır. Yaşam süresi  $\lambda$  parametresi ile Üstel dağılıma sahip ise, olasılık yoğunluk fonksiyonu

$$f(t) = \begin{cases} \sigma^{-1} e^{-(\sigma^{-1}t)} & t \geq 0, \sigma^{-1} > 0 \\ 0 & t < 0 \end{cases} \quad (3.10)$$

biçiminde tanımlanır. Üstel dağılımın yaşam fonksiyonu, tehlike fonksiyonu ve birikimli tehlike fonksiyonu sırasıyla aşağıdaki gibi elde edilebilir (Karasoy ve Tutkun, 2016):

$$\begin{aligned} S(t) &= e^{-\sigma^{-1}t} \\ h(t) &= \sigma^{-1}, \quad \sigma^{-1} > 0 \\ H(t) &= \sigma^{-1}t \end{aligned} \quad (3.11)$$

#### 3.2.2. Weibull Dağılımı

Weibull model, ilk olarak Waloddi Weibull (1939) tarafından önerilmiş ve iki pozitif parametre ile Üstel modelin genelleştirilmiş halidir. Modeldeki ikinci parametre tehlike fonksiyonunun farklı şekillerine ve modelin büyük esnekliğini sağlamaktadır. Modelin esnekliği, yaşam ve tehlike fonksiyonlarının basit oluşundan dolayı Weibull model deneysel çalışmalarda uygulama kolaylığı sağlamaktadır (Liu, 2012). Weibull dağılımının tehlike, yaşam, birikimli tehlike ve olasılık yoğunluk fonksiyonları sırasıyla,

$$\begin{aligned} h(t) &= \sigma^{-1} \gamma (\sigma^{-1} t)^{\gamma-1} \\ S(t) &= \exp\left[-(\sigma^{-1} t)^\gamma\right] \\ H(t) &= (\sigma^{-1} t)^\gamma \end{aligned}$$

$$f(t) = \sigma^{-1} \gamma (\sigma^{-1} t)^{\gamma-1} e^{-(\sigma^{-1} t)^\gamma}, \quad t \geq 0, \quad \gamma, \sigma^{-1} > 0 \quad (3.12)$$

şeklinde tanımlanabilir (Karasoy ve Tutkun, 2016).  $\sigma^{-1}$  ve  $\gamma$  sırasıyla dağılımın ölçek ve şekil parametreleridir.  $\gamma=1$  olduğunda Weibull dağılımı Üstel dağılıma dönüşmektedir (McCool, 2012).

### 3.2.3. Log-lojistik Dağılım

Yaşam süresinin logaritması ( $\log(T)$ ) lojistik dağılıma sahip ise, yaşam süresi Log-lojistik dağılıma sahiptir. Log-lojistik dağılımın tehlike, yaşam, birikimli tehlike ve olasılık yoğunluk fonksiyonları sırasıyla aşağıdaki gibidir:

$$\begin{aligned} h(t) &= \alpha \gamma t^{\gamma-1} (1 + \alpha t^\gamma)^{-1} \\ S(t) &= (1 + \alpha t^\gamma)^{-1} \\ H(t) &= \log [1 + \alpha t^\gamma] \\ f(t) &= (\alpha \gamma t^{\gamma-1}) (1 + \alpha t^\gamma)^{-2} \quad t \geq 0, \quad \alpha > 0, \quad \gamma > 0 \end{aligned} \quad (3.13)$$

(Klein ve Moeschberger, 1997; Steensmeier ve Jones, 2004; Hosmer, Lemeshow ve May, 2008).  $\alpha$  ve  $\gamma$  sırasıyla dağılımın ölçek ve şekil parametreleridir. Log-lojistik dağılım önce artan, daha sonra azalan ya da monoton azalan tehlike fonksiyonu için kullanılabilir (Karasoy ve Tutkun, 2016).

### 3.2.4. Log-normal Dağılım

McAlister (1879) tarafından önerilen Log-normal dağılım, logaritması alınan bir değişkenin normal dağılım göstermesi olarak tanımlanabilir (Karasoy ve Tutkun, 2016). Dağılımın tehlike, yaşam, birikimli tehlike ve olasılık yoğunluk fonksiyonları sırasıyla,

$$\begin{aligned} h(t) &= \exp \left\{ -(\log t - \mu)^2 2^{-1} \sigma^{-2} \right\} t^{-1} \sigma^{-1} (2\pi)^{-\frac{1}{2}} \left\{ 1 - \phi \left[ \frac{\log t - \mu}{\sigma} \right] \right\}^{-1} \\ S(t) &= 1 - \phi \left[ (\log t - \mu) \sigma^{-1} \right] \\ H(t) &= \log \left\{ 1 - \phi \left[ (\log t - \mu) \sigma^{-1} \right] \right\} \\ f(t) &= \exp \left\{ -(\log t - \mu)^2 2^{-1} \sigma^{-2} \right\} (t\sigma)^{-1} (2\pi)^{-\frac{1}{2}} \quad t > 0, \quad \sigma > 2 \end{aligned} \quad (3.14)$$

şeklinde tanımlanır (Evans, Hastings ve Peacock, 2000; Hosmer, Lemeshow ve May, 2008).  $\phi(z)$  standart normal dağılımın birikimli dağılım fonksiyonunu göstermektedir (Klein ve Moeschberger, 1997).

### 3.2.5. Gamma Dağılımı

İki parametrelili Gamma dağılımı ( $G(\lambda, \gamma)$ ) Üstel dağılımın özel bir durumundan oluşmaktadır. Gamma modellerinin yaşam ve tehlike fonksiyonları kapalı şekilde ifade edilemediğinden dolayı yaşam çözümlemesinde sınırlı olarak kullanılmaktadır (Liu, 2012). Gamma dağılımının tehlike, yaşam ve olasılık yoğunluk fonksiyonları sırasıyla aşağıdaki gibi tanımlanabilir:

$$h(t) = \frac{\lambda^\gamma t^{\gamma-1} e^{-\lambda t}}{\left( \int_0^\infty \frac{\lambda}{\Gamma(\gamma)} (\lambda x)^{\gamma-1} e^{-\lambda x} dx \right) \Gamma(\gamma)}$$

$$S(t) = \int_t^\infty \frac{\lambda}{\Gamma(\gamma)} (\lambda x)^{\gamma-1} e^{-\lambda x}$$

$$f(t) = \frac{\lambda^\gamma t^{\gamma-1} e^{-\lambda t}}{\Gamma(\gamma)}, \quad t > 0, \gamma > 0, \alpha > 0 \quad (3.15)$$

Gamma dağılımı,  $\gamma = 1$  olduğunda Üstel dağılıma dönüşmektedir. Tehlike fonksiyonu,  $\gamma > 1$  olduğu durumda monoton artan,  $\gamma = 1$  olduğunda sabit ve  $\gamma < 1$  olduğu durumda monoton azalan bir yapıya sahiptir (Rodriguez, 2010).

#### 4. BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİ

Boylamsal veriler tekrarlı ölçümlerden oluşurken, yaşam verileri aynı birimlerde ilgilenilen olay gerçekleşene kadar geçen süreyi içermektedir. Bu iki veri seti, tekrarlı ölçümlerin yaşam verileriyle ilişkili olduğu anlaşılana kadar ayrı ayrı analiz edilmiştir (McCrink, Marshall ve Cairns, 2013). Ancak, bu iki verinin ayrı modellenmesi sonucunda ölçüm hataları ve kayıp veri sorunları ortaya çıkmaktadır.

Bileşik modelleme genel olarak; boylamsal verilerin yaşam verileri üzerine etkilerini araştırmak (Faucett ve Thomas, 1996; Wulfsohn ve Tsiatis, 1997; Wang ve Taylor, 2001; Brown ve Ibrahim, 2003), iki verinin birlikte etkilerini araştırmak (Henderson, Diggle ve Dobson, 2000; Guo ve Carlin, 2004; Zeng ve Cai, 2005) ve bağımsız değişkenlerin bu iki veri seti üzerine etkilerini incelemek için kullanılmaktadır.

Boylamsal ve yaşam süresi verilerinin birlikte modellenebilmesi için, bileşik olabilirlik yöntemlerinden yararlanılmalıdır. Bileşik olabilirlik yapısı, karma-model (pattern-mixture) ve seçilmiş modeller (selection-model) gibi farklı model stratejileriyle elde edilmektedir (Sousa, 2011). Bu modeller rastgele etkiler eklenilerek genişletilebilir ve rastgele seçim modeli, rastgele karma model ve rastgele etkili model olarak adlandırılır (Pericleous, 2016). Sousa (2011) ve McCrink, Marshall ve Cairns (2013), farklı bileşik modelleme stratejilerini aşağıdaki gibi basitçe açıklamışlardır.  $Y$ ,  $T$  ve  $B$  sırasıyla boylamsal verileri, yaşam verilerini ve rastgele etkileri gösterdiğinde,

Seçilmiş Model;  $[Y, T, B] = [B][Y | B][T | Y]$ ,

Karma-Model;  $[Y, T, B] = [B][T | B][Y | T]$ ,

Rastgele Etkili Model;  $[Y, T, B] = [B][Y | B][T | B]$

biçiminde olmaktadır.

Karma modelde temel ilgi boylamsal kısım üzerine odaklanmışken, seçilmiş modelde ilgi DKEM ile modellenen boylamsal gözlemlere sahip yaşam süresi verilerine odaklanmıştır (Sousa, 2011; Pericleous, 2016). Rastgele etkili modellerde (paylaşılmış parametre modelleri olarak da adlandırılır) ise boylamsal ve yaşam süreci bağımsız paylaşılmış rastgele etkiler ile bağlanmaktadır (Sousa, 2011; McCrink, Marshall ve

Cairns, 2013). Bu tez çalışmasında bileşik modelin olabilirliği, iki süreci paylaşılmış rastgele etkilerle bağlayan karma model ile gösterilmiştir.

Literatürdeki temel bileşik model, boylamsal veriler için DKEM (boylamsal alt model) ve yaşam verileri için Cox regresyon modelinin (yaşam çözümlemesi alt modeli) paylaşılmış rastgele etkiler ile birbirine bağlanması ile elde edilmektedir (Wu ve ark., 2012). Ancak üçüncü bölümde de bahsedildiği gibi Cox regresyon modelinin uygulanabilmesi için orantılılık varsayımının sağlanması gerekmektedir. Varsayım sağlanmadığı durumlarda, yaşam çözümlemesi alt modeli, yaşam verilerinin belirli bir dağılıma sahip olduğu durumlar da parametrik regresyon yöntemleri kullanılarak oluşturulmalıdır.

#### 4.1. Yaşam Çözümlemesi Alt Modeli

$T_i^*$ ,  $i$ . birim ( $i = 1, \dots, n$ ) için gerçek yaşam süresini ve  $T_i$  gözlemlenmiş yaşam süresini gösterebilir.  $i$ . birim için  $C_i$  durdurulmuş süre olmak üzere  $\delta_i = I(T_i^* \leq C_i)$  olay göstergesi  $\delta_i = 0$  olduğunda durdurulmuş olayı ve  $\delta_i = 1$  olduğunda diğer durumları gösterebilir (Rizopoulos, 2012a; Wu ve ark., 2012).  $T_i^*$  ve  $C_i$ 'nin minimumu  $T_i$  olmak üzere,  $i$ . birim için gözlemlenmiş yaşam süresi verisi  $(T_i, C_i)$  şeklinde yazılabilir.  $y_i(t)$ ,  $i$ . birim için zamana bağlı bağımsız değişkenlerin  $t$  zamanında ölçülmüş gözlemlenmiş değerleri olsun. Gözlemlenmiş boylamsal veriler  $t_{ij}$  zamanında  $y_{ij}$  ( $i = 1, \dots, n$ ;  $j = 1, \dots, m_i$ ) ölçümlerinden elde edilmektedir (Rizopoulos, 2012a).

$m_i(t)$ ,  $t$  zamanındaki gözlemlenmemiş boylamsal verileridir ve ölçüm hataları içermektedir. Bir olay için  $m_i(t)$  ve tehlike arasındaki ilişkiyi incelemek için Cox regresyon modeli,

$$h_i(t | M_i(t), \alpha_i) = \lim_{\delta t \rightarrow 0} \left\{ \frac{P[t \leq T_i^* < t + \delta t | T_i^* \geq t, M_i(t), \alpha_i]}{\delta t} \right\}$$

$$h_i(t | M_i(t), \alpha_i) = h_0(t) \exp\{\gamma' \alpha_i + \alpha m_i(t)\}, \quad t > 0 \quad (4.1)$$

biçiminde yazılabilir (Rizopoulos, 2012a). Burada,  $M_i(t) = \{m_i(s), 0 \leq s < t\}$  yaşam süresi içerisinde ölçüm hatalarından arındırılmış boylamsal bağımsız değişkeni,  $h_0(\cdot)$  temel tehlike fonksiyonunu,  $\alpha_i$  temel bağımsız değişkenleri (tedavi göstergesi, hastalık

öyküsü gibi) ve  $\gamma$  bağımsız değişkenlere karşılık gelen regresyon katsayıları vektörünü göstermektedir.  $\alpha$ , boylamsal verilerin  $t$  zamanındaki yaşam sürelerine etkilerini göstermektedir (Rizopoulos, 2012a).  $\alpha = 0$  boylamsal ve yaşam verileri arasında ilişki olmadığını ifade etmektedir (Rizopoulos, Verbeke ve Lesaffre, 2009; Rizopoulos, 2012b). Eşitlik 4.1'deki tehlike modeli,  $t$  zamanında meydana gelen bir olayın tehlikesinin, zamana bağımlı bağımsız değişkenlerin ( $m_i(t)$ ) mevcut değerlerine bağlı olduğunu varsaymaktadır (Rizopoulos, 2012a). Ancak, yaşam fonksiyonu için bu varsayım geçerli değildir. Yaşam fonksiyonu ve birikimli tehlike fonksiyonu arasındaki ilişkiden, tüm ortak bağımsız değişkenlere ( $M_i(t)$ ) ait yaşam fonksiyonu,

$$S_i(t | M_i(t), \alpha_i) = P(T_i^* > t | M_i(t), \alpha_i)$$

$$S_i(t | M_i(t), \alpha_i) = \exp \left( - \int_0^t h_0(s) \exp \{ \gamma' \alpha_i + \alpha m_i(s) \} ds \right) \quad (4.2)$$

şeklinde elde edilebilir (Rizopoulos, 2012a).

Eşitlik 4.1'in çözümlenmesi için temel tehlike fonksiyonunun ( $h_0(\cdot)$ ) seçilmesi gerekmektedir. Orantılı tehlikeler varsayımının sağlandığı durumlarda bileşik model yaşam çözülmesi alt modeli için standart Cox regresyon modeli, orantılı tehlikeler varsayımının sağlanmadığı ve tehlike fonksiyonun bilinen bir olasılık dağılımına sahip olduğu durumlarda parametrik regresyon modelleri kullanılır (Rizopoulos, 2012a). Bu durumda, parametrik yaşam çözümlemesi alt modellerinden yararlanır.

#### 4.2. Parametrik Yaşam Çözümlemesi Alt Modelleri

Parametrik yaşam çözümlemesi alt modelleri, boylamsal alt modelin DKEM ile yaşam çözümlemesi alt modelinin orantılı tehlikeler varsayımını sağlamadığı ve yaşam verilerinin parametrik dağılımlara uygunluğuna göre Weibull, Üstel, Log-lojistik, Log-normal ve Gamma yaşam çözümlemesi alt modellerinin paylaşılmış parametre modelleri ile birleştirilmesiyle elde edilmektedir. Eşitlik 4.1'de ifade edilen yaşam çözümlemesi alt modelinin, parametrik modeller için yeniden tanımlanması gerekmektedir (Pericleous, 2016).

Zamanın fonksiyonu olmak üzere  $h(t)$ 'nin, aşağıdaki gibi iki önemli durumu vardır:

$$h(t); \begin{cases} \log(t), \text{ hızlandırılmış başarısızlık süresi modelleri} \\ t, \text{ diğer parametrik modeller} \end{cases}$$

i. birim için  $h(t)$  fonksiyonu,

$$h(t_i) = \mu_i(t_i) + \sigma \times \varepsilon_i \quad (4.3)$$

biçiminde tanımlanabilir (Pericleous, 2016). Burada  $\sigma$ , ölçek parametresi,  $\varepsilon_i$  uygun yaşam çözümlemesi dağılımından gelen hatalardır ve  $\mu_i(t_i)$ ,

$$\mu_i(t_i) = \alpha m_i(t) + M_i' \gamma \quad (4.4)$$

şeklindedir (Hosmer, Lemeshow ve May 2008; Pericleous, 2016).  $\gamma$ , zamandan bağımsız (time-independent) açıklayıcı değişkenler olan  $M_i'$ 'yi bağlayan regresyon katsayısıdır. i. birim için parametrik yaşam çözümlemesi alt modeli  $h_i(t)$ ,

$$h_i(t) = \alpha [x_i'(t)\beta + z_i'b_i] + M_i'\gamma + \sigma \times \varepsilon_i \quad (4.5)$$

biçiminde yazılabilir.

#### 4.2.1. Weibull Yaşam Çözümlemesi Alt Modeli

Eşitlik 3.12'de verilen Weibull dağılımına ait yaşam fonksiyonunda  $\lambda = \sigma^{-1}$  olarak yeniden tanımlanırsa,

$$S(t; \lambda, \gamma) = \exp\{-\lambda t^\gamma\} \quad (4.6)$$

elde edilir ve süreye logaritmik dönüşüm uygulanırsa ( $s = \log t$ ),

$$s = \log T = \mu + \sigma W \quad (4.7)$$

ifadesi yazılabilir. Burada,  $W$  uç değer dağılımını (extreme value distribution) göstermektedir ve sırasıyla olasılık yoğunluk ve yaşam fonksiyonu aşağıda verilmiştir:

$$\begin{aligned} f_w(w) &= \sigma^{-1} \exp\{w - \exp\{w\}\} \\ S_w(w) &= \exp\{-\exp\{w\}\} \end{aligned} \quad (4.8)$$

Log yaşam süresinin olasılık yoğunluk ve yaşam fonksiyonlarını elde etmek için  $W = \sigma^{-1}(s - \mu)$  yazılmalıdır. i. birim için Weibull dağılımının log yaşam çözümlemesi alt modeli,

$$\log T_i(t_i) = \mu_i(t_i) + \sigma \times \varepsilon_i \quad (4.9)$$

biçiminde ifade edilir ve  $\varepsilon_i$  uç değer dağılımını göstermektedir. Eşitlik 3.12 ve Eşitlik 4.4'den yararlanarak, Weibull yaşam çözümlemesi alt modelinin olasılık yoğunluk ve yaşam fonksiyonları sırasıyla aşağıdaki gibi yazılabilir (Pericleous, 2016):

$$f(T_i | \mathbf{b}_i; \theta, t) = (\text{texp}[\mu_i(t_i)])^{\exp(1/\sigma - (\text{texp}[\mu_i(t_i)])^{\exp(1/\sigma)})}$$

$$S(T_i | \mathbf{b}_i; \theta, t) = \exp\left[-(\text{texp}[\mu_i(t_i)])^{\exp(1/\sigma)}\right] \quad (4.10)$$

Boylamsal alt model için,  $Y_i(t_{ij})$  ölçümlerinin ve rastgele etkilerin olasılık yoğunluk fonksiyonu sırasıyla,

$$p(Y_i | \mathbf{b}_i; \theta) = (2\pi\sigma_e^2)^{-1/2} \exp[-(Y_i(t_{ij}) - m_i(t_{ij}))^2 (2\sigma_e^2)^{-1}], \quad (4.11)$$

$$p(\mathbf{b}_i; \theta) = (2\pi|G|)^{-1/2} \exp(-2^{-1} \mathbf{b}_i' G^{-1} \mathbf{b}_i) \quad (4.12)$$

biçiminde tanımlanabilir (Pericleous, 2016).

#### 4.2.2. Üstel Yaşam Çözümlemesi Alt Modeli

Üstel yaşam çözümlemesi alt modelini elde etmek için, boylamsal alt modelin DKEM ile ve yaşam çözümlemesi modelinin Üstel dağılım ile modellenmesi gerekmektedir. Üstel yaşam süresi alt modelinin i. birim için log yaşam süresi,

$$\log T_i(t_i) = \mu_i(t_i) + \sigma \times \varepsilon_i \quad (4.13)$$

biçiminde yazılabilir ve burada, hatalar Üstel dağılmaktadır. Eşitlik 3.11'de gösterilen Üstel dağılımın yaşam ve olasılık yoğunluk fonksiyonlarından yararlanılarak elde edilen, Üstel yaşam çözümlemesi alt modelinin olasılık yoğunluk ve yaşam fonksiyonları sırasıyla,

$$f(T_i | \mathbf{b}_i; \theta, t) = (\text{texp}[\mu_i(t_i)])^{\exp(-(\text{texp}[\mu_i(t_i)]))},$$

$$S(T_i | \mathbf{b}_i; \theta, t) = \exp\left[-(\text{texp}[\mu_i(t_i)])\right] \quad (4.14)$$

biçiminde elde edilebilir. Boylamsal alt modelin, boylamsal ölçümlerinin ve rastgele etkilerinin olasılık yoğunluk fonksiyonları ise sırasıyla Eşitlik 4.11 ve Eşitlik 4.12'deki gibi oluşturulmalıdır.

#### 4.2.3. Log-lojistik Yaşam Çözümlemesi Alt Modeli

Eşitlik 3.13'te verilen Log-lojistik dağılımın yaşam fonksiyonunun parametreleri  $\gamma = \sigma^{-1}$  ve  $\lambda = \exp\{-\sigma^{-1} \mu\}$  şeklinde yeniden tanımlanırsa ve sürenin logaritması alınırsa  $y = \log t$ ,

$$S(y; \lambda, \gamma) = (1 + \lambda \exp\{\gamma y\})^{-1} \quad (4.15)$$

olarak tanımlanır ve  $Y$  doğrusal modeli göstermektedir:



$$Y = \log T = \mu + \sigma \times W.$$

$W$  standart lojistik dağılımın rastgele değişkenini göstermektedir ve olasılık yoğunluk fonksiyonu ve yaşam fonksiyonu sırasıyla,

$$\begin{aligned} f_w(w) &= \exp(w)\sigma^{-1}(1 + \exp\{w\})^{-2}, \\ S_w(w) &= (1 + \exp\{w\})^{-1} \end{aligned} \quad (4.16)$$

biçiminde tanımlanabilir (Pericleous, 2016).

Log-lojistik yaşam çözümlemesi alt modelini elde etmek için, boylamsal alt model DKEM ile yaşam süresi alt modeli Log-lojistik dağılım kullanılarak modellenmiştir. Weibull dağılım için Eşitlik 4.9'da verilen denklemde  $\varepsilon_i$  Log-lojistik dağılım gösterdiği durumda Log-lojistik yaşam çözümlemesi alt modeli elde edilmektedir. Log-lojistik yaşam çözümlemesi alt modelinin olasılık yoğunluk ve yaşam fonksiyonları sırasıyla,

$$\begin{aligned} f(y_i; \mu_i, \sigma) &= \exp\{\sigma^{-1}(y_i - \mu_i)\}\sigma^{-1}(1 + \exp\{\sigma^{-1}(y_i - \mu_i)\})^{-2}, \\ S(y_i; \mu_i, \sigma) &= [1 + \exp\{\sigma^{-1}(y_i - \mu_i)\}]^{-1} \end{aligned} \quad (4.17)$$

şeklinde tanımlanabilir. Boylamsal alt model için, boylamsal ölçümlerin ve rastgele etkilerin olasılık yoğunluk fonksiyonları sırasıyla Eşitlik 4.11 ve Eşitlik 4.12'deki gibi elde edilmektedir.

#### 4.2.4. Log-normal Yaşam Çözümlemesi Alt Modeli

Log-normal yaşam çözümlemesi alt modeli için, diğer parametrik modeller gibi boylamsal alt model için DKEM ve yaşam çözümlemesi için Log-normal dağılım kullanılmalıdır. Log-normal yaşam süresi alt modelinin  $i$ . birim için log yaşam süresi,

$$\log T_i(t_i) = \mu_i(t_i) + \sigma \times \varepsilon_i \quad (4.18)$$

olarak tanımlanabilir ve  $\varepsilon_i$  normal dağılım göstermektedir (Pericleous, 2016). Eşitlik 3.14'te verilen Log-normal dağılımın yaşam ve olasılık yoğunluk fonksiyonundan yararlanılarak elde edilen, Log-normal yaşam çözümlemesi alt modelinin olasılık yoğunluk ve yaşam fonksiyonları sırasıyla,

$$\begin{aligned} f(y_i; \mu_i, \sigma) &= \exp\left[\frac{-(T_i(t_i)\sigma(2\pi)^{1/2})^{-1}(y_i - \mu_i)}{2\sigma^2}\right], \\ S(y_i; \mu_i, \sigma) &= \left[1 - \phi\left(\frac{(y_i - \mu_i)}{\sigma}\right)\right] \end{aligned} \quad (4.19)$$

biçiminde tanımlanabilir. Bileşik modelin boylamsal kısmı için boylamsal ölçümlerin ve rastgele etkilerin olasılık yoğunluk fonksiyonları ise sırasıyla Eşitlik 4.11 ve Eşitlik 4.12’de verilen eşitliklerle elde edilmektedir.

### 4.3. Boylamsal Alt Model

Tehlike modelindeki (Eşitlik 4.1) tehlike fonksiyonunun ( $h_i(t)$ ) t zamanında boylamsal verilere ( $m_i(t)$ ) bağlı olduğu varsayılmaktadır (Wu ve ark., 2012). Bununla birlikte, ölçüm hatalarına sahip boylamsal verilerde görülebilmektedir. Bu nedenle, boylamsal verilerin bir olayın tehlikesine etkisini araştırmak için, her birimin  $m_i(t)$  tahminlerine ihtiyaç duyulmaktadır. Mevcut boylamsal veriler ( $y_{ij} = \{y_i(t_{ij}), j=1, 2, \dots, m_i\}$ ) için DKEM kullanılarak tahminler elde edilebilmektedir (Rizopoulos, 2012a). Eşitlik 2.1’e benzer bir biçimde boylamsal alt model,

$$y_i(t) = x_i'(t)\beta + z_i'(t)b_i + \varepsilon_i(t),$$

$$y_i(t) = m_i(t) + \varepsilon_i(t),$$

$$m_i(t) = x_i'(t)\beta + z_i'(t)b_i$$

(4.20)

biçiminde yazılabilir. Burada,  $x_i(t)$  ve  $z_i(t)$  sırasıyla sabit etkiler ( $\beta$ ) ve rastgele etkiler ( $b_i$ ) için tasarım vektörleridir. Rastgele etkiler  $G$  varyans kovaryans matrisi ve 0 ortalama ile çok değişkenli normal dağılıma sahiptir (Fitzmaurice, Laird ve Ware, 2011). Hata terimleri ( $\varepsilon_i(t)$ ) 0 ortalama ve  $\sigma^2$  varyansla normal dağılıma sahiptir ve bağımsızdır (Rizopoulos, 2012a).

Daha önce belirtildiği gibi, yaşam çözümlemesi fonksiyonu (Eşitlik 4.2) boylamsal verilerin ölçüm hatalarından arındırılmış gerçek değerlerine dayanır ve bu nedenle  $S_i(t)$ ’nin doğru tahmini için  $M_i(t)$ ’nin iyi bir tahminini elde etmek önemlidir. Yani,  $x_i(t)$  ve  $z_i(t)$ ’deki zaman yapısının, temel bağımsız değişkenler ve varsayılan zaman yapısı ile arasındaki muhtemel etkileşim terimlerinin ayrıntılı bir biçimde belirlenmesi gerekmektedir. Örneğin, birimler doğrusal olmayan boylamsal ölçümler içeriyorsa,  $x_i(t)$  ve  $z_i(t)$  için yüksek mertebeden polinomlar veya splineler cinsinden ifade edilen, zaman fonksiyonlarının yüksek boyutlu vektörlerinin kullanılması önerilmektedir (Rizopoulos, 2012a).

#### 4.4. Bileşik Model Parametre Tahmini

Bileşik model parametre tahmini için genellikle iki aşamalı yaklaşım (Two-Stage Methods) ve bileşik olabilirlik yöntemi kullanılmaktadır (Wu ve ark., 2012).

##### 4.4.1. İki Aşamalı Yaklaşım

Bileşik model literatüründe birkaç farklı iki aşamalı yaklaşım kullanılmıştır (Self ve Pawitan, 1992; Tsiatis, DeGruttola ve Wulfsohn 1995; Wu ve ark., 2012). Self ve Pawitan (1992) birinci adımda en küçük kareler yöntemini kullanarak rastgele etkileri tahmin edip; ikinci adımda bu tahminleri Cox modelin kısmi olabilirliğinde yer alan  $m_i(t)$  'nin uygun değerlerini belirlemek için kullanmışlardır. Bileşik modelde parametre tahmini için Tsiatis, DeGruttola ve Wulfsohn (1995) Eşitlik 4.1 ve 4.3'ün birleştirilmesine dayanan bir diğer iki aşamalı yaklaşım yöntemini önermişlerdir.

İki aşamalı yaklaşım aşağıdaki gibi işlemektedir (Wu ve ark., 2012):

1. Boylamsal açıklayıcı değişkenler DKEM kullanılarak modellenebilir böylece açıklayıcı değişkenlerin birime özgü değerleri tahmin edilebilir.
2. Yaşam çözümlenmesi modeli, 1. adımda modellenen değerler kullanılarak tahmin edilebilir.

Bu yaklaşım, Cox regresyon modelinin parametre tahminlerindeki yanlılığı azaltmaktadır (Wulfsohn ve Tsiatis, 1997; Yu ve ark., 2004). Yaklaşımın uygulaması basittir ve mevcut yazılımlar kullanılarak parametre tahminlerine olanak sağlamaktadır (Wu ve ark., 2012). Buna rağmen, bu yöntem her bir modelin tahmin aşamasında yaşam süreci ve boylamsal süreçten eş zamanlı olarak bilgi kullanamaz (Wu ve ark., 2012; Yu ve ark., 2004). İlk aşamada sadece boylamsal sonuçların kullanılması, DKEM'de yanlı tahminler oluşturabilmektedir ve bunun sonucu olarak ikinci aşamada, yaşam çözümlenmesi parametre tahminlerinde yanlı ve etkisiz sonuçlar ortaya çıkabilmektedir (İbrahim, Chu ve Chen, 2010). Bu durum, Dafni ve Tsiatis (1998); Tsiatis ve Davidian (2001); Ye, Lin ve Taylor (2008a); Sweeting ve Thompson (2011) tarafından simülasyon çalışmaları ile gösterilmiştir. Bu nedenle, yanlılığı ortadan kaldırmak için bileşik modelleme literatüründe, yaklaşımlar yerine tam olasılık yaklaşım yöntemine odaklanılmıştır (Rizopoulos, 2012a).

#### 4.4.2. Bileşik Olabilirlik Yöntemi

Bileşik modelleme literatüründe temel tahmin yöntemi olarak yarı parametrik en çok olabilirlik yöntemi kullanılmaktadır (Wulfsohn ve Tsiatis, 1997; Henderson, Diggle ve Dobson, 2000; Hsieh, Tseng, ve Wang, 2006; Rizopoulos, 2012a; Wu ve ark., 2012). Bu yöntemde, boylamsal ve yaşam verilerinin bileşik olabilirliği kullanılmaktadır (Wu ve ark., 2012).

En çok olabilirlik tahminleri, gözlemlenmiş verinin ( $\{T_i, \delta_i, y_i\}$ ) bileşik dağılımına karşılık gelen log-olabilirlik fonksiyonundan elde edilebilmektedir. Bileşik dağılım tanımında, boylamsal ve yaşam verilerinin arasındaki ve tekrarlı gözlemler arasındaki ilişkinin zaman bağımsız (time-independent) rastgele etkiler ( $b_i$ ) tarafından açıklanacağı varsayılmaktadır (Rizopoulos, 2012a). Boylamsal ve yaşam verileri rastgele etkilerden bağımsızdır ve bileşik dağılımı,

$$p(T_i, \delta_i, y_i | b_i; \theta) = p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta), \quad (4.21)$$

$$p(y_i | b_i; \theta) = \prod_{j=1}^{m_i} p\{y_i(t_{ij}) | b_i; \theta\} \quad (4.22)$$

biçiminde yazılabilir. Burada,

$y_i$ : i. birimin boylamsal gözlemlerinin  $m_i \times 1$  boyutlu vektörü,

$\theta = (\theta'_t, \theta'_y, \theta'_b)$  tüm parametreler vektörü,

$\theta_t$ : yaşam verisi parametreleri,

$\theta_y$ : boylamsal veri parametreleri,

$\theta_b$ : rastgele etkilerin kovaryans matrisi parametrelerini

göstermektedir (Rizopoulos, 2010; Rizopoulos, 2012a).

Tekrarlı ölçümler ve gözlemlenmiş yaşam verilerinin i. birim için log-olabilirlik fonksiyonu aşağıdaki gibi yazılabilir (Wulfsohn ve Tsiatis, 1997; Rizopoulos, 2012a):

$$\begin{aligned} \log p(T_i, \delta_i, y_i; \theta) &= \log \int p(T_i, \delta_i, y_i, b_i; \theta) db_i \\ &= \log \int p(T_i, \delta_i | b_i; \theta_t, \beta) \left[ \prod_j p\{y_i(t_{ij}) | b_i; \theta_y\} \right] p(b_i; \theta_b) db_i \end{aligned} \quad (4.23)$$

Yaşam çözümlemesi kısmının  $p(T_i, \delta_i | b_i; \theta_t, \beta)$  koşullu yoğunluğu,

$$\begin{aligned}
p(T_i, \delta_i | b_i; \theta, \beta) &= h_i(T_i | M_i(T_i); \theta, \beta)^{\delta_i} S_i(T_i | M_i(T_i); \theta, \beta) \\
&= \left[ h_0(T_i) \exp\{\gamma' \alpha_i + \alpha m_i(T_i)\} \right]^{\delta_i} \exp\left[ -\int_0^{T_i} h_0(u) \exp\{\gamma' \alpha_i + \alpha m_i(u)\} du \right] \quad (4.24)
\end{aligned}$$

biçiminde yazılabilir. Burada,  $\alpha_i$  temel açıklayıcı değişkenleri,  $\gamma$  regresyon katsayılarını,  $m_i(t)$  t zamanında ölçülen boylamsal gözlemleri ve  $\beta$  boylamsal alt modeldeki karma etkiler parametrelerini göstermektedir.  $h_0(\cdot)$  Eşitlik 4.2’de verilen yaşam fonksiyonu veya bilinen bir dağılımın tehlike fonksiyonu veya parçalı-sabit model (piecewise-constant) olabilir (Rizopoulos, 2012a).

Eşitlik 4.5’ten, boylamsal gözlemler ve rastgele etkilerin bileşik yoğunluk fonksiyonu aşağıdaki gibi ifade edilebilir (Rizopoulos, 2012a):

$$\begin{aligned}
p(y_i | b_i; \theta) p(b_i; \theta) &= \prod_{j=1}^{m_i} p\{y_i(t_{ij}) | b_i; \theta_y\} f(b_i; \theta_b) \\
&= (2\pi)^{-m_i/2} \exp\left\{-\|y_i - X_i \beta - Z_i b_i\|^2 / 2\sigma^2\right\} \\
&\times (2\pi)^{-q_b/2} \det(G)^{-1/2} \exp\left\{-b_i' G^{-1} b_i / 2\right\} \quad (4.25)
\end{aligned}$$

Burada,  $\|x\| = \left\{ \sum_i x_i^2 \right\}^{1/2}$  Öklid vektör normunu ve  $q_b$  rastgele etkiler vektörünün boyutunu göstermektedir.

Tüm birimler için gözlemlenen verilerin log-olabilirliği,

$$\ell(\theta) = \sum_{i=1}^n \log p(T_i, \delta_i, y_i; \theta) \quad (4.26)$$

şeklinde tanımlanabilir (Wulfsohn ve Tsiatis, 1997; Rizopoulos, 2012a). Temel tehlike dışındaki tüm parametreler parametrik en çok olabilirlik yöntemiyle, temel tehlike  $h_0(\cdot)$  ise parametrik olmayan en çok olabilirlik yöntemi ile tahmin edilmektedir (Wulfsohn ve Tsiatis, 1997). Eşitlik 4.26’nın  $\theta$ ’ya göre maksimize edilebilmesi için EM (Expectation-Maximization) algoritması kullanılabilir. EM tekniğinde rastgele etkiler “eksik veri” olarak kabul edilmektedir (Rizopoulos, 2012a). Bazı araştırmacılar, bileşik modelin maksimizasyonu için “Gauss-Hermite Yaklaşımı” ve “Monte-Carlo Yaklaşımı” gibi iterasyon yöntemleri kullanmışlardır (Wulfsohn ve Tsiatis, 1997; Henderson, Diggle ve Dobson, 2000; Song, Davidian ve Tsiatis, 2002; Rizopoulos, 2012a).

#### 4.4.2.1. EM Algoritması

EM algoritması, eksik verilerin var olduğu durumlarda parametrelerin en çok olabilirlik tahminlerini elde etmek için kullanılabilir (Dempster, Laird ve Rubin, 1977; Bilmes, 1998). Bu algoritma genellikle veride eksik değer bulunduğu ve olabilirlik fonksiyonunun optimizasyonunun zor olduğu durumlarda kullanılmaktadır. Algoritma, E ve M adımlarından oluşmaktadır ve E-adımında koşullu olasılıklarla mevcut parametre tahminleri ve gözlemlenmiş verilerle eksik gözlemler hesaplanmaktadır; M-adımında ise birimci adımdan yararlanılarak koşullu olasılıklar hesaplanmaktadır (Bilmes, 1998; Borman, 2006).

$Y^o$  ve  $Y^m$ , tüm veri vektörü  $Y$ 'nin sırasıyla gözlemlenmiş ve eksik veri kısmını gösterebilir (Rizopoulos, 2012a).  $i$ . birim için,  $c$  ( $c = 0, 1, \dots$ ) iterasyonda tahmin edilen maksimizatör  $\theta^{(ic)}$  ve  $Q(\theta | \theta^{(ic)})$  bileşik log-olabilirliğin olasılığını gösterebilir. Sadece gözlemlenmiş veri kullanılarak, tüm veri modellerinin  $\theta$  parametrelerini tahmin etmek için EM algoritması aşağıdaki gibi işlemektedir (Rizopoulos, 2012a; Givens ve Hoeting, 2013):

(i) E-adımı: Gözlemlenmiş veriler varlığında, tüm verilerin log-olabilirliğinin beklenen değeri,

$$\begin{aligned} Q(\theta | \theta^{(ic)}) &= E\{\log p(y; \theta) | y^o; \theta^{(ic)}\} \\ &= \int \log p(y^m, y^o; \theta) p(y^m | y^o; \theta^{(ic)}) dy^m \end{aligned} \quad (4.27)$$

(ii) M-adımı:  $\theta$ 'ya göre  $Q(\theta | \theta^{(ic)})$  maksimize et ve

(iii) Yakınsama olana kadar E-adımına git.

EM algoritması sayısal olarak kararlıdır (stable) (Lange, 2010; Rizopoulos, 2012a). Bu algoritmada, gözlemlenen verinin log-olabilirliği her iterasyonda artmaktadır  $\log p(y^o; \theta^{(ic+1)}) \geq \log p(y^o; \theta^{(ic)})$  (Dempster, Laird ve Rubin, 1977). Ancak, verilerde çok fazla eksik değer bulunduğu durumlarda yakınsama oranı yavaş olmaktadır (Lange, 2010; Rizopoulos, 2012a).

#### E-Adımı

$i$ . birim için aşağıdaki model dikkate alınır,

$$h_i(t | M_i(t), \alpha_i) = h_0(t) \exp[\gamma' \alpha_i + \alpha \{x_i'(t)\beta + z_i'(t)b_i\}],$$

$$y_i(t) = x_i'(t)\beta + z_i'(t)b_i + \varepsilon_i(t) \quad (4.28)$$

$$b_i \sim N(0, G), \quad \varepsilon_i(t) \sim N(0, \sigma^2)$$

$\theta = (\theta'_t, \theta'_y, \theta'_b)'$ ,  $\theta_y = (\beta', \sigma^2)'$ ,  $\theta_b = \text{vech}(G)$  ve  $\theta_t = (\gamma', \alpha, \theta'_{h_0})'$  vektörleri ile tüm parametreler vektörünü ve  $\theta_{h_0}$ ,  $h_0(\cdot)$  temel tehlike fonksiyonundaki parametreleri göstermektedir. Gözlemlenmiş veri log-olabilirliğini maksimize edecek  $\hat{\theta}$  parametre değerlerini elde etmek için, tüm verilere ait log-olabilirliğin (Eşitlik 4.26) beklenen değerinin maksimize edilmesi gerekmektedir (Rizopoulos, 2012a).

$$Q(\theta | \theta^{(ic)}) = \sum_i \int_{-\infty}^{\infty} \log p(T_i, \delta_i, y_i, b_i; \theta) p(b_i | T_i, \delta_i, y_i; \theta^{(ic)}) db_i$$

$$= \sum_i \int \{ \log p(T_i, \delta_i | b_i; \theta_t, \beta) + \log p(y_i | b_i; \theta_y) + \log p(b_i; \theta_b) \}$$

$$\times p(b_i | T_i, \delta_i, y_i; \theta^{(ic)}) db_i \quad (4.29)$$

$Q(\theta | \theta^{(ic)})$  belirlemek için, “Gauss-Hermite İntegrasyon Yaklaşımı” ve “Monte-Carlo Yaklaşımı” uygulanmalıdır (Rizopoulos, 2012a).

### M-Adımı

Tüm verilerin log-olabilirliği üç parçadan oluşmaktadır (Rizopoulos, 2012a);

$$\log p(T_i, \delta_i, y_i, b_i; \theta) = \log p(T_i, \delta_i | b_i; \theta_t, \beta) + \log p(y_i | b_i; \theta_y) + \log p(b_i; \theta_b) \quad (4.30)$$

$\theta$ 'ya göre  $Q(\theta | \theta^{(ic)})$ 'yi maksimize etmek için, ilgili parametreye göre parçaların tek tek maksimize edilmesi gerekmektedir. Rastgele etkilerin kovaryans matrisi ve boylamsal verideki ölçüm hatalarının varyansı aşağıdaki gibi kapalı formda sırasıyla ifade edilebilir (Rizopoulos, 2012a):

$$G = \frac{\sum_i v b_i + b_i b_i'}{n}$$

$$\hat{\sigma}^2 = \frac{\sum_i \int (y_i - X_i \beta - Z_i b_i)' (y_i - X_i \beta - Z_i b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i}{N}$$

$$= \frac{\sum_i (y_i - X_i \beta)' (y_i - X_i \beta - 2Z_i b_i) + iz (Z_i' Z_i v b_i) + b_i' Z_i' Z_i b_i}{N} \quad (4.31)$$

Burada,

$N = \sum_{i=1}^n m_i$ , çalışmadaki toplam gözlem sayısı,

$$b_i = E(b_i | T_i, \delta_i, y_i; \theta^{(it)}) = \int b_i p(b_i | T_i, \delta_i, y_i; \theta^{(it)}) db_i$$

$$vb_i = Var(b_i | T_i, \delta_i, y_i; \theta^{(it)}) = \int (b_i - b_i)^2 p(b_i | T_i, \delta_i, y_i; \theta^{(it)}) db_i \text{ göstermektedir.}$$

Yukarıdaki kapalı formdan  $\beta$  ve yaşam çözümlemesi alt modelindeki  $\theta_t$ 'nin parametrelerinin çözümü elde edilemeyeceğinden dolayı M-adımında Newton-Raphson iterasyonları kullanılmaktadır (Rizopoulos, 2012a):

$$\begin{aligned} \beta^{(it+1)} &= \beta^{(it)} - \left\{ \frac{\partial S(\beta^{(it)})}{\partial \beta} \right\}^{-1} S(\beta^{(it)}), \\ \theta_t^{(it+1)} &= \theta_t^{(it)} - \left\{ \frac{\partial S(\theta_t^{(it)})}{\partial \theta_t} \right\}^{-1} S(\theta_t^{(it)}) \end{aligned} \quad (4.32)$$

Burada,  $\beta^{(it)}$  ve  $\theta_t^{(it)}$ , sırasıyla  $\beta$  ve  $\theta_t$ 'nin mevcut iterasyondaki değerleridir ve  $\partial S(\beta^{(it)})/\partial \beta$  ve  $\partial S(\theta_t^{(it)})/\partial \theta_t$  Hessian matrisinin karşılık gelen sütunlarıdır.  $\beta$  ve  $\theta_t$ 'nin skor vektörlerinin elemanları,

$$\begin{aligned} S(\beta) &= \frac{\sum_i X_i' \{y_i - X_i \beta - Z_i \tilde{b}_i\}}{\sigma^2} + \alpha \delta_i x_i(T_i) \\ &- \exp(\gamma' \alpha_i) \int \int_0^{T_i} h_0(s) \alpha x_i(s) \exp[\alpha \{x_i'(s) \beta + z_i'(s) b_i\}] \\ &\times p(b_i | T_i, \delta_i, y_i; \theta) ds db_i \end{aligned} \quad (4.33)$$

$$\begin{aligned} S(\gamma) &= \sum_i \alpha_i \left[ \delta_i - \exp(\gamma' \alpha_i) \int \int_0^{T_i} h_0(s) \exp[\alpha \{x_i'(s) \beta + z_i'(s) b_i\}] \right. \\ &\times p(b_i | T_i, \delta_i, y_i; \theta) ds db_i \left. \right] \end{aligned} \quad (4.34)$$

$$\begin{aligned} S(\alpha) &= \sum_i \delta_i \{x_i'(T_i) \beta + z_i'(T_i) \tilde{b}_i\} - \exp(\gamma' \alpha_i) \int \int_0^{T_i} h_0(s) \exp[\alpha \{x_i'(s) \beta + z_i'(s) b_i\}] \\ &\times p(b_i | T_i, \delta_i, y_i; \theta) ds db_i \end{aligned} \quad (4.35)$$



$$S(\theta_{h_0}) = \sum_i \delta_i \frac{\partial h_0(T_i; \theta_{h_0})}{\partial \theta'_{h_0}} - \exp(\gamma' \alpha_i) \int \int_0^{T_i} \frac{\partial h_0(s; \theta_{h_0})}{\partial \theta'_{h_0}} \exp[\alpha \{x'_i(s)\beta + z'_i(s)b_i\}] \times p(b_i | T_i, \delta_i, y_i; \theta) ds db_i \quad (4.36)$$

biçiminde ifade edilebilir (Rizopoulos, 2012a).

#### 4.4.2.2. İntegrasyon Yaklaşımı

Bileşik olabilirlik yapısındaki integrallerin analitik olarak hesaplaması oldukça zordur. Yaşam fonksiyonunun tanımındaki integral, kapalı formulu bir çözüme sahip değildir ve bu nedenle hesaplanması için 15-noktalı Gauss-Kronrod kuralı gibi sayısal bir yöntem kullanılmalıdır. İki modeli birbirine bağlayan rastgele etkiler, bileşik modellemeyi kolaylaştırırsa da rastgele etkilerin integrallerindeki zorluk, tahminleri hesaplamada oldukça zor bir yapıya neden olmaktadır (Rizopoulos, 2012b).

Bu modellerin olabilirliğinin hesaplanması için Gauss karesel (quadrature) kuralları (Pinheiro ve Bates, 1995; Lesaffre ve Spiessens, 2001; Pinheiro ve Chao, 2006) ve Monte Carlo örnekleme (McCulloch, 1997; Booth ve Hobert, 1999) gibi standart sayısal integrasyon teknikleri literatürde sıkça kullanılmıştır. Buna rağmen, bu yöntemlerde integral boyutları arttıkça hesaplama zorluğu çekmektedir ve rastgele etkilerin yapısı daha karmaşık hale gelmektedir (Rizopoulos, 2012b). Çok yüksek boyutlu rastgele etkilerin bulunduğu bileşik modellemede, standart Gauss-Hermite yaklaşımına alternatif olarak Ye, Lin ve Taylor (2008b) ve Rizopoulos, Verbeke ve Lesaffre (2009) tarafından önerilen Laplace yaklaşımı da kullanılabilir.

Rizopoulos (2012b) bileşik modelleme olabilirliğinin çözümü için alternatif olarak uyarlamalı Gauss-Hermite yaklaşımının yeniden ölçeklendirilmesine dayanan bir integrasyon yöntemi önermiştir. Bu yaklaşımda ilk önce boylamsal veriler DKEM ile tahmin edilmektedir ve her birim için verilen boylamsal değişkenlerle rastgele etkilerin sonsal dağılımının konum ve ölçek parametreleri hakkında kesin bilgi elde edilmektedir. Bu bilgi daha sonra bileşik modelin skor vektörü ve log-olabilirliğinin tanımındaki karmaşık integralleri yeniden ölçeklendirmek için kullanılmaktadır (Rizopoulos, 2012b; Turgal, 2016).

Uyarlamalı Gauss-Hermite yaklaşımı için skor vektörlerinin hesaplanmasında gerekli olan Gauss-Hermite yaklaşımının bilinmesi gerekmektedir (Turgal, 2016). Bileşik olabilirliğin (Eşitlik 4.26) skor vektörü aşağıdaki formda yeniden yazılırsa,

$$\begin{aligned} S(\theta) &= \frac{\partial \ell(\theta)}{\partial \theta'} = \sum_i \frac{\partial}{\partial \theta'} \log \int p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) db_i \\ &= \sum_i \int A(\theta, b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i \end{aligned} \quad (4.37)$$

biçiminde olur (Rizopoulos, 2012b).

Burada,  $A(\cdot)$ ,  $A(\theta, b_i) = \partial \{ \log p(T_i, \delta_i | b_i; \theta) + \log p(y_i | b_i; \theta) + \log p(b_i; \theta) \} / \partial \theta'$  tarafından elde edilen tüm verilerin skor vektörüdür. Skor vektörünün (Eşitlik 4.37) formülündeki integraller kapalı formda çözüme sahip olmadığı için sayısal yöntemler kullanılmalıdır. Literatürde en çok tercih edilen Gauss-Hermite kuralları, önceden belirlenmiş apsilerde integrant ölçümlerinin ağırlıklı toplamları tarafından integrali yaklaştırmaktadır (Press ve ark., 2007). Bu karesel kural altında ve rastgele etkilerin  $A(\cdot)$ 'nin her formu için skor vektörü aşağıdaki gibi hesaplanmaktadır (Rizopoulos, 2012b):

$$\begin{aligned} E \{ A(\theta, b_i) | T_i, \delta_i, y_i; \theta \} &= \int A(\theta, b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i \\ &\approx 2^{q/2} \sum_{t_1, \dots, t_q} \pi_t A(\theta, b_t \sqrt{2}) p(b_t \sqrt{2} | T_i, \delta_i, y_i; \theta) \exp(-\|b_t\|^2) \end{aligned} \quad (4.38)$$

Burada,  $q$  rastgele etkiler vektörünün boyutunu,  $\sum_{t_1, \dots, t_q} K$  karesel noktası için  $\sum_{t_1=1}^K \dots \sum_{t_q=1}^K$  ifadelerinin kısaltmasını ve  $b_t' = (b_{t_1}, \dots, b_{t_q})$ ,  $\pi_t$  ağırlıklarına karşılık gelen apsileri göstermektedir.  $K$  karesel noktaların sayısı arttıkça, yaklaşımın kalitesi artmaktadır. Buna rağmen Gauss-Hermite kuralı, her rastgele etki için apsilerin Kartezyen çarpımı üzerinde integrant hesaplamaları gerektirmesinden dolayı,  $q$  arttıkça hesaplama yükü artmaktadır (Rizopoulos, 2012b).

Gauss-Hermite yaklaşımını etkileyen bir diğer kritik nokta, integrantların ana kitlesinin konumuna göre karesel noktaların konumlarıdır. Yani,  $g(b) = A(\theta, b) p(b | T_i, \delta_i, y_i; \theta)$  sıfırdan uzak bir nokta etrafında konumlanmışsa veya  $g(b)$ 'deki yayılma  $\exp(-b'b)$  ağırlık fonksiyonundan oldukça farklıysa,  $g(b)$ 'nin büyük çoğunluğu karesel kuraldaki apsilerin konumlandığı yerde konumlanmayacağından dolayı, çok büyük  $K$  için bile

Gauss-Hermite kuralını doğrudan  $g(b)$ 'ye uygulamak çok zayıf yaklaşıma neden olabilmektedir (Pinheiro ve Bates, 1995). Bu problemi çözmek için önerilen uyarlanmış Gauss-Hermite kuralı, optimizasyon algoritmasının her bir iterasyonunda integrantı uygun bir biçimde ortalayıp ölçmektedir (Rizopoulos, 2012b):

$$E\{A(\theta, b_i) | T_i, \delta_i, y_i; \theta\} \approx 2^{q/2} \sum_{t_1, \dots, t_q} \pi_t A(\theta, \hat{r}_t \sqrt{2}) p(\hat{r}_t \sqrt{2} | T_i, \delta_i, y_i; \theta) \exp(-\|\hat{r}_t\|^2) \quad (4.39)$$

Burada,

$$\hat{r}_t = \hat{b}_i + \sqrt{2} \hat{B}_i^{-1} b_t$$

$$\hat{b}_i = \arg \max_b \{\log p(T_i, \delta_i, y_i, b; \theta)\}$$

$\hat{B}_i, \hat{H}_i = -\partial^2 \log p(T_i, \delta_i, y_i, b; \theta) / \partial b \partial b' |_{b=\hat{b}_i}$ 'nin Choleski faktörünü ifade etmektedir.

Bu dönüşümün kullanılmasıyla, integrant yaklaşık olarak  $N(0, 2^{-1}I)$  dağılımının yoğunluğu gibi davranmaktadır ve Gauss-Hermite kuralının ağırlık fonksiyonu bu dağılıma orantılı olduğundan dolayı optimal yaklaşım elde edilmektedir.

Uyarlanmış Gauss-Hermite kuralında aynı büyüklükte bir yaklaşım hatası elde etmek için standart Gauss-Hermite kuralından daha az karesel nokta kullanılması gerekmektedir. Ancak, mod  $\hat{b}_i$ 'nin konumunun ve her iterasyonda her birim için ikinci derece türev matrisi  $\hat{H}_i$ 'nin hesaplanması, işlem yükünü arttırmaktadır (Rizopoulos, 2012b).

Uyarlanmış Gauss-Hermite kuralındaki zorluk, mod  $\hat{b}_i$  ve  $\hat{H}_i$  matrislerini hesaplamak için kullanılan, rastgele etkilerin sonsal dağılımının  $p(b_i | T_i, \delta_i, y_i; \theta)$  özelliklerinden gelmektedir. Yeniden ölçeklendirilmiş uyarlanmış Gauss-Hermite kuralı uygulanarak, hesaplama yükü azaltılabilmektedir (Rizopoulos, 2012b).

Yoğunluk log-ölçeğinde yeniden yazılırsa,

$$\begin{aligned} \log p(b_i | T_i, \delta_i, y_i; \theta) &\propto \sum_{j=1}^{n_i} \log p\{y_i(t_{ij}) | b_i; \theta_y\} + \log p(b_i; \theta_b) \\ &+ \log p(T_i, \delta_i | b_i; \theta, \beta) \end{aligned} \quad (4.40)$$

biçiminde tanımlanabilir ve  $n_i$  arttıkça gözlemlendiğimiz yoğunluk ile aynıdır. Birinci terim  $\log p(y_i | b_i; \theta_y) = \sum_j \log p\{y_i(t_{ij}) | b_i; \theta_y\}$  DKEM'in logaritmasıdır.  $b_i$ 'de kareseldir ve çok değişkenli normal dağılımın şekline benzemektedir (Rizopoulos, 2012b). Özellikle, Bayesyen merkezi limit teoreminin (Cox ve Hinkley, 1974) bir varyasyonu kullanarak ve genel koşullar altında,  $n \rightarrow \infty$  elde edilebilir:

$$p(b_i | T_i, \delta_i, y_i; \theta) \xrightarrow{p} N(\tilde{b}_i, \tilde{H}_i^{-1}) \quad (4.41)$$

Burada,  $\tilde{b}_i = \arg \max_b \{\log p(y_i | b; \theta_y)\}$  ve  $\tilde{H}_i = -\partial^2 \log p(y_i | b; \theta_y) / \partial b \partial b' |_{b=\tilde{b}_i}$ 'yi göstermektedir (Rizopoulos, 2012b). Bu durum uygulamada,  $n_i$  arttıkça her bir birim için integrantın yeniden ölçeklendirme ve merkezlendirilmesinin, sadece boylamsal veriler için DKEM'den gelen bilgiler kullanılarak hesaplanabileceğini göstermektedir.

Yeniden ölçeklendirilmiş uyarlanmış Gauss-Hermite kuralında, ilk olarak DKEM tahmin edilmektedir ve daha sonra deneysel Bayes tahminleri  $\tilde{b}_i$  ve onların kovaryans matrisi  $\tilde{H}_i^{-1}$  hesaplanmaktadır (Rizopoulos, 2012b):

$$E\{A(\theta, b_i) | T_i, \delta_i, y_i; \theta\} \approx 2^{q/2} \sum_{t_1, \dots, t_q} \pi_t A(\theta, \tilde{r}_t \sqrt{2}) p(\tilde{r}_t \sqrt{2} | T_i, \delta_i, y_i; \theta) \exp(-\|\tilde{r}_t\|^2) \quad (4.42)$$

Bu yöntem, standart Gauss-Hermite kuralında kullanılan daha az karesel nokta kullanılması ve uyarlanmış Gauss-Hermite kuralının her iterasyonunda karesel noktaların zorunlu yer değiştirilmesini önlediğinden dolayı diğer yaklaşımlara göre daha avantajlıdır (Rizopoulos, 2012b).

#### 4.4. Bileşik Model Seçim Kriterleri

Bileşik model parametre tahmininin en çok olabilirlik yöntemiyle yapılabileceği 4.4.2. bölümde gösterilmiştir. Model testi için olabilirlik oran prosedürü dışında, sıfır hipotezini test etmek için alternatif yöntemler Rizopoulos (2012a) tarafından ifade edilmiştir;

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \quad (4.43)$$

- Olabilirlik Oran Testi (Likelihood Ratio Test, LRT)

$$LRT = -2 \left\{ \ell(\hat{\theta}_0) - \ell(\hat{\theta}) \right\} \quad (4.44)$$

$\hat{\theta}_0$  ve  $\hat{\theta}$ , sıfır ve alternatif hipotezler altında en çok olabilirlik tahmin edicileridir.

- Skor Testi

$$U = S'(\hat{\theta}_0) \{I(\hat{\theta}_0)\}^{-1} S(\hat{\theta}_0), \quad I(\hat{\theta}) = -\sum_{i=1}^n \frac{\partial S_i(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}} \quad (4.45)$$

$S(\cdot)$  ve  $I(\cdot)$ , alternatif hipotez altında test edilen modelin skor fonksiyonu ve gözlemlenmiş bilgi matrisini göstermektedir.

- Wald Testi

$$W = (\hat{\theta} - \theta_0)' I(\hat{\theta}) (\hat{\theta} - \theta_0) \quad (4.46)$$

Sıfır hipotezi için,  $p$  test edilen parametre sayısı olmak üzere, asimptotik olarak  $\chi_p^2$  dağılımı göstermektedirler (Diaz, 2014).

Bu testlerin genel dezavantajı, sadece iki iç içe geçmiş modelleri test etmek için kullanılmalarıdır (Diaz, 2014). İç içe olmayan modellerin karşılaştırılmasını yapmak için Akaike Bilgi Kriteri (Akaike's Information Criterion-AIC) ve Bayesci Bilgi Kriteri (Bayesian Information Criterion- BIC) kullanılmalıdır:

$$AIC = -2 \log L + 2p$$

$$BIC = -2 \log L + p \ln(n) \quad (4.47)$$

Burada,  $L$  modelin olabilirlik fonksiyonunu,  $n$  toplam gözlem sayısını ve  $p$  modeldeki bilinmeyen parametre sayısını göstermektedir.

## 5. UYGULAMA

Çalışmamızda farklı parametrik bileşik model yapılarını incelemek için literatürde yer alan PBS verisi kullanılmıştır (Mayo Clinic, Primary Biliary Cirrhosis, sequential data). Veriler, Ocak 1924-Mayıs 1984 yıllarında Mayo Klinikte görülen 312 hasta üzerinde PBS'nin ilerlemesini incelemek için toplanmıştır (Murtaugh ve ark., 1994). Farklı hastalıklar ve nedenlerden dolayı karaciğerde hasarlar oluşması ve zamanla karaciğerin sertleşmesi, küçülmesi ve işlevini yerine getirememesi gibi sorunlara neden olan PBS, karaciğer nakli dışında etkili bir tedavinin bulunmadığı kronik bir karaciğer hastalığıdır (Gordon ve ark., 1987; Esquivel ve ark., 1988). Karaciğer sirozunda tedavi, hastalığın evresine ve nedenine göre farklılık göstermektedir. Hastalığın şiddeti Child skoruna göre belirlenmektedir (Sonsuz, 2007). Child skoru A, B ve C şeklinde ve A'dan C'ye gidildikçe hastalığın seviyesi artmaktadır (Christensen ve ark., 1984; Danalıoğlu, 1995). Child skoruna asit, sarılık ve serum bilirubin gibi değerler ile karar verilmektedir ve bu değerlerin artması karaciğer sirozunun ilerlediğini göstermektedir.

PBS veri setindeki 312'i hastanın 158'i D-penisilin ilacını almış ve diğer 154 hasta placebo grubu olarak belirlenmiştir. Serum bilirubin seviyeleri belirli zaman aralıklarında (6 ay ve sonrasında her yıl) tekrarlı olarak ölçülmüştür ve toplam 1945 dengesiz boylamsal ölçümden oluşmaktadır. Çalışmanın temel amacı, serum bilirubin ve ölüm zamanı arasındaki ilişkiye tedavinin etkisini araştırmaktır.

Yaşam çözümlenmesi ve bileşik modelleme literatüründe, PBS verisi kullanılarak birçok çalışma yapılmıştır. Murtaugh ve ark. (1994) PBS verisini analiz etmek için Cox regresyon modelini kullanmışlardır. Pasha ve Dickson (1997) PBS'nin bağımsız değişkenleri için yaşam tahminlerini ve farklı açıklayıcı değişkenler kullanarak Cox regresyon model tahminlerini incelemişlerdir. Albert ve Shih (2010) kesikli yaşam süresi ve çok değişkenli boylamsal gözlemlerin bileşik modellemesi için önerdikleri yaklaşımı PBS veri seti üzerinde test etmişlerdir. Yan (2011) PBS verisinin Kaplan-Meier tahminlerini ve zaman bağımlı açıklayıcı değişken içeren Cox regresyon model tahminlerini incelemiştir. Crowther ve ark, (2013) sürekli boylamsal gözlemler ve yaşam verilerinin bileşik model tahminlerini elde etmek için STATA programında "stjm" komutunu geliştirmişlerdir. Bu komut kullanarak, boylamsal alt model için doğrusal karma etkili model ve yaşam verileri için Üstel, Weibull, Gompertz orantılı

tehlike modelleri ve esnek parametrik modellerin bileşik model tahminleri elde edilmektedir. Çalışmalarında, Weibull orantılı tehlike alt modeli ve doğrusal karma etkili modelin bileşik tahminlerini elde etmek için PBS verisi üzerinde bir uygulama yapmışlardır. Hsieh, Ding ve Wang (2013) boylamsal ve yaşam verilerinin tahminlerini elde etmek için yeni bir olabilirlik yaklaşımı önermişler ve önerdikleri yaklaşımı PBS verisinde test etmişlerdir. Mbohning, Bleakley ve Lavielle (2015) doğrusal olmayan karma etkili model ile bileşik olabilirlik üzerine çalışmışlardır. Thevaraja (2018) PBS verisinin yaşam çözümlemesi değerlerini SAS programı kullanarak elde etmiştir.

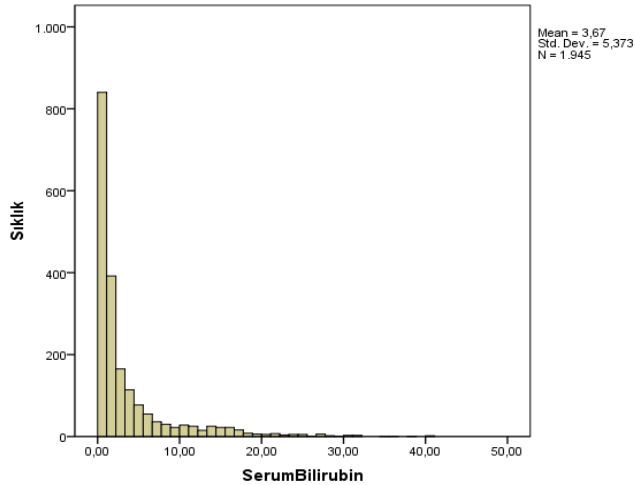
Çalışmada, yaşam çözümlemesi alt modeli için Cox regresyon modelinin kullanıldığı standart bileşik model ve farklı parametrik alt modellerin sonuçları incelenmiştir. Ayrıca iki sürecin ayrı analizleri yapılarak parametrik bileşik model sonuçlarıyla kıyaslaması yapılmıştır. Ayrı analizde, boylamsal gözlemler için doğrusal karma etkili model ve yaşam verileri için boylamsal gözlem olan serum bilirubin değerleri zamana bağlı bağımsız değişken olarak modele alınarak, Weibull parametrik yaşam çözümlemesi modeli kullanılmıştır. Model tahminlerini elde etmek için Stata programında; gsem, streg ve xtreg komutları ve R programında; lme, coxph ve jointModel paketlerinden yararlanılmıştır. Analiz sonuçlarından elde edilen p değerleri  $\alpha = 0.05$  anlamlılık düzeyinde kıyaslanmıştır.

Uygulamada, hastaların izlenmeye başlandığından ölüm gerçekleşene kadar geçen süre yaşam süresi (yıl) olarak alınmıştır. Burada başarısızlık ölüm olarak kabul edilmiştir. Diğer hastalar durdurulmuş olarak ifade edilmiştir. Çalışma sonunda 312 hastadan 140'ı (%44.9) ölmüş ve 172 (%55.1) hastada durdurma gözlemlenmiştir. Veri seti, her hasta için klinik, biyokimyasal ve demografik risk faktörlerini içermektedir. Demografik faktörler; hastaların yaşı ve cinsiyeti, Biyokimyasal faktörler; ilaç (D-penisilin ve placebo grubu), asit (karaciğer yetmezliği sebebiyle karın bölgesinde su birikmesi durumu-var/yok), hepatomegali (karaciğer büyümesi durumu-var/yok), deride bozukluk (derideki kan damarı bozuklukları-var/yok), ödem (el ve ayaklarda şişlik durumu-yok/ilaç kullanımına rağmen ödem/var) ve histolojik evreden oluşmaktadır. Çalışma kapsamında biyokimyasal özellik olarak serum bilirumin (mg/dl) değerleri alınmıştır. Kullanılan değişkenler ve düzeyleri Çizelge 5.1'de detaylı olarak verilmiştir.

**Çizelge 5.1.** Kullanılan açıklayıcı değişkenler ve düzeyleri

Değişken	Düzeyleyler	n	%
Cinsiyet	1- Kadın	276	88.5
	2- Erkek	36	11.5
İlaç	1- D-penisilin	158	50.6
	2- Placebo	154	49.4
Asit	1- Yok	247	79.2
	2- Var	41	13.1
Hepatomegali	1- Yok	131	42.0
	2- Var	157	50.3
Deride Bozukluk	1- Yok	188	60.3
	2- Var	100	32.1
Ödem	1- Yok	43	13.8
	2- İlaça Rağmen Ödem	64	20.5
	3- Var	205	65.7
Histolojik Evre	1- Evre 1	11	3.5
	2- Evre2	45	14.4
	3- Evre 3	102	32.7
	4- Evre 4	154	49.4

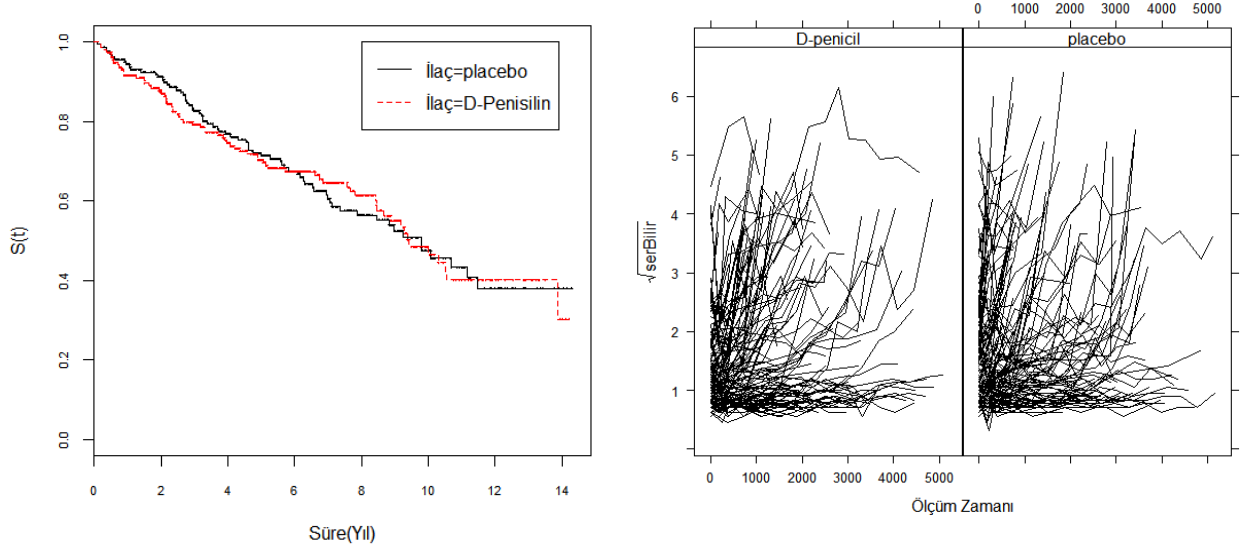
Boylamsal gözlem olan serum bilirubin değerleri, Şekil 5.1’de histogram grafiğinden de görüldüğü gibi sağa çarpık bir dağılım gösterdiğinden dolayı hesaplamalarda logaritması (log(serBilir)) alınarak kullanılmıştır.



**Şekil 5.1.** Serum bilirubin değerlerinin histogram grafiği.



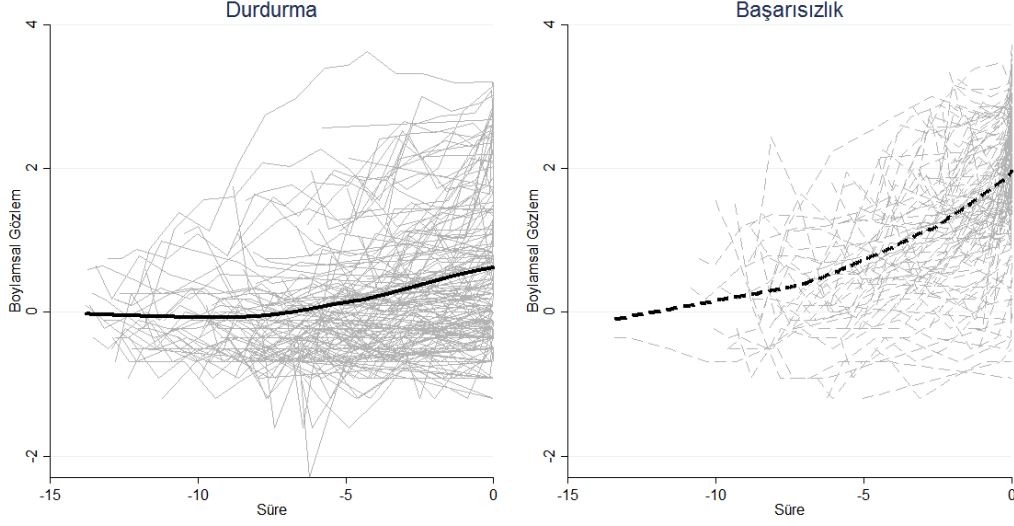
İlaç değişkeninin boylamsal ve yaşam verileri arasındaki ilişkiyi ayrıntılı incelemek için, ilaç değişkeninin boylamsal ölçüm grafikleri ve Kaplan-Meier tahminlerine ait grafikler Şekil 5.2’de verilmiştir.



**Şekil 5.2.** İlaç değişkeninin Kaplan-Meier ve boylamsal ölçüm grafikleri

Şekil 5.2’de verilen Kaplan-Meier grafiğine göre ilk 4 yılda D-penisilin ilacını kullananların kullanmayanlara göre yaşam olasılıklarının daha yüksek olduğu ancak 4 yıl sonrasında yaşam olasılıklarının benzerlik gösterdiği söylenebilir. İki tedavi yöntemine göre log(serBilir) değerlerinin grafiği incelendiğinde placebo grubunun serum bilirubin değerlerinin D-penisilin tedavisini olanlara göre daha yüksek olduğu ancak yine ölçüm değerlerinin birbirlerine yakın olduğu yorumu yapılabilir. İlaç değişkeninin düzeyleri arasında anlamlı bir farklılık olup olmadığını test etmek için log-rank testi uygulandığında,  $p=0.992$  olarak bulunmuştur. Log-rank test sonucuna göre de ilaç değişkeninin düzeyleri arasında anlamlı bir fark olmadığı söylenebilir.

Boylamsal ve yaşam verilerinin bileşik model yapısını inceleyebilmek için, durdurma ve ölen hastalar için log(serBilir) değerlerinin grafiği Şekil 5.3’te verilmiştir.

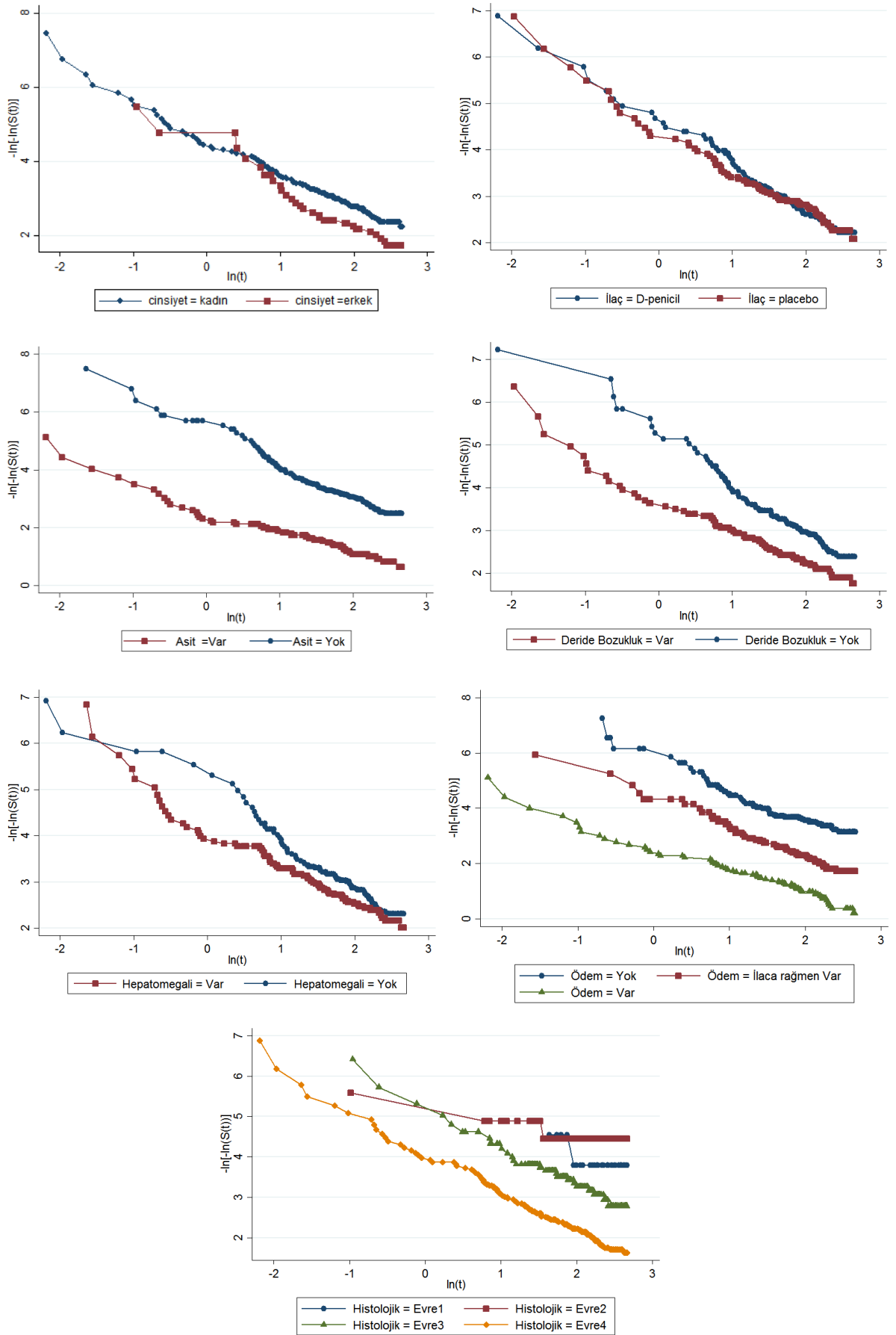


**Şekil 5.3.** Log(serBilir) değerlerinin zamansal grafikleri

Şekil 5.3 incelendiğinde, ölen hastaların durdurma gözlemlenen hastalara göre daha keskin bir eğilim sergilediği görülmektedir. Bu, boylamsal gözlem ile yaşam süresi arasında pozitif bir ilişkinin göstergesidir ve sonuç olarak boylamsal gözlem seviyelerindeki artışın ölüme doğrudan ilişkili olduğu görülmektedir.

### **5.1. Boylamsal ve Yaşam Verilerinin Bileşik Çözümlemesi**

4. bölümde belirtildiği gibi standart bileşik model yaşam çözümlemesi alt modeli için Cox regresyon ve boylamsal alt model için doğrusal karma etkili modelin paylaşılmış parametre modeli ile birleştirilmesine dayanmaktadır. Ancak yaşam çözümlemesi literatüründe ve 3. bölümde belirtildiği gibi Cox regresyon modeli uygulanmadan önce orantılı tehlikeler varsayımı kontrol edilmelidir. Orantılı tehlikeler varsayımının sağlanmadığı durumlarda, yaşam çözümlemesinde Cox regresyon modeli yerine, parametrik regresyon yöntemleri kullanılmalıdır. Aynı durum bileşik modelin yaşam çözümlemesi alt modeli için de geçerlidir. Bu nedenle bileşik model kurulmadan önce orantılı tehlikeler varsayımını test etmek için grafiksel yöntemler ve Schoenfeld artıkları kullanılmış ve sonuçlar sırasıyla Şekil 5.4 ve Çizelge 5.2’de verilmiştir.



Şekil 5.4. Değişkenlerin orantılı tehlikeler varsayımı için grafikler

**Çizelge 5.2.** Değişkenlerin orantılı tehlikeler varsayımı için Schoenfeld artıkları ile yaşam süresinin rankı arasındaki ilişki sonuçları

	<b>r</b>	<b>Ki-Kare</b>	<b>p</b>
Cinsiyet2	0.1349	2.73	0.0982
İlaç2	-0.0355	0.18	0.6755
Asit2	-0.1061	1.55	0.2132
Hepatome2	-0.0819	0.85	0.3563
DeriBoz2	-0.1125	1.65	0.1995
Ödem2	0.1405	3.11	0.0778
Ödem3	0.1584	3.82	0.0506
HistolojikEvre2	-0.0949	1.27	0.2605
HistolojikEvre3	-0.0250	0.09	0.7598
HistolojikEvre4	-0.0220	0.07	0.7885

Çizelge 5.2’de verilen sonuçlar incelendiğinde, tüm değişkenlerin orantılı tehlikeler varsayımını sağladığı ( $p > \alpha$ ) görülmüştür. Orantılı tehlikeler varsayımını incelemek için tek bir yöntem yetersiz olduğundan, varsayım kontrolü grafiksel yöntemlerle de incelenmiştir. Şekil 5.4’te değişkenlerin  $\ln(t)$ ’ye karşı  $-\ln(-\ln(S(t)))$  grafikleri incelendiğinde cinsiyet, ilaç, hepatomegali ve histolojik evre değişkenlerinin düzeyleri çakıştığından dolayı orantılı tehlikeler varsayımının sağlanmadığı söylenebilir. Varsayım sağlanmadığından dolayı, yaşam verileri için Cox regresyon modelinin kullanılması uygun değildir. Ancak yine de parametrik bileşik modellerle kıyaslayabilmek için yaşam çözümlmesi alt modeli için Cox regresyon modeli ve boylamsal alt model için doğrusal karma etkili modelin bileşik model tahminleri Çizelge 5.3’de verilmiştir.

**Çizelge 5.3.** Standart bileşik model sonuçları

		$\hat{\beta}$	Std. hata	z	p
Boylamsal Alt Model	Sabit Terim	0.4352	0.2714	1.6033	0.1089
	Yaş	-0.0038	0.0051	-0.7409	0.4588
	Cinsiyet2	0.2369	0.1704	1.3904	0.1644
	İlaç2	0.1283	0.1053	1.2185	0.2230
	Asit2	0.1043	0.0415	2.5118	<b>0.0120</b>
	Hepatome2	-0.0183	0.0247	-0.7427	0.4577
	DeriBoz2	0.0640	0.0287	2.2273	<b>0.0259</b>
	Ödem2	0.1798	0.0344	5.2289	<b>0.0000</b>
	Ödem3	0.2706	0.0537	5.0434	<b>0.0000</b>
	HistolojikEvre2	0.0340	0.0646	0.5256	0.5992
	HistolojikEvre3	0.1021	0.0669	1.5259	0.1270
	HistolojikEvre4	0.1642	0.1642	2.2939	<b>0.0218</b>
Cox Alt Model	Yaş	0.0306	0.0097	3.1460	<b>0.0017</b>
	Cinsiyet2	0.3957	0.2537	1.5596	0.1188
	İlaç2	0.2125	0.1840	1.1552	0.2480
	Asit2	0.2925	0.2292	1.2758	0.2020
	Hepatome2	-1.6163	0.2042	0.2042	<b>0.0000</b>
	DeriBoz2	-0.6230	0.1994	-3.1249	<b>0.0018</b>
	Ödem2	0.8477	0.2421	3.5012	<b>0.0005</b>
	Ödem3	2.0730	0.2692	7.7003	<b>0.0000</b>
	HistolojikEvre2	-1.7986	0.7371	-2.4401	<b>0.0147</b>
	HistolojikEvre3	-1.4441	0.5745	-2.5138	<b>0.0119</b>
HistolojikEvre4	-0.4095	0.5468	-0.7488	0.4540	
Assoct		1.3913	0.1161	11.9833	<b>0.0000</b>

Çizelge 5.3’de kurulan standart bileşik modele göre, boylamsal alt modelde asit, deride bozukluk, ödem ve histolojikevre4 değişkenleri istatistiksel olarak anlamlı ( $p < \alpha$ ) bulunmuştur. Cox regresyon alt modelinde ise yaş, hepatomegali, deride bozukluk, ödem, histolojikevre2 ve histolojikevre3 değişkenleri istatistiksel olarak anlamlı bulunmuştur. Cox alt modelindeki tahmini assoct değişkeni iki süreci birleştiren parametre tahminidir (ilişkilendirme parametresi) ve Eşitlik 4.1’de  $\alpha$  katsayımı göstermektedir. Analiz sonucunda ilişkilendirme parametresi istatistiksel olarak ( $p < \alpha$ ) anlamlı bulunması, boylamsal gözlem ile yaşam süreci arasında ilişki olduğunu göstermektedir. Serum bilirubin ölçümlerindeki artışın ölüm riskinde de artışa neden olacağı yorumu yapılabilir.

Ancak Şekil 5.3 ve Çizelge 5.2’de görüldüğü gibi orantılı tehlikeler varsayımı sağlanmadığından dolayı bileşik modelin yaşam çözümlemesi alt modeli için Cox regresyonun kullanılması uygun değildir. Bu nedenle, yaşam çözümlemesi alt modeli için Üstel, Weibull, Log-normal, Log-lojistik ve Gamma parametrik alt modellerine bakılmıştır ve sonuçlar sırasıyla Çizelge 5.4 - 5.8’de verilmiştir.

**Çizelge 5.4.** Üstel parametrik bileşik model sonuçları

		$\hat{\beta}$	Std. hata	z	p
Boylamsal Alt Model	Sabit Terim	0.6638	0.2923	2.27	<b>0.023</b>
	Yaş	-0.0082	0.0054	-1.52	0.129
	Cinsiyet2	0.4923	0.1769	2.78	<b>0.005</b>
	İlaç2	0.0939	0.1117	0.84	0.401
	Asit2	0.1416	0.0506	2.80	<b>0.005</b>
	Hepatome2	-0.0190	0.0288	-0.66	0.510
	DeriBoz2	0.1144	0.0334	3.42	<b>0.001</b>
	Ödem2	0.4018	0.0393	10.21	<b>0.000</b>
	Ödem3	0.6640	0.0569	11.67	<b>0.000</b>
	HistolojikEvre2	-0.1186	0.0741	-1.60	0.110
	HistolojikEvre3	0.0320	0.0762	0.42	0.674
	HistolojikEvre4	0.2302	0.0818	2.81	<b>0.005</b>
Üstel Alt Model	Yaş	-0.0514	0.0119	-4.31	<b>0.000</b>
	Cinsiyet2	-0.5668	0.3428	-1.65	0.098
	İlaç2	-0.3827	0.2351	-1.63	0.103
	Asit2	-0.4779	0.2260	-2.11	<b>0.034</b>
	Hepatome2	0.9773	0.1983	4.93	<b>0.000</b>
	DeriBoz2	0.1947	0.1961	0.99	0.321
	Ödem2	-0.9384	0.2323	-4.04	<b>0.000</b>
	Ödem3	-2.0893	0.2345	-8.91	<b>0.000</b>
	HistolojikEvre2	1.3061	0.9242	1.41	0.158
	HistolojikEvre3	0.3144	0.7577	0.41	0.678
HistolojikEvre4	-0.5661	0.7365	-0.77	0.442	
Assoct	-1.3978	0.1130	-12.37	<b>0.000</b>	

Çizelge 5.4’te yaşam çözümlemesi için Üstel alt model ile boylamsal veriler için doğrusal karma etkili modelin bileşik model tahminleri verilmiştir. Boylamsal alt modelde cinsiyet, asit, deride bozukluk, ödem ve histolojikevre4 değişkenleri, Üstel alt modelde ise yaş, asit, hepatomegali ve ödem değişkenleri istatistiksel olarak anlamlı bulunmuştur ( $p < \alpha$ ). İlişkilendirme parametresi assoct katsayısı istatistiksel olarak

anlamli bulunduğundan dolayı, serum bilirubin deęerleri arttikça yařam süresinin azaldığı söylenebilir.

**Çizelge 5.5.** Weibull parametrik bileşik model sonuçları

		$\hat{\beta}$	Std. hata	z	p
Boylamsal Alt Model	Sabit Terim	0.6535	0.2943	2.22	<b>0.026</b>
	Yař	-0.0079	0.0055	-1.42	0.155
	Cinsiyet2	0.4905	0.1781	2.75	<b>0.006</b>
	İlaç2	0.0945	0.1125	0.84	0.401
	Asit2	0.1408	0.0506	2.78	<b>0.005</b>
	Hepatome2	-0.0205	0.0288	-0.71	0.475
	DeriBoz2	0.1139	0.0334	3.41	<b>0.001</b>
	Ödem2	0.3977	0.0393	10.11	<b>0.000</b>
	Ödem3	0.6492	0.0569	11.40	<b>0.000</b>
	HistolojikEvre2	-0.1207	0.0741	-1.63	0.104
	HistolojikEvre3	0.0241	0.0763	0.32	0.752
	HistolojikEvre4	0.2189	0.0818	2.68	<b>0.007</b>
Weibull Alt Model	Yař	-0.0431	0.0092	-4.64	<b>0.000</b>
	Cinsiyet2	-0.4022	0.2682	-1.50	0.134
	İlaç2	-0.3059	0.1809	-1.69	0.091
	Asit2	-0.3434	0.1627	-2.11	<b>0.035</b>
	Hepatome2	0.7130	0.1491	4.78	<b>0.000</b>
	DeriBoz2	0.1397	0.1392	1.00	0.315
	Ödem2	-0.6588	0.1725	-3.82	<b>0.000</b>
	Ödem3	-1.5090	0.1957	-7.71	<b>0.000</b>
	HistolojikEvre2	0.9656	0.6554	1.47	0.141
	HistolojikEvre3	0.2690	0.5366	0.50	0.616
	HistolojikEvre4	-0.3743	0.5217	-0.72	0.473
Assoc	-1.1569	0.0943	-12.27	<b>0.000</b>	

Yařam çözümlemesi için Weibull alt model ile boylamsal verilerin doğrusal karma etkili modelle bileşik model sonuçları Çizelge 5.5'te verilmiştir. Buna göre boylamsal alt modelde cinsiyet, asit, deride bozukluk, ödem ve histolojik evre4 deęişkenleri, Weibull alt modelde yař, asit, hepatomegali ve ödem deęişkenleri istatistiksel olarak anlamlı bulunmuştur ( $p < \alpha$ ). Assoc katsayısı da anlamlı bulunduğundan dolayı, serum bilirubin ölçümlerindeki artışın yařam süresinde kısaltmaya neden olacağı söylenebilir.

**Çizelge 5.6.** Log-normal parametrik bileşik model sonuçları

		$\hat{\beta}$	Std. hata	z	p
Boylamsal Alt Model	Sabit Terim	0.6712	0.2932	2.29	<b>0.022</b>
	Yaş	-0.0082	0.0055	-1.50	0.135
	Cinsiyet2	0.4922	0.1774	2.77	<b>0.006</b>
	İlaç2	0.0960	0.1121	0.86	0.391
	Asit2	0.1384	0.0507	2.73	<b>0.006</b>
	Hepatome2	-0.0184	0.0289	-0.64	0.524
	DeriBoz2	0.1128	0.0335	3.37	<b>0.001</b>
	Ödem2	0.3993	0.0394	10.13	<b>0.000</b>
	Ödem3	0.6605	0.0571	11.57	<b>0.000</b>
	HistolojikEvre2	-0.1216	0.0741	-1.64	0.101
	HistolojikEvre3	0.0232	0.0763	0.30	0.761
	HistolojikEvre4	0.2162	0.0818	2.64	<b>0.008</b>
Log- normal Alt Model	Yaş	-0.0395	0.0095	-4.15	<b>0.000</b>
	Cinsiyet2	-0.5272	0.2785	-1.89	0.058
	İlaç2	-0.2811	0.1907	-1.47	0.140
	Asit2	-0.3950	0.2014	-1.96	<b>0.050</b>
	Hepatome2	0.7002	0.1687	4.15	<b>0.000</b>
	DeriBoz2	-0.0210	0.1606	-0.13	0.896
	Ödem2	-0.7349	0.1809	-4.06	<b>0.000</b>
	Ödem3	-1.7067	0.2270	-7.52	<b>0.000</b>
	HistolojikEvre2	1.2052	0.5783	2.08	<b>0.037</b>
	HistolojikEvre3	0.3813	0.4556	0.84	0.403
	HistolojikEvre4	-0.1955	0.4417	-0.44	0.658
Assoct		-1.1330	0.1000	-11.33	<b>0.000</b>

Çizelge 5.6’da verilen Log-normal bileşik model sonuçlarına göre, boylamsal alt modelde cinsiyet, asit, deride bozukluk, ödem ve histolojikevre4 değişkenleri, Log-normal alt modelde ise yaş, asit, hepatomegali, ödem, histolojikevre2 ve assoct değişkenleri istatistiksel olarak anlamlı bulunmuştur ( $p < \alpha$ ). Serum bilirubin ölçümleri ile yaşam süresi arasında ilişki bulunmaktadır. Serum bilirubin değerlerindeki artışın yaşam süresinde kısalmaya neden olacağı söylenebilir.



**Çizelge 5.7.** Log-lojistik parametrik bileşik model sonuçları

		$\hat{\beta}$	Std. hata	z	p
Boylamsal Alt Model	Sabit Terim	0.6571	0.2941	2.23	<b>0.025</b>
	Yaş	-0.0079	0.0055	-1.44	0.151
	Cinsiyet2	0.4913	0.1780	2.76	<b>0.006</b>
	İlaç2	0.0955	0.1124	0.85	0.396
	Asit2	0.1383	0.0507	2.73	<b>0.006</b>
	Hepatome2	-0.0199	0.0288	-0.69	0.489
	DeriBoz2	0.1138	0.0334	3.40	<b>0.001</b>
	Ödem2	0.3980	0.0394	10.11	<b>0.000</b>
	Ödem3	0.6534	0.0570	11.46	<b>0.000</b>
	HistolojikEvre2	-0.1215	0.0741	-1.64	0.101
	HistolojikEvre3	0.0233	0.0762	0.31	0.760
	HistolojikEvre4	0.2171	0.0818	2.65	<b>0.008</b>
Log-lojistik Alt Model	Yaş	-0.0419	0.0094	-4.46	<b>0.000</b>
	Cinsiyet2	-0.4663	0.2748	-1.70	0.090
	İlaç2	-0.2817	0.1858	-1.52	0.129
	Asit2	-0.3838	0.1796	-2.14	<b>0.033</b>
	Hepatome2	0.7241	0.1592	4.55	<b>0.000</b>
	DeriBoz2	0.0715	0.1513	0.47	0.636
	Ödem2	-0.6852	0.1745	-3.93	<b>0.000</b>
	Ödem3	-1.5498	0.2077	-7.46	<b>0.000</b>
	HistolojikEvre2	0.9881	0.6268	1.58	0.115
	HistolojikEvre3	0.3057	0.5153	0.59	0.553
HistolojikEvre4	-0.3343	0.5011	-0.67	0.505	
Assoct		-1.1499	0.0972	-11.83	<b>0.000</b>

Çizelge 5.7’de verilen Log-lojistik parametrik bileşik model sonuçları incelendiğinde, boylamsal alt model için cinsiyet, asit, deride bozukluk, ödem ve histolojikevre4 değişkenlerinin, yaşam çözümlemesi alt modelinde ise yaş, asit, hepatomegali ve ödem değişkenlerinin istatistiksel olarak anlamlı olduğu görülmektedir ( $p < \alpha$ ). İlişkilendirme parametresine göre, serum bilirubin ölçümlerinde artışın yaşam süresinde kısaltmaya neden olacağı söylenebilir.

**Çizelge 5.8.** Gamma parametrik bileşik model sonuçları

		$\hat{\beta}$	Std. hata	z	p
Boylamsal Alt Model	Sabit Terim	0.6560	0.2941	2.23	<b>0.026</b>
	Yaş	-0.0078	0.0055	-1.43	0.151
	Cinsiyet2	0.4911	0.1780	2.76	<b>0.006</b>
	İlaç2	0.0949	0.1124	0.84	0.398
	Asit2	0.1401	0.0506	2.77	<b>0.006</b>
	Hepatome2	-0.0203	0.0288	-0.71	0.480
	DeriBoz2	0.1136	0.0334	3.40	<b>0.001</b>
	Ödem2	0.3982	0.0393	10.12	<b>0.000</b>
	Ödem3	0.6518	0.0570	11.45	<b>0.000</b>
	HistolojikEvre2	-0.1209	0.0741	-1.63	0.103
	HistolojikEvre3	0.0242	0.0763	0.32	0.750
	HistolojikEvre4	0.2187	0.0818	2.67	<b>0.008</b>
Gamma Alt Model	Yaş	-0.0429	0.0095	-4.53	<b>0.000</b>
	Cinsiyet2	-0.4345	0.2733	-1.59	0.112
	İlaç2	-0.3058	0.1853	-1.65	0.099
	Asit2	-0.3547	0.1707	-2.08	<b>0.038</b>
	Hepatome2	0.7297	0.1548	4.71	<b>0.000</b>
	DeriBoz2	0.1198	0.1461	0.82	0.412
	Ödem2	-0.6870	0.1755	-3.91	<b>0.000</b>
	Ödem3	-1.5713	0.1996	-7.87	<b>0.000</b>
	HistolojikEvre2	1.0084	0.6478	1.56	0.120
	HistolojikEvre3	0.2992	0.5301	0.56	0.572
HistolojikEvre4	-0.3546	0.5155	-0.69	0.492	
Assoct		-1.1665	0.7272	9.23	<b>0.000</b>

Gamma alt model ile boylamsal alt modelin paylaşılmış parametre ile bileşik model sonuçları Çizelge 5.8’de verilmiştir. Boylamsal alt modelde cinsiyet, asit, deride bozukluk, ödem ve histolojikevre4 değişkenleri, Gamma alt model de ise yaş, asit, hepatomegali, ödem ve ilişkilendirme parametresi istatistiksel olarak anlamlı bulunmuştur ( $p < \alpha$ ). Boylamsal ölçüm olan serum bilirubin değerlerindeki artışın yaşam süresinde kısaltmaya neden olacağı söylenebilir.

Parametrik bileşik modeller arasında en iyi modeli belirlemek için AIC ve BIC değerlerine bakılmıştır. Modellerin AIC ve BIC değerleri Çizelge 5.9’de verilmiştir.

**Çizelge 5.9.** Parametrik bileşik modeller için AIC ve BIC değerleri

<b>Model</b>	<b>AIC</b>	<b>BIC</b>
Üstel	4738.236	4894.281
<b>Weibull</b>	<b>4721.314</b>	<b>4882.982</b>
Log-normal	4742.735	4904.352
Log-lojistik	4727.06	4888.678
Gamma	4722.846	4884.464

Çizelge 5.9’da verilen model karşılaştırma kriterlerine göre, yaşam çözülmesi için Weibull alt model ve boylamsal gözlemler için doğrusal karma etkili modelin paylaşılmış parametre modeli ile birleştirilmesiyle elde edilen modelin en iyi model olduğu söylenebilir. Boylamsal alt model sonuçlarına göre, kadınların yıllık serum bilirubin değerlerinin erkeklere göre 0.49 birim arttığı, karaciğer yetmezliği sebebiyle karın bölgesinde su birikintisi olanların yıllık serum bilirubin değerlerinin olmayanlara göre 0.14 birim arttığı, deride kan bozukluğu olanların olmayanlara göre yıllık serum bilirubin değerlerinin 0.11 birim arttığı, ilaç kullanımına rağmen ellerinde ayaklarında şişlik olanların olmayanlara göre yıllık serum bilirubin değerlerinin 0.40 birim arttığı, ellerinde ve ayaklarında şişlik olanların olmayanlara göre yıllık serum bilirubin değerlerinin 0.65 birim arttığı ve hastalığın 4. evresinde olanların hastalığın 1. evresinde olanlara göre yıllık serum bilirubin değerlerinin 0.22 birim arttığı yorumları yapılabilir. Weibull alt model sonuçlarına göre yaş ilerledikçe yaşam süresinin kısaldığı, karaciğer yetmezliği sebebiyle karında su birikintisi olanların olmayanlara göre yaşam sürelerinin  $\exp(0.3434)=1.41$  kat daha kısa olduğu, karaciğer büyümesi olanların olmayanlara göre yaşam sürelerinin  $\exp(0.7130)=2.04$  ve ilaç kullandığı halde ellerinde ve ayaklarında şişlik olanların olmayanlara göre yaşam süresinin  $\exp(0.6588)=1.93$  kat daha kısa, ellerinde ve ayaklarında şişlik olanların olmayanlara göre yaşam sürelerinin  $\exp(1.5090)=4.52$  kat daha kısa olduğu söylenebilir. Assoc kat sayısına göre serum bilirubin değerleri ile yaşam süresi arasında güçlü bir ilişki bulunmaktadır. Serum bilirubin değerlerindeki artışın yaşam süresinde kısaltmaya neden olduğu yani ölüm riskinde artışa neden olduğu söylenebilir.

## 5.2. Boylamsal ve Yaşam Verilerinin Ayrı Çözümlemesi

Bileşik modelden elde edilen sonuçları kıyaslayabilmek için, PBS veri setine boylamsal analiz için kullanılan doğrusal karma etkili model uygulanmıştır ve model sonuçları Çizelge 5.10'da verilmiştir.

**Çizelge 5.10.** Doğrusal karma etkili model sonuçları

	$\hat{\beta}$	Std. hata	z	p
Sabit Terim	-0.0679	0.3925	-0.17	0.863
Yaş	-0.0132	0.0049	-2.67	<b>0.008</b>
Cinsiyet2	0.4760	0.1539	3.09	<b>0.002</b>
İlaç2	-0.0166	0.0978	-0.17	0.866
Asit2	1.1179	0.2717	4.11	<b>0.000</b>
Hepatome2	0.7778	0.1539	5.05	<b>0.000</b>
DeriBoz2	0.4243	0.1468	2.89	<b>0.000</b>
Ödem2	0.2944	0.1786	1.65	0.100
Ödem3	0.4256	0.2645	1.61	0.109
HistolojikEvre2	0.0515	0.3805	1.35	0.177
HistolojikEvre3	0.6385	0.3459	1.85	0.066
HistolojikEvre4	0.7478	0.3483	2.15	<b>0.033</b>

Çizelge 5.10'da verilen doğrusal karma etkili model sonuçlarına göre; yaş, cinsiyet, asit, hepatomegali, deride bozukluk, ödem ve histolojik evre4 değişkenleri istatistiksel olarak anlamlı bulunmuştur.

Çizelge 5.9'da verilen bileşik model karşılaştırmaları sonuçlarına göre en iyi model, Weibull parametrik bileşik model olduğundan dolayı, sonuçları kıyaslayabilmek için yaşam verilerinin ayrı analizi Weibull parametrik model ile yapılmıştır ve sonuçlar Çizelge 5.11'de verilmiştir.

**Çizelge 5.11.** Weibull parametrik model sonuçları

	$\hat{\beta}$	Std. hata	z	p
Sabit Terim	7.8327	0.9469	8.27	<b>0.000</b>
Yaş	-0.0309	0.0100	-3.08	<b>0.002</b>
Cinsiyet2	-0.4469	0.2399	-1.86	0.062
İlaç2	-0.1999	0.1761	-1.13	0.256
Asit2	-0.7177	0.2333	-3.08	<b>0.002</b>
Hepatome2	0.5186	0.1986	2.61	<b>0.009</b>
DeriBoz2	-0.0140	0.1945	-0.07	0.943
Ödem2	-0.8790	0.2462	-3.57	<b>0.000</b>
Ödem3	-1.7537	0.2927	-5.99	<b>0.000</b>
HistolojikEvre2	0.6030	0.9179	0.66	0.511
HistolojikEvre3	-0.3688	0.7476	-0.49	0.622
HistolojikEvre4	-1.0932	0.7361	-1.49	0.138
Log(serBilir)(t)	-0.0001	0.0000	-2.92	<b>0.004</b>

Çizelge 5.11’de verilen Weibull parametrik model sonuçlarına göre; yaş, asit, hepatomegali, ödem ve zamana bağımlı serum bilirubin değişkenleri istatistiksel olarak anlamlı bulunmuştur ( $p < \alpha$ ). Weibull parametrik modelin tehlike oranlarının ve parametre tahminlerinin standart hatalarının Weibull parametrik birleşik model sonuçlarından daha yüksek olduğu görülmektedir. Weibull parametrik modele göre, el ve ayaklarında ilaç kullanımına rağmen şişlik olanların olmayanlara göre yaşam süreleri  $\exp(0.8790) = 2.41$  kat daha kısayken, Weibull parametrik bileşik alt model sonucuna göre el ve ayaklarında ilaç kullanımına rağmen şişlik olanların olmayanlara göre yaşam süreleri  $\exp(0.6588) = 1.93$  kat kısadır. Ek olarak, zamana bağımlı serum bilirubin değerleri ile yaşam süresi arasındaki ilişki Weibull parametrik birleşik modele göre daha zayıf bulunmuştur.

Ayrıca Cox regresyon ile doğrusal karma etkili modelin paylaşılmış parametre modeli ile birleştirilmesiyle elde edilen standart bileşik model sonucuna göre yaşam süresi ile serum bilirubin değerleri arasındaki ilişki, Weibull parametrik birleşik modele göre daha yüksek bulunmuştur.

## 6. SONUÇLAR

Bu çalışmada, boylamsal ve yaşam verilerinin bileşik modellemesi ve parametrik bileşik model yapıları verilmiş, yaşam süresinin orantılı tehlikeler varsayımını sağlamadığı durumda kullanılabilir olan parametrik bileşik model yöntemleri anlatılmıştır.

Aynı çalışma kapsamında elde edilen boylamsal ve yaşam verilerinin ayrı analizi sonucunda ölçüm hataları ve kayıp veri sorunları ortaya çıkmaktadır. Bu sorunları ortadan kaldırmak ve yansız ve etkin tahminler elde etmek için bileşik modelleme üzerine çalışmalar son zamanlarda giderek artmaktadır. Genel olarak kullanılan standart bileşik model yapısı, boylamsal gözlemlerin doğrusal karma etkili model ve yaşam verilerinin Cox regresyon modeliyle paylaşılmış parametre modelleri yardımıyla birleştirilmesine dayanmaktadır. Ancak yaşam verilerine Cox regresyon modelinin uygulanabilmesi için orantılı tehlikeler varsayımının sağlanması gerekmektedir. Varsayımın sağlanmadığı durumlarda tabakalandırılmış Cox regresyon, genelleştirilmiş Cox regresyon veya yaşam süresinin bilinen bir dağılıma uyması durumunda parametrik regresyon yöntemleri kullanılmalıdır. Varsayım sağlanmadığı durumlarda yansız ve güvenilir sonuçlar elde edebilmek için bileşik model yaşam çözülmesi alt modeli için Üstel, Weibull, Log-normal, Log-lojistik ve Gamma parametrik alt modelleri tercih edilebilir.

Uygulamada öncelikle standart bileşik model yapısı incelenmiştir. Bu model kurulmadan önce orantılı tehlikeler varsayımı Schoenfeld artıkları ve grafiksel yöntemle test edilmiş ve orantılı tehlikeler varsayımının sağlanmadığı görülmüştür. Ancak parametrik bileşik modelleme sonuçları ile kıyaslama yapabilmek için Cox regresyon ve doğrusal karma etkili modelin bileşik analiz sonuçları verilmiştir. Daha sonra yaşam çözümlemesi için Üstel, Weibull, Log-normal, Log-lojistik ve Gamma alt modelleri ve doğrusal karma etkili modelin paylaşılmış parametre modeli ile birleştirilmesiyle elde edilen parametrik bileşik model sonuçları verilmiştir. Bu modellerden en iyi modele karar vermek için bilgi kriterlerine bakılmış ve Weibull parametrik bileşik model en iyi model olarak bulunup, sonuçlar ayrıntılı olarak yorumlanmıştır. Daha sonra boylamsal ve yaşam verileri için ayrı ayrı doğrusal karma etkili model ve Weibull parametrik regresyon modeli kurulmuştur. Weibull parametrik bileşik modelle kıyaslayabilmek için Weibull parametre modeline boylamsal gözlem olan serum bilirubin değeri zamana

bağlı değişken olarak eklenmiştir. Model sonuçları incelendiğinde yaşam verileri için ayrı analiz sonuçlarında, parametre standart hatalarının ve tehlike oranlarının bileşik model yaşam çözümlemesi kısmından daha büyük olduğu görülmüştür. Ayrıca, serum bilirubin değerleri ile yaşam süresi arasındaki ilişki ayrı analizde bileşik analize göre daha düşük olarak bulunmuştur.

Yaşam verileri için Cox regresyon modeli kullanılarak elde edilen standart bileşik model sonuçlarıyla Weibull parametrik bileşik model sonuçları kıyaslandığında, tehlike oranlarının ve parametre standart hatalarının standart bileşik modelde daha yüksek olduğu görülmüştür. Ayrıca serum bilirubin değerleri ile yaşam süresi arasındaki ilişki Weibull parametrik model de daha yüksek bulunmuştur.

Sonuç olarak, boylamsal gözlem ve yaşam verilerinin birlikte olduğu çalışmalarda ayrı analiz yapmak uygun değildir. İki veri arasındaki ilişkiyi doğru şekilde analiz etmek için bileşik modelleme kullanılmalıdır. Ancak, Cox regresyon modeli ile doğrusal karma etkili modelin bileşik modellemesi kullanılmadan önce orantılı tehlikeler varsayımının incelenmesi gerekmektedir. Orantılı tehlikeler varsayımı sağlanmadığında ve verilerin bilinen dağılımlara uygunluğu söz konusu olduğu durumlarda, yansız ve etkin parametre tahminleri için bileşik modelin yaşam çözümlemesi alt modeli parametrik regresyon yöntemleri ile yapılmalıdır.

## KAYNAKLAR

- Ahmed, F. E., Vos, P. W. and Holbert, D., Modeling Survival in Colon Cancer: A Methodological Review, *Molecular Cancer*, 6 (2007) 15.
- Albert P. S. and Shih J. H., An Approach For Jointly Modeling Multivariate Longitudinal Measurements and Discrete Time-To-Event Data, *The Annals of Applied Statistics*, 4 (2010) 1517-1532.
- Allison, P. D., *Survival Analysis Using SAS: A Practical Guide*, SAS Institute Inc., Vol. 2, North Carolina, USA, 2010.
- Asimow, L. A. and Maxwell, M. M., *Probability and Statistics with Applications: A Problem Solving Text*, ACTEX Publications Inc., Vol. 2, New Hartford, 2015.
- Ata, N., Karasoy, D. ve Sözer, M. T., Orantısız Hazardlar için Parametrik ve Yarı Parametrik Yaşam Modelleri, *İstatistikçiler Dergisi*, 1 (2008) 125-134.
- Ballinger, G. A., Using Generalized Estimating Equations for Longitudinal Data Analysis, *Organizational Research Methods*, 7 (2004) 127-150.
- Bates, D., Machler, M., Bolker, B. M., and Walker, S. C., Fitting Linear Mixed-Effects Models Using lme4, *Journal of Statistical Software*, 67 (2015) 1-48.
- Bewick, V., Cheek, L., and Ball, J., *Statistics Review 12: Survival Analysis*. *Critical Care*, 8 (2004) 389-394.
- Bilmes, J. A., A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Technical Report ICSI-TR-97-021, University of Berkeley, 1998.
- Brown, E. R. and Ibrahim, J. G., A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data. *Biometrics*, 59 (2003) 221-228.
- Borman, S., The Expectation Maximization Algorithm, A short tutorial, Retrieved from, 2006  
[http://www.csd.uwo.ca/faculty/olga/Courses/Fall2006/Papers/EM\\_algorithm.pdf](http://www.csd.uwo.ca/faculty/olga/Courses/Fall2006/Papers/EM_algorithm.pdf)
- Booth, J. and Hobert, J., Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm, *Journal of The Royal Statistical Society, Series B*, 61 (1999) 265–285.



- Chernick, M. R. and Friis R. H., *Introductory Biostatistics for The Health Sciences: Modern Application Including Bootstrap*, John Wiley and Sons, Vol. 1, Hoboken, New Jersey, **2003**.
- Christensen, E., Schlichting, P., Fauerholdt L., Gluud, C., Andersen, P. K., Juhl, E., Poulsen, H. and Tygstrup, N., Prognostic Value of Childturtcotte Criteria in Medically Treated Cirrhosis, *Hepatology* 4 (**1984**) 430-435.
- Collett D., *Modelling Survival Data in Medical Research*, Chapman and Hall/ CRC Press, Vol. 3, U.S.A, **2003**.
- Cox, D. R., *Regression Models and Life-Tables (With Discussion)*. *Journal of the Royal Statistical Society, Series B*, 34 (**1972**) 187-220.
- Cox, D. and Hinckley, D., *Theoretical Statistics*, Chapman and Hall/ CRC Press, Vol. 1, U.S.A, **1974**.
- Crowder, M., On the use of a working correlation matrix in using generalised linear models for repeated measures, *Biometrika*, 82 (**1995**) 407–410.
- Crowther, M. C., Abramsa, K. R. and Lambert P. C., Flexible Parametric Joint Modelling of Longitudinal and Survival Data, *Statistics in Medicine*, 31 (**2012**) 4456-4471.
- Crowther, M. C., Abrams K. R. and Lambert P. C., Joint Modeling of Longitudinal and Survival Data, *The Stata Journal*, 13 (**2013**) 165-184.
- Dafni, U. G. and Tsiatis, A. A., Evaluating Surrogate Markers of Clinical Outcome when Measured with Error, 54 (**1998**) 1445:1462.
- Danalioğlu, A., Karacığer Sirozu Nedeniyle Ex Olmuş 100 Vakanın Retrospektif Değerlendirilmesi, *Uzmanlık Tezi*, SSK Göztepe Hastanesi 2. Dahiliye Kliniği, İstanbul, **1995**.
- DeGruttola, V. and Tu, X. M., Modelling Progression of CD4-Lymphocyte Count and its Relationship to Survival Time, *Biometrics*, 50 (**1994**) 1003-1014.
- Dempster, A., Laird, N. and Rubin, D., Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Series B*, 39 (**1977**) 1-38.
- Dobson, A. and Henderson, R., Diagnostics for Joint Longitudinal and Dropout Time Modeling, *Biometrics*, 59 (**2003**) 741-751.

- Doğanay, B., Uzunlamasına Çalışmaların Analizinde Karma Etki Modelleri, Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara, **2007**.
- Diaz, L. C., Joint Modelling for Longitudinal and Time-to-Event Data. Application to Liver Transplantation Data, Master Thesis, University of Santiago de Compostela, Spain, **2014**.
- Diggle, P., Heagerty, P., Liang, K. Y., and Zeger, S. L., Analysis of Longitudinal Data. Oxford University Press. Vol. 2, New York, **2002**.
- Efendi, A., Molenberghs, G., Njagi E. D. and Dendale, P., A Joint Model for Longitudinal Continuous and Time-To-Event Outcomes with Direct Marginal Interpretation, *Biometrical Journal*, 55 (**2013**) 572–588.
- Elashoff, R., Li, G. and Li, N., A Joint Model for Longitudinal Measurements and Survival Data in The Presence of Multiple Failure Types, *Biometrics*, 64 (**2008**) 762–771.
- Esquivel, C. O., Thiel D. H. V., Demetris, A. J., Bernardos, A., Iwatsuki, S., Markus, B., Gordon, R. D., et al., Transplantation for Primary Biliary Cirrhosis, *Gastroenterology*, 94 (**1988**) 1207-1216.
- Evans M., Hastings N. and Peacock B., *Statistical Distributions*, Wiley and Sons, Vol. 3, New York, **2000**.
- Faucett, C. and Thomas, D., Simultaneously Modelling Censored Survival Data and Repeatedly Measured Covariates: A Gibbs Sampling Approach, *Statistics in Medicine*, 15 (**1996**) 1663–1685.
- Fitzmaurice, G. M., Laird, N. M. and Ware, J. H., *Applied Longitudinal Analysis*, John Wiley and Sons, Vol. 2, New York, **2011**.
- Fox, J., *Cox Proportional-Hazards Regression for Survival Data, Appendix to An R and SPLUS Companion to Applied Regression*, SAGE Publications, Vol. 1, California, (**2008**).
- Fox, J. and Weisberg, S., *Cox Proportional-Hazards Regression for Survival Data in R, an Appendix to An R Companion to Applied Regression*, SAGE Publications, Vol. 2, California, (**2011**).
- Ghisletta, P. and Spini, D., An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals, *Journal of Educational and Behavioral Statistics*, 29 (**2004**) 421–437.

- Givens, G. H. and Hoeting, J. A., Computational Statistics, John Wiley and Sons, Vol. 2, Hoboken, New Jersey, **2013**.
- Gordon, R. D., Iwatsuki, S., Tzakis, A. G., Esquivel, C. O., Todo, S., Makowka, L. and Starzl, T. E., The Denver-Pittsburgh Liver Transplant Series, National Library of Medicine National Institutes of Health, (**1987**) 43-50.
- Guo, X. and Carlin, B. P., Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages, The American Statistician, 58 (**2004**) 16-24.
- Harville, D. A., Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems, Journal of The American Statistical Association, 72 (**1977**) 320-338.
- Hedeker D. And Gibbons R.D., Longitudinal Data Analysis, John Wiley and Sons, Vol. 1, Hoboken, New Jersey, **2006**.
- Henderson, R., Diggle, P. and Dobson, A., Joint Modeling of Longitudinal Measurements and Event Time Data, Biostatistics, 1 (**2000**) 465-480.
- Hickey, G. L., Philipson, P., Jorgensen, A. and Dona, R. K., Joint Modelling of Time-to-Event and Multivariate Longitudinal Outcomes: Recent Developments and Issues, BMC Medical Research Methodology, 16 (**2016**) 1-15.
- Hickey, G. L., Philipson, P., Jorgensen, A. and Dona R. K., joineRML: A Joint Model and Software Package for Time-To-Event and Multivariate Longitudinal Outcomes, BMC Medical Research Methodology, 18 (**2018**) 1:24.
- Hsieh, F., Tseng, Y. K. and Wang, J. L. Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited, Biometrics, 62 (**2006**) 1037-1043.
- Hsieh, F., Ding, J. and Wang, J. L., Method of Sieves to Jointly Model Survival and Longitudinal Data, Statistica Sinica, 23 (**2013**) 1181-1213.
- Hosmer D. W., Lemeshow S. and May S., Applied Survival Analysis: Regression Modeling of Time-to-Event Data, John Wiley and Sons, Vol. 2, Hoboken, New Jersey, **2008**.
- Hu, W. H., Li, G. and Li, N., A Bayesian Approach to Joint Analysis of Longitudinal Measurements and Competing Risks Failure Time Data, Statistics in Medicine, 28 (**2009**) 1601- 1619.

- Ibrahim, J. G., and Molenberghs, G., Missing Data Methods in Longitudinal Studies: A Review, *An Official Journal of the Spanish Society of Statistics and Operations Research*, 18 (2009) 1–43.
- Ibrahim, J. G., Chu, H. and Chen, L. M., Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data, *Journal of Clinical Oncology*, 28 (2010) 2796-2801.
- Kalkan, S. B., *Boylamsal Verilerde Semiparametrik Karma Etki Modelleri ve Bir Uygulama, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 2014.*
- Karasoy, D. ve Tutkun, N. A., *Yaşam Çözümlemesi, 1. Baskı, Nobel Akademik Yayıncılık, 2016.*
- Klein, J. P. and Moeschberger, M. L., *Survival Analysis Techniques for Censored and Truncated Data, Springer-Verlag, Vol. 2, New York, 1997.*
- Kleinbaum, D. G. and Klein, M., *Survival analysis: A Self-Learning Text, Springer, Vol. 3, New York, 2012.*
- Laird, N. M. and Ware, J. H., Random-Effects Models for Longitudinal Data, *Biometrics*, 38 (1982) 963-974.
- Lange, K., *Numerical Analysis for Statisticians, Springer, Vol. 2, New York, 2010.*
- Lesaffre, E. and Spiessens, B., On The Effect of the Number of Quadrature Points in a Logistic Random-Effects Model: An Example, *Journal of the Royal Statistical Society, Series C*, 50 (2001) 325–335.
- Lewis, T. H., *Complex Survey Data Analysis with SAS, CRC Press Taylor and Francis Group, Vol. 1, Virginia, 2017.*
- Li, N., Elashoff, R. M., Li, G., and Saver, J., Joint Modeling of Longitudinal Ordinal Data and Competing Risks Survival Times and Analysis of the NINDS RT-PA Stroke Trial, *Statistics in Medicine*, 29 (2009) 546-557.
- Liang, K. and Zeger, S. L., Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, 73 (1986) 13-22.
- Lima, C. P. and Taylor, J. M. G., Development and Validation of a Dynamic Prognostic Tool for Prostate Cancer Recurrence Using Repeated Measures of Posttreatment PSA: A Joint Modeling Approach, *Biostatistics*, 10 (2009) 535-549.

- Lima, C. P., Joly, P., Dartigues, J. and Gadda, H. J., Joint Modelling of Multivariate Longitudinal Outcomes and a Time-To-Event: A Nonlinear Latent Class Approach, *Computational Statistics and Data Analysis*, 53 (2012) 1142-1154.
- Liu, X., *Survival Analysis: Models and Applications*, John Wiley and Sons, Vol. 1, Hoboken, New Jersey, 2012.
- Liu, F. and Li, Q., A Bayesian Model for Joint Analysis of Multivariate Repeated Measures and Time to Event Data in Crossover Trials, *Statistical Methods in Medical Research*, 25 (2014) 2180-2192.
- Mayo Clinic, Primary Biliary Cirrhosis, sequential data, <https://www.mayo.edu/research/documents/pscseqhtml/doc-10027141>.
- Mbohning, C., Bleakley, K. and Lavielle, M., Joint Modeling Of Longitudinal And Repeated Time-To-Event Data using Nonlinear Mixed-Effects Models and the SAEM algorithm, *Journal of Statistical Computation and Simulation*, 85 (2015) 1512-1528.
- McCrink, L. M., Marshall, A. H. and Cairns, K. J., Advances in Joint Modelling: A Review of Recent Developments with Application to the Survival of End Stage Renal Disease Patients, *International Statistical Review*, 81 (2013) 249-269.
- McCool, J., *Using the Weibull Distribution: Reliability, Modelin and Inference*, John Wiley and Sons, Vol. 1, Hoboken, New Jersey, 2012.
- McCulloch, C., Maximum likelihood algorithms for generalized linear mixed models, *Journal of the American Statistical Association*, 92 (1997) 162–170.
- Mondal, P. K., *Joint Modeling of Longitudinal Measurements and Survival Data with Competing Risks: Application to Hiv/Aids Study*, Doctora Thesis, University of Saskatchewan, Canada, 2017.
- Murtaugh, P., Dickson, E. R., Dam, G. V., Malinchoc, M., Grambsch, P. M., Langworthayn A. and Gips, C. H., Primary Biliary Cirrhosis: Prediction of Short-Term Survival Based On Repeated Patient Visits, *Hepatology*, 20 (1994) 126–134.
- Pasha, T. M. and Dickson, E. R., *Survival Algorithms and Outcome Analysis in Primary Biliary Cirrhosis*, *Seminars in Liver Disease*, 17 (1997) 147-158.
- Pauler, D. K. and Finkelstein, D. M., Predicting Time to Prostate Cancer Recurrence Based on Joint Models for Non-Linear Longitudinal Biomarkers and Event Time Outcomes, *Statistics in Medicine*, 21 (2002) 3897-3911.

- Pericleous, P., Parametric Joint Modelling for Longitudinal and Survival Data, Doctoral Thesis, School of Computing Sciences University of East Anglia, UK, **2016**.
- Persson, I., Essays on the Assumption of Proportional Hazards in Cox Regression. Doctoral Thesis, Uppsala University, Sweden, **2002**.
- Pinheiro, J. and Bates, D., Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model, *Journal of Computational and Graphical Statistics*, 4 (**1995**) 12–35.
- Pinheiro, J. and Chao, E., Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models, *Journal of Computational and Graphical Statistics*, 15 (**2006**) 58–81.
- Pintilie, M., *Competing Risks: A Practical Perspective*, John Wiley and Sons, Vol. 1, New York, **2006**.
- Press, W., Teukolsky, S., Vetterling, W. and Flannery, B., *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Vol. 3, New York, **2007**.
- Rizopoulos, D., Verbeke, G. and Molenberghs, G., Shared Parameter Models under Random Effects Misspecification, *Biometrika* 95 (**2008**) 63-74.
- Rizopoulos, D., Verbeke, G. and Lesaffre, E., Fully Exponential Laplace Approximations for the Joint Modelling of Survival and Longitudinal Data, *Journal of the Royal Society Statistical Methodology*, 71 (**2009**) 637-654.
- Rizopoulos, D., JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data, *Journal of Statistical Software*, 35 (**2010**) 1-33.
- Rizopoulos, D., Verbeke, G. and Molenberghs, G., Multiple-Imputation-Based Residuals and Diagnostic Plots for Joint Models of Longitudinal and Survival Outcomes, *Biometrics*, 66 (**2010**) 20-29.
- Rizopoulos, D., *Joint Models for Longitudinal and Time-to-Event Data with Applications in R*, CRC Press Taylor and Francis Group, Vol. 1, Boca Raton, **2012a**.
- Rizopoulos, D., Fast fitting of Joint Models for Longitudinal and Event Time Data Using a Pseudo-Adaptive Gaussian Quadrature Rule, *Computational Statistics and Data Analysis*, 56 (**2012b**) 491-501.
- Rizopoulos, D., The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC, *Journal of Statistical Software*, 72 (**2016**) 1 45.

- Rodriguez, G., Parametric Survival Models, Technical report, Princeton University, New Jersey, **2010**.
- San, E. C. H., Joint Modelling of Survival and Longitudinal Data Under Nested Case Control Sampling, Master Thesis, National University of Singapore, Singapore, **2013**.
- Self, S. and Pawitan, Y., Modeling a Marker of Disease Progression and Onset of Disease, *AIDS Epidemiology: Methodological Issues*, Boston, Birkhuser, 231-255, **1992**.
- Singh, R. and Mukhopadhyay, K., Survival Analysis in Clinical Trials: Basics and must Know Areas, *Perspectives in Clinical Research*, 2 (**2011**) 145–148.
- Song, X., Davidian, M. and Tsiatis, A., A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-To-Event Data, *Biometrics*, 58 (**2002**) 742-753.
- Sonsuz, A., Karacığer Sirozunda Hasta Takibi ve Klinik Sorunlar, *İ.Ü. Cerrahpaşa Tıp Fakültesi Sürekli Tıp Eğitimi Etkinlikleri*, 58 (**2007**) 99-112.
- Sousa, I., A Review on Joint Modelling of Longitudinal Measurements and Time-to-Event, *REVSTAT-Statistical Journal*, 9 (**2011**) 57-81.
- Steensmeier J. M. and Jones B. S., *Event History Modeling: A Guide For Social Scientists*, Cambridge University Press, Vol. 1, Avenue of the Americans, New York, **2004**.
- Su, Y. R. and Wang, J. L., Modeling Left-Truncated and Right-Censored Survival Data with Longitudinal Covariates, *The Annals of Statistics*, 40 (**2012**) 1465-1488.
- Sullivan, L. M., *Essentials of Biostatistics in Public Health*, Jones and Bartlett Learning, Vol. 2, Sudbury, **2012**.
- Sweeting, M. J. and Thompson, S. G., Joint Modelling of Longitudinal and Time-to-Event Data with Application to Predicting Abdominal Aortic Aneurysm Growth and Rupture, *Biometrical Journal*, 53 (**2011**) 750-763.
- Tang, A. M. and Tang, N. S., Semiparametric Bayesian Inference on Skew-Normal Joint Modeling of Multivariate Longitudinal and Survival Data, *Statistics in Medicine*, 34 (**2015**) 824-843.

- Thevaraja, M., Survival Analysis for Primary Biliary Cirrhosis Data by using SAS, 2018. [https://www.researchgate.net/publication/323915012\\_Survival\\_Analysis\\_or\\_Primary\\_Biliary\\_Cirrhosis\\_Data\\_by\\_using\\_SAS](https://www.researchgate.net/publication/323915012_Survival_Analysis_or_Primary_Biliary_Cirrhosis_Data_by_using_SAS)
- Tseng, Y., Hsieh, F. and Wang, J., Joint Modelling of Accelerated Failure Time and Longitudinal Data, *Biometrika*, 92 (2005) 587-603.
- Tsiatis, A. A., DeGruttola, V. and Wulfsohn, M. S., Modeling the Relationship of Survival to Longitudinal Data Measured with Error: Applications to Survival and CD4 Counts in Patients with AIDS, *Journal of the American Statistical Association*, 90 (1995) 27-37.
- Tsiatis, A. A. and Davidian, M., A Semiparametric Estimator for the Proportional Hazards Model with Longitudinal Covariates Measured with Error, *Biometrika*, 88 (2001) 447-458.
- Turgal, E., Sağlık Verileri ile Boylamsal Verilerin Birleşik Modellenmesi ve Sağlık Alanına Bir Uygulama, Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, 2016.
- Turgal, E., Erdoğan, B. D. ve Köse, S. K., Tekrarlı Gözlem ve Sağlık Verilerinin Birleşik Model Performans Değerlendirmesi: Bir Simülasyon Çalışması, XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi, Antalya, 25-28 Ekim 2017, Belek Antalya, 2017, 10-11.
- Ülgen E. ve Asar, Ö., Tekrarlı Gözlem ve Sağlık Verilerinin Bileşik Modellenmesi, *Türkiye Klinikleri Biyoistatistik Dergisi* 9 (2017) 156-166.
- Verbeke, G. and Molenberghs, G., *Linear Mixed Models for Longitudinal Data*, Springer-Verlag, Vol. 1, New York, 2000.
- Vittinghoff, E., Shiboski, S. C., Glidden, D. V. and McCulloch, C. E., *Regression Methods in Biostatistics: Linear, Logistics, Survival and Repeated Measures Models*, Springer-Verlag, Vol. 1, New York, 2005.
- Wang, Y. and Taylor, J. M. G., Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immundeficiency Syndrome, *Journal of the American Statistical Association*, 96 (2001) 895-905.
- Wang, M., Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments, *Advances in Statistics*, 2014 (2014) 1-11.
- Wedderburn, R. W., Quasi-Likelihood Functions, Generalized Linear Models and The Gauss-Newtonmethod, *Biometrika*, 61 (1974) 439-447.



- Williamson, P. R., Kolamunnage-Dona, R., Philipson, P. and Marson, A. G., Joint Modeling of Longitudinal and Competing Risks Data, *Statistics in Medicine*, 27 (2008) 6426-6438.
- Wu, L., Liu, W., Yi, G. Y. and Huang, Y., Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods and Issues, *Journal of Probability and Statistics*, 2012 (2012) 1:17.
- Wulfsohn, M. S. and Tsiatis, A. A., A Joint Model for Survival and Longitudinal Data Measured with Error, *Biometrics*, 53 (1997) 330-339.
- Yan, Y. F., Survival Analysis Final Report: Data Analysis for the Sequential Primary Biliary Cirrhosis Data, 2011. <http://www.stat.ncu.edu.tw/teacher/Tsengyk/final.pdf>
- Ye, W., Lin, X. and Taylor, J. M. G., Semiparametric Modeling of Longitudinal Measurements and Time-To-Event Data: A Two-Stage Regression Calibration Approach, *Biometrics*, 64 (2008a) 1238-1246.
- Ye, W., Lin, X. and Taylor, J. M. G., A Penalized Likelihood Approach to Joint Modeling of Longitudinal Measurements and Time-To-Event Data, *Statistics and Its Interface*, 1 (2008b) 33-45.
- Yu, M., Law, N. J., Taylor, J. M. G. and Sandler, H. M., Joint Longitudinal-Survival-Cure Models and their Application to Prostate Cancer, *Statistica Sinica*, 14 (2004) 835-862.
- Zeng, D. and Cai, J., Asymptotic Results For Maximum Likelihood Estimators in Joint Analysis of Repeated Measurements and Survival Time, *Annals of Statistics*, 33 (2005) 2132-2163.

## **EKLER**

### **EK 1 - Tezden Türetilmiş Bildiriler**

DİL, E. and KARASOY, D., Parametric Joint Modelling for Longitudinal and Survival Data, 11. International Statistics Days Conference, Muğla, 2018.



HACETTEPE ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ  
FEN BİLİMLER ENSTİTÜSÜ  
İSTATİSTİK ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 15/02/2019

Tez Başlığı / Konusu: BOYLAMSAL YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİ

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 53 sayfalık kısmına ilişkin, 15/02/2019 tarihinde ~~şahsım~~/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 7 'dir.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/dâhil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

15.02.2019

Adı Soyadı: Elif Dil  
Öğrenci No: N17123726  
Anabilim Dalı: İstatistik  
Programı: Yüksek Lisans  
Statüsü:  Y.Lisans  Doktora  Bütünleşik Dr.

**DANIŞMAN ONAYI**

UYGUNDUR.

Prof. Dr. Duru Karasoy

## **ÖZGEÇMİŞ**

Adı Soyadı : Elif DİL  
Doğum yeri : Eminönü  
Doğum tarihi : 13/11/1993  
Medeni hali : Evli  
Yazışma adresi : Necip Fazıl Kısakürek Mh. Esenyurt/ İstanbul  
Telefon : 0555 878 81 93  
Elektronik posta adresi : elifdil@hacettepe.edu.tr  
Yabancı dili : İngilizce - 72.5

## **EĞİTİM DURUMU**

Lisans : 2011 - 2015 Sinop Üniversitesi İstatistik  
Yüksek Lisans : 2017 - 2019 Hacettepe Üniversitesi İstatistik

## **İŞ TECRÜBESİ**

2010 – 2011 THY

