

**BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK  
BİLEŞİK MODELLEMESİNDE PARAMETRE TAHMİN  
YÖNTEMLERİ**

**PARAMETER ESTIMATION METHODS IN  
PARAMETRIC JOINT MODELLING OF  
LONGITUDINAL AND SURVIVAL DATA**

**ELİF YILDIRIM**

**PROF. DR. DURU KARASOY**

**Tez Danışmanı**

Hacettepe Üniversitesi

Lisansüstü Eğitim - Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı için Öngördüğü

DOKTORA TEZİ olarak hazırlanmıştır.

2024



Canım annem ve babama,

## ÖZET

# BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİNDE PARAMETRE TAHMİN YÖNTEMLERİ

**Elif YILDIRIM**

**Doktora, İstatistik Bölümü**

**Tez Danışmanı: Prof. Dr. Duru KARASOY**

**Aralık 2024, 91 Sayfa**

Özellikle biyomedikal ve klinik araştırmalarda, boylamsal ve yaşam verileri sıklıkla birlikte elde edilmektedir. Genellikle bu iki veri ayrı analiz edildiğinde, yaşam verileri için Cox regresyon modeli ve parametrik regresyon modelleri kullanılırken, boylamsal veriler için doğrusal karma etkili model ve genelleştirilmiş tahmin denklemleri kullanılmaktadır. Ancak boylamsal ve yaşam verilerinin ilişkili olduğu durumlarda, birimlerden farklı zaman noktalarında elde edilen boylamsal ölçümlerde, aynı birimin yaşam süresi içerisinde takibi gerçekleştirilememesinden kaynaklı rastgele olmayan kayıp veri oluşmaktadır. Bu nedenle bazı araştırmacılar, bu iki verinin ayrı analizine alternatif olarak rastgele olmayan kayıp verileri de sürece dahil eden bileşik modellemeyi önermişlerdir.

Bileşik modellemede temel amaç boylamsal ve yaşam verileri arasındaki ilişkiyi tahmin etmek ve bu iki veri üzerine bağımsız değişkenlerin etkisini araştırmaktır. Bileşik model temel olarak boylamsal alt model ve yaşam süresi alt modelinden oluşmaktadır. Literatürde, farklı yapılarıdaki boylamsal ve yaşam verilerinin birleşik modellemesi üzerine birçok model geliştirilmiştir. Ayrıca, farklı model yapılarının yanı sıra, bileşik

model parametre tahminlerini elde etmek için iki - aşamalı yaklaşım, olabilirlik yaklaşımı ve Bayesci yaklaşımlar gibi farklı yöntemler geliştirilmiştir. Olabilirlik yaklaşımında, bileşik model olabilirlik fonksiyonunda yer alan integrallerin kapalı formda çözümü olmadığından Laplace, Gauss Hermite ve uyarlanmış Gauss Hermite gibi farklı sayısal yaklaşımlar önerilmiştir.

Bileşik modelleme ve parametre tahmin yaklaşımları üzerine yapılan çalışmaların çoğunda boylamsal veriler için doğrusal karma etkili model ile yaşam verileri için Cox regresyon modelinin, paylaşılmış rastgele etkiler ile birleştirilmesinden elde edilen bileşik model yapısı kullanılmaktadır. Ancak, yaşam verilerinin orantılı tehlikeler varsayımını sağlamadığı ve belirli bir dağılıma uygun olduğu durumlarda, daha yansız ve güvenilir tahminler elde edebilmek için parametrik bileşik modellemeler kullanılmalıdır.

Bu kapsamda bu tez çalışmasında, yaşam verilerinin orantılı tehlikeler varsayımını sağlamadığı ve belirli bir dağılıma uygun olduğu durumlarda kullanılan parametrik bileşik model parametre tahminlerini hesaplamak için Bayesci iki - aşamalı yaklaşım önerilmiştir. Önerilen yaklaşım, literatürde sıklıkla kullanılan Gauss Hermite ve uyarlanmış Gauss Hermite yaklaşımları ile kıyaslanmıştır. Yöntemlerin performansını incelemek için farklı örneklem genişliklerinde, durdurma oranlarında ve parametre değerlerinde simülasyon senaryoları gerçekleştirilmiştir. Simülasyon sonuçlarına göre, önerilen Bayesci iki - aşamalı yaklaşımın tüm senaryolarda Gauss Hermite ve uyarlanmış Gauss Hermite yaklaşımlarına kıyasla daha iyi sonuçlar verdiği görülmüştür. Riskin yüksek olduğu durumlarda ise tüm yöntemlerin riskin düşük ve orta düzey olduğu durumlara kıyasla daha yanlı tahminler verdiği tespit edilmiştir. Aynı zamanda tüm yöntemlerin örneklem genişliği arttıkça ve durdurma oranı azaldıkça daha az yanlı tahminler verdiği ve grup değişkenini birleştirme parametresine kıyasla daha az yanlı tahmin ettiği görülmüştür.

Önerilen yaklaşımın uygulanabilirliğini göstermek amacıyla R programında yer alan aort kapakçığı değiştirme ameliyatı veri seti ve AIDS veri setinden, simülasyon çalışmasında dikkate alınan parametrik bileşik model yapısına uygun olarak rastgele örneklem seçilmiş ve yöntemlerin performansları kıyaslanmıştır. Her iki uygulama sonucunda en iyi model tahminleri simülasyon sonuçlarına benzer şekilde Bayesci iki - aşamalı yaklaşımdan elde edilmiştir.

**Anahtar Kelimeler:** Parametrik bileşik model, Hızlandırılmış başarısızlık süresi modelleri, Bayesci iki - aşamalı yaklaşım, Gauss Hermite, uyarlanmış Gauss Hermite

Bu tez çalışması Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) - ARDEB 1002 - A kapsamında desteklenmiştir (Proje No: 124F132).



## **ABSTRACT**

# **PARAMETER ESTIMATION METHODS IN PARAMETRIC JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA**

**Elif YILDIRIM**

**Doctor of Philosophy, Department of Statistics**

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

**December 2024, 91 Pages**

Especially in biomedical and clinical research, longitudinal and survival data are often obtained together. Generally, when these two data are analyzed separately, while Cox regression model and parametric regression models are used for survival data, linear mixed effect model and generalized estimating equations are used for longitudinal data. However, in cases where longitudinal and survival data are related, non-random missing data occurs due to the inability to follow up the same unit during survival in longitudinal measurements obtained from units at different time points. For this reason, some researchers have suggested joint modeling that includes non - random missing data as an alternative to separate analysis of these two data.

The main purpose of joint modelling is to estimate the relationship between longitudinal and survival data and to investigate the effect of independent variables on these two data. The joint model basically consists of a longitudinal sub-model and a survival sub-model. In the literature, many models have been developed for joint modelling of longitudinal and survival data with different structures. In addition to different model structures, different methods such as two - stage approach, likelihood approach and Bayesian approaches have been developed to obtain parameter estimates of the joint model. In the likelihood approach, since the integrals in the joint model likelihood function do not have

a closed-form solution, different numerical approaches such as Laplace, Gauss Hermite and adapted Gauss Hermite have been proposed.

In most of the studies on joint modeling and parameter estimation approaches, the joint model structure obtained by combining the linear mixed - effect model for longitudinal data and the Cox regression model for survival data with shared random effects is used. However, in cases where survival data does not satisfy the proportional hazards assumption and are suitable for a certain distribution, parametric joint models should be used to obtain more unbiased and reliable estimates.

In this context, in this thesis, a Bayesian two - stage approach is proposed to calculate the parameter estimates of the parametric joint model used in cases where the survival data does not satisfy the proportional hazards assumption and is suitable for a certain distribution. The proposed approach was compared with the Gauss Hermite and adaptive Gauss Hermite approaches, which are frequently used in literature. Simulation scenarios were performed with different sample sizes, censoring rates and parameter values to examine the performance of the methods. According to the simulation results, it was seen that the proposed Bayesian two - stage approach gave better results compared to the Gauss Hermite and adapted Gauss Hermite approaches in all scenarios. In cases where the risk is high, it was determined that all methods gave more biased estimates compared to cases where the risk is low and medium. It was also observed that all methods gave less biased estimates as the sample size increased and the censoring rate decreased and estimated the group variable less biased compared to the association parameter.

To demonstrate the applicability of the proposed approach, a random sample was selected from the aortic valve replacement surgery dataset and the AIDS dataset in the R program in accordance with the parametric joint model structure considered in the simulation study and the performances of the methods were compared. As a result of both applications, the best model estimations was obtained from the Bayesian two-stage approach, similar to the simulation results.

**Keywords:** Parametric joint modelling, Accelerated failure time models, Bayesian two - stage approach, Gauss Hermite, adapted Gauss Hermite

This thesis was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) - ARDEB 1002 - A (Project No: 124F132).



## TEŞEKKÜR

Akademik birikimi ve insani değerleriyle bana yol gösteren, lisansüstü eğitimim boyunca üzerimde çok fazla emeği olan, beni her zaman destekleyen ve öğrencisi olmaktan gurur duyduğum saygı değer hocam Prof. Dr. Duru KARASOY'a,

Tez izleme komiteleri boyunca, değerli önerileri ve rehberlikleriyle tezimin şekillenmesine büyük katkı sağlayan, her adımda beni destekleyen kıymetli hocalarım Prof. Dr. Yasemin YAVUZ ve Prof. Dr. Nihal ATA TUTKUN'a ve tezimin başarısına önemli katkılarda bulunan değerli jüri üyelerim Prof. Dr. Mehtap AKÇİL OK ve Prof. Dr. Nursel KOYUNCU'ya,

Tez sürecimin tamamlanmasında büyük destek sağlayan ve büyük özveride bulunan Prof. Dr. Aleksandar S. NASTIC'e,

Hayatımı her zaman güzelleştiren, her konuda desteğini hissettiğim ve bana her zaman motivasyon veren Hasan KOCADAYIOĞULLARI'na ve akademik çalışmalarımda fikir ve önerileri ile yanımda olan ve her zaman motivasyonumu yükseltmemeye yardımcı olan Eda Selin ILIKKAN BAŞDOĞAN'a,

Hayatımın her aşamasında bana sonsuz sevgi ve güven ile destek veren ve bu süreci kolaylıkla bitirmemde büyük emekleri olan canım annem, babam ve abime,

SONSUZ TEŞEKKÜRLERİMİ SUNARIM.

# İÇİNDEKİLER

ÖZET .....	i
ABSTRACT .....	iv
TEŞEKKÜR.....	vii
İÇİNDEKİLER.....	viii
ÇİZELGELER DİZİNİ .....	xi
ŞEKİLLER DİZİNİ .....	xii
KISALTMALAR.....	xiii
1. GİRİŞ.....	1
2. BOYLAMSAL VERİ .....	8
2.1. Genelleştirilmiş Doğrusal Modeller .....	8
2.2. Doğrusal Karma Etkili Model .....	10
2.2.1. Doğrusal Karma Etkili Modelde Parametre Tahmini .....	12
2.3. Varyans Kovaryans Yapısı .....	13
2.3.1. Rastgele Kesişim Modeli .....	13
2.3.2. Rastgele Kesişim ve Rastgele Eğim Modeli .....	14
3. YAŞAM ÇÖZÜMLEMESİ.....	16
3.1. Cox Regresyon Modeli .....	19
3.2. Orantılı Tehlikeler Varsayımının İncelenmesi .....	21
3.2.1. Log-Log Grafiği .....	21
3.2.2. Schoenfeld Artıkları .....	22
3.3. Yaşam Çözümlemesinde Kullanılan Parametrik Modeller .....	23
3.3.1. Weibull Orantılı Tehlikeler Modeli .....	24
3.3.2. Hızlandırılmış Başarısızlık Süresi Modelleri .....	25
4. BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİ .....	28

<b>4.1. Yaşam Süresi Alt Modeli</b> .....	30
<b>4.1.1. Parametrik Yaşam Süresi Alt Modelleri</b> .....	31
<b>4.1.2. Üstel Hızlandırılmış Başarısızlık Süresi Alt Modeli</b> .....	32
<b>4.1.3. Weibull Hızlandırılmış Başarısızlık Süresi Alt Modeli</b> .....	32
<b>4.1.4. Log-Lojistik Hızlandırılmış Başarısızlık Süresi Alt Modeli</b> .....	33
<b>4.1.5. Log-Normal Hızlandırılmış Başarısızlık Süresi Alt Modeli</b> .....	33
<b>4.2. Boylamsal Alt Model</b> .....	34
<b>4.3. Bileşik Model Parametre Tahmini</b> .....	35
<b>4.3.1. İki - Aşamalı Yaklaşım</b> .....	36
<b>4.3.2. Gauss Hermite Yaklaşımı</b> .....	37
<b>4.3.3. Uyarlanmış Gauss Hermite Yaklaşımı</b> .....	39
<b>5. BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİNDE BAYESÇİ YAKLAŞIM</b> .....	41
<b>5.1. Bayesci Yaklaşım</b> .....	41
<b>5.2. MCMC Yöntemi</b> .....	42
<b>5.3. MCMC Yönteminin Performansı</b> .....	43
<b>5.4. Bayesci İki-Aşamalı Yaklaşım</b> .....	44
<b>6. SİMÜLASYON ÇALIŞMASI VE UYGULAMA</b> .....	46
<b>6.1. Simülasyon Çalışması</b> .....	46
<b>6.1.1. Parametrik Bileşik Model için Veri Simülasyonu</b> .....	46
<b>6.1.2. Simülasyon Algoritması</b> .....	48
<b>6.1.3. Simülasyon Çalışması Sonuçları</b> .....	49
<b>6.2. Uygulama</b> .....	66
<b>6.2.1. Uygulama 1</b> .....	66
<b>6.2.2. Uygulama 2</b> .....	71
<b>7. SONUÇ VE TARTIŞMA</b> .....	76
<b>8. KAYNAKLAR</b> .....	80

<b>EKLER</b> .....	88
<b>EK 1 - Tezden Türetilmiş Yayınlar</b> .....	88
<b>EK 2 - Tezden Türetilmiş Bildiriler</b> .....	89
<b>EK 3 - Tez Çalışması Orijinallik Raporu</b> .....	90
<b>ÖZGEÇMİŞ</b> .....	91



## ÇİZELGELER DİZİNİ

Çizelge 3.1. Yaşam çözümlenmesinde yaygın olarak kullanılan parametrik dağılımlar...	24
Çizelge 6.1. Senaryo 1’de $\alpha$ parametresi için ortalama mutlak yan ve HKO değerleri....	50
Çizelge 6.2. Senaryo 1’de $\gamma$ parametresi için ortalama mutlak yan ve HKO değerleri.....	53
Çizelge 6.3. Senaryo 2’de $\alpha$ parametresi için ortalama mutlak yan ve HKO değerleri....	57
Çizelge 6.4. Senaryo 2’de $\gamma$ parametresi için ortalama mutlak yan ve HKO değerleri.....	59
Çizelge 6.5. Senaryo 3’te $\alpha$ parametresi için ortalama mutlak yan ve HKO değerleri.....	63
Çizelge 6.6. Senaryo 3’te $\gamma$ parametresi için ortalama mutlak yan ve HKO değerleri...	65
Çizelge 6.7. Aort kapakçığı değiştirme ameliyatı veri setinde nitel değişkenlere ilişkin betimsel istatistikler.....	67
Çizelge 6.8. Aort kapakçığı değiştirme ameliyatı veri setinde sürekli değişkenlere ilişkin betimsel istatistikler.....	68
Çizelge 6.9. Aort kapakçığı değiştirme ameliyatı veri seti için Weibull HBS parametrik bileşik modelin parametre tahmin değerleri.....	70
Çizelge 6.10. AIDS veri setinde nitel değişkenlere ilişkin betimsel istatistikler.....	72
Çizelge 6.11. AIDS veri setinde sürekli değişkenlere ilişkin betimsel istatistikler.....	72
Çizelge 6.12. AIDS veri seti için Weibull HBS parametrik bileşik modelin parametre tahmin değerleri.....	74

## ŞEKİLLER DİZİNİ

Şekil 6.1. Senaryo 1 için ölçeklendirilmiş Schoenfeld artıkları testinden elde edilen p - değerleri.....	50
Şekil 6.2. Senaryo 1’de $\alpha$ parametresi için kestirim değerleri.....	51
Şekil 6.3. Senaryo 1’de $\alpha$ parametresi için HKO değerleri.....	51
Şekil 6.4. Senaryo 1’de $\gamma$ parametresi için kestirim değerleri.....	53
Şekil 6.5. Senaryo 1’de $\gamma$ parametresi için HKO değerleri.....	54
Şekil 6.6. Senaryo 2 için ölçeklendirilmiş Schoenfeld artıkları testinden elde edilen p - değerleri.....	56
Şekil 6.7. Senaryo 2’de $\alpha$ parametresi için kestirim değerleri.....	57
Şekil 6.8. Senaryo 2’de $\alpha$ parametresi için HKO değerleri.....	58
Şekil 6.9. Senaryo 2’de $\gamma$ parametresi için kestirim değerleri.....	60
Şekil 6.10. Senaryo 2’de $\gamma$ parametresi için HKO değerleri.....	60
Şekil 6.11. Senaryo 3 için ölçeklendirilmiş Schoenfeld artıkları testinden elde edilen p - değerleri.....	62
Şekil 6.12. Senaryo 3’te $\alpha$ parametresi için kestirim değerleri.....	63
Şekil 6.13. Senaryo 3’te $\alpha$ parametresi için HKO değerleri.....	64
Şekil 6.14. Senaryo 3’te $\gamma$ parametresi için kestirim değerleri.....	65
Şekil 6.15. Senaryo 3’te $\gamma$ parametresi için HKO değerleri.....	66
Şekil 6.16. hs değişkeni için a) log – logS grafiği b) Dönüştürülmüş zamana karşı ölçeklendirilmiş Schoenfeld artıklarının grafiği.....	69
Şekil 6.17. Yaşam süresi (fuyrs) için Q-Q grafiği.....	69
Şekil 6.18. İlaç değişkeni için a) log – logS grafiği b) Dönüştürülmüş zamana karşı ölçeklendirilmiş Schoenfeld artıklarının grafiği.....	73
Şekil 6.19. Yaşam süresi için Q-Q grafiği.....	73

## KISALTMALAR

DKEM	Doğrusal Karma Etkili Model
GDM	Genelleştirilmiş Doğrusal Model
HBS	Hızlandırılmış Başarısızlık Süresi
MCMC	Markov Chain Monte Carlo
EM	Expectation Maximization
GH	Gauss Hermite
AGH	Adaptive Gauss Hermite
BTS	Bayesian Two – Stage (Bayesci İki – Aşamalı Yaklaşım)
EÇO	En Çok Olabilirlik
HKO	Hata Kareler Ortalaması
Lvmi	sol ventrikül kitle indeksi (left ventricular mass index)
ddC	Didanosine
ddI	Zalcitabine
AIDS	Acquired Immune Deficiency Syndrome (Edinilmiş Bağışıklık Eksikliği Sendromu)
AIC	Akaike Bilgi Kriteri
BIC	Bayesci Bilgi Kriteri



# 1. GİRİŞ

Aynı birimlerden belirli zaman aralıklarında tekrarlı olarak ölçüm yapılması ile elde edilen boylamsal (longitudinal) veriler ile ve herhangi bir olayın gerçekleşmesine kadar geçen süreden oluşan yaşam (survival) verileri ile özellikle biyoloji, çevre ve sağlık - medikal alanlarında sıklıkla karşılaşılmaktadır (Basharat, 2019). Bu iki farklı veri yapısını analiz etmek için literatürde birçok yöntem bulunmaktadır. Boylamsal verilerin analizinde en çok kullanılan yöntemler doğrusal karma etkili modeller (DKEM) ve geliştirilmiş tahmin denklemleriyken (Mondal, 2017), yaşam verileri için Cox regresyon modeli ve parametrik regresyon modelleridir (Mao, 2016).

Bir birimin farklı zaman noktalarında tekrarlı ölçümleri alındığında, aynı birimin yaşam süresi içerisinde takibi gerçekleştirilemediği zaman drop out olarak adlandırılan rastgele olmayan kayıp veriler ortaya çıkmaktadır. Boylamsal ve yaşam verileri ayrı ayrı analiz edildiğinde ortaya çıkan bu rastlantısal olmayan kayıp veri, tahmin sürecine dahil edilmeyeceğinden bazı araştırmacılar tarafından bu iki verinin bileşik modellenmesi önerilmiştir (Guo ve Carlin, 2004).

Bileşik modelleme, boylamsal olarak ölçülen içsel kovaryantlara veya ölçüm hatalı boylamsal değişkenlere sahip yaşam verilerini; ölçüm hataları veya kayıp gözlemlere sahip boylamsal verileri ve bir birimin yaşam süresi üzerine boylamsal değişkenlerin etkisini araştırmada, boylamsal ve yaşam verilerini aynı anda analiz etmek için kullanılan istatistiksel yöntemler bütünüdür (Tsiatis, DeGruttola ve Wulfsohn, 1995; Wulfsohn ve Tsiatis 1997; McCrink, Marshall ve Cairns, 2013). Bu yöntem, özellikle biyomedikal ve epidemiyolojik çalışmalarda bireylerden ölçülen tekrarlı gözlemlerin yaşam sürelerine etkisini incelemek için sıklıkla kullanılmaktadır. Örneğin, AIDS çalışmalarında zaman içerisinde tekrarlı olarak ölçülen ve bağışıklık sisteminin bir göstergesi olarak kullanılan CD4 hücre sayımlarının, hastalığın nüksetmesi veya bireylerin hastalıktan ölene kadar geçen süre üzerine etkilerinin araştırılması (Self ve Pawitan, 1992; DeGruttola ve Tu, 1994; Tsiatis, DeGruttola ve Wulfsohn, 1995); kanser çalışmalarında zaman içerisinde tekrarlı olarak kaydedilen ölçümlerin hastalığın nüksetme süresine etkisinin araştırılması (Ibrahim, Chen ve Sinha, 2004; Yu ve ark., 2004; Chi ve Ibrahim, 2006) ve böbrek hastalarında, böbrek fonksiyonlarındaki zaman içindeki değişimlerin hastaların yaşam

sürelerine etkisinin araştırılması (Rizopoulos ve Ghosh, 2011; Hu, Li ve Greene, 2016) gibi çeşitli alanlarda kullanılmaktadır.

Bileşik modelleme boylamsal ve yaşam süresi alt modellerinden oluşmakta ve bu iki alt modeli birleştirmek için farklı rastgele etki yapıları kullanılmaktadır (McCrink, Marshall ve Cairns, 2013; Proust-Lima ve ark., 2014). Genellikle literatürde boylamsal alt model için DKEM ve yaşam süresi alt modeli için Cox regresyon modeli kullanılsa da son yirmi yılda artan ilgiyle birlikte farklı bileşik model yapıları üzerine çalışmalar olduğu görülmektedir. Tek değişkenli boylamsal ve yaşam verilerinin bileşik modellemesi 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 gerçekleştirilmiştir. Self ve Pawitan (1992), tek değişkenli bileşik model parametre tahminlerini hesaplamak için iki - aşamalı yaklaşımı geliştirmişlerdir. Buna göre, birinci aşamada boylamsal alt modelden en küçük kareler yöntemi ile rastgele etkileri tahmin etmişler ve elde ettikleri bu tahminleri ikinci aşamada Cox regresyon modeli olabilirlik fonksiyonunda kullanarak, boylamsal gözlemlerin tahminlerini hesaplamışlardır.

Birimlerin yaşam süreleri içerisinde birden fazla olay yaşama riskini içeren yarışan riskler durumu ile boylamsal verilerin bileşik modellemesi üzerine ilk çalışmalar Elashoff, Li ve Li (2008), Williamson ve ark. (2008) ve Hu, Li ve Li (2009) tarafından gerçekleştirilmiştir. Bu çalışmalarda bileşik model genellikle boylamsal alt model için DKEM ve yarışan riskler içeren yaşam verileri için orantılı sebebe - özel tehlikeler (proportional cause-specific hazards) alt modelinin çeşitli rastgele etkiler ile birleştirilmesiyle elde edilmiştir. Ancak yarışan riskler üzerine yapılan bazı çalışmalar, yarışan riskler için standart regresyon modelinin (orantılı nedene özgü tehlikeler modeli) kullanılması durumunda, belirli bir başarısızlık türünün sebebe - özel tehlike fonksiyonu üzerindeki etkisinin, kümülatif insidans fonksiyonu üzerindeki etkisinden çok farklı olabileceğini göstermiştir (Lunn ve McNeil, 1995; Fine ve Gray, 1999; Putter, Fiocco ve Geskus, 2007). Bu kapsamda, Deslandes ve Chevret (2010) boylamsal veriler için DKEM ve yarışan riskler durumundan oluşan yaşam verileri için orantılı alt dağılım tehlikeler alt modelinin (proportional subdistribution hazards submodel) paylaşılmış rastgele etkiler ile birleştirilmesine dayanan yeni bir bileşik model yapısı önermişlerdir. Önerdikleri modelin parametre tahminlerini elde etmek için Gibbs örnekleme Markov Zinciri Monte Carlo (MCMC) tekniğini kullanmışlardır. Modelin performansını incelemek için simülasyon

çalışması gerçekleştirmişler ve sebebe – özel tehlikeler alt modeli ile karşılaştırmışlardır. Teixeira ve ark. (2019) yarışan riskler içeren yaşam verileri ve boylamsal verilerin bileşik modellemesi için R programındaki sistematik uygulamaları kullanarak farklı parametrelendirmeleri incelemişler ve model parametrelerinin farklı yorumlarını göstermişlerdir. Thomadakis ve ark. (2023) yarışan riskler içeren yaşam verileri ve boylamsal gözlemlerin bileşik modellemesi için kümülatif insidans fonksiyonları kullanarak esnek bir model önermişlerdir. Bileşik modelde başarısızlık nedenlerinin yanlış sınıflandırılmasını dikkate alan çift örnekleme yöntemi ile modeli oluşturmuşlar ve model parametrelerini elde etmek için Bayesci yaklaşımı uygulamışlardır.

Özellikle klinik çalışmalarda çoklu boylamsal veriler, soldan veya sağdan kesilmiş veya kesikli yaşam verileri gibi farklı veri yapıları ile sıklıkla karşılaşılmaktadır. Literatürde yaşam ve boylamsal verilerin farklı yapıları için geliştirilmiş çeşitli bileşik model yapıları bulunmaktadır. Brown, Ibrahim ve DeGruttola (2005), çok değişkenli boylamsal veriler ve yaşam verilerinin bileşik modellemesi için çok değişkenli boylamsal verilerin kübik B-spline modeli ve yaşam verileri için Cox regresyon modeline dayanan yeni bir bileşik model yapısı önermişlerdir. Model parametre tahminlerini elde etmek için MCMC yaklaşımını kullanmışlar ve koşullu tahmini ordinat ve sapma bilgi kriterlerini kullanarak gerçekleştirdikleri simülasyon çalışmasıyla kübik B-spline modeli düğüm sayılarını belirlemişlerdir. Albert ve Shih (2010) kesikli yaşam verileri ve çok değişkenli boylamsal ölçümlerin bileşik modellemesinde yeni bir regresyon kalibrasyon yöntemi önermişlerdir. Çalışmalarında geleneksel iki - aşamalı yaklaşımda meydana gelen bilgilendirici kayıp veri nedeniyle oluşan yanlılığı azaltmak için ilk aşamada çoklu boylamsal gözlemlerin koşullu dağılımını tahmin etmede ikili değişkenli DKEM kullanmışlar ve ikinci aşamada bu tahminleri kullanarak iki veri arasındaki ilişkiyi tahmin etmişlerdir. Su ve Wang (2012) boylamsal veriler için DKEM ve soldan kesilmiş yaşam verileri için Cox regresyon modelinin bileşik modellemesinde, soldan kesilmiş veri varlığında ortaya çıkan tahmin yanlılığını azaltmak için modifiye edilmiş tam olabilirlik yaklaşımını önermişlerdir. Efendi ve ark. (2013) sürekli boylamsal veriler ve tekrarlı yaşam verileri için marjinalleştirilmiş bileşik model yapısını önermişlerdir. Alam ve ark. (2021) ise sürekli ve sıralı boylamsal veriler ve yaşam verilerinin paylaşılmış rastgele etkiler ile birleştirilmesine dayanan farklı bileşik model yapısı önermişlerdir. Çalışmalarında, sürekli boylamsal veriler için DKEM, sıralı boylamsal veriler için orantılı odds karma modeli ve yaşam verileri için hızlandırılmış başarısızlık süresi (accelerated failure time -

HBS) modellerini dikkate almışlardır. Önerdikleri bileşik model parametre tahminlerini maksimum olabilirlik yaklaşımı ile elde etmişler ve farklı örneklem genişliklerinde modellerin performanslarını incelemişlerdir.

Literatürde farklı yapılardaki boylamsal ve yaşam verilerinin bileşik modellemesi üzerine çok sayıda çalışma bulunmasına rağmen, orantılı tehlikeler varsayımının sağlanmadığı yaşam verileri durumunda kullanılacak alternatif modeller üzerine geliştirilen bileşik model yapılarıyla ilgili az sayıda çalışma bulunmaktadır. Yaşam verilerinin orantılı tehlikeler varsayımını sağlamadığı ve belirli bir dağılıma uygun olduğu durumlarda HBS modelleri, tüm açıklayıcı değişkenleri sürece dahil edecek şekilde formalize edildiğinden ve orantılı modellere göre daha az kısıtlayıcı olduğundan orantılı tehlikeler modellerine alternatif olarak kullanılmaktadır (Zhang ve Wu, 2019; Khan ve Basharat, 2022). Tseng, Hsieh ve Wang (2005) ölçüm hatalı boylamsal veriler için DKEM ve orantılı tehlikeler varsayımını sağlamayan yaşam verileri için HBS modellerinin paylaşılmış rastgele etkiler ile birleştirilmesine dayanan bileşik model yapısını önermişlerdir. Önerdikleri modelin parametre tahminlerini hesaplamak için Monte Carlo EM algoritması yaklaşımını kullanmışlar ve model performansını, simülasyon çalışması ve bir uygulama verisi üzerinde incelemişlerdir. Pericleous (2016) çalışmasında boylamsal gözlemler için DKEM ve yaşam verileri için HBS modellerinin, paylaşılmış rastgele etkiler ile birleştirilmesine dayanan parametrik bileşik modelleri incelemiştir. Weibull HBS alt modeli ile oluşturduğu bileşik modelin parametre tahminleri için Gauss Hermite yaklaşımını kullanmış ve yöntemin performansını farklı örneklem genişliklerinde oluşturduğu simülasyon senaryolarında incelemiştir. Yıldırım ve Karasoy (2020) DKEM ve parametrik yaşam verilerinin bileşik modellemesi için Stata programında “gsem” kodunu geliştirmişlerdir. Khan ve Basharat (2022) boylamsal gözlemler için DKEM ve yaşam verileri için HBS modellerinden oluşan bileşik model yapısını incelemişlerdir. Weibull ve Log-Lojistik dağılımlarına uygun veri üreterek gerçekleştirdikleri simülasyon çalışmasında, HBS modellerinin performansını değerlendirmişlerdir.

Bileşik modellerin karmaşık yapısından dolayı araştırmacıların bir diğer odak noktası, model parametrelerini tahmin etmek olmuştur. Bileşik model parametre tahminleri genellikle olabilirlik fonksiyonunun maksimize edilmesine dayanan en çok olabilirlik yaklaşımı ile elde edilmektedir. Ancak bileşik modelin olabilirlik fonksiyonu karmaşık bir yapıya sahiptir ve en çok olabilirlik yaklaşımı ile parametre tahminlerinin elde

edilmesi için olabilirlik fonksiyonundaki yüksek katlı integrallerin hesaplanması gerekmektedir. Olabilirlik fonksiyonunun karmaşık yapısı ve kapalı çözümü olmadığından dolayı literatürde farklı parametre tahmin yaklaşımları önerilmiştir. Faucett ve Thomas (1996) Gibbs örnekleme MCMC tekniğini kullanarak bileşik model parametre tahminlerini elde etmişlerdir. Wulfsohn ve Tsiatis (1997) boylamsal ve yaşam verilerinin bileşik model parametrelerini eş zamanlı olarak tahmin edebilmek için tam olabilirlik yaklaşımını kullanmışlardır. Song, Davidian ve Tsiatis (2002) boylamsal ve yaşam verilerini birleştiren rastgele etkilerin normallik varsayımını gerektirmeyen yarı parametrik bir model kullanarak, bileşik model parametre tahminlerini elde etmek için EM (Expected-Maximization) algoritması yaklaşımından yararlanmışlardır. Li ve ark. (2009) yarışan riskler ve sıralı boylamsal gözlemlerden oluşan bileşik model parametre tahminlerini hesaplamak için EM algoritmasını kullanmışlardır. Wu, Liu ve Hu (2010) bileşik model parametre tahminlerini hesaplamak için Laplace yaklaşımı ve Monte Carlo EM yöntemini kullanmışlar ve yöntemlerin performansını, simülasyon çalışması ve gerçek bir veri seti üzerinde incelemişlerdir. Rizopoulos (2012b) bileşik model olabilirliğinin çözümü için uyarlanmış Gauss-Hermite yaklaşımının yeniden ölçeklendirilmesi ile elde edilen yeni bir integrasyon yaklaşımı önermiştir. Rizopoulos (2016) sürekli ve kategorik boylamsal veriler ile yaşam verilerinin bileşik modellemesine ek olarak çeşitli model yapılarında parametre tahminlerini elde etmek için MCMC algoritmalarını kullanarak Bayesci tahminleri hesaplayan “JMBayes” isimli R paketini geliştirmiştir. Bu paket ayrıca, iki sürecin dinamik tahminlerini hesaplamaya ve tahminlerin kalibrasyonu ile ayrımı için doğrulama işlemlerini gerçekleştirmeye olanak sağlamaktadır. Furgal, Sen ve Taylor (2019) bileşik modelde parametre tahminlerini elde etmek için R, SAS ve Stata programlarında kullanılan klasik ve Bayesci modeller için bir derleme çalışması gerçekleştirerek, her prosedürün açıklaması, tahmin teknikleri, girdi ve veri gereksinimlerini detaylı olarak incelemişlerdir. Ayrıca yazılımların uygulamaları ve güçlü ve zayıf yönlerini simülasyon çalışmasıyla karşılaştırmışlardır. Atlı (2020) DKEM ile Cox regresyon modelinin bileşik modellemesinde parametre tahminlerini elde etmek için klasik yaklaşım ile Bayesci yaklaşımı karşılaştırmıştır. Çalışmada farklı örneklem genişlikleri ve varyans büyüklüklerinde tahmin yöntemlerinin performansını incelemiştir. Leiva-Yamaguchi ve Alvares (2020), boylamsal ve yaşam süresi verilerinin ortak modellenmesinde daha yansız ve tutarlı parametre tahminleri elde etmek için iki - aşamalı yeni bir yaklaşım önermişlerdir. Önerilen yaklaşımlarını Bayes çerçevesindeki bileşik olabilirlik yaklaşımı ve standart iki - aşamalı yaklaşımlarla karşılaştırmışlar ve

yöntemlerin performansını incelemişlerdir. Çalışmalarında yaşam verileri için Cox regresyon, boylamsal veriler için ise DKEM kullanmışlardır. Lin ve Luo (2022) yaşam verileri ve çoklu boylamsal verilerinin bileşik model parametre tahminlerini elde etmek için yeni bir yapay sinir ağı mimarisi önermişler ve simülasyon çalışması ile önerdikleri yaklaşımın performansını incelemişlerdir. James ve Philipson (2022) bir veya daha fazla boylamsal veri ile yaşam verilerinin çok boyutlu bileşik modellemesinde klasik yaklaşımlarla parametre tahminlerinin elde edilmesindeki zorluğa değinerek alternatif bir EM algoritması yaklaşımı önermişlerdir.

Literatürde, farklı bileşik model yapılarında daha yansız ve tutarlı sonuçlar elde etmek için parametre tahmin yaklaşımları üzerine çalışmalar bulunsa da yaşam verilerinin orantılı tehlikeler varsayımını sağlamadığı ve bilinen bir dağılıma uygun olduğu durumlarda kullanılan parametrik bileşik modellerin parametre tahminleri üzerine bir çalışma bulunmamaktadır. Bu nedenle bu çalışmada öncelikle parametrik bileşik model parametre tahminlerini hesaplamak için Bayesci iki – aşamalı (BTS) yaklaşım önerilmiştir. Daha sonra parametrik bileşik model parametre tahminlerini elde etmek için literatürde kullanılan klasik yaklaşımlar ve önerilen BTS yaklaşımının hangi örneklem büyüklüğünde ve hangi durdurma oranlarında daha iyi sonuçlar verdiği simülasyon çalışması ve gerçek veri seti üzerinde araştırılmıştır. Bir boylamsal gözlem ve orantılı tehlikeler varsayımını sağlamayan ve Weibull dağılımına uygun yaşam verisinden oluşan bileşik modelin parametre tahminleri için yöntemlerin işlevselliği araştırılmış ve en doğru yöntemin belirlenmesine katkı sağlanmıştır.

Çalışmanın ikinci bölümünde boylamsal veriler ve bu verilerin analizi üzerine temel bilgiler sunulmuştur. Bileşik modelin boylamsal alt modelinde sıklıkla kullanılan ve parametrik regresyon yöntemlerinden olan DKEM ve parametre tahmin yöntemleri detaylı olarak incelenmiştir.

Üçüncü bölümde yaşam çözümlemesi ve yaşam çözümlemesinde kullanılan temel fonksiyonlar detaylı olarak incelenmiştir. Bileşik modelin yaşam süresi alt modelinde sıklıkla kullanılan ve yarı parametrik bir yaklaşım olan Cox regresyon modeli ve orantılı tehlikeler varsayımının sağlanmadığı ve yaşam verilerinin belirli bir dağılıma uygun olduğu durumlarda kullanılan parametrik regresyon modelleri detaylı olarak anlatılmıştır.

Çalışmanın dördüncü bölümünde öncelikle boylamsal veriler için DKEM ve yaşam süresi verileri için Cox regresyon modelinin paylaşılmış rastgele etkiler ile birleştirilmesine dayanan standart bileşik model incelemiştir. İkinci aşamada, yaşam süresi verilerinin orantılı tehlikeler varsayımını sağlamadığı durumlarda alternatif olarak kullanılan parametrik bileşik model yöntemleri detaylı olarak tanıtılmıştır.

Beşinci bölümde parametrik bileşik modellerin parametrik bileşik model tahminlerini elde etmek için kullanılan klasik yaklaşımlar ve tez kapsamında önerilen BTS yaklaşım detaylı olarak sunulmuştur. Ek olarak, bileşik model parametreleri için önsel dağılım kavramı ve MCMC yöntemi Gibbs algoritması incelenmiştir.

Çalışmanın altıncı bölümünde ise simülasyon çalışması ve gerçek veri uygulamasına yer verilmiştir. Simülasyon çalışmasının algoritması verilerek, farklı regresyon katsayıları, örneklem büyüklüğü ve durdurma oranlarında Weibull parametrik bileşik modele uygun veri üretilmiştir. Parametrik bileşik model parametre tahminleri, Gauss Hermite (GH) ve uyarlanmış Gauss Hermite (Adaptive Gauss Hermite - AGH) yaklaşımları ve çalışmada önerilen BTS yaklaşım ile tahmin edilerek, sonuçlar kıyaslanmıştır. Daha sonra, incelenen modele uygun ve literatürde yer alan iki farklı veri seti üzerinde önerilen yöntemin uygulanabilirliği gösterilmiştir. Çalışmanın son bölümünde ise simülasyon ve gerçek veri seti uygulamasından elde edilen sonuçlar üzerinde genel değerlendirmeler yapılmıştır.

## 2. BOYLAMSAL VERİ

Boylamsal veriler, belirli bir takip süresi boyunca aynı birimlere (haneler, ülkeler veya bireyler vb.) ait özelliklerin tekrarlı ölçümünü ifade etmektedir (Sousa, 2011). Boylamsal çalışmalardaki amaç, bağımlı değişkenin zaman içerisindeki değişimini ve bu değişime neden olan faktörleri incelemektir. Boylamsal verilerde belirli birimlerden alınan tekrarlı gözlemler aynı birimde ilişkili ancak birimler arasında bağımsız olduğundan, klasik regresyon veya zaman serileri analizi yerine özel istatistiksel yöntemler kullanılmaktadır (Wu, 2009; Fitzmaurice, Laird ve Ware, 2011).

Boylamsal verilerde birim içi ve birimler arası olmak üzere iki tip değişim bulunmaktadır. Birim içindeki değişimin modellenmesi boylamsal gözlemin zaman içerisindeki değişimini görmeyi sağlarken, birimler arasındaki değişimin modellenmesi birimler arasındaki farkların anlaşılmasını sağlamaktadır (Diggle, Sousa ve Chetwynd, 2008; Wu, 2009). Boylamsal veri analizi için parametrik, yarı parametrik ve parametrik olmayan regresyon yöntemleri gibi farklı yöntemler bulunmakta ve boylamsal veri türüne göre analiz yöntemleri belirlenmektedir. Örneğin; sürekli ve yaklaşık olarak normal dağılıma sahip boylamsal ölçümler için DKEM tercih edilirken, kesikli ve normal dağılıma sahip olmayan boylamsal ölçümler için geliştirilmiş doğrusal karma etkili modeller tercih edilmektedir (Alkathami, 2021). Ek olarak, Liang ve Zeger (1986) kesikli boylamsal verilerin analizi için geliştirilmiş tahmin denklemlerini önermişlerdir. Boylamsal veriler için yaygın olarak kullanılan bu regresyon yöntemleri, temel olarak klasik geliştirilmiş doğrusal modellere (GDM) boylamsal verilerdeki korelasyonu dahil etmek için rastgele etkiler eklenmesiyle elde edilmektedir (Alkathami, 2021).

### 2.1. Genelleştirilmiş Doğrusal Modeller

GDM, bağımlı değişken  $Y$ 'nin ortalamasının, açıklayıcı değişken  $X$ 'in etkilerini temsil eden  $\beta$  parametrelerinin doğrusal bir fonksiyonu olduğunu varsayan ve en yaygın kullanılan istatistiksel modeldir (Alkathami, 2021).

$Y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ , bir birimin  $n_i$  gözleminden oluşan bir vektör ve  $X_i$ , buna karşılık gelen açıklayıcı değişken matrisi olsun. Buna göre GDM aşağıdaki şekilde yazılabilir:

$$Y_i = X_i\beta + \epsilon_i, \quad i = 1, 2, \dots, N \quad (2.1)$$

Burada  $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{in_i})'$  regresyon modelinde rastgele hatalar vektörüdür.

Bağımlı değişken  $Y_i$ , bağımlı değişken ile açıklayıcı değişken arasındaki doğrusal ilişkileri tanımlayan sistematik bileşen ( $X_i\beta$ ) ve hataların dağılımından oluşmaktadır. Genel olarak  $Y_i$ 'nin ortalama  $\mu_i$  ve kovaryans matrisi  $V_i$  ile çok değişkenli normal dağılıma uygun olduğu varsayılır. Yani  $Y_i \sim MVN(\mu_i, V_i)$ 'dir ve  $\mu_i = E(Y_i | X_i) = X_i\beta$  ve  $V_i = Cov(Y_i | X_i)$ 'dir.

Boylamsal çalışmalarda, aynı birimden tekrarlı olarak elde edilen ölçümler genellikle bağımsız değildir. Tekrarlanan gözlemler arasındaki bu bağımlılık, kovaryans matrisi  $V_i$  ile yakalanabilir ve bu durumda köşegen dışı elemanlar sıfıra eşit olmayacaktır. Genel olarak kovaryans matrisi  $V_i$ ,  $i$ . birimin gözlemleri arasındaki farklı ilişki düzeylerini içermektedir ve kovaryantlar açıklayıcı değişkenlere bağlı olabilmektedir (Alkhatami, 2021). Çoğu boylamsal çalışmada, gözlemler birimler arasında değişiklik gösterdiğinden, her birimin kendine özgü parametrelere sahip olmasına izin verilmektedir. Bu kapsamda birime özgü parametreler, kitle parametrelerine rastgele etkiler veya bireysel etkiler olarak eklenerek elde edilmektedir (Wu, 2009). Bir modele eklenen rastgele etkiler, sadece gözlemlerdeki heterojenliği değil aynı zamanda her birim veya küme içindeki tekrarlı ölçümler arasındaki korelasyonu da içermektedir.

Boylamsal veri analizinde sıklıkla kullanılan istatistiksel modeller, klasik regresyon yöntemlerinden aşağıdaki gibi elde edilebilmektedir;

- DKEM; doğrusal regresyon modellerine rastgele etkiler eklenerek elde edilebilir,
- Genelleştirilmiş DKEM; GDM'lere rastgele etkiler eklenerek elde edilebilir,
- Frality modeller; yaşam süresi modellerine rastgele etkiler eklenerek elde edilebilir.

Bu tez kapsamında, parametrik bileşik modelin boylamsal alt modelinde DKEM kullanıldığından gelecek bölümde DKEM, parametre tahmin yöntemleri ve kovaryans matris yapıları detaylı olarak incelenmiştir.

## 2.2. Doğrusal Karma Etkili Model

Boylamsal çalışmalarda, belirli bir zaman içerisinde birimlerden elde edilen tekrarlı gözlemler arasında bağımsızlık varsayımı geçerli olmayacağından dolayı bu verilerin analizinde klasik regresyon yöntemi kullanılmamalıdır. Bu nedenle, boylamsal veri analizinde Laird ve Ware (1982) tarafından önerilen ve birimler arası ve birimler içindeki değişimi içeren DKEM yaygın olarak kullanılmaktadır. DKEM'in temel varsayımı, kitledeki her birimin zaman içinde kendi ortalama yanıt eğrilerine sahip olması gerektiğidir. Burada ortalama yanıt, bir birimden diğerine rastgele değişen regresyon katsayıları ile açıklayıcı değişkenlerin doğrusal bir fonksiyonunu ifade etmektedir (Diggle ve ark., 2002; Fitzmaurice, Laird ve Ware, 2011). Örneğin, kontrol ve tedavi grubundaki hastaların zaman içerisinde kan basıncı değerlerinin ölçüldüğünü varsayalım. Her bir hastanın ortalama kan basıncı doğası gereği birbirinden farklı olacaktır. Bu nedenle, tedavinin etkisini değerlendirmek için her hastaya ilişkin ayrı bir kesişme (intercept) noktasına sahip ve her hastanın ilaca verdiği yanıt bir bireyden diğerine değişeceğinden, her hasta için ayrı eğime sahip bir model kurulmalıdır (Molenberghs ve Verbeke, 2000).

DKEM, modele rastgele etkiler eklenerek doğrusal regresyon analizinin bir uzantısı olarak düşünülebilir (Alkathami, 2021). Bu nedenle hem sabit hem de rastgele etkileri birleştirdiğinden dolayı karma modeller olarak kabul edilmektedir (Wu, 2009; Fitzmaurice, Laird ve Ware, 2011; Rizopoulos, 2012a). Burada sabit etkiler bir değişkenin düzeylerinin önceden belirlendiğini gösterirken, rastgele etkiler değişkenin düzeylerinin daha geniş bir kitleden rastgele olarak seçildiğini göstermektedir (Doğanay, 2007). DKEM'de sabit etkiler tüm birimlerde ortak olup kitle etkisini ifade ederken, rastgele etkiler birimler arasında farklılık göstermekte ve birimlerin kitle ortalamasından sapmalarını ifade etmektedir (Vittinghoff ve ark., 2005; Mondal, 2017).

DKEM, boylamsal ölçümlerin her bir birim için aynı zaman noktalarında ve aynı sayıda ölçüm yapılmasını ifade eden dengeli boylamsal verilerin yanı sıra, her bir birim için farklı zaman noktalarında ve farklı sayıda ölçüm yapılmasıyla elde edilen dengesiz verilerde de kullanılabilirdiğinden boylamsal veri analizinde sıklıkla tercih edilmektedir (Wu, 2009; Fitzmaurice, Laird ve Ware, 2011; Rizopoulos, 2012a).

$t_{ij}$  zamanında  $i$ . birim için  $y_{ij}$ ,  $i = 1, 2, \dots, N; j = 1, 2, \dots, n_{ij}$ .  $Y$  sürekli boylamsal gözlemlerin değerlerini gösterebilir. Belirli zaman noktalarında her birim için tekrarlı olarak ölçülen boylamsal gözlem için DKEM Eşitlik (2.2)'deki gibi yazılabilir (Laird ve Ware, 1982);

$$y_{ij} = x_{1ij}\beta_1 + \dots + x_{pij}\beta_p + z_{1ij}b_{i1} + \dots + z_{qij}b_{iq} + \epsilon_i \quad (2.2)$$

$$b_{ik} \sim N(0, \sigma_k^2), \text{Cov}(b_{ik}, b'_{ik}) = \sigma_{kk}'$$

$$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

Ek olarak, DKEM'in matris formu aşağıdaki gibi yazılabilir:

$$y_i = X_i\beta + Z_i b_i + \epsilon_i, \quad i = 1, 2, \dots, N$$

$$b_i \sim N(0, D), \quad \epsilon_i \sim N(0, R_i) \quad (2.3)$$

Burada,  $y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$   $i$ . birim için  $n_i$  boyutlu boylamsal gözlem vektörü,  $\beta$  sabit etkiler parametreleri için  $p$  boyutlu vektör,  $X_i$  sabit etkilere karşılık gelen  $(n_i \times p)$  boyutlu tasarım matrisi,  $b_i$  birimler arası değişimi karakterize eden  $q$  boyutlu vektör,  $Z_i$  rastgele etkilerin  $(n_i \times q)$  boyutlu tasarım matrisini ve  $\epsilon_i$  birim içi değişimi karakterize eden rastgele hataların  $n_i$  boyutlu vektörünü temsil etmektedir. Eşitlik (2.3)'te verilen modelde, rastgele etkiler  $b_i$  ve rastgele hata  $\epsilon_i$ 'nin 0 ortalama ve sırasıyla  $D$  ve  $R_i$  varyans-kovaryans matrisi ile çok değişkenli normal dağılıma sahip olduğu varsayılır. Kovaryans matrisi  $R_i$ 'nin  $\sigma_\epsilon^2 I_{n_i}$  diyagonal matrisi ile temsil edildiği varsayılır ( $I_{n_i}$ ,  $n_i \times n_i$  boyutlu birim matristir). Birim içi hata  $\epsilon_i$ 'nin diğer değişkenlerden bağımsız olduğu ve rastgele etkiler parametresi  $b_i$ 'nin hata terimi  $\epsilon_i$ 'den bağımsız olduğu varsayılır ( $\text{Cov}(\epsilon_i, b_i) = 0$ ) (Wu, 2009; Fitzmaurice, Laird ve Ware, 2011).

Boylamsal gözlem  $y$ 'nin rastgele etkiler parametresi  $b$ 'ye göre koşullu ortalaması ve varyansı sırasıyla Eşitlik (2.4) ve Eşitlik (2.5)'teki gibi yazılabilir (Atlı, 2020; Alkhatami, 2021).

$$E(y|b) = X\beta + Zb,$$

$$E(y) = E(X\beta + Zb + \epsilon) = X\beta \quad (2.4)$$

$$\begin{aligned} \text{Var}(y) &= V = \text{Var}(X\beta + Zb + \epsilon) \\ &= Z\text{Var}(b)Z' + \text{Var}(\epsilon) \end{aligned}$$

$$= ZDZ' + \sigma_{\epsilon}^2 I_{n_i} = ZDZ' + R \quad (2.5)$$

Buna göre, boylamsal gözlemin koşullu ve marjinal dağılımları sırasıyla aşağıdaki gibi yazılabilir;

$$y_i | b_i \sim N(X_i \beta + Z_i b_i, \sigma_{\epsilon}^2 I_{n_i}),$$

$$y_i \sim N(X_i \beta, V_i), \quad V_i = Z_i D Z_i' + \sigma_{\epsilon}^2 I_{n_i}.$$

### 2.2.1. Doğrusal Karma Etkili Modelde Parametre Tahmini

Literatürde DKEM'in parametre tahminlerini elde etmek için EM algoritması (Laird ve Ware, 1982), Newton-Raphson algoritması (Jennrich ve Schluchter, 1986), en çok olabilirlik (EÇO) ve kısıtlanmış EÇO yöntemi (Wu, 2009; Rizopoulos, 2012a) gibi farklı yöntemler kullanılmaktadır. Olabilirlik fonksiyonunun maksimize edilmesi ile elde edilen EÇO tahminleri, modeldeki gözlemler bağımsız olduğunda her birim için yoğunluk fonksiyonları ile basitçe hesaplanırken, modelde tekrarlı gözlemler olduğunda gözlemler arasında bağımlılık yapısı olacağından bileşik olabilirlik fonksiyonu kullanılarak hesaplanmaktadır (Fitzmaurice, Laird ve Ware, 2011). DKEM'in parametre tahminlerini elde etmek için en çok tercih edilen ve bu çalışmanın boylamsal alt modeli için dikkate alınan EÇO yöntemi ile parametre tahminleri aşağıdaki gibi elde edilebilir.

Eşitlik (2.3)'te verilen DKEM'de boylamsal gözlem  $y$ 'nin  $X\beta$  ortalama ve  $V$  varyans kovaryans matrisi ile normal dağılıma sahip olduğu Eşitlik 2.4 ve 2.5'te gösterilmiştir. Buna göre, boylamsal gözleme ilişkin olasılık yoğunluk ve olabilirlik fonksiyonu sırasıyla,

$$p(y|\beta, V) = (2\pi)^{-\frac{1}{2}} |V|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta)\right)$$

$$L(\theta|y) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta)\right) \quad (2.6)$$

şeklinde yazılabilir (Fitzmaurice, Laird ve Ware, 2011; Atlı, 2020). EÇO yöntemiyle parametre tahminlerini elde etmek için Eşitlik (2.6)'da verilen olabilirlik fonksiyonunun  $\beta$ 'ya göre maksimize edilmesi gerekmektedir. Bu nedenle olabilirlik fonksiyonunun logaritması alınarak hesaplanan log-olabilirlik fonksiyonu aşağıdaki gibi yazılabilir (Diggle ve ark., 2002; Fitzmaurice, Laird ve Ware, 2011).

$$\ell(\theta|y) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|V|) - \frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta) \quad (2.7)$$

Varyans kovaryans matrisi  $V$ 'nin bilindiği varsayımında Eşitlik (2.7)'nin sabit etkiler parametresi  $\beta$ 'ya göre birinci türevi alınıp sıfıra eşitlendiğinde,  $\beta$ 'nın EÇÖ tahmin edici aşağıdaki gibi elde edilir.

$$\left. \frac{\partial \ell}{\partial \beta} \right|_{\beta=\hat{\beta}} = X'V^{-1}y - X'V^{-1}X\hat{\beta} = 0$$

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \quad (2.8)$$

Bu formül  $V$ 'nin bilindiği varsayımla uygulanır ancak uygulamada  $V$  genellikle bilinmez ve bu nedenle veriden tahmin edilerek hesaplanır. Varyans kovaryans matrisinin bilinmediği durumlarda  $\hat{\beta}$ 'yi hesaplamak için öncelikle  $V$ 'nin EÇÖ tahmini elde edilir ve Eşitlik (2.8)'de  $V$  yerine  $\hat{V}$  yazılarak  $\hat{\beta}$  hesaplanır (Rizopoulos, 2012a).

### 2.3. Varyans Kovaryans Yapısı

DKEM'de verilerdeki birimler arası farklılıklar ve zaman içerisindeki değişimler rastgele etkiler ile temsil edilmektedir. DKEM'de kovaryans yapıları ve rastgele etkilerin modele nasıl dahil edildiğine bağlı olarak “rastgele kesişim” ve “rastgele kesişim ve rastgele eğim” modelleri sıklıkla kullanılmaktadır (Fitzmaurice, Laird ve Ware, 2011; Alkhatami, 2021).

#### 2.3.1. Rastgele Kesişim Modeli

DKEM, birimler arasında rastgele farklılıklar gösteren, birime özgü rastgele etkileri içeren karma bir modeldir. Bu model, tüm birimlerin boylamsal gözlemlerinin zaman içinde tam olarak aynı gelişimi sergilediği, ancak birimlerin başlangıç düzeylerinde farklılık gösterdiğini varsaymaktadır. Bu tür modellere, birime özgü sabit terimlerin bulunduğu, ancak eğimin sabit kaldığı “rastgele kesişim modeli” denilmektedir. Modelde, boylamsal gözlemler, doğrusal regresyon modeline, birim düzeyindeki rastgele etkilerin sabit terime eklenmesiyle genişletilirken, eğim parametresine rastgele etkiler dahil edilmemektedir. Buna göre, rastgele kesişim modeli aşağıdaki gibi yazılabilir (Fitzmaurice, Laird ve Ware, 2011; Rizopoulos, 2012a).

$$y_i = \beta_0 + \beta_1 t_{ij} + b_{i0} + \epsilon_{ij}$$

$$b_{i0} \sim N(0, \sigma_b^2), \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n_i \quad (2.9)$$

Burada, i. birim için kesişim (birime özgü parametre)  $\beta_0 + b_{i0}$  olarak tanımlanabilir ve bu parametre, birimler arasındaki rastgele farklılıkları göstermektedir. Yani, i. birimin j. ölçüm zamanındaki yanıtının, kitle ortalaması  $\beta_0 + \beta_1 t_{ij}$ 'den bireysel etki  $b_{i0}$  ve birim içi ölçüm hatası  $\epsilon_{ij}$  kadar saptığı varsayılmaktadır ve  $b_{i0}$  ve  $\epsilon_{ij}$ 'nin rastgele ve birbirlerinden bağımsız olduğu kabul edilmektedir (Fitzmaurice, Laird ve Ware, 2011). Bu modele ilişkin boylamsal gözlem  $Y_{ij}$ 'nin koşullu ve marjinal ortalamaları,

$$E(Y_{ij}|b_i) = x'_{ij}\beta + b_{i0}$$

$$E(Y_{ij}) = x'_{ij}\beta$$

şeklinde yazılabilir. Burada  $x'_{ij} = (1, t_{ij})'$  ve  $\beta = (\beta_0, \beta_1)'$ 'dir. Eşitlik (2.9) için marjinal varyans zaman içinde sabit  $\sigma_b^2 + \sigma_\epsilon^2$  olarak tanımlanırken,  $(y_{ij}, y_{ij}')$  gözlem çifti arasındaki marjinal kovaryans  $\sigma_b^2$ 'ye eşittir. Buna göre rastgele kesişim modelinde boylamsal gözlemler için marjinal kovaryans yapısı,

$$\text{Var}(Y_i) \equiv V_i = \sigma_b^2 J_{n_i} + \sigma_\epsilon^2 I_{n_i}$$

olarak yazılabilir. Burada  $I_{n_i}$ ,  $n_i \times n_i$  boyutunda birim matrisi ifade etmektedir. Marjinal kovaryans  $\sigma_b^2$  verildiğinde, herhangi iki gözlemin tekrarlı ölçümleri arasındaki korelasyon ise,

$$\text{Cor}(Y_{ij}, Y_{ik}) = \rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2}, \quad j \neq k, \quad j, k = 1, 2, \dots, n_i$$

şeklinde ifade edilir. (Alkhatami, 2021). Rastgele birim etkisi  $b_{i0}$ , tekrarlı gözlemler arasında korelasyon oluşturmaktadır. Rastgele etkinin mevcut olmaması ( $\sigma_b^2 = 0$ ) korelasyonun olmadığını gösterirken, birim içi değişimin birimler arasındaki değişimden çok daha küçük olması ( $\sigma_\epsilon^2 \ll \sigma_b^2$ ) korelasyonun yüksek olduğunu göstermektedir (Fitzmaurice, Laird ve Ware, 2011).

### 2.3.2. Rastgele Kesişim ve Rastgele Eğim Modeli

Rastgele kesişim modeli basit yapısı nedeniyle tercih sebebi olsa da tekrarlı gözlemler arasında sabit korelasyon varsayımı uygulamada genellikle geçerlilik göstermemektedir (Rizopoulos, 2012a). Teorikte, her birimin yanıt eğrisinin yalnızca başlangıç seviyesinde değil, takip eden zaman noktalarındaki yanıt değişimlerinde de farklılık göstermesi beklenmektedir. Eşitlik (2.9)'da verilen modelin bir uzantısı olan ve birimler arasındaki

herhangi bir anda yanıtın rastgele değişmesine izin veren model, rastgele kesişim ve rastgele eğim modeli olarak adlandırılır ve,

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} + \epsilon_{ij}$$

$$(b_{i0}, b_{i1}) \sim N(0, D), \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n_i \quad (2.10)$$

şeklinde tanımlanır. Burada,  $y_{ij}$  yanıt,  $t_{ij}$  zaman değişkeni ve  $\epsilon_{ij}$  rastgele hata terimini göstermektedir. Eşitlik (2.10), matris ve vektör notasyonu ile aşağıdaki gibi yazılabilir:

$$y_i = X_i \beta + Z_i b_i + \epsilon_i, \quad i = 1, 2, \dots, N,$$

Burada,

$$X_i = Z_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, b_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix}, d = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}, \quad j = 1, 2, \dots, n_i$$

şeklinde. Buna göre, aynı birime ilişkin herhangi bir çift boylamsal gözlem için marjinal kovaryans yapısı,

$$\text{cov}(Y_{ij}, Y_{ik}') = [1 \quad t_{ij}] D \begin{bmatrix} 1 \\ t_{ij}' \end{bmatrix} + \sigma_\epsilon^2 = d_{11} + (t_{ij} + t_{ij}') d_{12} + t_{ij} t_{ij}' d_{22} + \sigma_\epsilon^2$$

ve marjinal varyans  $\text{Var}(Y_{ij}) = d_{11} + 2t_{ij} d_{12} + t_{ij}^2 d_{22} + \sigma_\epsilon^2$  şeklinde tanımlanabilir (Alkathami, 2021).  $Y_i$ 'nin marjinal ortalama ve kovaryansı,  $t_{ij}$  ölçüm zamanlarının bir fonksiyonudur ve  $Y_i$ 'nin marjinal dağılımı  $Y_i \sim N(X_i \beta, Z_i D Z_i' + \sigma_\epsilon^2)$  olarak tanımlanır. Rastgele kesişim ve eğim modelindeki kovaryans yapısının temel özelliği, değişimin birimler arasındaki farklılıklardan mı yoksa birim içindeki zaman içerisinde olan değişimden mi kaynaklandığını ayırt edebilmesidir (Wu, 2009; Fitzmaurice, Laird ve Ware, 2011).

### 3. YAŞAM ÇÖZÜMLEMESİ

Yaşam çözümlemesi, ilgilenilen bir olay meydana gelene kadar geçen sürenin analizinde kullanılan yöntemler bütünüdür (Lawless, 2011). Burada bağımlı değişken, olay gerçekleşene kadar geçen süredir ve bu süreye genellikle yaşam süresi, olay süresi veya başarısızlık süresi denilmektedir. Yaşam çözümlemesinde başarısızlık süresi; ölene kadar geçen süre, bir tümörün meydana gelme süresi veya kaza süresi gibi klinik ve epidemiyolojik çalışmalardan tanımlanabileceği gibi sosyoloji, pazarlama, ekonomi, sanayi gibi alanlarda boşanma veya evlilik süresi, işsizlerin işe kabul edilme süresi veya bir makinenin arızalanma süresi olarak da tanımlanabilir (Collett, 2015).

Yaşam süresi verilerinin analizinde kullanılan istatistiksel modelleri diğer bilinen yöntemlerden ayıran bazı özellikler bulunmaktadır (Wu ve ark., 2012; Rizopoulos, 2012a; Collett, 2015). Bu özelliklerin ilki, yaşam süreleri pozitif olacağından dolayı genellikle sağa çarpık bir dağılıma sahip olmalarıdır. İkinci özellik, araştırma birimleri çalışmaya farklı zamanlarda katılabileceğinden dolayı birimlerin takip sürelerinin eşit olmamasıdır. Bir diğer özellik ise yaşam sürelerinde durdurulmuş (censored) verilerin olmasıdır. Durdurulmuş veri, tüm birimlerin ilgilenilen olay zamanının tamamen gözlemlenemediği durumları ifade etmekte ve takip kaybı (lost to follow-up), birimlerin herhangi bir nedenden dolayı çalışmadan ayrılması (drop out) ve çalışmanın sona ermesi sebepleri ile ortaya çıkmaktadır (Kleinbaum ve Klein, 2012; Collett, 2015).

Rizopoulos (2012a) yaşam çözümlemesinde, durdurulmuş veri mekanizmasının iki sınıfı olduğunu belirtmiştir. Durdurma mekanizmasının ilk sınıfı, durdurma zamanları ve gerçek yaşam sürelerinin konumlarıyla ilişkidir ve sağ, sol ve aralıklı olmak üzere üç temel durdurma türü bulunmaktadır. Sağdan durdurulmuş veriler, birimlerde çalışma süresi sonuna kadar çalışmaya katılan birimin ilgilenilen olayı yaşamamış olması, takip sırasında ilgili birimin takipten çıkması veya çalışmaya katılan birimin başka bir olaydan dolayı (yarışan risk) çalışmadan ayrılması sebepleri ile ilgilenilen olayın gözlemlenememesinden kaynaklanmaktadır. Soldan durdurulmuş veriler, ilgilenilen olayın gözlemlendiği ancak kesin olay zamanının bilinmediği durumlar olarak tanımlanmaktadır. Diyabet hastası bir kişinin diyabet teşhisi konulduğu yaşı kesin olarak bilinmesine rağmen hastalığın kesin zamanının belirsiz olması, soldan durdurulmuş veriye örnek olarak verilebilir. Aralıklı durdurulmuş veriler ise, ilgilenilen olayın

gözlemlendiği ancak birimlerin sadece belirli zaman aralıklarında gözlemlendiği ve ilgilenilen olayın bilinen bir zaman aralığı içinde gerçekleştiği durumlardan oluşmaktadır.

Durdurma sınıfının ikinci mekanizması bir birimin durdurma olasılığının başarısızlık süresine bağlı olup olmamasına dayanmakta ve üç temel varsayım olarak ele alınmaktadır. Birinci varsayım “bağımsız durdurmadır” ve yaşam süresi ile durdurma süresinin bağımsız olduğu varsayılmaktadır. Buna göre, bir birimin gerçek yaşam süresi bu birimin durdurulma mekanizmasından bağımsızdır ve durdurulma süresi yaşam süresinden küçük olduğunda iki süre arasında bir ilişki bulunmamaktadır. İkinci varsayım “rastgele durdurmadır” ve durdurulmuş birimlerin başarısızlık (olay gerçekleşmesi) olasılığı, risk altında kalan birimlerle aynıdır. Üçüncü varsayım ise “bilgi içermeyen durdurmadır” ve bu durumda, durdurmanın nedeni, birimlerin ilgilenilen olay dışında başka bir nedenden kaynaklanmaktadır (Alkathami, 2021).

Yaşam çözümlemesinde, yaşam süresi dağılımını belirlemek ve modellemek için tehlike (hazard), yaşam ve olasılık yoğunluk fonksiyonları kullanılmaktadır. Bu fonksiyonlar arasında ilişki olduğundan dolayı herhangi bir fonksiyon bilindiğinde diğer fonksiyonlar kolaylıkla elde edilebilmektedir (Lee ve Wang, 2003; Bewick, Cheek ve Ball, 2004).

Bir birimin yaşam süresi,  $T$  rastlantı değişkeni ile ifade edilsin. Bir birimin  $t$  zamanından daha uzun süre yaşama olasılığı  $S(t)$  yaşam fonksiyonu ile ifade edilir ve,

$$S(t) = P(T > t) = \int_t^{\infty} f(x)dx, \quad 0 < t < \infty$$

olarak gösterilir. Bu fonksiyon monoton azalan bir fonksiyondur ve  $t = 0$  başlangıç noktasında  $S(t) = 1$ ,  $t = \infty$  olduğunda ise  $S(t) = 0$  olacak şekilde tanımlanır (Bewick, Cheek ve Ball, 2004; Kleinbaum ve Klein, 2012).

Yaşam sürelerinin birikimli dağılım fonksiyonu, bilinen bir olayın  $t$  zamanından önce meydana gelme olasılığını gösterir ve

$$F(t) = P(T < t)$$

şeklinde tanımlanır.  $F(t)$  monoton artan bir fonksiyondur ve 0 ile 1 arasında değer alır. Buna göre yaşam fonksiyonu  $S(t)$ ,

$$S(t) = 1 - F(t)$$

şeklinde birikimli dağılım fonksiyonu ile ifade edilebilir.

Yaşam süresinin birikimli dağılım fonksiyonunun türevi alınarak, olasılık yoğunluk fonksiyonu aşağıdaki gibi hesaplanabilir ve  $[t, t + \Delta t)$  aralığı gibi kısa bir zaman diliminde birimlerin olaya maruz kalma olasılığını gösterir:

$$f(t) = \frac{\partial F(t)}{\partial t} = -\frac{\partial S(t)}{\partial t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

Yaşam süresine ilişkin olasılık yoğunluk fonksiyonu negatif olmayan bir fonksiyon olup  $f(t) \geq 0$  koşulunu sağlamaktadır.

Tehlike fonksiyonu  $h(t)$  ise  $t$  zamanına kadar hayatta kalan bir birimin, sonraki bir zamanda ilgilenilen olayı gerçekleştirme riskini ifade etmektedir. Başka bir ifadeyle, tehlike fonksiyonu,  $t$  zamanına kadar olayın gerçekleşmediği varsayıldığında, ilgilenen olayın  $t + \Delta t$  gibi kısa bir zaman aralığında meydana gelme riskini göstermektedir.  $h(t)$ ,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad t \geq 0$$

biçiminde verilir.  $h(t)$  anlık risk fonksiyonu olarak adlandırılır ve zaman içerisinde artış, azalış gösterebilir ya da sabit kalabilir. Tehlike fonksiyonu, yaşam fonksiyonunun aksine, başarısızlık olasılığına odaklanır ve olayın gerçekleşme riskini ifade etmektedir.

Yaşam çözümlemesinde genellikle tehlike fonksiyonu üzerinde durulmaktadır ve açıklayıcı değişkenlerin tehlike fonksiyonu üzerine etkisini belirlemek ve birimlerin tehlike fonksiyonlarını değerlendirmek için parametrik olmayan, yarı parametrik ve parametrik modeller kullanılmaktadır (Collet, 2015).

Parametrik olmayan modeller, özellikle tek bir örneklemin yaşam sürelerini tahmin etmek veya iki ya da daha fazla grubun karşılaştırılmasında yaygın olarak kullanılmaktadır. Parametrik olmayan testlerden, birimlerin yaşam veya tehlike fonksiyonlarını tahmin etmek için genellikle Kaplan Meier tahmin edicisi, iki veya daha fazla grubun yaşam fonksiyonlarını karşılaştırmak için ise log-rank testi kullanılmaktadır (Collett, 2015). Yarı parametrik modeller verinin dağılımı hakkında herhangi bir varsayımda bulunmazken, açıklayıcı değişkenlerin yaşam süresine etkisini parametrik olarak incelemektedir. Cox regresyon modeli en bilinen yarı parametrik modeldir ve temel tehlike fonksiyonu belirtilmemişken, açıklayıcı değişkenlerin tehlike oranı üzerine etkisini parametrik olarak incelemektedir (Cox, 1972). Parametrik modeller ise yaşam süresinin belirli bir dağılıma sahip olduğunu varsaymaktadır (Rizopoulos, 2012a; Collett, 2015). Literatürde sıklıkla

kullanılan parametrik modeller, orantılı tehlikeler modelleri ve HBS modelleridir (Lawless, 2011).

Tez çalışması kapsamında standart bileşik model yapısından bahsedildikten sonra yaşam süresi alt model için HBS modeller dikkate alındığından, bu bölümde Cox regresyon modeli ve parametrik regresyon modelleri detaylı olarak incelenmiştir.

### 3.1. Cox Regresyon Modeli

Yaşam çözümlenmesinde temel amaç bir veya daha fazla açıklayıcı değişkenin, birimlerin yaşam sürelerine etkisini araştırmaktır (Singh ve Mukhopadhyay, 2011). Birimler, yaş, cinsiyet, eğitim düzeyi, kan basıncı, sigara kullanımı, fiziksel aktivite seviyesi gibi demografik ve biyolojik değişkenler açısından farklılık gösterebilmekte ve bu değişkenler yaşam süreleri üzerinde sabit veya zamanla değişen (zamana bağlı) etkilere sahip olabilmektedir (Alkhathami, 2021). Bu kapsamda, Cox (1972) tarafından geliştirilen ve orantılı tehlikeler modeli olarak da adlandırılan Cox regresyon modeli yaşam verilerinin analizinde sıklıkla kullanılmaktadır.

Cox regresyon modeli  $i$ . birim için,

$$h_i(t|x_i) = h_0(t) \exp(\gamma'x_i) \quad (3.1)$$

şeklinde tanımlanabilir (Rizopoulos, 2012a). Burada  $h_0(t)$  tanımlanmamış ve negatif olmayan temel tehlike fonksiyonunu ( $x_i = (x_1, \dots, x_p)'$  olan teorik bir birim tehlike fonksiyonu),  $x$  açıklayıcı değişkenler vektörünü ve  $\gamma$  regresyon katsayıları vektörünü göstermektedir. Cox regresyon modeli, temel tehlike fonksiyonu için dağılımsal bir varsayımda bulunmazken, açıklayıcı değişkenlerin tehlike oranı üzerindeki etkilerini parametrik olarak birleştirmekte ve bu nedenle yarı-parametrik bir yöntem olarak adlandırılmaktadır (Fox ve Weisberg, 2011). Ek olarak,  $h_0(t)$  temel tehlike fonksiyonu parametrik formda ya da keyfi bir negatif olmayan fonksiyon olarak tanımlanabilmektedir. Yarı parametrik Cox modeli, keyfi negatif olmayan bir fonksiyon varsayarken, parametrik orantılı tehlikeler modelleri Weibull, Üstel gibi parametrik formlarda tanımlanabilmektedir (Lawless, 2011).

Cox regresyon modelinin parametre tahminlerini elde etmek için genellikle Cox (1972) tarafından önerilen kısmi olabilirlik yaklaşımı kullanılmaktadır (Lewis, 2017).  $n$  birimlik bir veri seti,  $k$  adet başarısızlık süresi ve  $n - k$  sağdan durdurulmuş yaşam sürelerinden

oluşsun ve her başarısızlık süresinde yalnızca bir bireyde başarısızlık gözlemlensin (veriler bağımsız). Başarısızlık sürelerinin sıralı olduğu varsayımına dayanarak,  $t_{(1)} < t_{(2)} < \dots < t_{(k)}$  sıralı başarısızlık süreleri ve bu sürelerle karşılık gelen açıklayıcı değişkenler  $x_{(1)}, x_{(2)}, \dots, x_{(k)}$  olarak tanımlansın. Buna göre,  $R(t_{(i)})$  risk kümesi  $t_{(i)}$  anına kadar başarısızlık yaşamamış veya durdurulmamış birimlerden oluşur ve belirli bir  $t_{(i)}$  anındaki bir birimin başarısızlığı için,  $R(t_{(i)})$  risk kümesine koşullu olarak, başarısızlığın gözlemlenen birimde olma olasılığı,

$$\frac{\exp(\gamma'x_i)}{\sum_{l \in R(t_{(i)})} \exp(\gamma'x_l)} \quad (3.2)$$

şeklinde tanımlanabilir. Buna göre Eşitlik (3.1)'deki Cox regresyon modeli için kısmi olabilirlik fonksiyonu aşağıdaki gibi tanımlanabilir (Cox, 1972, Collett, 2015).

$$L(\gamma) = \prod_{i=1}^k \frac{\exp(\gamma'x_i)}{\sum_{l \in R(t_{(i)})} \exp(\gamma'x_l)} \quad (3.3)$$

Verilerin  $n$  gözlemlenmiş yaşam süresi ve bu sürelerin  $t_1, t_2, \dots, t_n$  ile gösterildiği varsayalım.  $\delta_i$  olay göstergesi,  $i$ . birim için yaşam süresi  $t_i$  durdurulmuş ise 0, değilse 1 değerini alsın ( $i = 1, 2, \dots, n$ ). Buna göre Eşitlik (3.3)'te verilen denklem Eşitlik (3.4)'deki gibi yeniden tanımlanabilir:

$$L(\gamma) = \left\{ \prod_{i=1}^n \frac{\exp(\gamma'x_i)}{\sum_{l \in R(t_i)} \exp(\gamma'x_l)} \right\}^{\delta_i} \quad (3.4)$$

Eşitlik (3.4)'te verilen kısmi olabilirlik fonksiyonuna karşılık gelen log olabilirlik fonksiyonu ise aşağıdaki gibi elde edilebilir (Collett, 2015);

$$\log L(\gamma) = \sum_{i=1}^n \delta_i \{ \gamma'x_i - \log \sum_{l \in R(t_i)} \exp(\gamma'x_l) \} \quad (3.5)$$

Eşitlik (3.5)'in  $\gamma$  parametresine göre birinci derece türevi alınırsa,  $\gamma$  parametrelerinin EÇO tahminleri

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

biçiminde elde edilebilir.  $\gamma$  parametresinin EÇO tahmin edicisini hesaplamak için genellikle Newton-Raphson gibi iteratif yaklaşımlar kullanılmaktadır.

### 3.2. Orantılı Tehlikeler Varsayımının İncelenmesi

Yarı parametrik veya parametrik modellerden hangisinin kullanılacağına karar verebilmek için öncelikle orantılı tehlikeler varsayımının kontrol edilmesi gerekmektedir. Bu varsayımın kontrolü için genellikle grafiksel yöntemler, uyum iyiliği testleri ve zamana bağlı açıklayıcı değişken içeren prosedürler kullanılmaktadır (Kleinbaum ve Klein, 2012).

Orantılı tehlikeler varsayımı, bir birim için tehlikenin herhangi bir başka birimin tehlikesi ile orantılı olduğunu ve bu orantılılık katsayısının zamandan bağımsız olduğunu varsayar. Orantılı tehlikeler varsayımını kontrol etmek için en yaygın kullanılan grafiksel yöntemler;  $\log(-\log)$  yaşam eğrileri, beklenen ve gözlenen yaşam eğrilerinin karşılaştırılması ve Arjas grafiklerinden oluşurken, en yaygın kullanılan uyum iyiliği testleri ise Schoenfeld artıkları ile yaşam sürelerinin ranklarının korelasyon testidir (Klein ve Moeschberger, 2003; Kleinbaum ve Klein, 2012; Karasoy ve Tutkun, 2016).

Tez çalışması kapsamında orantılı tehlikeler varsayımı  $\log(-\log)$  grafiği ve ölçeklendirilmiş Schoenfeld artıkları testi ile incelendiğinden, bu iki test detaylı olarak anlatılmıştır.

#### 3.2.1. Log-Log Grafiği

Belirli bir kategorik açıklayıcı değişken  $x_i$  için orantılı tehlikeler varsayımını grafiksel olarak kontrol etmek için  $\log(-\log)$  grafiği kolaylıkla kullanılmaktadır. Orantılı tehlikeler varsayımı sağlandığında, Bölüm 3'te verilen yaşam ve birikimli tehlike fonksiyonları ve Eşitlik (3.1)'de verilen Cox regresyon modeline uyumlu olarak aşağıdaki eşitlik yazılabilir:

$$S(t|x) = \exp(-H(t|x)) = \exp(-H_0(t)\exp(\gamma'x)) = S_0(t)^{\exp(\gamma'x)} \quad (3.6)$$

Eşitlik (3.6)'nın iki kere logaritması alınırsa,

$$\log(-\log(S(t|x))) = \gamma'x + \log(-\log(S_0(t)))$$

elde edilir. Burada, yaşam fonksiyonları 0 ile 1 arasında değer aldığından ve ilk logaritma negatif olduğundan dolayı eksi işareti eklenmiştir.

$x_1$  ve  $x_2$  iki farklı birime veya gruba karşılık gelen  $x$  vektörünün iki farklı özelleştirilmiş  $\log - \log$  eğrileri aşağıdaki gibi hesaplanabilir.

$$\gamma'x_1 + \log(-\log(S_0(t)))$$

$$\gamma'x_2 + \log(-\log(S_0(t)))$$

Buna göre iki gözlem kümesi için  $\log(-\log)$  grafiği,

$$\log(-\log(S(t|x_2))) - \log(-\log(S(t|x_1))) = \gamma'(x_2 - x_1)$$

formülü ile kolayca hesaplanabilir. Bu grafikte, iki veya daha fazla birim veya kategorinin tehlike eğrilerinin çakışmaması orantılı tehlikeler varsayımının sağlandığını gösterirken, eğrilerin çakışması varsayımın sağlanmadığını göstermektedir. Bu nedenle bu grafiksel yaklaşımın bazı dezavantajları bulunmaktadır. Bu dezavantajlar, birimlerin tehlike fonksiyonları kesişse bile varsayımın sağlanmasının mümkün olması ve  $\log(-\log)$  grafiğinin sadece kategorik değişkenlerde kullanılmasından dolayı sürekli değişkenlerin kategorik değişkene dönüştürülmesi gerektiğidir (Kleinbaum ve Klein, 2012; Ekman, 2017).

### 3.2.2. Schoenfeld Artıkları

Schoenfeld (1982), orantılı tehlikeler varsayımını kontrol etmek için artıklara dayanan bir yöntem önermiştir. Bu yöntemde, her bir birim için tek bir artık yerine her açıklayıcı değişken ve başarısızlık süresi için ayrı bir artık bulunmaktadır.  $k$ . açıklayıcı değişken ve  $i$ . başarısızlık süresi için Schoenfeld artıkları aşağıdaki gibi tanımlanabilir:

$$r_{(i)k}^S = x_{(i)k} - \sum_{j \in R_i} (x_{jk} \times p(\hat{Y}_k, x_{jk})) \quad (3.7)$$

Burada,

$$p(\hat{Y}_k, x_{jk}) = \frac{\exp(\hat{Y}_k x_{jk})}{\sum_{j \in R_i} \exp(\hat{Y}_k x_{jk})}$$

$t_{(i)}$  zamanında  $i$ . birimin risk kümesinden seçilme olasılığını göstermektedir. Eşitlik (3.7)'de verilen Schoenfeld artıkları,  $t_{(i)}$  zamanında gerçekleşen başarısızlık olayı için ilgili bireyin açıklayıcı değişken değerleri  $x_{ik}$ 'nin, aynı zaman dilimindeki risk seti için açıklayıcı değişkenlerin beklenen değerinden çıkarılmasıyla elde edilmektedir. Burada beklenen değer, risk setindeki her bir birimin  $t_{(i)}$  zamanında gerçekten başarısızlığı yaşama olasılığıyla ağırlıklandırılmış açıklayıcı değişken değerlerinden oluşmaktadır. Bu

testte incelenen hipotez; “orantılı tehlikeler varsayımı sağlanmaktadır” şeklindedir ve test sonucunda her bir değişken için hesaplanan p değeri  $< 0.05$  ise, test edilen değişkenin varsayımı sağlamadığı yorumu yapılmaktadır (Kleinbaum ve Klein, 2012; Karasoy ve Tutkun, 2016).

Schoenfeld artıkları aslında Eşitlik (3.5)’te verilen Cox regresyona ilişkin kısmi olabilirlik ve türev fonksiyonunun bireysel terimleridir. Bu nedenle, belirli bir açıklayıcı değişken için varsayım geçerli olduğunda, Schoenfeld artıklarının başarısızlık zamanlarına karşı çizilen grafiğinde, noktaların başarısızlık zamanlarından bağımsız olarak sıfır etrafında merkezlenmiş bir desene sahip olması beklenmektedir (Ekman, 2017).

### **3.3. Yaşam Çözümlemesinde Kullanılan Parametrik Modeller**

Cox regresyon modelinde, yaşam süreleri için belirli bir olasılık dağılımı varsayımı bulunmazken, bazı uygulamalarda yaşam verilerinin Üstel, Weibull, Gompertz, Log-Lojistik, Log-Normal ve Gamma dağılımı gibi parametrik bir dağılıma sahip olduğu görülmektedir (Wienke, 2010). Bu durumda, parametrik modeller, Cox regresyon modeline kıyasla daha az yanlı sonuçlar vermektedir (Alkathami, 2021). Yaşam sürenin parametrik modellemesinde genellikle orantılı tehlikeler modelleri ve HBS modelleri kullanılmaktadır (Lawless, 2011). Parametrik orantılı tehlikeler modelleri ve Cox regresyon modeli, zaman boyunca tehlikenin sabit olduğu orantılı tehlikeler varsayımının geçerli olduğunu varsaymaktadır. Bu varsayımın sağlanmadığı durumlarda, orantılı tehlikeler modellerine alternatif olarak HBS modelleri kullanılmaktadır (Wei, 1992; Collett, 2015; Basharat, 2019).

Parametrik orantılı tehlikeler ve HBS modellerinde yaygın olarak kullanılan bazı dağılımlar için olasılık yoğunluk ve yaşam fonksiyonları Çizelge 3.1’de verilmiştir (Lawless, 2011). Bu dağılımlar yaşam verilerinin analizinde, regresyon modellerini formalize etmek için önemli bir rol oynamaktadır. Bu tez kapsamında yaşam verisi alt modelinde Weibull HBS modeli dikkate alındığından, sonraki bölümde Weibull orantılı tehlikeler ve HBS modelleri detaylı olarak anlatılmıştır.

Çizelge 3.1. Yaşam çözümlemesinde yaygın olarak kullanılan parametrik dağılımlar

Dağılım	Parametre	S(t)	f(t)
Üstel	$\lambda > 0$	$\exp(-\lambda t)$	$\lambda \exp(-\lambda t)$
Weibull	$\lambda, \kappa > 0$	$\exp[-(\lambda t)^\kappa]$	$\kappa \lambda (\lambda t)^{(\kappa-1)} \exp(-(\lambda t)^\kappa)$
Log-Lojistik	$\lambda, \kappa > 0$	$[1 + (\lambda t)^\kappa]^{-1}$	$\kappa \lambda (\lambda t)^{(\kappa-1)} [1 + (\lambda t)^\kappa]^{-2}$
Log-Normal	$\mu \in \mathbb{R}, \sigma > 0$	$1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$	$\frac{1}{(2\pi)^{\frac{1}{2}} \sigma} \exp\left[-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right]$

$\Phi$ , standart normal dağılımın birikimli dağılım fonksiyonudur.

### 3.3.1. Weibull Orantılı Tehlikeler Modeli

Yaşam çözümlemesinde Weibull parametrik modelde, tehlike fonksiyonu çeşitli formlar aldığından (artan, azalan vb.) verilerin modellenmesinde esneklik sağlamaktadır. Aynı zamanda, yaşam süresi verilerinin normal dağılıma uygun olmamasından kaynaklı ortalama ve standart sapma yerine medyan ve yüzdeler gibi betimleyici istatistiklerin hesaplanmasına da olanak sağlamaktadır. Tüm bu özelliklere ek olarak, orantılı tehlikeler ve HBS modellerine de uygun özellikte olduğu için yaygın olarak tercih edilmektedir (Alkhatami, 2021).

Yaşam süresi verileri  $\lambda$  ölçek (scale) parametresi ve  $\kappa$  şekil (shape) parametresi ile Weibull dağılıma sahip olduğunda, Eşitlik (3.1)'de tanımlanan Cox regresyon modelinin temel tehlike fonksiyonu  $h_0(t) = \lambda t^{\kappa-1}$  olarak tanımlanacaktır. Buna göre, Weibull orantılı tehlikeler modeline ilişkin tehlike fonksiyonu,

$$h_i(t) = \lambda \kappa t^{\kappa-1} \exp(\gamma'x_i), \quad i = 1, 2, \dots, n$$

ve yaşam fonksiyonu

$$S_i(t) = \exp\{-\lambda t^\kappa \exp(\gamma'x_i)\} \quad (3.8)$$

olarak tanımlanır (Collett, 2015). Buna göre, Weibull orantılı tehlikeler modelinin parametre tahminlerini elde etmek için olabilirlik fonksiyonu,

$$L(\lambda, \kappa, \gamma) = \prod_{i=1}^n [h_i(t_i)]^{\delta_i} S_i(t_i)$$

$$\begin{aligned}
&= \prod_{i=1}^n [\lambda \kappa t_i^{\kappa-1} \exp(\gamma' x_i)]^{\delta_i} \exp \{-\lambda t_i^{\kappa} \exp(\gamma' x_i)\} \\
&= \prod_{i=1}^n [\kappa t_i^{\kappa-1} \exp(\gamma' x_i)]^{\delta_i} \exp \{-t_i^{\kappa} \exp(\gamma' x_i)\}, \quad (\lambda = e^{\gamma_0})
\end{aligned} \tag{3.9}$$

şeklinde tanımlanabilir. Eşitlik (3.9)'un logaritması alınarak, log-olabilirlik fonksiyonu ise aşağıdaki gibi elde edilebilir.

$$L(\lambda, \kappa, \gamma) = \sum_{i=1}^n \delta_i \log(\kappa) + (\kappa - 1) \sum_{i=1}^n \delta_i \log(t_i) + \sum_{i=1}^n \delta_i \gamma' x_i - \sum_{i=1}^n t_i^{\kappa} \exp(\gamma' x_i)$$

Modeldeki bilinmeyen parametreleri  $(\lambda, \kappa, \gamma)$  tahmin etmek için Newton-Raphson yaklaşımı kullanılmaktadır.

### 3.3.2. Hızlandırılmış Başarısızlık Süresi Modelleri

Cox ve Weibull orantılı regresyon modellerinde,  $h_i(t)/h_0(t)$  tehlike oranının zaman içerisinde sabit olduğu varsayılrsa da uygulamalarda bu varsayım her zaman geçerli olmamaktadır. Bu durumlarda, orantılı tehlikeler modellerine alternatif olarak HBS modelleri kullanılmalıdır. Ayrıca, orantılı tehlikeler modelleri açıklayıcı değişkenlerin tehlike fonksiyonu üzerindeki etkisini modellerken, HBS modelleri açıklayıcı değişkenler ile yaşam süreleri arasında doğrudan bir ilişki olduğunu varsaymaktadır (George, Seals ve Aban, 2014).

HBS modelleri, yaşam sürelerinin logaritması olarak aşağıdaki gibi ifade edilebilir:

$$\log T = -\gamma' x_i + \epsilon \tag{3.10}$$

T, Weibull, Log-Lojistik veya Log-Normal dağılımlara sahipse, Eşitlik (3.10)'daki  $\epsilon$  sırasıyla uç değer dağılımı (extreme value distribution), standart lojistik dağılım ve standart normal dağılıma sahip olacaktır (Basharat, 2019).

Zamandan bağımsız açıklayıcı değişkenler için T'nin olasılık yoğunluk, yaşam ve tehlike fonksiyonları sırasıyla aşağıdaki gibi ifade edilebilir (Kalbfleisch ve Prentice, 2002).

$$f_i(t|x) = f_0(\text{texp}(-\gamma' x_i)) \exp(-\gamma' x_i)$$

$$S_i(t|x) = S_0(\text{texp}(-\gamma' x_i))$$

$$h_i(t|x) = h_0(\text{texp}(-\gamma' x_i)) \exp(-\gamma' x_i), \quad i = 1, 2, \dots, n \tag{3.11}$$

Burada  $\gamma' x_i$ , modelin hızlandırma faktörü (acceleration factor) olarak da isimlendirilen doğrusal bileşeni ve  $f_0(\cdot)$ ,  $S_0(\cdot)$  ve  $h_0(\cdot)$  sırasıyla temel olasılık yoğunluk, yaşam ve tehlike fonksiyonlarını göstermektedir (Basharat, 2019; Alkhatami, 2021). Hızlandırma

faktörü, tehlike oranının, açıklayıcı değişkenlerin tehlike üzerindeki etkisini değerlendirilmesine izin verirken aynı zaman da açıklayıcı değişkenlerin yaşam süreleri üzerindeki etkilerinin de incelenmesine olanak tanımaktadır (Kleinbaum ve Klein, 2012). Parametrik HBS modelleri için, aşağıda verilen log-lineer form yaygın olarak kullanılmaktadır:

$$\log t_i = \mu + \gamma'x_i + \sigma\epsilon_i \quad (3.12)$$

Burada  $\mu$  ve  $\sigma$  sırasıyla kesişim ve ölçek parametreleri,  $\epsilon_i$  ise hata terimini göstermektedir. Eşitlik (3.12)'deki HBS modelinin log-lineer formülasyonundaki  $\epsilon_i$  için parametrik dağılımlar yaşam süresi  $t_i$ 'nin farklı dağılımlarına yol açmaktadır. Uygulamada en yaygın kullanılan parametrik HBS modelleri, Weibull, Log-Lojistik ve Log-Normal dağılımlarına dayanmaktadır.  $\epsilon_i$ 'nin tehlike ve yaşam fonksiyonlarından,  $t_i$ 'nin yaşam ve tehlike fonksiyonları sırasıyla aşağıdaki gibi elde edilebilir (Alkhathami, 2021):

$$\begin{aligned} S_i(t) &= S_{\epsilon_i} \left( \frac{\log t - \mu - \gamma'x_i}{\sigma} \right) \\ h_i(t) &= \frac{1}{\sigma t} h_{\epsilon_i} \left( \frac{\log t - \mu - \gamma'x_i}{\sigma} \right) \end{aligned} \quad (3.13)$$

Ek olarak, açıklayıcı değişken  $x$ 'in zamana bağlı olduğu durumlarda,

$$\psi(t) = \int_0^t \exp(-\gamma'x(s)) ds \quad (3.14)$$

eşitliği tanımlanabilir. Bu dönüşüm, açıklayıcı değişkenler tarafından tanımlanan dönüştürülmüş bir zaman ölçeği olarak dikkate alınabilir ve HBS modellerine ilişkin olasılık yoğunluk, yaşam ve tehlike fonksiyonları aşağıdaki gibi yeniden tanımlanabilir (Cox ve Oakes, 1984).

$$\begin{aligned} f_i(t|x) &= f_0(\psi(t)) \exp(-\gamma'x_i) \\ S_i(t|x) &= S_0(\psi(t)) \\ h_i(t|x) &= h_0(\psi(t)) \exp(-\gamma'x_i), \quad i = 1, 2, \dots, n \end{aligned} \quad (3.15)$$

Parametrik HBS modellerinin parametre tahminlerini elde etmek için genellikle olabilirlik yaklaşımı kullanılmaktadır. HBS modelleri için olabilirlik fonksiyonu,

$$L(\gamma) = \prod_{i=1}^n \{f_i(t_i)\}^{\delta_i} \{S_i(t_i)\}^{1-\delta_i} \quad (3.16)$$

şeklinde tanımlanabilir ve olabilirlik fonksiyonunun logaritmasının birinci derece türevi alınarak ve Newton-Raphson gibi iteratif yaklaşımlar kullanılarak parametre tahminleri elde edilebilir (Alkathami, 2021).

### **Weibull Hızlandırılmış Başarısızlık Süresi Modeli**

Yaşam süresi verileri  $\lambda$  ölçek parametresi ve  $\kappa$  şekil parametresi ile Weibull dağılıma sahip olduğunda Weibull HBS modeli için tehlike fonksiyonu,

$$\begin{aligned} h_i &= \lambda \kappa (\exp(-\gamma'x_i)t)^{\kappa-1} \exp(-\gamma'x_i) \\ &= \lambda \kappa \exp(-\kappa\gamma'x_i) t^{\kappa-1} \end{aligned} \quad (3.17)$$

olarak tanımlanır. Burada temel tehlike fonksiyonu  $h_0(t) = \lambda \kappa t^{\kappa-1}$ 'dir. Eşitlik (3.12)'de verilen log-linear HBS modeli cinsinden Weibull HBS modeli için yaşam fonksiyonu aşağıdaki gibi yazılabilir:

$$S_i(t) = \exp \left\{ - \exp \left( \frac{\log t - \mu - \gamma'x_i}{\sigma} \right) \right\} \quad (3.18)$$

Buna göre, Weibull dağılımının ölçek parametresi  $\lambda_i = \exp\{-(\mu + \gamma'x_i)/\sigma\}$  ve şekil parametresi  $1/\sigma$  olmak üzere,  $S_i(t) = \exp(-\lambda_i t^{1/\sigma})$  olarak tanımlanabilir.

Weibull HBS ve orantılı tehlikeler modelleri arasındaki ilişkiden, Eşitlik (3.8)'deki Weibull orantılı tehlikeler modeli,  $\epsilon_i$ 'nin Gumbel dağılımına sahip olduğu varsayımında,  $\log t_i = \frac{1}{\kappa}(-\log \lambda - \gamma'x_i + \epsilon_i)$  dönüşümüyle Eşitlik (3.12)'de verilen log-linear model cinsinden ifade edilebilir (Alkathami, 2021).

## 4. BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİ

Tekrarlı ölçümlerden oluşan boylamsal veriler ve ilgilenen olay gerçekleşene kadar geçen sürelerden oluşan yaşam verileri genellikle birlikte elde edilmektedir. Örneğin, AIDS hastalarından tekrarlı olarak ölçülen CD4 hücre sayılarını ele alalım. Burada amaç, CD4 sayılarının aktif tedavi yöntemleriyle yaşam süreleri üzerinde önemli bir etkiye sahip olup olmadığı belirlemek olacaktır. Bu çalışmada, tekrarlı olarak ölçülen CD4 ölçümleri ile ölüm süreleri arasındaki ilişki, araştırmanın birincil amacıdır. Boylamsal ve yaşam verileri için metodolojiler ayrı ayrı iyi bir şekilde geliştirilmiş olmasına rağmen, Bölüm 2 ve Bölüm 3'te detaylı olarak verilen ayrı modeller bu araştırma sorusu için yetersiz kalmaktadır. Özellikle, boylamsal ölçümler ile yaşam süresi verileri arasındaki ilişkiyi incelemek için genişletilmiş Cox regresyon modeli kullanılabilmesine rağmen, bu model içsel (endojen) açıklayıcı değişkenler olduğu durumlarda yanlış parametre tahminleri ve istatistiksel çıkarımlara yol açmaktadır (Tsiatis, DeGruttola ve Wulfsohn, 1995; Wulfsohn ve Tsiatis 1997).

Her bir birim için boylamsal olarak elde edilen ölçümler genellikle ölçüm hataları içerdiğinden ve tüm birimler için boylamsal ölçümlerin tam bilgisi, aralıklı ölçümler veya durdurulma nedeni ile sağlanamadığından dolayı, yeni modelleme çerçevelerine ihtiyaç duyulmuştur. Bu tür uygulama gereksinimleri nedeniyle, son yirmi yılda iki modeli paylaşılan bilgi aracılığıyla eş zamanlı olarak uyarlayan bileşik modelleme üzerine çalışmalar giderek artmıştır. İki modelin ayrı tahmin edilmesi prosedürü ile karşılaştırıldığında, bileşik modelleme, boylamsal ve yaşam verileri arasındaki ilişkiyi dikkate alarak model tahminlerindeki yanlışlığı iyileştirmekte ve daha güvenilir çıkarımlar sağlamaktadır (Alkathami, 2021).

Bileşik modelleme, genellikle boylamsal ölçümlerin yaşam süresi verileri üzerindeki etkilerini incelemek (Faucett ve Thomas, 1996; Wulfsohn ve Tsiatis, 1997; Wang ve Taylor, 2001; Brown ve Ibrahim, 2003), bu iki veri türünün karşılıklı etkileşimlerini araştırmak (Henderson, Diggle ve Dobson, 2000; Guo ve Carlin, 2004; Zeng ve Cai, 2005) ve bağımsız değişkenlerin hem boylamsal hem de yaşam süresi verileri üzerindeki etkilerini analiz etmek amacıyla (Tsiatis, DeGruttola ve Wulfsohn, 1995; McCrink, Marshall ve Cairns, 2013) kullanılmaktadır. Genel olarak bileşik model yapısı, yaşam

süresi modeli, boylamsal model ve her iki modeli paylaşılan gizli bir yapıyla birbirine bağlayan bileşenlerden oluşmaktadır (Proust-Lima ve ark. 2014).

Literatürde bileşik modelin paylaşılmış gizli sınıflar (shared latent classes) ve paylaşılmış rastgele etkiler (shared random effects) ile birleştirilmesine dayanan iki farklı model yapısı bulunmaktadır (Henderson, Diggle ve Dobson, 2000; Proust-Lima ve ark. 2014). Paylaşılmış gizli sınıflarda, araştırma birim kitlelerinin heterojen olduğu kabul edilmekte ve olayın aynı risk ve aynı işaretli eğriyi paylaşan homojen gizli alt grupları içerdiği varsayılmaktadır (Proust-Lima ve ark. 2014). Paylaşılmış rastgele etkilerde ise araştırma birim kitlelerinin heterojen olduğu varsayılmakta ve hem boylamsal hem de yaşam süreleri, araştırma birimleri üzerindeki boylamsal eğrilerde değişkenliği gösteren bireysel rastgele etkilerin bir fonksiyonu ile birbirine bağlanmaktadır (Wulfsohn ve Tsiatis, 1997; Tsiatis ve Davidian, 2004).

Bileşik model yapısı için paylaşılmış rastgele etkilerin farklı parametrelendirilmelerine dayanan rastgele karma etkiler (random pattern mixture) ve rastgele seçim (random selection) modelleri de boylamsal ve yaşam süresi verilerini birleştirmek için kullanılmaktadır. Boylamsal veriler, yaşam verileri ve rastgele etkiler modeli sırasıyla L, T ve R ile gösterilsin. Buna göre rastgele karma etkili modeller;  $[R][T|R][L|T]$ , rastgele seçilmiş modeller;  $[R][L|R][T|L]$  ve rastgele etkiler modeli  $[R][L|R_1][T|R_2]$  olarak ifade edilebilir. Rastgele etkiler modeli hem boylamsal ölçümlerin hem de yaşam süresi verilerinin, belirli bir iki değişkenli dağılımla belirtilen gözlemlenmemiş bir rastgele etkiye bağlı olduğunu varsaydığından,  $R_1$  ve  $R_2$  şeklinde iki rastgele etki ile tanımlanmakta ve boylamsal gözlemler ve yaşam süreleri arasındaki ilişki bu iki rastgele etkinin korelasyon yapısı tarafından belirlenmektedir (Sousa, 2011). Karma modelde temel odak boylamsal verilere yönelirken, seçilmiş modelde ilgi, yaşam süresi verilerindedir ve boylamsal ölçümlerdeki korelasyondan yararlanılarak çıkarımlar yapılmaktadır (Sousa, 2011; Pericleous, 2016). Rastgele etkili modellerde ise, boylamsal ve yaşam süresi süreçleri, bağımsız olarak paylaşılan rastgele etkiler aracılığıyla birleştirilmektedir (Sousa, 2011; McCrink, Marshall ve Cairns, 2013). Bu çalışmada bileşik model, iki sürecin paylaşılmış rastgele etkilerin özel ve daha esnek bir versiyonu aracılığıyla bağlandığı bir karma model yapısı çerçevesinde ele alınmıştır (Henderson, Diggle ve Dobson, 2000; Wang, 2021).

Literatürde sıklıkla kullanılan bileşik model yapısı, boylamsal alt model için DKEM ve yaşam süresi alt modeli için Cox regresyon modelinin, paylaşılmış rastgele etkiler modeliyle birleştirilmesinden oluşmaktadır (Wu, Liu ve Hu, 2010). Ancak, orantılı tehlikeler varsayımını sağlamayan ve belirli bir dağılıma uygun olan yaşam verileri için parametrik regresyon modelleri Cox regresyona kıyasla daha doğru tahminler verdiği için bu durumda yaşam süresi alt modeli için HBS veya orantılı tehlikeler modelleri kullanılmalıdır (Wei, 1992; Collett, 2015; Basharat, 2019). Bu kapsamda bu tez çalışmasında, boylamsal veriler için DKEM ve yaşam süresi alt modeli için HBS modellerinin paylaşılmış rastgele etkiler modeli ile birleştirilmesine dayanan parametrik bileşik model dikkate alınmış ve bu modelin parametre tahminleri için sıklıkla kullanılan klasik yaklaşımlar bu bölümde detaylı olarak incelenmiştir.

#### 4.1. Yaşam Süresi Alt Modeli

$n$  birim için  $T^* = T_1^*, T_2^*, \dots, T_n^*$  gerçek yaşam sürelerini gösterebilir. Yaşam sürelerinin sağdan durdurulmuş olduğu varsayımında ve  $C_i$ ,  $i$ . birim için potansiyel durdurulmuş süreleri göstermek üzere, gözlenen yaşam süresi  $T_i = \min(T_i^*, C_i)$  olarak tanımlanır. Buna göre, olay göstergesi  $\delta_i = I(T_i^* \leq C_i)$  şeklinde tanımlanabilir ve yaşam süresi sağdan durdurulmuş ise  $\delta_i = 0$ , diğer durumlarda  $\delta_i = 1$  değerini alır. Ayrıca,  $i$ . birim için  $t_{ij}$  ( $j = 1, 2, \dots, n_i$ ) zamanında,  $y_{ij}$  ( $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, n_i$ ) boylamsal sürekli ölçümleri gösterebilir. Boylamsal ölçümler ve  $t$  zamanındaki bir olayın tehlikesi arasındaki ilişkiyi incelemek için Cox regresyon alt modeli,

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

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

şeklinde tanımlanabilir (Rizopoulos, 2012a; Alkhatami, 2021). Burada,  $m_i$  gizli Gauss değişkeni,  $h_0(\cdot)$  temel risk fonksiyonu ve  $x_i$ ,  $\gamma$  regresyon katsayılar vektörüne karşılık gelen, temel açıklayıcı değişken (tedavi grubu, cinsiyet vb.) vektörüdür.  $\alpha$ , boylamsal ölçümlerin tahminlerinden elde edilen birleştirme katsayısıdır ve boylamsal modeldeki temel sonuçların bir olayın riskine olan etkisini ölçmek için kullanılmaktadır (Alkhatami, 2021). Boylamsal ve yaşam süresi birbirleri ile ilişkili olmadıklarında  $\alpha = 0$  olmaktadır (Rizopoulos, Verbeke ve Lesaffre, 2009; Rizopoulos, 2012b).

Tehlike fonksiyonunun aksine yaşam fonksiyonu yapısı gereği sadece belirli aralıktaki değerlerle değil, birimin tüm değerlerine bağlı olarak aşağıdaki gibi ifade edilmektedir (Alkathami, 2021):

$$S_i(t|m_i(t), x_i) = P(T_i > t|m_i(t), x_i) \\ = \exp \left\{ - \int_0^t h_0(u) \exp(\gamma x_i' + \alpha m_i(u)) du \right\} \quad (4.2)$$

Hem tehlike hem de yaşam fonksiyonları  $h_0(\cdot)$  temel risk fonksiyonu tarafından ifade edilmektedir. Buna göre  $h_0(\cdot)$ 'nin doğru belirlenmesi daha yansız ve güvenilir sonuçlar elde edebilmek için oldukça önemlidir. Orantılı tehlikeler varsayımının sağlanmadığı ve yaşam verilerinin belirli bir dağılıma uygun olduğu durumlarda parametrik regresyon yöntemleri Cox regresyon modeline göre daha güvenilir sonuçlar verdiğinden, bu gibi durumlarda bileşik modelin yaşam süresi alt modeli için de parametrik regresyon modelleri kullanılmalıdır.

#### 4.1.1. Parametrik Yaşam Süresi Alt Modelleri

Parametrik yaşam süresi modelleri, orantılı tehlikeler varsayımının sağlanmadığı ve yaşam verilerinin belirli bir dağılıma uygun olduğu durumlarda, yaşam süresi alt modeli için Cox regresyon modeline alternatif olarak Üstel, Weibull, Log-Lojistik ve Log-Normal HBS modellerinin DKEM ile paylaşılmış rastgele etkiler modelinin esnek bir versiyonu yardımıyla birleştirilmesi ile elde edilmektedir (Basharat, 2019).

$t$  zamanında bir olayın tehlike sürecini tanımlamak için HBS modeli,

$$h_i(t) = h_0(g_i(t)) \exp(-\gamma x_i' - \alpha m_i(t)), \quad t > 0 \quad (4.3)$$

olarak tanımlanabilir (Cox and Oakes, 1984). Burada  $h_0(\cdot)$  temel tehlike fonksiyonu ve  $x_i$ ,  $\gamma$  regresyon katsayılar vektörüne karşılık gelen temel açıklayıcı değişken vektörüdür.  $x_i$ , boylamsal modelin açıklayıcı değişken vektörüyle ortak yapıya sahip olabilir veya olmayabilir.  $g_i(t) = \int_0^t \exp(-\gamma x_i' - \alpha m_i(u)) du$  şeklinde tanımlanmaktadır ve integralin çözümü için Gauss-Kronrod veya Gauss-Legendre kuralı gibi integrasyon yöntemleri kullanılabilir. Buna göre HBS modelleri için olasılık yoğunluk ve yaşam fonksiyonları aşağıdaki gibi sırasıyla yazılabilir:

$$f_i(t) = f(g_i(t)) \exp(-\gamma x_i' - \alpha m_i(t))$$

$$S_i(t) = S_0(g_i(t)) \quad (4.4)$$

Burada  $f_0(\cdot)$  ve  $S_0(\cdot)$  sırasıyla temel olasılık yoğunluk ve temel yaşam fonksiyonlarını göstermektedir. Yaşam verilerinin dağılımına göre bu fonksiyonlar Çizelge 3.1'deki temel fonksiyonlara göre özelleştirilebilir (Basharat, 2019).

#### 4.1.2. Üstel Hızlandırılmış Başarısızlık Süresi Alt Modeli

Yaşam sürelerinin orantılı tehlikeler varsayımını sağlamadığı ve Üstel dağılıma uygun olduğu varsayıldığında, Eşitlik (4.3)'e göre ve  $m_i(t) = z_i'(t)b_i$  olmak üzere Üstel HBS modelinin  $t_i$  zamanında tehlike fonksiyonu,

$$h_i(t_i) = \lambda \exp(-\gamma x_i' - \alpha m_i(t)) \quad (4.5)$$

şeklinde tanımlanabilir. Burada  $h_0(g_i(t_i)) = \lambda$ 'dır. Benzer şekilde, yaşam fonksiyonu aşağıdaki gibi ifade edilebilir.

$$S_i(t_i) = S_0(g_i(t_i)) = \exp(-\lambda(g_i(t_i))) \quad (4.6)$$

Burada  $g_i(t_i) = \int_0^{t_i} \exp(-\gamma x_i' - \alpha z_i'(u)b_i) du$  biçimindedir. Eşitlik (4.5) ve Eşitlik (4.6) kullanılarak,  $b_i$  ve  $\theta$  verildiğinde  $(t_i, \delta_i)$  için olasılık yoğunluk fonksiyonu,

$$f(t_i, \delta_i | b_i, \theta) = \{\lambda \exp(-\gamma x_i' - \alpha m_i(t))\}^{\delta_i} \times \exp(-\lambda(g_i(t_i))) \quad (4.7)$$

şeklinde tanımlanabilir. Üstel dağılım tek parametrelidir bir dağılım olduğundan en basit dağılımdır.  $\lambda$ , dağılımın ölçek parametresidir ve dağılımın şekil hakkında bilgisi olmadığından, yaşam verilerinin tehlike şekillerini tanımlamak için yeterince esnek bir yapı sunmamaktadır (Basharat, 2019).

#### 4.1.3. Weibull Hızlandırılmış Başarısızlık Süresi Alt Modeli

Yaşam verilerinin Weibull dağılımına uygun olduğu durumlarda Weibull HBS modeline ilişkin  $t_i$  zamanında tehlike fonksiyonu,

$$h_i(t_i) = \lambda \kappa (\lambda g_i(t_i))^{\kappa-1} \exp(-\gamma x_i' - \alpha m_i(t)) \quad (4.8)$$

olarak elde edilir. Burada  $h_0(g_i(t_i)) = \lambda \kappa (\lambda g_i(t_i))^{\kappa-1}$  biçimindedir. Weibull HBS modeline ilişkin yaşam ve  $b_i$  ve  $\theta$  verildiğinde  $(t_i, \delta_i)$  için olasılık yoğunluk fonksiyonları sırasıyla aşağıdaki gibi ifade edilebilir:

$$S_i(t_i) = S_0(g_i(t_i)) = \exp[-(\lambda g_i(t_i))^\kappa]$$

$$f(t_i, \delta_i | b_i, \theta) = \left\{ \lambda \kappa (\lambda g_i(t_i))^{\kappa-1} \exp(-\gamma x_i' - \alpha m_i(t)) \right\}^{\delta_i} \times \exp[-(\lambda g_i(t_i))^\kappa] \quad (4.9)$$

Weibull dağılımı Üstel dağılımın genelleştirilmiş bir formudur ve  $\lambda > 0$  ve  $\kappa > 0$  sırasıyla ölçek ve şekil parametrelerini göstermek üzere dağılımın iki parametresi bulunmaktadır. Weibull modeli farklı veri tiplerini modelleyebilme (çarpık veya simetrik dağılım) özelliğine sahip olduğundan yaşam çözümlemesinde sıklıkla kullanılmaktadır. Weibull dağılımından elde edilen model, monoton tehlike şekillerini içerdiğinden dolayı esnek bir yapıya sahiptir (Basharat, 2019).

#### 4.1.4. Log-Lojistik Hızlandırılmış Başarısızlık Süresi Alt Modeli

Yaşam verilerinin Log-Lojistik dağılıma uygun olduğu durumlarda Log-Lojistik HBS modeline ilişkin  $t_i$  zamanında tehlike fonksiyonu,

$$h_i(t_i) = \frac{\lambda \kappa (\lambda g_i(t_i))^{\kappa-1}}{1 + (\lambda g_i(t_i))^\kappa} \exp(-\gamma x_i' - \alpha m_i(t_i)) \quad (4.10)$$

şeklinde tanımlanabilir. Burada  $h_0(g_i(t_i)) = \kappa \lambda (\lambda g_i(t_i))^{\kappa-1} / [1 + (\lambda g_i(t_i))^\kappa]$  biçimindedir. Log-Lojistik HBS modeline ilişkin yaşam ve  $b_i$  ve  $\theta$  verildiğinde  $(t_i, \delta_i)$  için olasılık yoğunluk fonksiyonları sırasıyla aşağıdaki gibi ifade edilebilir:

$$S_i(t_i) = S_0(g_i(t_i)) = \exp[1 + (\lambda g_i(t_i))^\kappa]^{-1}$$

$$f(t_i, \delta_i | b_i, \theta) = \left\{ \frac{\lambda \kappa (\lambda g_i(t_i))^{\kappa-1}}{1 + (\lambda g_i(t_i))^\kappa} \exp(-\gamma x_i' - \alpha m_i(t_i)) \right\}^{\delta_i} \times \exp[1 + (\lambda g_i(t_i))^\kappa]^{-1} \quad (4.11)$$

Log-lojistik dağılımı  $\lambda > 0$  ve  $\kappa > 0$  sırasıyla ölçek ve şekil parametreleri olmak üzere iki parametreye sahiptir ve özellikle tek modlu (monoton olmayan-unimodal) fonksiyonları modellemek için kullanılmaktadır. Log-Lojistik dağılımının tehlike fonksiyonu  $\kappa \leq 1$  olduğu durumlarda monoton azalan ve  $\kappa > 1$  olduğunda tek modlu bir yapıya sahiptir. Bu modeller, ölüm oranlarının belirli bir zaman noktasından sonra zirveye ulaştığı ve daha sonra yavaşça azaldığı durumlarda yaygın olarak kullanılmaktadır (Bennett, 1983; Basharat, 2019).

#### 4.1.5. Log-Normal Hızlandırılmış Başarısızlık Süresi Alt Modeli

Yaşam verilerinin Log-Normal dağılıma uygun olduğu durumlarda Log-Normal HBS modeline ilişkin  $t_i$  zamanında tehlike fonksiyonu,

$$h_i(t_i) = \frac{\frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\log(g_i(t_i))-\mu}{\sigma}\right)^2\right]}{1-\Phi\left(\frac{\log(g_i(t_i))-\mu}{\sigma}\right)} \exp(-\gamma x'_i - \alpha m_i(t_i)) \quad (4.12)$$

şeklinde ifade edilebilir. Burada  $\Phi(x) = \int_{-\infty}^x 1/\sqrt{2\pi} \exp(-u^2/2) du$  olmak üzere normal dağılımın birikimli dağılım fonksiyonunu ifade etmektedir. Log-Normal HBS modeline ilişkin yaşam ve  $b_i$  ve  $\theta$  verildiğinde,  $(t_i, \delta_i)$  için olasılık yoğunluk fonksiyonları sırasıyla aşağıdaki gibi ifade edilebilir:

$$h_i(t_i) = S_0(g_i(t_i)) = 1 - \Phi\left(\frac{\log(g_i(t_i))-\mu}{\sigma}\right) \quad (4.13)$$

$$f(t_i, \delta_i | b_i, \theta) = \left\{ \frac{\frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\log(g_i(t_i))-\mu}{\sigma}\right)^2\right]}{1-\Phi\left(\frac{\log(g_i(t_i))-\mu}{\sigma}\right)} \exp(-\gamma x'_i - \alpha m_i(t_i)) \right\}^{\delta_i} \\ \times 1 - \Phi\left(\frac{\log(g_i(t_i))-\mu}{\sigma}\right)$$

Log-Normal model, tek modlu (monoton olmayan) tehlike fonksiyonuna sahip bir diğer popüler dağılımdır. Dağılımın  $-\infty < \mu < \infty$  ve  $\sigma > 0$  olmak üzere iki parametresi bulunmaktadır. Log-Lojistik ve Log-Normal dağılımların tehlike fonksiyonları aynı şekilde sahiptir (Lawless, 2011; Basharat, 2019).

## 4.2. Boylamsal Alt Model

Eşitlik (4.3)'te verilen tehlike fonksiyonu, boylamsal gözlemlerin gerçek gözlemlenmemiş değerlerine bağlıdır (Wu ve ark., 2012). Buna rağmen,  $t_{ij}$  ( $i = 1, 2, \dots, n, j = 1, 2, \dots, n_i$ ) zamanında boylamsal gözlemler  $y_{ij}$  her birimde hatalar ile toplanmaktadır. Bu nedenle, bir birimin tehlikesine boylamsal gözlemlerin etkisini ölçmek için  $m_i(t)$ 'nin tahmin edilmesi gerekmektedir (Rizopoulos, 2012a).  $t_{ij}$  zamanında sürekli boylamsal gözlem  $y_{ij}$  tahminlerini elde etmek için DKEM kullanılarak boylamsal alt model aşağıdaki gibi yazılabilir.

$$y_{ij} = \mu_i(t_{ij}) + m_i(t_{ij}) + \epsilon_{ij} \\ = x'_i(t)\beta + z'_i(t)b_i \quad (4.14)$$

Burada,  $x'_i(t)$  ve  $z'_i(t)$  sırasıyla  $\beta$  sabit etkiler ve  $b_i$  rastgele etkilere karşılık gelen tasarım vektörüdür.  $m_i(t)$  rastgele etkiler arasındaki doğrusal kombinasyonu ifade etmektedir.  $b_i$

rastgele etkilerdir ve  $b_i \sim N(0, \Sigma_b)$  olduğu varsayılmaktadır.  $\epsilon_i$ , 0 ortalama ve  $\sigma^2$  varyans ile normal dağılıma sahiptir ve  $\text{Cov}(\epsilon_i, \epsilon_j) = 0, i \neq j$  olduğu varsayılır. Ayrıca, rastgele etkiler ve hata teriminin birbirinden bağımsız olduğu varsayılmaktadır ( $\text{Cov}(\epsilon_i, b_i) = 0$ ) (Wu, 2009; Fitzmaurice, Laird ve Ware, 2011).

Boylamsal ve yaşam süresi alt modelleri arasındaki bağlantı, her bir birime ait olan ve her iki alt modelde hesaplanan  $m_i(t)$  gizli sıfır ortalamalı ikili Gauss sürecidir. Boylamsal süreçteki  $m_i(t) = U_{i1} + U_{i2}t$  olarak tanımlanır ve  $U_{i1}$  ve  $U_{i2}$  birime özgü rastgele etkileri ifade etmektedir. Ayrıca,  $U_{i1} \sim N(0, \sigma_1^2)$  ve  $U_{i2} \sim N(0, \sigma_2^2)$  normal dağılıma sahiptir ve  $\text{Cov}(U_{i1}, U_{i2}) = \sigma_{u1}, \sigma_{u2}$  şeklinde tanımlanmaktadır. Boylamsal alt modelden elde edilen  $m_i(t)$  değerleri, birleştirme parametresi  $\alpha$  ile yaşam süresi alt modeline eklenerek boylamsal gözlemlerin yaşam verileri üzerinde etkisi araştırılır (Henderson, Diggle ve Dobson, 2000; Powney, 2015; Khan ve Basharat, 2022).

#### 4.3. Bileşik Model Parametre Tahmini

Bileşik modelin parametre tahminlerini elde etmek için çeşitli prosedürler geliştirilmiştir. Boylamsal ve yaşam süresi verilerinin bileşik modellemesinde genellikle kullanılan üç yaklaşım; iki - aşamalı yaklaşım, Bayesci MCMC yöntemi ve bileşik olabilirlik yaklaşımıdır (Rizopoulos ve Ghosh, 2011; Alkhatami, 2021).

Bileşik model parametre tahminlerini elde etmek için kullanılan olabilirlik yaklaşımı, çoklu integrallerin çözümünden kaynaklı bazı zorluklara sahiptir.  $p(t_i, \delta_i | b_i, \theta)$  için zamana göre tek boyutlu integral ve  $b_i$  rastgele etkilere göre çok boyutlu integral çözümü yapılması gerekmektedir. Ancak bu integrallerin çözümü analitik olarak olmadığından ve rastgele etkilere göre integrallerin boyutları arttığında hesaplamada yaşanan zorluklardan dolayı, GH yaklaşımı (Wulfsohn ve Tsiatis, 1997; Henderson ve ark., 2000; Song, Davidian ve Tsiatis, 2002; Rizopoulos, Verbeke ve Molenberghs, 2008), AGH yaklaşımı (Rizopoulos, 2012b), Laplace yaklaşımı (Ye, Lin ve Taylor, 2008; Rizopoulos, Verbeke ve Lesaffre, 2009) ve Bayesci MCMC yaklaşımı (Xu ve Zeger 2001; Guo ve Carlin 2004) gibi sayısal yöntemler kullanılmaktadır.

Bu tez çalışmasında, Bayesci yaklaşımla boylamsal veriler DKEM ile modellendikten sonra, bu tahminler bilgilendirici önsel yardımıyla HBS modellerine eklenerek parametrik bileşik modelleme için iki - aşamalı yaklaşımın bir uzantısı önerilmiştir.

Önerilen BTS yaklaşımının performansı tam olabilirlik yaklaşımlarından GH ve AGH yaklaşımları ile kıyaslandığından, bu iki klasik yaklaşım detaylı olarak incelenmiştir.

#### 4.3.1. İki - Aşamalı Yaklaşım

Bileşik model parametrelerini tahmin etmek için, literatürde farklı iki - aşamalı yaklaşımlar önerilmiştir (Self ve Pawitan, 1992; Tsiatis, DeGruttola ve Wulfsohn, 1995; Wu ve ark., 2012). Self ve Pawitan (1992), ilk adımda rastgele etkilerin tahminlerini elde etmek için en küçük kareler yöntemini dikkate almışlar ve ikinci adımda elde ettikleri bu tahminleri Cox regresyon modelinin kısmi olabilirlik fonksiyonunda yer alan  $m_i(t)$  değerlerini hesaplamak için kullanmışlardır. Tsiatis, DeGruttola ve Wulfsohn (1995) ölçüm hatası içeren boylamsal ölçümleri içeren Cox regresyon modelinin parametre tahminlerindeki yanlılığı azaltmayı amaçlayan iki - aşamalı bir prosedür önermişlerdir. Buna göre, ilk aşamada ölçüm hatası içermeyen boylamsal ölçümlerin tahmini için DKEM kullanmışlar ve ikinci aşamada DKEM'den elde edilen tahminleri, açıklayıcı değişken olarak Cox regresyon modeline eklemişlerdir. Literatürde sıklıkla kullanılan iki - aşamalı yaklaşımın çalışma prensibi aşağıdaki gibidir (Wu ve ark., 2012):

- Birinci aşama: Boylamsal ölçümler, ölçümlerin birime özgü değerlerini tahmin etmek için DKEM kullanılarak modellenir.
- İkinci aşama: Birinci aşamada modellenmiş değerler, zamana bağlı açıklayıcı değişken olarak kullanılarak yaşam süresi modeli elde edilir.

İki - aşamalı yaklaşım hesaplama açısından basit olmasına rağmen boylamsal süreci modellerken, yaşam süresi modelinde boylamsal ölçümlerdeki ihmal edilemeyen kayıp veri sürecini modele dahil edemeyeceğinden dolayı yaşam süreleri bilgisini ihmal ederek yanlılığa ve verimlilik kaybına neden olabilir (Faucett ve Thomas, 1996; Albert ve Shih, 2010; Sweeting ve Thompson, 2011). Literatürde iki - aşamalı yaklaşım parametre tahminlerinde oluşan bu sorunu inceleyen birçok çalışma bulunmaktadır (Self ve Pawitan, 1992; Tsiatis, DeGruttola ve Wulfsohn, 1995; Ye, Lin ve Taylor, 2008; Albert ve Shih, 2010; Leiva-Yamaguchi ve Alvares, 2020).

Parametre tahminlerinde oluşan bu sorunları çözmek için, iki - aşamalı yaklaşıma alternatif olarak bileşik olabilirlik fonksiyonu (Wulfsohn ve Tsiatis, 1997; Tsiatis ve Davidian, 2004; Tseng, Hsieh ve Wang, 2005; Rizopoulos, 2012b) ve Bayesci MCMC yaklaşımları (Brown ve Ibrahim, 2003; Brown, Ibrahim ve DeGruttola 2005; Rizopoulos

ve Ghosh, 2011) önerilmiştir. Bileşik olabilirlik yaklaşımına dayalı ilk çalışmalar, tek boylamsal ölçüm ve yaşam sürelerinin bileşik modellenmesine dayanmaktadır. Ancak, boylamsal gözlemlerin çok değişkenli olduğu veya yaşam verilerinin daha karmaşık yapılarından oluştuğu bileşik modellerde, bileşik olabilirlik yaklaşımında çok boyutlu integrallerden kaynaklı hesaplama zorluğu yaşanmaktadır (Guler ve ark., 2017). Bu nedenle daha karmaşık yapıya sahip bileşik modeller için bazı araştırmacılar iki - aşamalı yaklaşımın farklı versiyonlarını önermişlerdir (Guler ve ark., 2017; Leiva-Yamaguchi ve Alvares, 2020).

#### 4.3.2. Gauss Hermite Yaklaşımı

Analitik olarak çözülemeyen integrallerin yaklaşık olarak hesaplanabilmesi için sıklıkla Gauss karelemesi yöntemi kullanılmaktadır. GH kareleme yöntemi aşağıdaki gibi formalize edilebilir (Naylor ve Smith, 1982):

$$\int_{-\infty}^{\infty} e^{-x^2} f(x) dx \approx \sum_{q=1}^m w_q f(x_q) \quad (4.15)$$

Burada,  $x_q$  ve  $w_q$  kuadratik nodlar (nodes) ve ağırlıklardır ve  $w_q = 2^{m-1} m! \sqrt{\pi} / m^2 [H_{m-1}(x_q)]^2$  ve  $x_q$ ,  $H_m(x)$  Hermite polinomunun  $q$ . kökünü ifade etmektedir.  $\int_{\mathbb{R}^L} q(\theta) \exp\{h(\theta)\} d\theta$  formunda bir integral tanımlansın. Burada  $\theta$ ,  $L$  boyutlu bir vektör ve  $\hat{\theta}$ ,  $h$ 'in modu olduğu varsayımında  $H = -\frac{\partial^2 h(\hat{\theta})}{\partial \theta \partial \theta'}$  ile gösterilsin. Ek olarak,  $B'B = H$  ifadesiyle  $H$ 'in Cholesky faktörizasyonunu gösterebilirsin.

GH ağırlık fonksiyonu  $N(0, 2^{-1}I)$  dağılımına orantılıdır ve GH kuralını kullanmak için bu dağılıma uygun bir fonksiyon seçilmektedir.  $\theta$ ,  $h(\theta)$  log-yoğunluk fonksiyonu ile rastgele bir niceliği göstermek üzere, bu dağılımın birinci dereceden normal dağılıma yaklaşımı  $N(\hat{\theta}, H^{-1})$  olarak ifade edilebilir. Burada  $\theta$ ,  $A = 2^{-\frac{1}{2}} B(\theta - \hat{\theta})$  olarak parametrize edilirse,  $A$  değişkeni yaklaşık olarak  $N(0, 2^{-1}I)$  dağılımına sahip olacaktır.

Buna göre yukarıda, tanımlanan integral  $2^{\frac{L}{2}} |B|^{-1} \int \exp(-A'A) f(A) dA$  olarak ifade edilebilir. Burada  $f(A) = q(2^{1/2} B^{-1} A + \hat{\theta}) \exp(h(2^{1/2} B^{-1} A \hat{\theta}) + AA')$  biçimindedir ve integral  $Q$ -noktalı GH kuralı kullanılarak yaklaşık olarak  $2^{L/2} |B|^{-1} \sum_{q_1=1}^Q w_{q_1} f(\theta_{q_1})$  yazılabilir (Pericleous, 2016).

Bileşik modelin olabilirliğinde yer alan integrallerin kapalı formda çözümü olmadığından, parametre tahminlerini elde etmek için GH yaklaşımı gibi iterasyon yöntemler sıklıkla kullanılmaktadır (Rizopoulos, 2012b; Pericleous, 2016). Bileşik modelde maksimum olabilirlik tahmini yaşam süresi ve boylamsal gözlemler  $\{T_i, \delta_i, y_i\}$ 'nin, bileşik olabilirliğine karşılık gelen log-olabilirlik fonksiyonunun maksimize edilmesine dayanmaktadır. Bileşik dağılımın tanımlanması için zamandan bağımsız rastgele etkiler vektörü  $b_i$ 'nin hem boylamsal hem de yaşam süresi süreçleri ile ilişki olduğu varsayılır. Buna göre, rastgele etkilerin hem boylamsal ve yaşam süresi verileri arasındaki ilişkiyi hem de boylamsal modeldeki tekrarlı ölçümler arasındaki korelasyonu (koşullu bağımsızlık) hesaplamaya dahil ettiği söylenebilir. i. birim için zamandan bağımsız rastgele etkiler ( $b_i$ ) tarafından açıklanan boylamsal ve yaşam süresi verileri için log-olabilirlik fonksiyonu,

$$\begin{aligned} \log p(T_i, \delta_i, y_i; \theta) &= \log \int p(T_i, \delta_i, y_i, b_i; \theta) db_i \\ &= \log \int p(T_i, \delta_i | b_i; \theta_t, \gamma) \left[ \prod_j p\{y_i(t_{ij}) | b_i; \theta_y\} \right] p(b_i; \theta_b) db_i \end{aligned} \quad (4.16)$$

biçiminde yazılabilir. Burada,  $\theta = (\theta'_t, \theta'_y, \theta'_b)$  tüm parametre vektörü,  $\theta_t$ ,  $\theta_y$  ve  $\theta_b$  sırasıyla yaşam süresi verileri, boylamsal verileri ve rastgele etkilere ilişkin kovaryans matris parametrelerini göstermektedir.  $p(T_i, \delta_i | b_i; \theta_t, \gamma)$  yaşam süresi alt modelinin koşullu olasılık yoğunluk fonksiyonunu,  $p(y_i | b_i; \theta)p(b_i; \theta)$  ise boylamsal ölçümler ve rastgele etkilerin bileşik olasılık yoğunluk fonksiyonunu göstermekte ve parametrik bileşik modeller için sırasıyla aşağıdaki gibi tanımlanmaktadır:

$$\begin{aligned} p(T_i, \delta_i | b_i; \theta_t, \gamma) &= h_i(T_i | m_i(T_i); \theta_t, \gamma)^{\delta_i} S_i(T_i | m_i(T_i); \theta_t, \gamma) \\ &= [h_0(g_i(T_i)) \exp\{-\gamma x'_i - \alpha m_i(t)\}]^{\delta_i} \times S_0(g_i(T_i)) \end{aligned} \quad (4.17)$$

$$\begin{aligned} p(y_i | b_i; \theta)p(b_i; \theta) &= \prod_j p\{y_i(t_{ij}) | b_i; \theta\} p(b_i; \theta_b) \\ &= (2\pi\sigma^2)^{-n_i/2} \exp\{-\|y_i - x'_i\beta - z'_i b_i\|^2 / 2\sigma^2\} \\ &\quad \times (2\pi)^{-q_b/2} \det(D)^{-1/2} \exp(-b'_i D^{-1} b_i / 2) \end{aligned} \quad (4.18)$$

Burada  $q_b$  rastgele etkiler vektörünün boyutunu,  $\|x\| = \{\sum_i x_i^2\}^{1/2}$  Öklit vektör normunu ve  $h_0(\cdot)$  yaşam süresinin bilinen dağılımına ilişkin temel tehlike fonksiyonunu ifade etmektedir (Rizopoulos, 2012a).

Eşitlik (4.16)'da tanımlanan bileşik olabilirliğin skor vektörü aşağıdaki formda yeniden tanımlansın:

$$\begin{aligned} S(\theta) &= \frac{\partial \ell(\theta)}{\partial \theta'} = \sum_i \frac{\partial}{\partial \theta'} \log \int p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) db_i \\ &= \sum_i \int A(\theta, b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i \end{aligned} \quad (4.19)$$

Burada,  $A(\cdot)$ ,  $A(\theta, b_i) = \partial \{ \log p(T_i, \delta_i | b_i; \theta) + \log p(y_i | b_i; \theta) + \log p(b_i; \theta) \} / \partial \theta'$  tüm verilerden elde edilen skor vektörünü ifade etmektedir. Bu vektörün formüllerinde yer alan integraller kapalı formda çözüme sahip olmadığından, önceden belirlenmiş apsislerde integrant hesaplamalarının ağırlıklı toplamına göre integrali yaklaştıran (Eşitlik 4.15) ve sayısal bir yöntem olan GH kuralı tarafından integral çözümü yapılabilmektedir (Press ve ark., 2007). Rastgele etkilerin  $A(\cdot)$  fonksiyonunun herhangi bir formu için ve karesel kural altında, skor vektörü Eşitlik (4.20)'deki gibi hesaplanmaktadır (Rizopoulos, 2012b):

$$\begin{aligned} E\{A(\theta, b_i) | T_i, \delta_i, y_i; \theta\} &= \int A(\theta, b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i \\ &\approx 2^{\frac{q}{2}} \sum_{t_1, \dots, t_q} \pi_t A(\theta, b_t \sqrt{2}) p(b_t \sqrt{2} | T_i, \delta_i, y_i; \theta) \exp(-\|b_t\|^2) \end{aligned} \quad (4.20)$$

Burada  $q$  rastgele etkiler vektörünün boyutunu,  $K$  kareleme (quadrature) noktalarının sayısını göstermek üzere  $\sum_{t_1, \dots, t_q}$ ,  $\sum_{t_1=1}^K \dots \sum_{t_q=1}^K$  ifadesinin kısaltılmış versiyonunu ve  $b'_t = b_{t_1}, \dots, b_{t_q}$ ,  $\pi_t$  ağırlıkları için karşılık gelen apsisleri ifade etmektedir.

GH yaklaşımının kalitesi  $K$  kareleme noktalarının sayısı arttıkça artsa da bileşik olabilirlik için GH kuralı, her rastgele etki için apsislerin kartezyen çarpımı üzerinden integrant hesaplamaları gerektirdiğinden,  $q$  rastgele etkiler vektörünün boyutu arttıkça hesaplama yükü artmaktadır (Rizopoulos, 2012b).

### 4.3.3. Uyarlanmış Gauss Hermite Yaklaşımı

Bileşik modelin parametre tahminleri için GH kuralının kalitesini,  $K$  kareleme noktalarının sayısı ve  $q$  rastgele etkiler vektörünün boyutları etkilese de integrantın ana kütlelerinin konuma göre kareleme noktalarının yerleşimi de etkilemektedir. Eğer,  $g(b) = A(\theta, b)p(b | T_i, \delta_i, y_i; \theta)$  sıfırdan uzak bir noktada konumlanmışsa veya  $g(b)$ 'deki yayılma, ağırlık fonksiyonu  $\exp(-b'b)$ 'den oldukça farklıysa, standart GH yaklaşımını doğrudan  $g(b)$ 'ye uygulamak, büyük  $K$  değerlerinde bile kareleme kurallarındaki

apsisler,  $g(b)$ 'nin kütlesinin çoğunun bulunduğu yerde konumlanamayacağından dolayı kötü sonuçlar verebilir (Pinheiro ve Bates, 1995; Rizopoulos, 2012b). Bu sorunu çözmek için integrantın her optimizasyon algoritması adımı merkezlendiği ve ölçeklendiği AGH kuralı önerilmiştir (Rizopoulos, 2012a):

$$E\{A(\theta, b_i) | T_i, \delta_i, y_i; \theta\} \approx 2^{\frac{q}{2}} \sum_{t_1, \dots, t_q} \pi_t A(\theta, \hat{r}_t \sqrt{2}) p(\hat{r}_t \sqrt{2} | T_i, \delta_i, y_i; \theta) \exp(-\|\hat{r}_t\|^2) \quad (4.21)$$

Burada,

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

$$\hat{b}_i = \operatorname{argmax}_b \{\log p(T_i, \delta_i, y_i, b; \theta)\}$$

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

Bu dönüşüm kullanılarak, integrant yaklaşık olarak  $N(0, 2^{-1}I)$  dağılımının yoğunluğu gibi davranır ve GH ağırlık fonksiyonu bu yoğunluğa orantılı olduğundan, optimum yaklaşım elde edilebilir. Bu nedenle, AGH yaklaşımı, standart GH yaklaşımı ile kıyaslandığında aynı büyüklükte bir yaklaşım hatası elde etmek için genellikle çok daha az kareleme gerektirmektedir. Ancak,  $\hat{b}_i$ 'nin modunun konumu ve her bir iterasyonda her birimin ikinci derece türev matrisi  $\hat{H}_i$ 'nin hesaplanması gereksinimi, AGH yaklaşımının hesaplama yükünü önemli derecede arttırmaktadır (Rizopoulos, 2012b).

## 5. BOYLAMSAL VE YAŞAM VERİLERİNİN PARAMETRİK BİLEŞİK MODELLEMESİNDE BAYESÇİ YAKLAŞIM

Literatürde bileşik model parametre tahminlerini elde etmek için Bayesci yaklaşım sıklıkla kullanılmaktadır. Bayesci MCMC yaklaşımı genellikle bileşik olabilirlik yönteminde çoklu (multiple) boylamsal veriler ile yaşam verilerinin bileşik model tahmini için kullanılırken (Brown ve Ibrahim, 2003; Brown ve ark., 2005; Rizopoulos ve Ghosh, 2011), iki - aşamalı yaklaşımda farklı boyutlardaki boylamsal ve yaşam verilerinin bileşik modellemesinde daha az yanlış parametre tahminleri elde etmek için kullanılmıştır (Murawska, Rizopoulos ve Lesaffre, 2012; Guler ve ark., 2017; Mauff ve ark., 2020; Leiva-Yamaguchi ve Alvares, 2020).

Bu bölümde, HBS modelleri ile boylamsal verilerinin parametrik bileşik modellemesinde parametre tahminlerini elde etmek için önerilen BTS yaklaşımı detaylı olarak incelenmiştir. Bu nedenle öncelikle Bayesci yaklaşım ve MCMC yönteminden kısaca bahsedilerek, BTS yaklaşımda kullanılan önsel dağılımlar tanıtılmıştır.

### 5.1. Bayesci Yaklaşım

Bayesci yaklaşım, parametrelerin olasılık dağılımlarına dayalı olarak modelin tahmin edilmesi süreci olarak tanımlanmaktadır. Bu yaklaşımda parametreler rastgele değişken olarak kabul etmektedir ve parametreler hakkında mevcut bilgi, olasılık dağılımlarıyla ifade edilmektedir (Van de Schoot ve ark., 2014; Alkhathami, 2021).

Bayesci yaklaşım, parametrelerin rastgele değişkenler olarak kabul edildiği durumlarda istatistiksel modelleme için oldukça etkili bir yöntemdir.  $\theta$  parametresi veya parametre vektörü  $n$  gözlemlenmiş  $y$  verisinden tahmin edilmek istenir. Buna göre, verinin  $\theta$  parametresine bağlı olarak koşullu olasılığını açıklayan istatistiksel modeli kullanarak elde edilirken Bayesci yaklaşımın temel unsurları aşağıdaki gibi tanımlanabilir (Bolstad ve Curran, 2016):

- Önsel dağılım:  $\theta$  için belirlenen olasılık dağılımı  $p(\theta)$  olarak ifade edilmektedir. Önsel dağılım, veriler değerlendirilmeden önce parametrenin ortalaması, yayılımı ve çarpıklığını tanımlamaktadır. Araştırmacı tarafından belirlenen önsel dağılımın seçimi özellikle küçük örneklerde çok önemliken, büyük örneklerde sonsal dağılım üzerinde etkisi daha azdır. Bayesci yaklaşımda önsel dağılım

seçimi için bilgilendirici ve bilgilendirici olmayan önseller olmak üzere iki farklı grup bulunmaktadır. Bilgilendirici olmayan önseller, veri hakkında bilgi taşımayan önsel dağılımlar olarak tanımlanır ve sonsal dağılım üzerinde minimal etkiye sahiptir. Düzgün (Uniform, flat), belirsiz ve Jeffreys önsel dağılımları bilgilendirici olmayan önsel dağılımlardır. Bilgilendirici önsel dağılımlar ise model parametrelerinin mevcut ön bilgileri veya uzman görüşleri dikkate alınarak oluşturulan ve olabirlik fonksiyonu üzerinde etkisi olmayıp sonsal dağılım üzerinde etkisi olan önsellerdir (Öztürk ve Cengiz, 2017).

- Olabirlik fonksiyonu: Gözlemlenen verilerin  $\theta$  parametresine bağlı olarak dağılımını açıklayan  $p(y|\theta)$  olabirlik fonksiyonu olarak ifade edilmektedir.
- Sonsal dağılım:  $\theta$  parametresinin önsel dağılımı  $p(\theta)$  ve örneklem olabirlik fonksiyonu  $p(y|\theta)$ 'nin çarpılmasıyla elde edilen bileşik olabirlik fonksiyonu  $p(\theta, y)$ 'nin,  $y$  bilindiği durumda  $\theta$ 'nın koşullu sonsal dağılımı  $p(\theta|y) = p(\theta)p(y|\theta)/p(y)$  olarak ifade edilebilir. Burada  $p(y)$ ,  $y$ 'nin marjinal dağılımını göstermekte olup, sürekli verilerde  $p(y) = \int p(\theta)p(y|\theta)d\theta$  ve kesikli verilerde  $p(y) = \sum_{\theta} p(\theta)p(y|\theta)$  olarak hesaplanmaktadır. Burada integralin sonlu olması durumunda,  $y$ 'nin marjinal olasılığı sonsal dağılım için ek bir bilgi sağlamayacağından dolayı, sonsal dağılım  $p(\theta|y) \propto p(y|\theta)p(\theta)$  olarak ifade edilebilir (Zhao, Yu ve Tong, 2008).

## 5.2. MCMC Yöntemi

MCMC yöntemi, analitik olarak hesaplanması zor olan sonsal dağılımlardan örneklem seçilerek ilgilenilen sonsal nicelikleri hedeflenen dağılımdan hesaplamak için kullanılan bir simülasyon yöntemidir. Bu yaklaşım, Markov zincirine dayalı olup, her bir adımda zincirin bir önceki adımındaki sonuçlara bağlı olarak güncellemesi esasına dayanmakta ve beklenen değeri yaklaşık olarak hesaplamak için kullanılmaktadır. Bu nedenle yöntem doğru olarak uygulandığında, simülasyon ilerledikçe zincir hedeflenen dağılıma yakınsamaktadır (Meyn ve Tweedie, 2012). MCMC yönteminde, sonsal dağılımlardan örneklem seçilmesi işlemi için genellikle Metropolis-Hastings ve Gibbs örnekleme algoritmaları kullanılmaktadır (Givens ve Hoeting, 2005).

Bu tez çalışmasında Bayesci çıkarıma için MCMC yöntemi Gibbs örnekleme algoritması kullanılacağından, Gibbs örnekleme algoritması detaylı olarak ele alınmıştır.

## Gibbs Örneklemesi

Gibbs örneklemesi, uygulama kolaylığından dolayı özellikle çok boyutlu problemlerde yaygın olarak kullanılan bir yöntemdir. Bu algorithmada, ilk olarak parametre vektörü farklı boyutlarda bileşenlere ayrıştırılır ve daha sonra bu bileşenlerin her biri tek tek güncellenir. Belirli bileşenler için tam koşullu olasılıklarından Markov zinciri örneği üretilir ve bunlardan örnek oluşturulur.

$\theta = \theta_1, \theta_2, \dots, \theta_p$  parametre vektörü,  $p(\theta)$  önsel dağılım ve  $p(\theta|y)$  bileşik sonsal dağılımı göstermek üzere, Gibbs örnekleme algoritmasının çalışma prensibi aşağıdaki gibi tanımlanabilir (Zhao, Yu ve Tong, 2008):

- $t$ , iterasyon sayısını göstermek üzere,  $t = 0$  için rastgele başlangıç değeri seç  $\theta^0 = \theta_1^0, \theta_2^0, \dots, \theta_p^0$ .
- $(t + 1)$ . iterasyon için  $\theta$ 'nin her bir  $k$  bileşenini aşağıdaki gibi türet:  
 $P(\theta_1 | \theta_2^{(t)}, \dots, \theta_t^{(t)}, y)$  yardımıyla  $\theta_1^{(t+1)}$ 'yi örnekle.  
 $P(\theta_2 | \theta_1^{(t+1)}, \dots, \theta_t^{(t)}, y)$  yardımıyla  $\theta_2^{(t+1)}$ 'yi örnekle.  
:  
 $P(\theta_k | \theta_1^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, y)$  yardımıyla  $\theta_k^{(t+1)}$ 'yi örnekle.
- Önceden belirlenen maksimum iterasyon sayısı  $T$  olmak üzere, eğer  $t < T$  ise 2. adıma tekrar dönülür, aksi halde işlem sonlandırılır.

### 5.3. MCMC Yönteminin Performansı

Markov zincirleri çalıştırıldığında, bu zincirlerin performansını değerlendirmek oldukça önemlidir. MCMC sonuçlarını değerlendirmek için en az iki zincirin çalıştırılması gerekmektedir. Buna göre, Markov zincirlerinin performansını değerlendirmek için bazı önemli özellikler aşağıda verilmiştir.

#### MCMC Karıştırma (MCMC Mixing)

MCMC algoritmasının verimliliği karıştırma yeteneğine bağlıdır. Bir zincirin karıştırma özelliği ise başlangıç değerlerini ne kadar çabuk unuttuğu ve bir zincirin hedef dağılımın kendisini ve şeklini ne kadar çabuk keşfedebildiği ile ilişkilidir.

### **Zincir Sayısı (Number of Chains)**

MCMC algoritmasında zincir sayısı konusunda farklı görüşler bulunmaktadır. Gelman ve Rubin (1992) birden fazla uzun zincir önerirken, Geyer (1992) tek bir uzun zincirin yeterli olabileceğini önermiştir. Zincir seçimindeki temel fikir, zincirin hedef dağılımın moduna ulaşarak burada çok uzun süre kalabilmesidir. Uygulamada MCMC algoritmasında genellikle iki veya daha fazla zincir çalıştırılarak, bu zincirlerin en az birinin hedef dağılıma yakınsaması ve dağılım ile ilgili tüm özellikleri keşfetmesi beklenmektedir.

### **Yanma ve Durma Süresi (Burn-in and Stopping Time)**

Bir Markov zinciri uzun süre çalıştırdıktan sonra da başlangıç değerine bağımlı olabilir. Seçilen başlangıç değerleri sonsal dağılımın modundan farklıysa, bu bağımlılık zincirin yavaş yakınsamasına neden olmaktadır (Cowles ve Carlin, 1996). Uygulamada, başlangıç S iterasyonları, zinciri başlangıç değerinden bağımsız hale getirmek ve hedef dağılıma daha hızlı yakınsamak için yanma süresi (burn-in time) olarak tanımlanmaktadır. Ek olarak, bir zincir, iyi bir karışım elde etmek için zinciri uzun bir süre çalıştırdıktan sonra belirli bir zamanda durdurulmalıdır. Durdurma süresine karar vermek zor olsa da Gilks, Richardson ve Spiegelhalter (1995) birkaç uzun zincir çalıştırmak ve her zincirden gelen sonsal ortalamaları veya medyanları karşılaştırmanın, durdurma süresi belirlemenin bir yolu olduğunu belirtmişlerdir.

## **5.4. Bayesci İki-Aşamalı Yaklaşım**

Bileşik model olabilirliğinde karmaşık yapı nedeniyle ve tahmin sürecini hızlandırmak amacıyla, bileşik modelin parametre tahminlerini hesaplamak için sıklıkla iki - aşamalı yaklaşım kullanılmaktadır. Literatürde bileşik model parametre tahminlerini elde etmek için sıklıkla kullanılan iki - aşamalı yaklaşım Bölüm 4'te detaylı olarak anlatılmıştır. Bu tez çalışmasında, Tsiatis, DeGruttola ve Wulfsohn (1995) tarafından geliştirilen iki - aşamalı yaklaşımın, HBS ve boylamsal verilerin parametrik bileşik model tahminlerini elde etmek için farklı bir versiyonu olan BTS yaklaşımı dikkate alınmıştır (Yıldırım ve Karasoy, 2024).

Buna göre önerilen yaklaşımda, ilk aşamada, DKEM yardımı ile boylamsal alt modelin ( $\hat{\beta} = E(\beta|y)$ ) parametreleri ve yaşam süresi alt modeli için paylaşılmış rastgele etkilerin farklı bir versiyonu olan gizli sıfır ortalamalı ikili Gauss süreci ile ( $\hat{b} = E(b|y)$ ) hesaplanmaktadır. İkinci aşamada ise ilk aşamadan elde edilen  $\hat{m}_i(t)$  tahminlerini Eşitlik

(4.3)'te yerine yazarak HBS alt modellerinin sonsal dağılımları hesaplanmaktadır. Tüm parametre ve hiper parametreler için önsel dağılımlar, Leiva-Yamaguchi ve Alvares (2020) referans alınarak belirlenmiştir. Buna göre  $\beta, \alpha, \gamma \sim N(0, 100)$ ,  $\sigma \sim \text{half - Cauchy}(0, 5)$  ve  $V$ ,  $K \times K$  boyutlu bir matris olmak üzere  $\Sigma \sim \text{inverse - Wishart}(V, r)$  şeklinde alınmıştır (Yıldırım ve Karasoy, 2024).



## 6. SİMÜLASYON ÇALIŞMASI VE UYGULAMA

Bu bölümde öncelikle Weibull HBS modeli ile DKEM'in gizli sıfır ortalamalı ikili Gauss süreci ile birleştirilmesinden elde edilen parametrik bileşik model dikkate alınmıştır. Bu modelin parametre tahminlerini hesaplamak için GH, AGH ve BTS yaklaşımları kullanılmıştır. Yöntemlerin performanslarını karşılaştırmak için uygulanan simülasyon çalışması hakkında detaylı bilgi verilmiş ve sonuçlar detaylı olarak karşılaştırılmıştır. Daha sonra önerilen parametre tahmin yaklaşımının uygulanabilirliğini göstermek amacıyla literatürde yer alan iki veri setinde uygulama gerçekleştirilmiştir. Birinci uygulamada, aort kapakçığı değiştirme ameliyatı geçiren hastaların, 1991-2001 yılları arasında elde edilen veri setinde ve ikinci uygulamada AIDS hastalarına ilişkin veri setinde parametre tahmin yöntemleri karşılaştırılarak, sonuçlar detaylı olarak değerlendirilmiştir.

### 6.1. Simülasyon Çalışması

Parametrik bileşik model parametre tahminlerini hesaplamak için önerilen BTS yaklaşımının performansı, literatürde sıklıkla kullanılan GH ve AGH yaklaşımları ile farklı parametre değerleri, örneklem genişlikleri ve durdurma oranlarında gerçekleştirilen simülasyon çalışmaları ile karşılaştırılmıştır. Simülasyon çalışması için kullanılan parametrik bileşik modelin alt modelleri,  $t$  zamanında  $i$ . birim için Eşitlik (4.14) ve Eşitlik (4.8) yardımıyla sırasıyla aşağıdaki gibi yeniden tanımlansın:

$$y_i(t) = x_i'(t)\beta + z_i'(t)b_i + \epsilon_i(t) = \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_i + \beta_2 x_i + \epsilon_i(t)$$

$$h_i(t_i) = \lambda \kappa (\lambda g_i(t_i))^{k-1} \exp(-\gamma x_i' - \alpha m_i(t)) \quad (6.1)$$

Burada  $(b_{0i}, b_{1i}) \sim N(0, \Sigma)$  ve  $\epsilon_i(t) \sim N(0, \sigma^2)$  biçimindedir.  $x$  her iki alt modelde de ortak olan, 0.5 olasılık dağılımına sahip Bernoulli dağılımından türetilmiş açıklayıcı değişkeni göstermektedir. Ayrıca,  $g_i(t_i) = \int_0^{t_i} \exp(-\gamma x_i' - \alpha z_i'(u)b_i) du$ 'nun çözümü için 5-noktalı Gauss-Legendre Kuadratik kuralı kullanılmıştır.

#### 6.1.1. Parametrik Bileşik Model için Veri Simülasyonu

Parametre tahmin yaklaşımlarının performansını incelemek için öncelikle Eşitlik (6.1)'de verilen modellerle uyumlu olarak ve farklı parametre değerleri kullanılarak oluşturulan farklı senaryolarda veri üretilmiştir. Buna göre Senaryo 1 için parametre değerleri;

boylamsal alt model için,  $\beta_0 = 2, \beta_1 = 1.5, \beta_2 = 0.5, \sigma = 0.8, \Sigma_{11} = 0.8, \Sigma_{12} = \Sigma_{21} = -0.025$  ve  $\Sigma_{22} = 0.15$  olarak ve Weibull HBS alt modeli için  $\gamma = -0.3, \alpha = 0.6, \lambda = 0.75$  ve  $\kappa = 1.5$  olarak, senaryo 2 için parametre değerleri;  $\beta_0 = 2, \beta_1 = 1.5, \beta_2 = 0.5, \sigma = 0.2, \Sigma_{11} = 0.5, \Sigma_{12} = \Sigma_{21} = -0.15$  ve  $\Sigma_{22} = 0.15$  olarak ve Weibull HBS alt modeli için  $\gamma = 0.8, \alpha = -2.5, \lambda = 0.25$  ve  $\kappa = 1.5$  olarak ve senaryo 3 için parametre değerleri; boylamsal alt model için senaryo 2 ile aynı olmak üzere, Weibull HBS alt modeli için  $\gamma = 1.3, \alpha = 1.5, \lambda = 0.25$  ve  $\kappa = 1.5$  olarak dikkate alınmıştır. Ayrıca, boylamsal veriler için zaman  $t_{ij} = 0, 2, 4, 6, 8$  ( $t_{\min} = 0$  ve  $t_{\max} = 8$  olmak üzere 5 zaman noktası), örneklem genişliği  $n = 30, 100, 500$  ve durdurma oranı  $c = 0.10, 0.30, 0.60$  olarak belirlenmiştir.

Weibull HBS alt modelinde i. birim için gerçek yaşam süresi  $T^*$ , u uniform dağılımından türetilen rastgele sayı olmak üzere, Eşitlik (6.1)'de verilen yaşam süresi alt modelinden elde edilen yaşam fonksiyonunun ters dönüşüm fonksiyonu yardımıyla  $T^* = S_i^{-1}(u)$  hesaplanmıştır. c durdurma oranını belirlemek için kullanılan kontrol parametresi olmak üzere her birim için durdurma zamanı  $C_i \sim \text{Uniform}(t_{\min}, t_{\max})$  dağılımından ve durdurma oranına temellenerek üretilmiştir. Ek olarak, gözlemlenmiş yaşam süresi  $T_i = \min(T_i^*, C_i)$  ve  $\delta_i = 0$  durdurulmuş olayları ve  $\delta_i = 1$  başarısız olayları göstermek üzere olay göstergesi  $\delta = I(T_i^* \leq C_i)$  olarak hesaplanmıştır.

Parametrik bileşik modelde parametre tahmin yaklaşımlarını incelemek için geliştirilen bu simülasyon senaryolarında, orantılı tehlikeler varsayımını sağlamayan yaşam verilerini üretmek oldukça önemlidir. Bu nedenle, üretilen her veriden sonra R programı “survival” paketi kullanılarak, ölçeklendirilmiş Schoenfeld artıkları testi ve  $\log(-\log(S))$  grafiği yardımıyla yaşam verilerinin orantılı tehlikeler varsayımını sağlayıp sağlamadığı test edilmiştir. Her iterasyonda üretilen verinin orantılı tehlikeler varsayımını sağlamadığını ( $p < \alpha = 0.05$ ) göstermek için p - değerleri kaydedilmiş ve her senaryo için grafiksel olarak gösterilmiştir.

Boylamsal ölçüm  $y_i(t_i)$  ise beş zaman noktasında dengeli olarak, i. birim için Eşitlik (6.1)'deki boylamsal alt modele uygun olarak hesaplanmıştır. Rastgele etkiler  $b_i \sim N(0, \Sigma)$  ve rastgele hata terimi  $\epsilon_i \sim N(0, \sigma^2)$  dağılımından türetilmiştir.

### 6.1.2. Simülasyon Algoritması

DKEM ve Weibull HBS modellerinden bileşik olarak üretilen veri yapısı ve parametre tahmin yaklaşımlarının hesaplanmasına ilişkin adımlar detaylı olarak aşağıda listelenmiştir (Yıldırım ve Karasoy, 2024).

1. Rastgele etkiler ile birleştirilmiş boylamsal ve Weibull HBS verilerini üret.

- Tüm parametrelerin, durdurma oranlarının ve örneklem genişliklerinin başlangıç değerlerini tanımla.
- Her birim için boylamsal zaman noktalarını  $t_i = 0, 2, 4, 6, 8$  olarak ayarla.
- Grup değişkeni  $x_i \sim \text{Bernoulli}(0.5)$ , rastgele etkileri  $(b_{0i}, b_{1i}) \sim N(0, \Sigma)$  ve rastgele hata terimini  $\epsilon_i(t_i) \sim N(0, \sigma^2)$  başlangıç parametrelerine uygun olarak belirlenen dağılımlardan üret.
- $y_i(t_i)$  boylamsal ölçümleri Eşitlik (6.1)'de verilen boylamsal alt modele uygun olarak üret.
- Gerçek yaşam süreleri  $T_i^*$ 'yi,  $\hat{m}_i(t)$  tahminlerini kullanarak Eşitlik (6.1)'daki Weibull HBS alt model yardımıyla ve durdurma süresi  $C_i \sim \text{Uniform}(t_{\min}, t_{\max})$  dağılımına uygun olarak üret.
- Gözlemlenmiş yaşam sürelerini  $T_i = \min(T_i^*, C_i)$  ve olay göstergesini  $\delta = I(T_i^* \leq C_i)$  olarak tanımla.
- Üretilen her bileşik veri için hesaplanan ölçeklendirilmiş Schoenfeld artıkları testinden elde edilen p - değerlerini kaydet.

2. Üretilen parametrik bileşik model yapısına uygun veriyi kullanarak önerilen BTS, GH ve AGH yaklaşımlarından parametre tahminleri elde et.

- Bölüm 4.3.2 ve Bölüm 4.4.3'te tanımlanan GH ve AGH yaklaşımları ile parametre tahminlerini elde et.
- Bölüm 5.4'te tanımlanan BTS yaklaşımı ile parametre tahminlerini elde et.
- Yaklaşımların performansını karşılaştırmak için elde edilen sonuçlardan mutlak yan ve hata kareler ortalaması (HKO) değerlerini hesapla.

Özellikle iki - aşamalı yaklaşımda bileşik model yaşam süresi alt modelinde yer alan parametre tahminleri birleştirme yapısından dolayı daha yanlı elde edildiğinden (Leiva-Yamaguchi ve Alvares, 2020), bu tez çalışmasında yaşam süresi alt modelinde yer alan

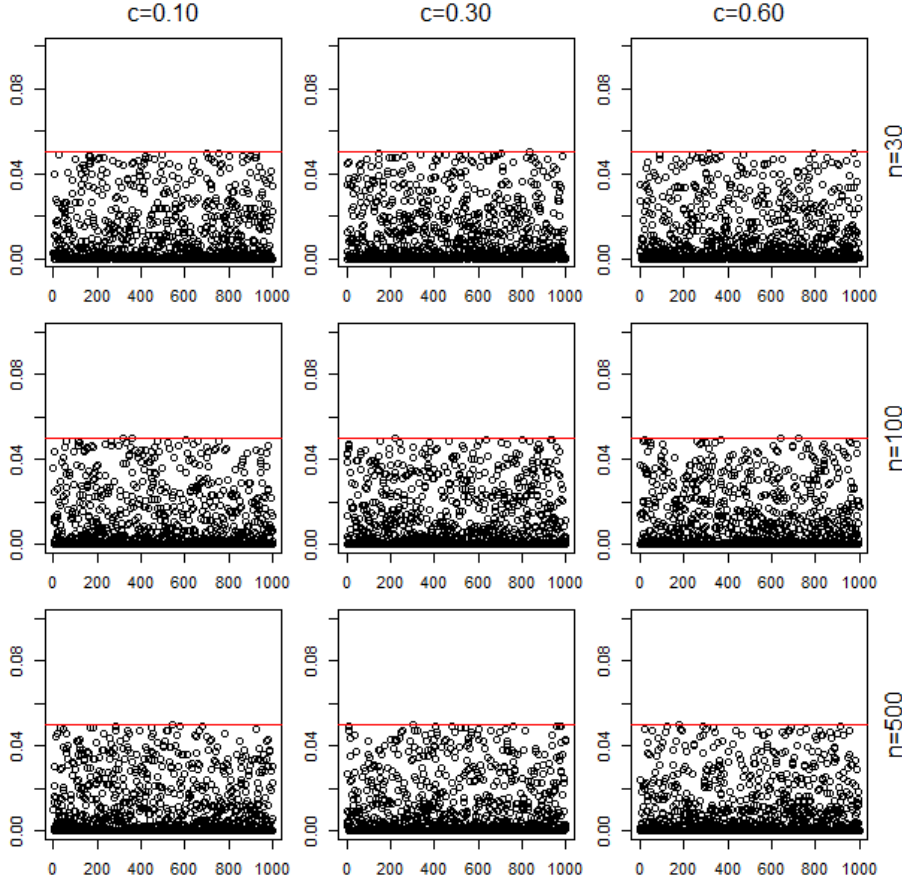
birleştirme katsayısı ( $\alpha$ ) ve açıklayıcı değişken katsayısının ( $\gamma$ ) tahminleri dikkate alınmıştır.

Parametrik bileşik modellerde parametre tahminlerini elde etmek için alternatif olarak önerilen BTS yaklaşımı, Bölüm 5.4'te tanımlanan önsel dağılımlara uygun olarak R programı "rstan" paketinde geliştirilen kod yardımıyla hesaplanmıştır. Önerilen BTS yaklaşımı için MCMC yapılandırılması, boylamsal alt model için 1000 ısınma ile 2000 yineleme ve Weibull HBS alt modeli için 500 ısınma ile 1000 yineleme olarak ayarlanmıştır. Ek olarak, literatürde bileşik modellerin parametre tahminleri için kullanılan GH ve AGH yaklaşımları için R programında "JM" paketi kullanılmıştır. Üç yaklaşımın performansını incelemek için oluşturulan farklı simülasyon senaryoları için 1000 tekrar alınmıştır.

### 6.1.3. Simülasyon Çalışması Sonuçları

Farklı durdurma ve örneklem genişliklerinde yöntemlerin performansını karşılaştırmak için her simülasyon senaryosu için  $HKO = \frac{1}{1000} \sum_1^{1000} (par - \widehat{par}_1)^2$  ve  $|Yan| = \sum_1^{1000} |par - \widehat{par}_1|$  değerleri hesaplanmıştır.

Senaryo 1 için orantılı tehlikeler varsayımına ilişkin farklı durdurma oranları ve örneklem genişliklerinde elde edilen p - değerleri Şekil 6.1'de verilirken,  $\alpha$  parametresi için elde edilen parametre tahmin değerleri Çizelge 6.1 ve Şekil 6.2 - Şekil 6.3'te verilmiştir. Şekil 6.2'de kırmızı yatay çizgi  $\alpha = 0.6$  değerini göstermektedir.

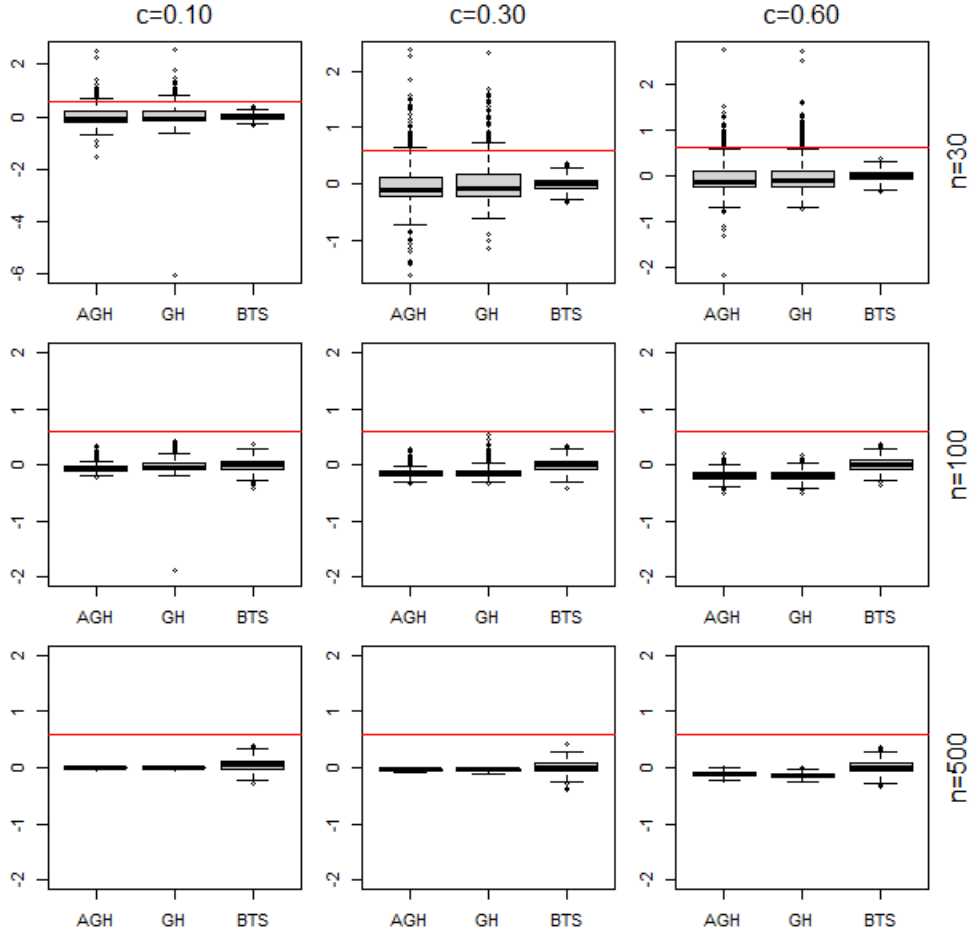


Şekil 6.1. Senaryo 1 için ölçeklendirilmiş Schoenfeld artıkları testinden elde edilen p - değerleri

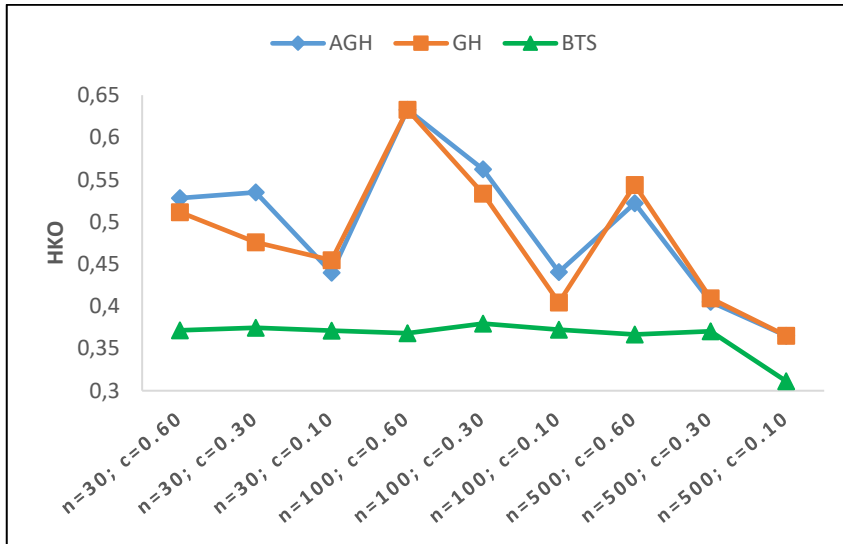
Şekil 6.1’de kırmızı yatay çizgi  $\alpha = 0.05$  değerini göstermektedir. Buna göre, tüm örneklem genişlikleri ve durdurma oranlarında orantılı tehlikeler varsayımı sonuçlarından elde edilen p - değerlerinin  $\alpha = 0.05$ ’ten küçük olduğu görülmektedir. Sonuç olarak, üretilen yaşam verilerininin orantılı tehlikeler varsayımını sağlamadığı söylenebilir.

Çizelge 6.1. Senaryo 1’de  $\alpha$  parametresi için ortalama mutlak yan ve HKO değerleri

		AGH		GH		BTS	
		Yan	HKO	Yan	HKO	Yan	HKO
n=30	c=0.60	0.6716	0.5280	0.6638	0.5114	<b><u>0.5993</u></b>	<b><u>0.3713</u></b>
	c=0.30	0.6680	0.5351	0.6365	0.4757	<b><u>0.6027</u></b>	<b><u>0.3745</u></b>
	c=0.10	0.6066	0.4399	0.6061	0.4544	<b><u>0.5993</u></b>	<b><u>0.3712</u></b>
n=100	c=0.60	0.7909	0.6329	0.7817	0.6328	<b><u>0.5969</u></b>	<b><u>0.3682</u></b>
	c=0.30	0.7461	0.5622	0.7232	0.5335	<b><u>0.6056</u></b>	<b><u>0.3795</u></b>
	c=0.10	0.6603	0.4406	0.6262	0.4046	<b><u>0.6009</u></b>	<b><u>0.3724</u></b>
n=500	c=0.60	0.7215	0.5222	0.7359	0.5436	<b><u>0.5958</u></b>	<b><u>0.3668</u></b>
	c=0.30	0.6364	0.4054	0.6396	0.4095	<b><u>0.5988</u></b>	<b><u>0.3705</u></b>
	c=0.10	0.6041	0.3649	0.6042	0.3650	<b><u>0.5492</u></b>	<b><u>0.3112</u></b>



Şekil 6.2. Senaryo 1’de  $\alpha$  parametresi için kestirim değerleri



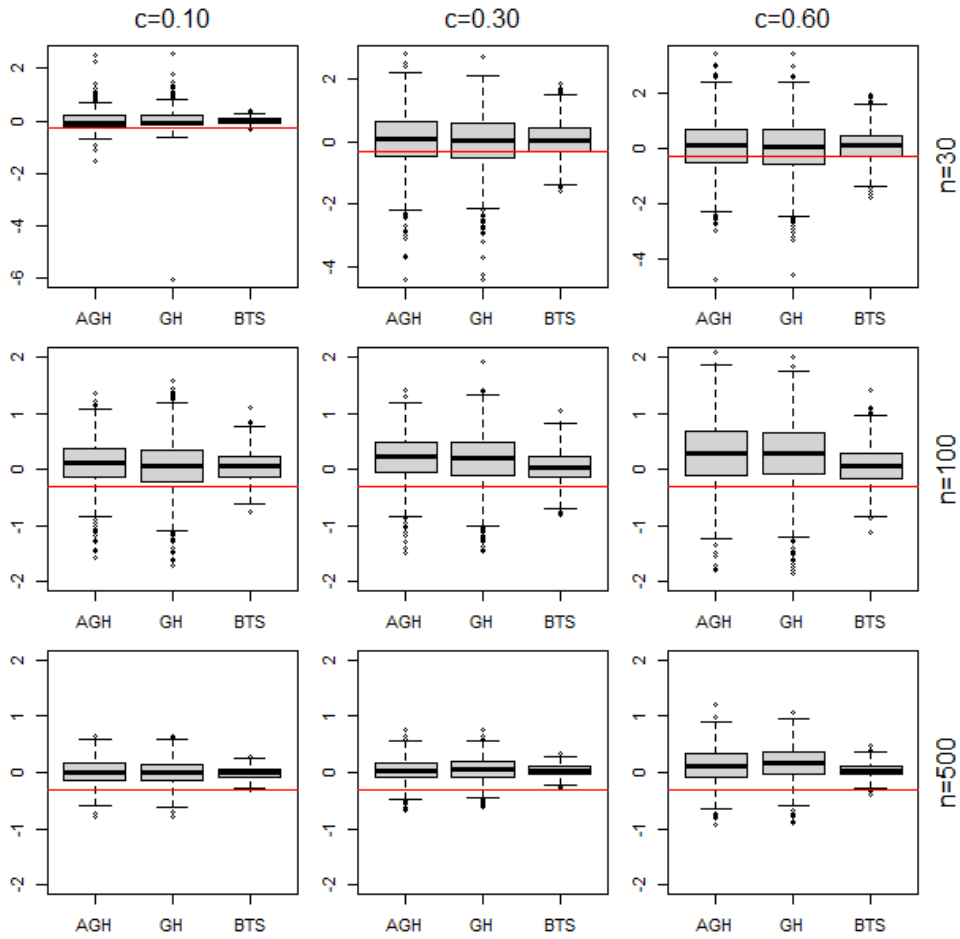
Şekil 6.3. Senaryo 1’de  $\alpha$  parametresi için HKO değerleri

Senaryo 1’de  $\alpha$  parametresi için Çizelge 6.1 ve Şekil 6.2 - Şekil 6.3 incelendiğinde,  $n = 30$  olduğu durumda, tüm durdurma oranlarında BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahip olduğu görülmektedir.  $n = 30$  ve  $c = 0.10$  durumunda ise BTS yaklaşımının  $c = 0.60$  ve  $c = 0.30$  durumlarına kıyasla nispeten daha küçük HKO ve mutlak yan değerine sahip olduğu tespit edilmiştir. Aynı zamanda,  $c = 0.60$  ve  $c = 0.30$  durdurma oranlarında GH yaklaşımının AGH yaklaşımına kıyasla daha küçük HKO ve mutlak yan değerlerine sahip olduğu görülürken,  $c = 0.10$  olduğu durumda ise AGH yaklaşımının GH yaklaşımına kıyasla nispeten daha küçük HKO değerine sahip olduğu görülmüştür.  $n = 100$  durumu için  $n = 30$  durumuna benzer şekilde, tüm durdurma oranlarında BTS yaklaşımı GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahiptir. BTS yaklaşımının  $c = 0.60$  durdurma oranında,  $c = 0.10$  ve  $c = 0.30$  durdurma oranlarına kıyasla daha küçük mutlak yan ve HKO değerlerine ulaştığı görülmüştür. Ek olarak,  $n = 100$  örneklem genişliğinde, tüm durdurma oranlarında GH yaklaşımının AGH yaklaşımına kıyasla daha küçük mutlak yan ve HKO değerlerine sahip tespit edilmiştir.  $n = 500$  olduğu durumda ise diğer durumlara benzer şekilde tüm durdurma oranlarında BTS yaklaşımı, GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahiptir.  $n = 500$  ve  $c = 0.10$  durumunda ise BTS yaklaşımının  $c = 0.60$  ve  $c = 0.30$  olduğu durumlara kıyasla daha iyi sonuçlar verdiği görülmüştür. Aynı zamanda,  $n = 500$  olduğu durumda tüm durdurma oranlarında AGH yaklaşımının mutlak yan ve HKO değerlerinin, GH yaklaşımına kıyasla nispeten daha küçük olduğu tespit edilmiştir. Tüm durumlar kıyaslandığında, BTS yaklaşımının tüm örneklem genişlikleri ve durdurma oranlarında GH ve AGH yaklaşımlarına kıyasla daha iyi sonuçlar verdiği ve özellikle  $n = 500$  ve  $c = 0.10$  olduğu durumda diğer örneklem genişlikleri ve durdurma oranlarına kıyasla en küçük HKO ve mutlak yan değerine sahip olduğu görülmüştür. AGH ve GH yaklaşımlarının da  $n = 500$  ve  $c = 0.10$  olduğu durumda diğer durumlara kıyasla daha küçük HKO ve mutlak yan değerlerine sahip olduğu tespit edilmiştir.

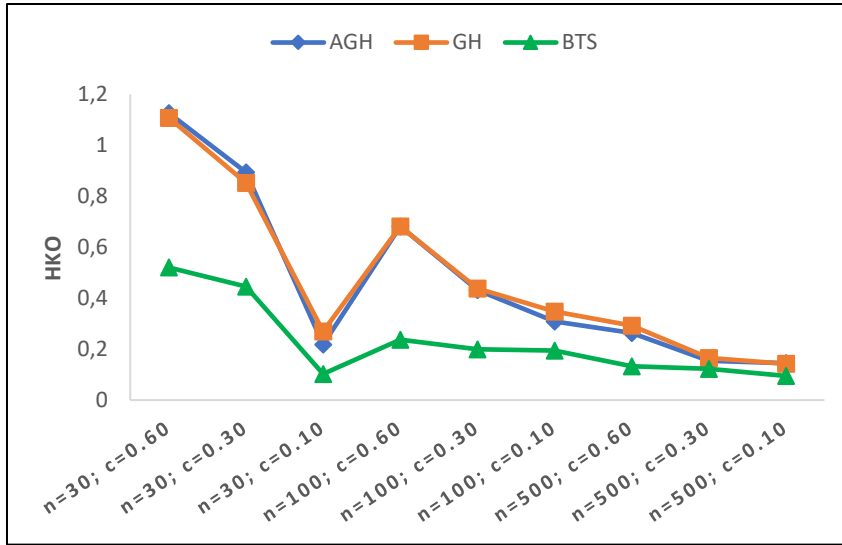
Senaryo 1’de grup parametresi  $\gamma$  için elde edilen parametre tahmin değerleri Çizelge 6.2 ve Şekil 6.4 - Şekil 6.5’te verilmiştir. Şekil 6.4’te kırmızı yatay çizgi  $\gamma = -0.3$  değerini göstermektedir.

Çizelge 6.2. Senaryo 1’de  $\gamma$  parametresi için ortalama mutlak yan ve HKO değerleri

		AGH		GH		BTS	
		Yan	HKO	Yan	HKO	Yan	HKO
n=30	c=0.60	0.8524	1.1267	0.8465	1.1084	<b><u>0.5755</u></b>	<b><u>0.5205</u></b>
	c=0.30	0.7510	0.8941	0.7285	0.8528	<b><u>0.5300</u></b>	<b><u>0.4458</u></b>
	c=0.10	0.3408	0.2181	0.3652	0.2701	<b><u>0.3008</u></b>	<b><u>0.1024</u></b>
n=100	c=0.60	0.6851	0.6818	0.6915	0.6818	<b><u>0.4038</u></b>	<b><u>0.2371</u></b>
	c=0.30	0.5620	0.4313	0.5617	0.4386	<b><u>0.3749</u></b>	<b><u>0.1993</u></b>
	c=0.10	0.4688	0.3093	0.4850	0.3473	<b><u>0.3751</u></b>	<b><u>0.1947</u></b>
n=500	c=0.60	0.4409	0.2630	0.5754	0.2924	<b><u>0.3411</u></b>	<b><u>0.1323</u></b>
	c=0.30	0.3437	0.1541	0.3598	0.1650	<b><u>0.3361</u></b>	<b><u>0.1228</u></b>
	c=0.10	0.3244	0.1449	0.3215	0.1424	<b><u>0.2930</u></b>	<b><u>0.0949</u></b>



Şekil 6.4. Senaryo 1’de  $\gamma$  parametresi için kestirim değerleri



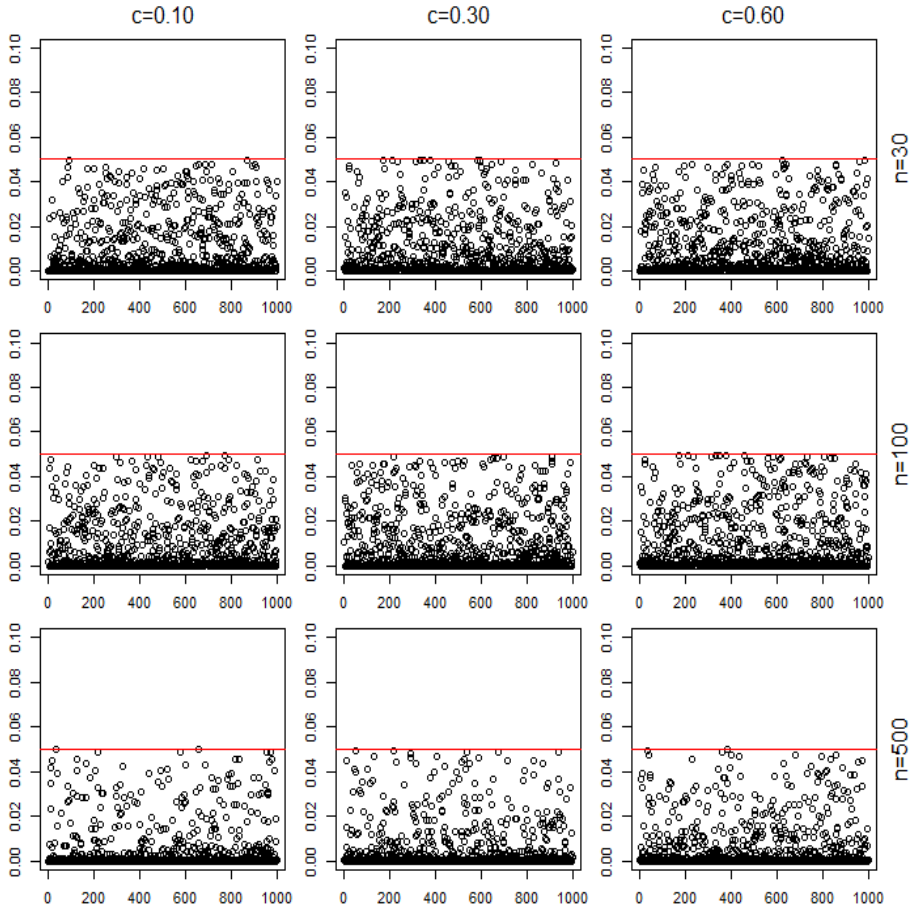
Şekil 6.5. Senaryo 1’de  $\gamma$  parametresi için HKO değerleri

Senaryo 1’de  $\gamma$  parametresi için Çizelge 6.2 ve Şekil 6.4 - Şekil 6.5 incelendiğinde,  $n = 30$  örneklem genişliğinde tüm durdurma oranlarında BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahip olduğu görülmektedir.  $n = 30$  ve  $c = 0.10$  durumunda ise BTS yaklaşımının  $c = 0.60$  ve  $c = 0.30$  durumlarına kıyasla daha küçük HKO ve mutlak yan değerine sahip olduğu tespit edilmiştir. Aynı zamanda,  $c = 0.60$  ve  $c = 0.30$  durdurma oranlarında GH yaklaşımının AGH yaklaşımına kıyasla daha küçük HKO ve mutlak yan değerlerine sahip olduğu görülürken,  $c = 0.10$  olduğu durumda ise AGH yaklaşımının GH yaklaşımına kıyasla nispeten daha küçük mutlak yan ve HKO değerine sahip olduğu görülmüştür.  $n = 100$  durumu için  $n = 30$  durumuna benzer şekilde, tüm durdurma oranlarında BTS yaklaşımı GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahiptir. BTS yaklaşımının  $c = 0.10$  durdurma oranında,  $c = 0.30$  ve  $c = 0.60$  durdurma oranlarına kıyasla daha küçük HKO değerlerine ulaştığı görülmüştür. Ek olarak,  $n = 100$  örneklem genişliğinde tüm durdurma oranlarında AGH yaklaşımının GH yaklaşımına kıyasla daha küçük HKO değerlerine sahip olduğu görülmüştür.  $n = 500$  olduğu durumda ise diğer durumlara benzer şekilde tüm durdurma oranlarında BTS yaklaşımı, GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahipken,  $c = 0.60$  ve  $c = 0.30$  durdurma oranlarında AGH yaklaşımının GH yaklaşımına kıyasla daha küçük mutlak yan ve HKO değerlerine sahip olduğu tespit edilmiştir. Tüm durumlar kıyaslandığında, BTS yaklaşımının tüm örneklem genişliklerinde ve durdurma oranlarında GH ve AGH yaklaşımlarına kıyasla daha iyi

sonular verdiđi ve zellikle  $n = 500$  ve  $c = 0.10$  olduđu durumda diđer rneklem geniřlikleri ve durdurma oranlarına kıyasla en kk HKO ve mutlak yan deđerine sahip olduđu grlmřtr. AGH ve GH yaklařımlarının da  $n = 500$  ve  $c = 0.10$  olduđu durumda diđer durumlara kıyasla daha kk HKO ve mutlak yan deđerlerine sahip olduđu tespit edilmiřtir.

Tm rneklem geniřlikleri ve durdurma oranları karřılařtırıldıđında hem  $\alpha$  hem de  $\gamma$  parametresi iin BTS yaklařımının AGH ve GH yaklařımlarına kıyasla daha yansız parametre tahminleri verdiđi grlrken, tm yaklařımların  $\gamma$  parametresini  $\alpha$  parametresine kıyasla daha iyi tahmin ettiđi sylenebilir. GH ve AGH yaklařımları kıyaslandıđında  $\alpha$  parametresi iin  $n = 30$  ve  $n = 100$  rneklem geniřliklerinde tm durdurma oranları iin GH yaklařımının AGH yaklařımına kıyasla nispeten daha iyi tahminler verdiđi grlrken  $n = 500$  iin AGH yaklařımı GH yaklařımına kıyasla nispeten daha iyi parametre tahmin deđerlerine sahiptir.  $\gamma$  parametresi iin ise  $n = 30$  durumunda GH yaklařımı AGH yaklařımına kıyasla daha iyi sonulara sahipken,  $n = 100$  ve  $n = 500$  iin AGH yaklařımı GH yaklařımına kıyasla daha iyi sonulara sahiptir.

Senaryo 2 için orantılı tehlikeler varsayımına ilişkin farklı durdurma oranları ve örneklem genişliklerinde elde edilen p - değerleri Şekil 6.6’da verilirken,  $\alpha$  parametresi için elde edilen değerler Çizelge 6.3 ve Şekil 6.7 - Şekil 6.8’de verilmiştir. Şekil 6.7’de kırmızı yatay çizgi  $\alpha = -2.5$  değerini göstermektedir.

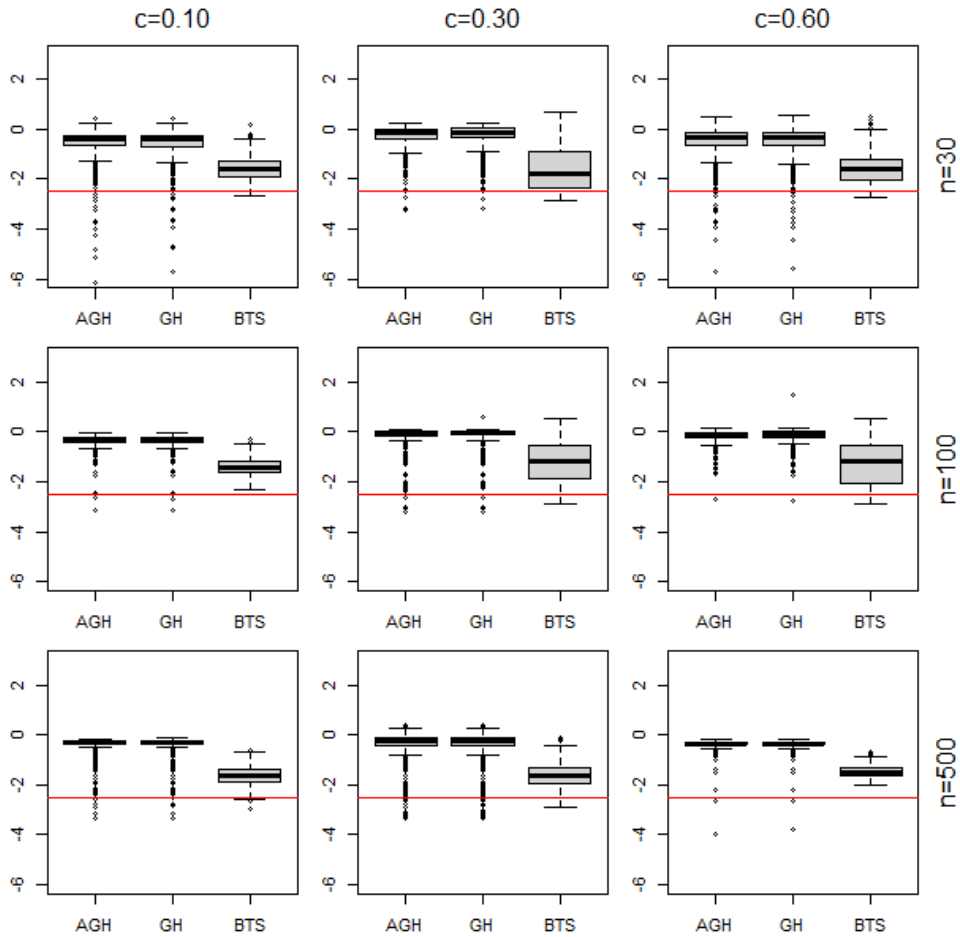


Şekil 6.6. Senaryo 2 için ölçeklendirilmiş Schoenfeld artıkları testinden elde edilen p - değerleri

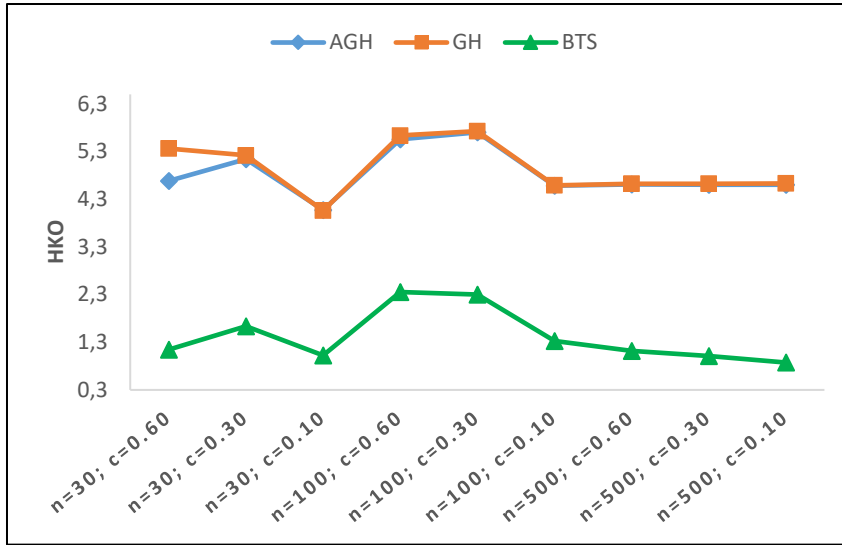
Şekil 6.6’da kırmızı yatay çizgi  $\alpha = 0.05$  değerini göstermektedir. Buna göre, tüm örneklem genişlikleri ve durdurma oranlarında elde edilen p - değerlerinin  $\alpha = 0.05$ ’ten küçük olduğu görülmektedir. Sonuç olarak, üretilen yaşam verilerinin orantılı tehlikeler varsayımını sağlamadığı söylenebilir.

Çizelge 6.3. Senaryo 2’de  $\alpha$  parametresi için ortalama mutlak yan ve HKO değerleri

		AGH		GH		BTS	
		Yan	HKO	Yan	HKO	Yan	HKO
n=30	c=0.60	1.9987	4.6835	1.9679	5.3650	<b><u>0.8989</u></b>	<b><u>1.1442</u></b>
	c=0.30	2.2258	5.1391	2.2479	5.2203	<b><u>0.9194</u></b>	<b><u>1.6341</u></b>
	c=0.10	1.9358	4.0753	1.9354	4.0617	<b><u>0.9038</u></b>	<b><u>1.0215</u></b>
n=100	c=0.60	2.3442	5.5561	2.3617	5.6371	<b><u>1.2581</u></b>	<b><u>2.3533</u></b>
	c=0.30	2.3593	5.7039	2.3646	5.7285	<b><u>1.2882</u></b>	<b><u>2.3002</u></b>
	c=0.10	2.1299	4.5855	2.1312	4.5913	<b><u>1.1009</u></b>	<b><u>1.3301</u></b>
n=500	c=0.60	2.1389	4.6084	2.1436	4.6275	<b><u>1.0308</u></b>	<b><u>1.1217</u></b>
	c=0.30	2.0428	4.6058	2.0500	4.6266	<b><u>0.8788</u></b>	<b><u>1.0099</u></b>
	c=0.10	2.1103	4.6071	2.1165	4.6329	<b><u>0.8733</u></b>	<b><u>0.8741</u></b>



Şekil 6.7. Senaryo 2’de  $\alpha$  parametresi için kestirim değerleri



Şekil 6.8. Senaryo 2’de  $\alpha$  parametresi için HKO değerleri

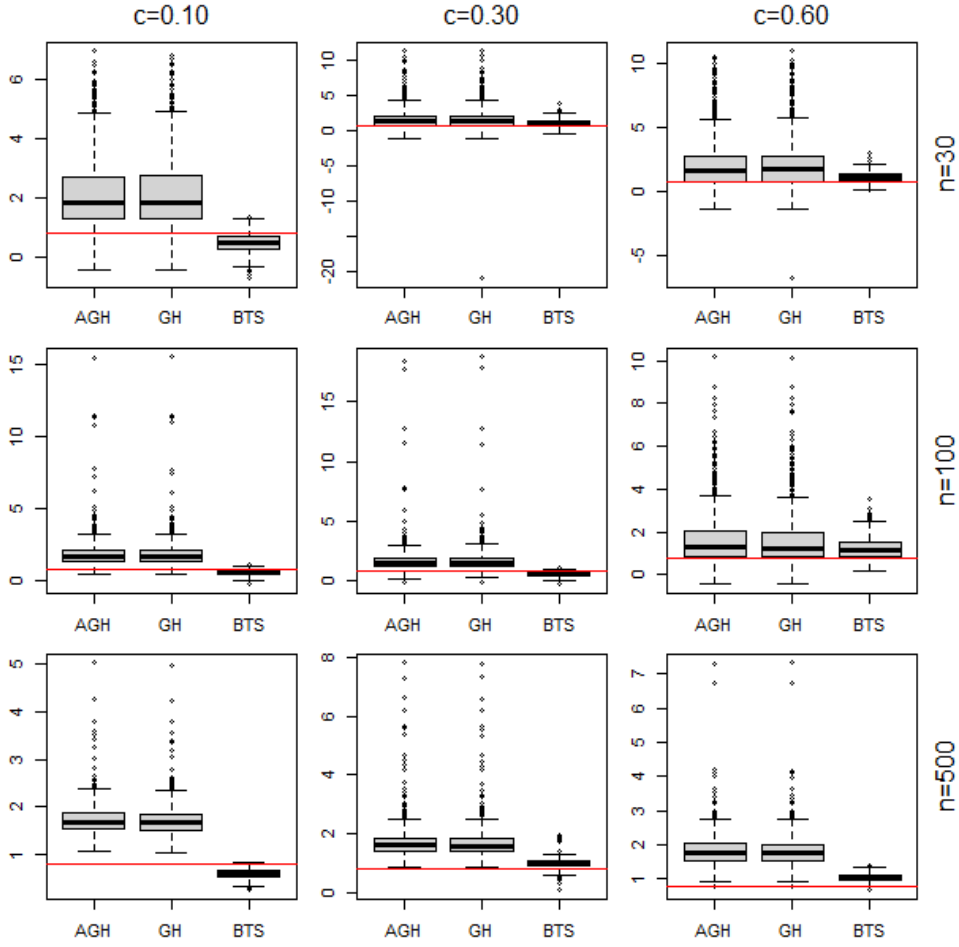
Senaryo 2’de  $\alpha$  parametresi için Çizelge 6.3 ve Şekil 6.7 - Şekil 6.8 incelendiğinde,  $n = 30$  örneklem genişliğinde tüm durdurma oranlarında BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahip olduğu görülmektedir.  $n = 30$  ve  $c = 0.10$  durumunda ise BTS yaklaşımının  $c = 0.60$  ve  $c = 0.30$  durumlarına kıyasla nispeten daha küçük HKO değerine sahip olduğu tespit edilmiştir. Aynı zamanda,  $n = 30$  ve  $c = 0.10$  durumunda GH yaklaşımı AGH yaklaşımına kıyasla daha küçük HKO ve mutlak yan değerlerine sahipken,  $c = 0.60$  ve  $c = 0.30$  olduğu durumlarda AGH yaklaşımı GH yaklaşımına kıyasla nispeten daha küçük HKO değerine sahiptir.  $n = 100$  durumu için  $n = 30$  durumuna benzer şekilde, tüm durdurma oranlarında BTS yaklaşımı GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahiptir. BTS yaklaşımının  $c = 0.10$  durdurma oranında,  $c = 0.60$  ve  $c = 0.30$  durdurma oranlarına kıyasla daha küçük mutlak yan ve HKO değerlerine ulaştığı görülmüştür. Ek olarak  $n = 100$  durumunda, tüm durdurma oranlarında AGH yaklaşımı GH yaklaşımına kıyasla daha küçük mutlak yan ve HKO değerlerine sahiptir.  $n = 500$  olduğu durumda ise diğer durumlara benzer şekilde tüm durdurma oranlarında BTS yaklaşımı, GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahiptir.  $n = 500$  ve  $c = 0.10$  durumunda ise BTS yaklaşımının,  $c = 0.60$  ve  $c = 0.30$  olduğu durumlara kıyasla daha iyi sonuçlar verdiği tespit edilmiştir. Aynı zamanda,  $n = 500$  olduğu durum için tüm durdurma oranlarında AGH yaklaşımının mutlak yan ve HKO değerlerinin, GH yaklaşımına kıyasla nispeten daha küçük olduğu sonucuna ulaşılmıştır.

Tüm durumlar kıyaslandığında, BTS yaklaşımının tüm örneklem genişlikleri ve durdurma oranlarında GH ve AGH yaklaşımlarına kıyasla daha iyi sonuçlar verdiği ve özellikle  $n = 500$  örneklem genişliğinde tüm durdurma oranlarında, diğer örneklem genişlikleri ve durdurma oranlarına kıyasla en küçük HKO ve mutlak yan değerine ulaştığı görülmüştür. Aynı zamanda riskin yüksek olduğu bu senaryo durumunda, yöntemlerin  $\alpha$  parametresini Senaryo 1'e kıyasla daha yanlı tahmin ettiği söylenebilir.

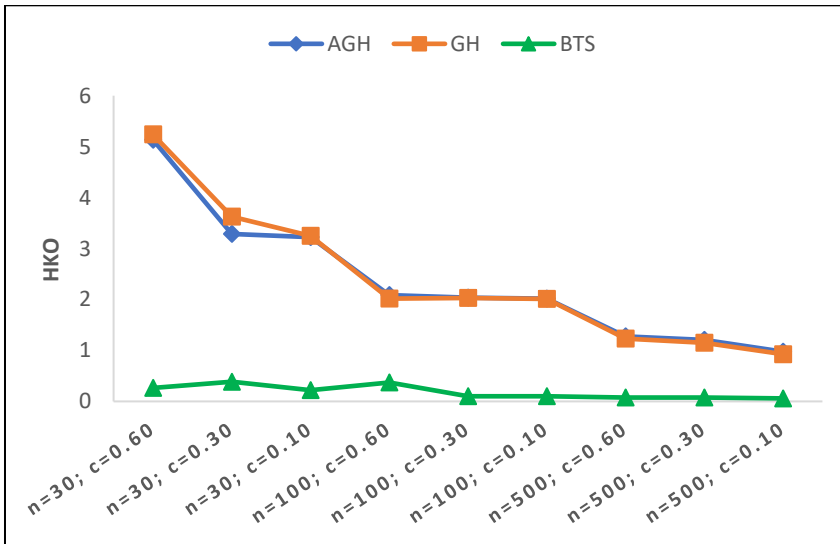
Senaryo 2'de grup parametresi  $\gamma$  için elde edilen parametre tahmin değerleri Çizelge 6.4 ve Şekil 6.9 - Şekil 6.10'da verilmiştir. Şekil 6.9'da kırmızı yatay çizgi  $\gamma = 0.8$  değerini göstermektedir.

Çizelge 6.4. Senaryo 2'de  $\gamma$  parametresi için ortalama mutlak yan ve HKO değerleri

		AGH		GH		BTS	
		Yan	HKO	Yan	HKO	Yan	HKO
n=30	c=0.60	1.2751	5.1349	1.2725	5.2380	<b><u>0.3464</u></b>	<b><u>0.2642</u></b>
	c=0.30	0.9403	3.2851	0.8656	3.6262	<b><u>0.2573</u></b>	<b><u>0.3865</u></b>
	c=0.10	1.3251	3.2253	1.3304	3.2469	<b><u>0.3473</u></b>	<b><u>0.2221</u></b>
n=100	c=0.60	0.8046	2.0885	0.7678	2.0167	<b><u>0.3927</u></b>	<b><u>0.3736</u></b>
	c=0.30	0.8834	2.0380	0.8708	2.0282	<b><u>0.2686</u></b>	<b><u>0.1038</u></b>
	c=0.10	1.0422	2.0162	1.0364	2.0131	<b><u>0.2685</u></b>	<b><u>0.1037</u></b>
n=500	c=0.60	1.0254	1.2743	1.0041	1.2327	<b><u>0.2555</u></b>	<b><u>0.0765</u></b>
	c=0.30	0.9237	1.2052	0.8910	1.1489	<b><u>0.2006</u></b>	<b><u>0.0762</u></b>
	c=0.10	0.9312	0.9762	0.9044	0.9266	<b><u>0.2229</u></b>	<b><u>0.0591</u></b>



Şekil 6.9. Senaryo 2’de  $\gamma$  parametresi için kestirim değerleri

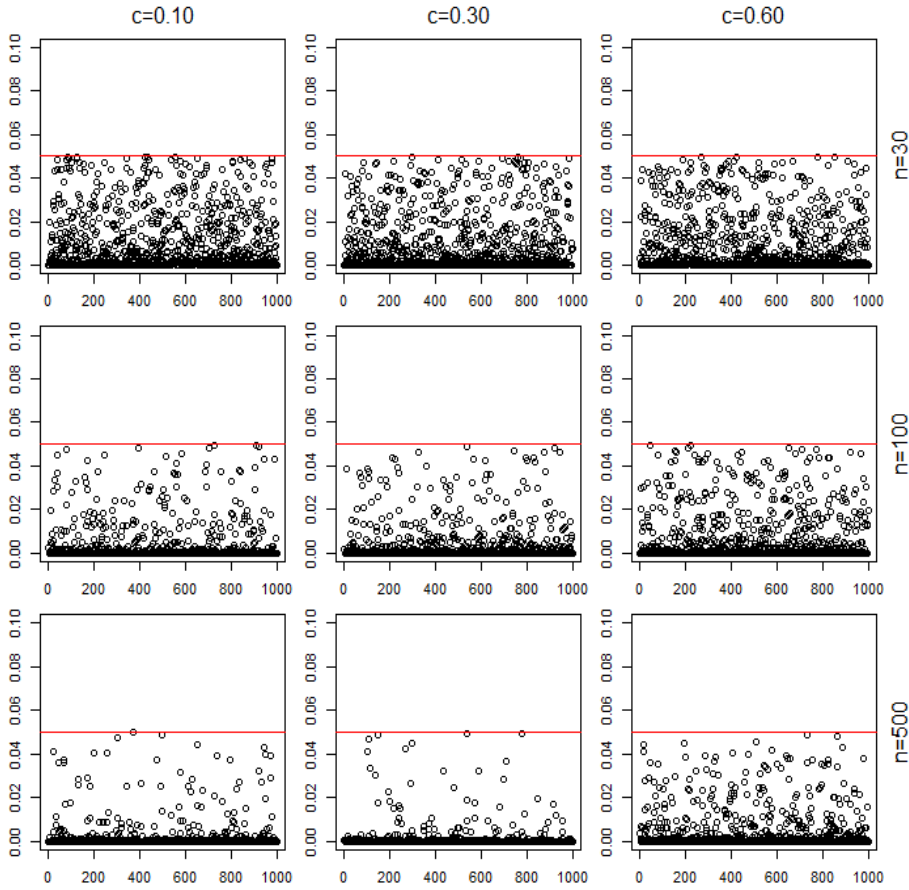


Şekil 6.10. Senaryo 2’de  $\gamma$  parametresi için HKO değerleri

Senaryo 2’de  $\gamma$  parametresi için Çizelge 6.4 ve Şekil 6.9 - Şekil 6.10 incelendiğinde,  $n = 30$  örneklem genişliğinde tüm durdurma oranlarında BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahip olduğu görülmektedir.  $n = 30$  ve  $c = 0.10$  durumunda ise BTS yaklaşımının  $c = 0.60$  ve  $c = 0.30$  durumlarına kıyasla daha küçük HKO değerine sahip olduğu tespit edilmiştir. Aynı zamanda,  $n = 30$  olduğu durumda tüm durdurma oranlarında AGH yaklaşımının GH yaklaşımına kıyasla daha küçük HKO ve mutlak yan değerlerine sahip olduğu görülmüştür.  $n = 100$  durumu için  $n = 30$  durumuna benzer şekilde, tüm durdurma oranlarında BTS yaklaşımı GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahiptir. BTS yaklaşımının  $c = 0.10$  durdurma oranında,  $c = 0.30$  ve  $c = 0.60$  durdurma oranlarına kıyasla daha küçük mutlak yan ve HKO değerlerine ulaştığı görülmüştür. Ek olarak,  $n = 100$  durumunda tüm durdurma oranlarında GH yaklaşımının AGH yaklaşımına kıyasla daha küçük mutlak yan ve HKO değerlerine sahip olduğu görülmüştür.  $n = 500$  olduğu durumda ise diğer durumlara benzer şekilde tüm durdurma oranlarında BTS yaklaşımı, GH ve AGH yaklaşımlarına kıyasla daha küçük mutlak yan ve HKO değerlerine sahipken, tüm durdurma oranlarında GH yaklaşımının AGH yaklaşımına kıyasla daha küçük mutlak yan ve HKO değerlerine sahip olduğu tespit edilmiştir. Tüm durumlar kıyaslandığında, BTS yaklaşımının tüm örneklem genişliklerinde ve durdurma oranlarında GH ve AGH yaklaşımlarına kıyasla daha iyi sonuçlar verdiği ve özellikle  $n = 500$  ve  $c = 0.10$  olduğu durumda diğer örneklem genişlikleri ve durdurma oranlarına kıyasla en küçük HKO değerine sahip olduğu görülmüştür. AGH ve GH yaklaşımlarının da  $n = 500$  ve  $c = 0.10$  olduğu durumda diğer durumlara kıyasla daha küçük HKO değerine sahip olduğu tespit edilmiştir.

Tüm örneklem genişlikleri ve durdurma oranları karşılaştırıldığında hem  $\alpha$  hem de  $\gamma$  parametresi için BTS yaklaşımının AGH ve GH yaklaşımlarına kıyasla daha az yanlış parametre tahminleri verdiği görülürken, tüm yaklaşımların  $\gamma$  parametresini  $\alpha$  parametresine kıyasla daha iyi tahmin ettiği söylenebilir. Riskin yüksek olduğu bu senaryo için, GH ve AGH yaklaşımları kıyaslandığında  $\alpha$  parametresi için tüm örneklem genişliklerinde ve durdurma oranlarında AGH yaklaşımının GH yaklaşımına kıyasla nispeten daha iyi tahminler verdiği görülmüştür.  $\gamma$  parametresi için ise  $n = 30$  durumunda AGH yaklaşımı GH yaklaşımına kıyasla daha iyi sonuçlara sahipken,  $n = 100$  ve  $n = 500$  için GH yaklaşımı AGH yaklaşımına kıyasla daha iyi sonuçlara sahiptir.

Senaryo 3 için orantılı tehlikeler varsayımına ilişkin farklı durdurma oranları ve örneklem genişliklerinde elde edilen p - değerleri Şekil 6.11’de verilirken,  $\alpha$  parametresi için elde edilen parametre tahmin değerleri Çizelge 6.5 ve Şekil 6.12 - Şekil 6.13’te verilmiştir. Şekil 6.12’de kırmızı yatay çizgi  $\alpha = 1.5$  değerini göstermektedir.

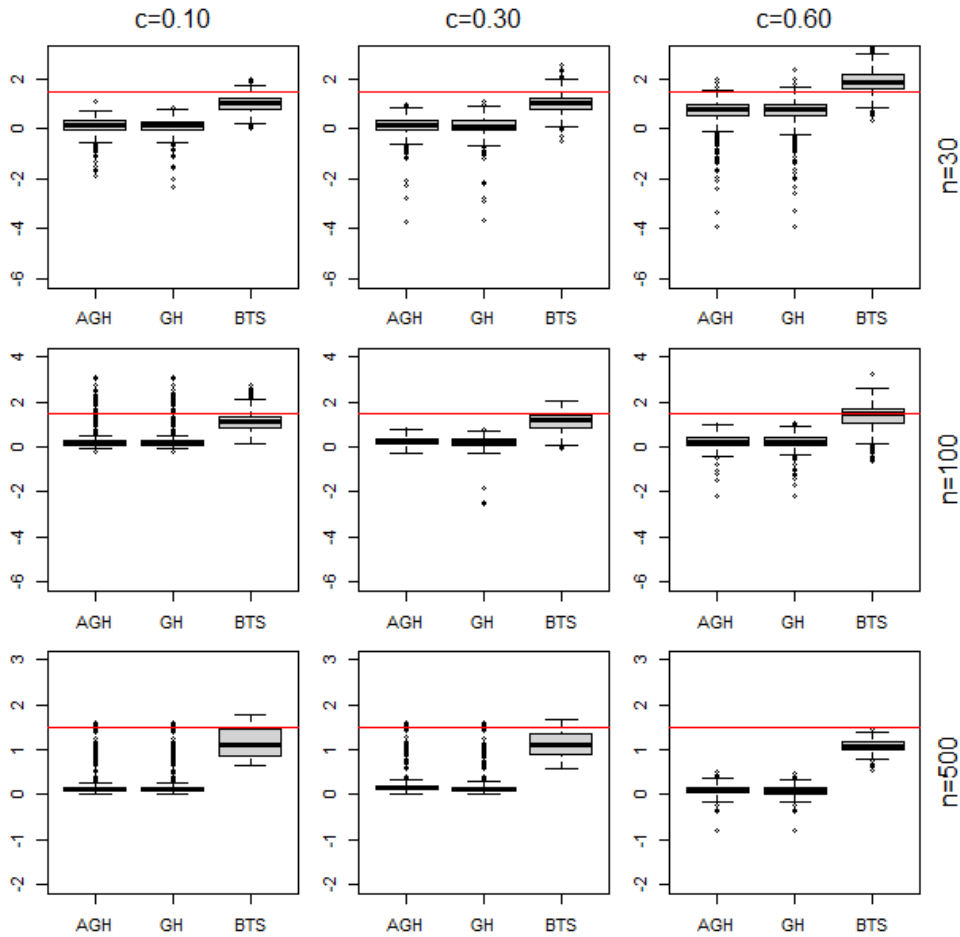


Şekil 6.11. Senaryo 3 için ölçeklendirilmiş Schoenfeld artıkları testinden elde edilen p - değerleri

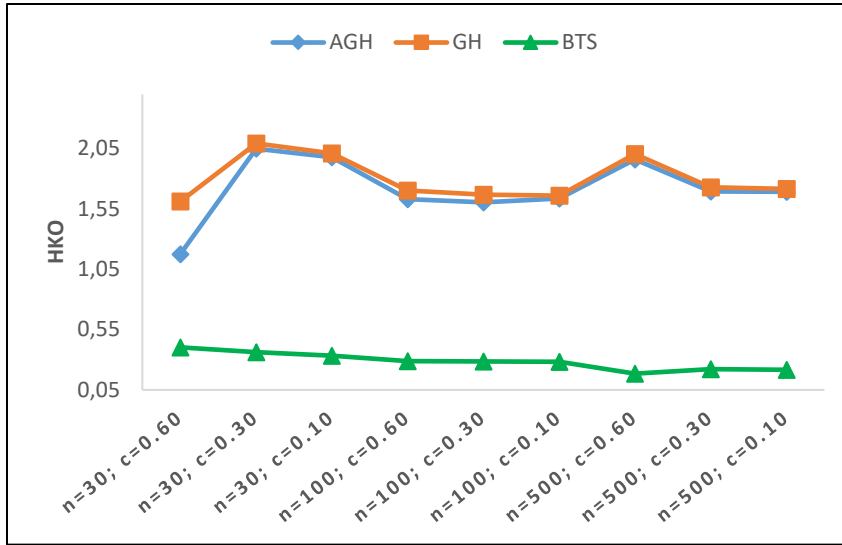
Şekil 6.11’de kırmızı yatay çizgi  $\alpha = 0.05$  değerini göstermek üzere, tüm örneklem genişlikleri ve durdurma oranlarında elde edilen p - değerleri  $\alpha = 0.05$ ’ten küçüktür. Buna göre, üretilen yaşam verilerinin orantılı tehlikeler varsayımını sağlamadığı söylenebilir.

Çizelge 6.5. Senaryo 3'te  $\alpha$  parametresi için ortalama mutlak yan ve HKO değerleri

		AGH		GH		BTS	
		Yan	HKO	Yan	HKO	Yan	HKO
n=30	c=0.60	0.8556	1.1749	0.8304	1.6118	<b><u>0.4157</u></b>	<b><u>0.4028</u></b>
	c=0.30	1.3842	2.0496	1.3963	2.0935	<b><u>0.4715</u></b>	<b><u>0.3636</u></b>
	c=0.10	1.3748	1.9779	1.3849	2.0111	<b><u>0.4881</u></b>	<b><u>0.3330</u></b>
n=100	c=0.60	1.2467	1.6321	1.2708	1.7024	<b><u>0.1598</u></b>	<b><u>0.2891</u></b>
	c=0.30	1.2531	1.6050	1.2718	1.6685	<b><u>0.3573</u></b>	<b><u>0.2848</u></b>
	c=0.10	1.1769	1.6361	1.1939	1.6606	<b><u>0.3664</u></b>	<b><u>0.2842</u></b>
n=500	c=0.60	1.3963	1.9615	1.4122	2.0049	<b><u>0.4130</u></b>	<b><u>0.1849</u></b>
	c=0.30	1.2657	1.6941	1.2817	1.7306	<b><u>0.3803</u></b>	<b><u>0.2235</u></b>
	c=0.10	1.2596	1.6931	1.2683	1.7167	<b><u>0.3528</u></b>	<b><u>0.2178</u></b>



Şekil 6.12. Senaryo 3'te  $\alpha$  parametresi için kestirim değerleri



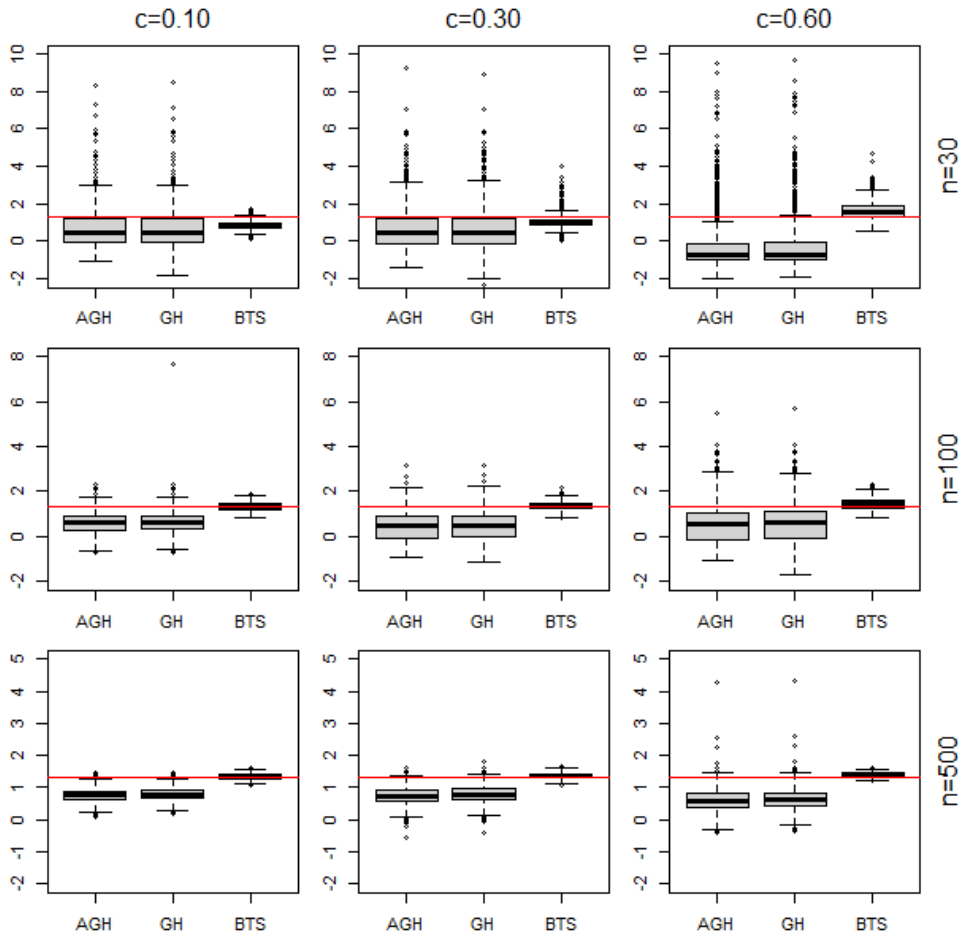
Şekil 6.13. Senaryo 3'te  $\alpha$  parametresi için HKO değerleri

Riskin orta düzey olduğu bu senaryoda  $\alpha$  parametresi için Çizelge 6.5 ve Şekil 6.12 - Şekil 6.13 incelendiğinde, tüm örneklem genişlikleri ve durdurma oranlarında BTS yaklaşımının AGH ve GH yaklaşımlarına kıyasla daha az yanlı tahminler verdiği tespit edilmiştir. Aynı zamanda, BTS yaklaşımının  $n = 30$  ve  $c = 0.60$  durumunda ve AGH ve GH yaklaşımlarının tüm örneklem genişlikleri ve durdurma oranlarında, riskin yüksek olduğu Senaryo 2'ye kıyasla daha az yanlı tahminlere sahip olduğu ve riskin düşük olduğu Senaryo 1'e kıyasla daha yanlı tahminlere sahip olduğu görülmektedir. Tüm senaryolar kendi içlerinde kıyaslandığında ise BTS yaklaşımının  $n = 500$  ve  $c = 0.60$  durumunda daha küçük HKO'ya sahip olduğu ve GH ve AGH yaklaşımlarının ise  $n = 30$  ve  $c = 0.60$  durumunda daha küçük HKO'ya sahip olduğu söylenebilir.

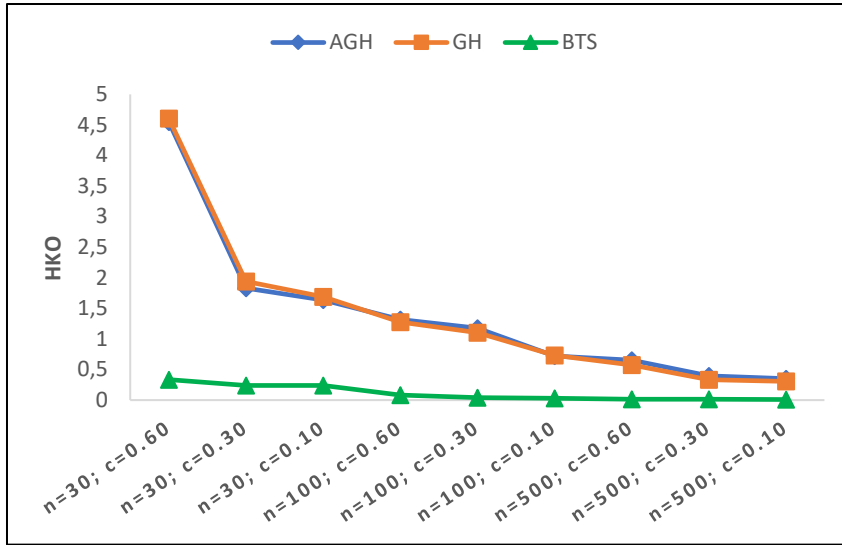
Senaryo 3'te grup parametresi  $\gamma$  için elde edilen parametre tahmin değerleri Çizelge 6.6 ve Şekil 6.14 - Şekil 6.15'te verilmiştir. Şekil 6.14'te kırmızı yatay çizgi  $\gamma = 1.3$  değerini göstermektedir.

Çizelge 6.6. Senaryo 3'te  $\gamma$  parametresi için ortalama mutlak yan ve HKO değerleri

		AGH		GH		BTS	
		Yan	HKO	Yan	HKO	Yan	HKO
n=30	c=0.60	1.5333	4.5505	1.4937	4.6077	<b><u>0.3165</u></b>	<b><u>0.3351</u></b>
	c=0.30	0.6305	1.8322	0.6421	1.9417	<b><u>0.2867</u></b>	<b><u>0.2387</u></b>
	c=0.10	0.6401	1.6384	0.6439	1.6876	<b><u>0.4376</u></b>	<b><u>0.2413</u></b>
n=100	c=0.60	0.7707	1.3196	0.7342	1.2782	<b><u>0.1541</u></b>	<b><u>0.0825</u></b>
	c=0.30	0.8840	1.1743	0.8497	1.1014	<b><u>0.0537</u></b>	<b><u>0.0394</u></b>
	c=0.10	0.6984	0.7214	0.6730	0.7299	<b><u>0.0094</u></b>	<b><u>0.0326</u></b>
n=500	c=0.60	0.7139	0.6519	0.6640	0.5739	<b><u>0.0973</u></b>	<b><u>0.0136</u></b>
	c=0.30	0.5707	0.3983	0.5206	0.3342	<b><u>0.0696</u></b>	<b><u>0.0138</u></b>
	c=0.10	0.5544	0.3508	0.5167	0.3061	<b><u>0.0441</u></b>	<b><u>0.0095</u></b>



Şekil 6.14. Senaryo 3'te  $\gamma$  parametresi için kestirim değerleri



Şekil 6.15. Senaryo 3’te  $\gamma$  parametresi için HKO değerleri

Riskin orta düzey olduğu Senaryo 3’te  $\gamma$  parametresi için Çizelge 6.6 ve Şekil 6.14 - Şekil 6.15 incelendiğinde,  $\alpha$  parametresine benzer şekilde, tüm örneklem genişlikleri ve durdurma oranlarında BTS yaklaşımının AGH ve GH yaklaşımlarına kıyasla daha az yanlı tahminler verdiği görülmektedir. Diğer senaryolara benzer şekilde, tüm yaklaşımlar  $\gamma$  parametresini  $\alpha$  parametresine kıyasla daha az yanlı tahmin etmiştir. Tüm yaklaşımların örneklem genişliği arttıkça ve durdurma oranı azaldıkça daha az yanlı tahminler verdiği görülmektedir. Aynı zamanda, riskin orta düzey olduğu bu senaryoda tüm yaklaşımlarda, Senaryo 2’ye kıyasla daha az yanlı tahminler elde edilmiştir.

## 6.2. Uygulama

Bu bölümde, parametrik bileşik modelde parametre tahminlerini hesaplamak için önerilen yaklaşımın uygulanabilirliğini göstermek amacıyla literatürde yer alan iki veri seti dikkate alınmıştır. GH, AGH ve BTS yaklaşımlarının performansı bu veri setleri üzerinde incelenerek, modeller karşılaştırılmıştır.

### 6.2.1. Uygulama 1

Tez çalışmasında önerilen yaklaşımın performansını incelemek için ilk olarak R programı “joineR” paketinde yer alan aort kapakçığı değiştirme ameliyatı veri seti kullanılmıştır (Philipson ve ark., 2018). Çalışmaya katılan 256 hasta, 1991 ve 2001 yılları arasında en az bir yıl ve tek bir cerrah tarafından takip edilmiş ve ekokardiyografik değerleri yıllık olarak ölçülmüştür. Çalışmanın amacı aort pozisyonlarına yerleştirilen farklı tipteki

kapakçıkların hastaların yaşam sürelerine etkilerini incelemektir. Çalışmada, iki veya daha fazla ameliyat geçiren hastalar, her hastanın sadece bir kere analiz edilmesini sağlamak amacıyla ikinci ameliyat zaman noktasından itibaren durdurulmuş gözlem olarak dikkate alınmıştır.

Ekokardiyografik ölçüm olarak takipteki hastalardan, sol ventrikül hipertrofinin ölçümü olarak kullanılan sol ventrikül kitle indeksi (lvmi - left ventricular mass index) değerleri hesaplanmıştır. Sol ventrikül hipertrofisi, ani kardiyak ölümler, kalp yetmezliği ve kalp krizi gibi hastalıkları değerlendirmek için kullanılan bir risk faktörüdür. Lvmi ise sol ventrikül hipertrofinin kolay hesaplanabilen, güvenilir bir ölçüsü olarak sıklıkla kullanılmaktadır (Çetin ve ark., 2019).

Çalışmada kalp kapakçığı türünün hastaların yaşam süreleri üzerine etkisini araştırmak amacıyla, stentsiz domuz dokusu ve homogreft kapakçığı ameliyatı geçiren hastalardan düzensiz olarak lvmi değerleri ölçülmüş ve hastaların ölene kadar geçen süresi üzerine etkileri incelenmiştir. Lvmi ölçümleri dengesiz boylamsal gözlemlerden oluştuğundan dolayı, bu tez çalışmasında dikkate alınan modele uygun olarak veriyi değerlendirebilmek için %79,07 durdurma oranına sahip bir alt örneklem seçilerek, yeni bir veri seti oluşturulmuştur. Oluşturulan bu yeni veri seti, 43 hastanın 3 farklı zaman noktasında ölçülen (yıl) dengeli lvmi ölçümlerinden oluşmaktadır. Yaşam süresi (fuyrs), hastaların ameliyat tarihinden ölene kadar geçen süreyi ifade ederken, açıklayıcı değişken olarak simülasyon modelinde tek kategorik değişken dikkate alındığından, hastaların geçirmiş oldukları kalp kapakçığı ameliyat türünü gösteren hs değişkeni dikkate alınmıştır.

Çalışmada kullanılan nitel değişkenlere ilişkin betimsel istatistikler Çizelge 6.7’de, sürekli değişkenlere ilişkin betimsel istatistikler Çizelge 6.8’de verilmiştir.

Çizelge 6.7. Aort kapakçığı değiştirme ameliyatı veri setinde nitel değişkenlere ilişkin betimsel istatistikler

Değişken	Düzye	Frekans	Yüzde
Durum	0 (durdurma)	102	79.07
	1 (ölüm)	27	20.93
hs	0 (stentsiz)	69	53.48
	1 (homogreft)	60	46.51

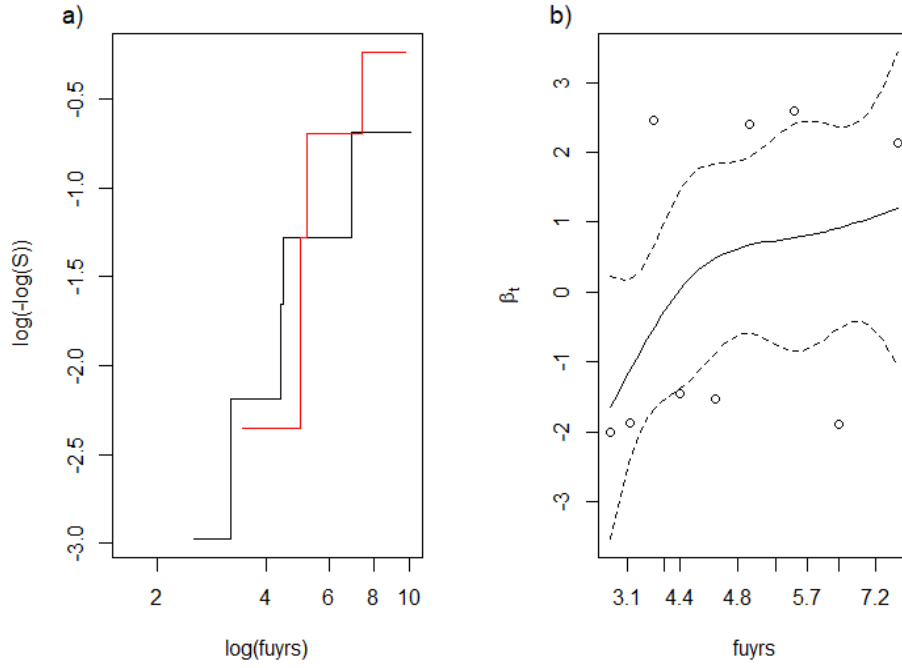
Çizelge 6.7’de verilen betimsel istatistiklere göre, hastaların %20.93’ünün öldüğü gözlemlenirken, %70.07’sinin ya takipten çıktığı ya da takip süresi içerisinde olayı (ölüm) yaşamadığı söylenebilir. Ayrıca ameliyat olan hastaların %53.48’ine stentsiz domuz doku tipi kapak implant edilirken, %46.51’ine homogreft tipi kapak implant edilmiştir.

Çizelge 6.8. Aort kapakçığı değiştirme ameliyatı veri setinde sürekli değişkenlere ilişkin betimsel istatistikler

Değişken	Min	Maks	Ortalama	Ortanca	Std. Sapma
log.lvmi (boylamsal ölçüm)	4.0499	6.1701	5.0044	4.9216	0.3966
Yıl (boylamsal ölçüm zamanı)	0.0082	8.6109	2.0674	1.9945	1.6876
fuyrs (yaşam süresi)	1.6109	10.1123	4.7098	4.0931	2.1748

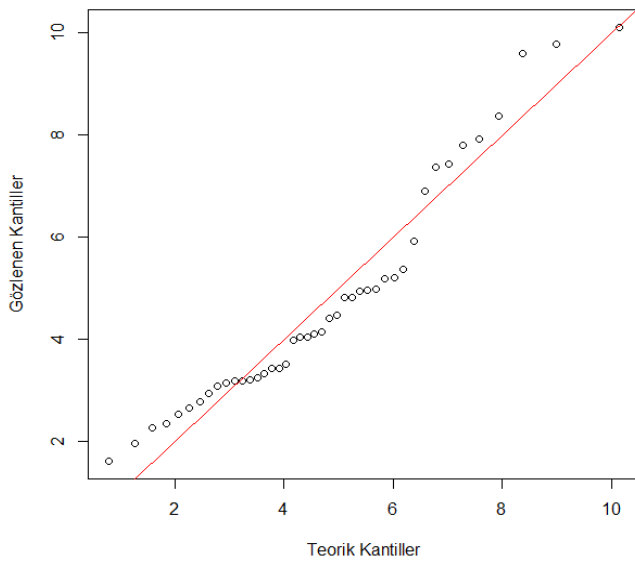
Çizelge 6.8’e göre, çalışmada yer alan hastaların ekokardiyografik ölçüm olarak hesaplanan log.lvmi değeri ortalama  $5.0044 \text{ gm/m}^2 \pm 0.3966$ ’dir. Aort kalp kapakçığı değiştirme ameliyatı sonrası hastaların ortalama takip süresi  $2.0674 \pm 1.6876$  ve ortalama yaşam süresi  $4.7098 \pm 2.1748$ ’dir.

Açıklayıcı değişken hs’nin, boylamsal ölçümler ve ölüm riski üzerine etkisini araştırmak için bileşik model uygulanmadan önce orantılı tehlikeler varsayımı incelenmiştir. Ölçeklendirilmiş Schoenfeld artıkları test sonuçlarına göre, hs değişkeni için  $p = 0.032 < \alpha = 0.05$  olarak elde edilmiştir. Buna göre, orantılı tehlikeler varsayımının sağlanmadığı tespit edilmiştir. Ayrıca, Şekil 6.16a’da verilen  $\log(-\log(S))$  grafiği ve Şekil 6.16b’de verilen dönüştürülmüş zamana karşı ölçeklendirilmiş Schoenfeld artıklarının grafiği ile de varsayım test edilmiş ve orantılı tehlikeler varsayımının hs değişkeni için sağlanmadığı görülmüştür.



Şekil 6.16. hs değişkeni için a)  $\log(-\log(S))$  grafiği b) Dönüştürülmüş zamana karşı ölçeklendirilmiş Schoenfeld artıklarının grafiği

Orantılı tehlikeler varsayımı kontrol edildikten sonra yaşam sürelerinin Weibull dağılıma uygun olup olmadığını test etmek için Kolmogorov-Smirnov testi ve Şekil 6.17’de verilen Q-Q grafiği kullanılmıştır. Kolmogorov-Smirnov testi sonucuna göre,  $D = 0.134$  ve  $p = 0.387 > \alpha = 0.05$  olarak hesaplanmış ve buna göre verilerin Weibull dağılımına uygun olduğu tespit edilmiştir.



Şekil 6.17. Yaşam süresi (fuyrs) için Q-Q grafiği

Açıklayıcı değişken  $hs$ , orantılı tehlikeler varsayımını sağlamadığından ve yaşam süresi Weibull dağılımına uygun olduğundan,  $hs$ 'nin boylamsal ve yaşam verileri üzerindeki etkisi ve boylamsal ölçümler ile yaşam verileri arasındaki ilişkiyi incelemek için DKEM ile Weibull HBS modellerinin birleştirilmesinden oluşan parametrik bileşik model kullanılmıştır. Üç farklı parametre tahmin yaklaşımından elde edilen sonuçlar Çizelge 6.9'da verilmiştir. Yaklaşımların performanslarını incelemek için  $AIC = 2k - 2\ln L$  ve  $BIC = k\ln(n) - 2\ln L$  değerleri kullanılmıştır.

Çizelge 6.9. Aort kapakçığı değiştirme ameliyatı veri seti için Weibull HBS parametrik bileşik modelin parametre tahmin değerleri

		GH	AGH	BTS
		Katsayı (Std. hata)	Katsayı (Std. hata)	Katsayı (Std. hata)
Boylamsal alt model	yıl	0.0158 (0.0096)	0.0125 (0.0163)	<b>0.0220</b> <b>(0.0076)</b>
	hs	-0.0448 (0.0573)	0.0164 (0.0813)	<b>-0.0521</b> <b>(0.0011)</b>
Weibull HBS alt model	hs	-0.9896 (0.9049)	0.8546 (1.6384)	<b>-0.4419</b> <b>(0.1225)</b>
	$\alpha$	0.06514 (0.0096)	-0.0433 (0.0256)	<b>0.1225</b> <b>(0.0074)</b>
AIC		217.3215	405.9841	<b>80.3352</b>
BIC		236.6947	425.3573	<b>85.6189</b>

Çizelge 6.9 incelendiğinde, AIC ve BIC kriterlerine göre HBS yaklaşımının, GH ve AGH yaklaşımlarına kıyasla daha iyi sonuçlar verdiği görülmektedir. BTS yaklaşımından elde edilen Weibull HBS alt modelinde bileştirme katsayısı  $\alpha$ 'ya göre,  $\log.lvmi$  değerleri arttığında yaşam süresinin arttığı ve açıklayıcı değişken  $hs$ 'ye göre stentsiz domuz dokusu türü ameliyat olanların, homogreft türünde ameliyat olanlara kıyasla yaşam süresinin  $\exp(0.4419) = 1.56$  kat daha kısa olduğu söylenebilir. Elde edilen bu sonuçlar simülasyon sonuçları ile kıyaslandığında, küçük örneklerde  $c = 0.60$  durdurma oranında BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla daha az yanlış tahminler verdiği ve özellikle düşük riskte GH yaklaşımının AGH yaklaşımına kıyasla daha az yanlış

tahminler verdiđi görülmüştü. Simülasyon sonuçlarına benzer şekilde aort kapakçığı deđiştirme ameliyatı veri seti içinde BTS yaklaşımının, GH ve AGH yaklaşımlarına kıyasla ve GH yaklaşımının ise AGH yaklaşımına kıyasla daha iyi sonuçlar verdiđi görülmüştür.

### 6.2.2. Uygulama 2

Önerilen yaklaşımın performansını incelemek için ikinci uygulama olarak R programı “JM” paketinde yer alan ve ilk olarak Goldman ve ark. (1996) tarafından tanımlanan AIDS veri seti kullanılmıştır.

Orijinal çalışma, AIDS hastalığını önlemek ve tedavi etmek için sıklıkla kullanılan zidovudin ilaç tedavisine karşı başarısız olan veya tedaviye tolerans göstermeyen 467 HIV enfeksiyonlu hastada, didanozin (ddI) ve zalsitabin (ddC) olmak üzere iki alternatif antiretroviral ilacın etkinliğini ve güvenliđini karşılaştırmak amacıyla gerçekleştirilmiştir. Hastalara rastgele ddI ve ddC ilaçları verilmiş ve çalışma başlangıcında, 2., 6., 12. ve 18. aylarda hastaların bađışıklık sistemlerinin sađlık durumlarını gösteren CD4 hücre sayıları ölçülmüştür. CD4 hücre sayımı ölçümleri, hastaların yalnızca hastaneyi ziyaret ettikleri zamanlarda ölçüldüđünden düzensiz boylamsal gözlemler olarak elde edilmiştir. Çalışmanın sonunda 188 hasta ölmüş ve durdurma oranı %59.7 olarak hesaplanmıştır. Çalışmanın temel amacı, CD4 hücre sayıları hesaplandıktan sonra ilaç tedavilerinin yaşam süresi üzerine etkisini araştırmaktır.

Tez çalışmasında dikkate alınan modele uygun olarak, önerilen yaklaşımın performansını incelemek için orijinal veriden çalışmanın başlangıcında, 2. ve 6. aylarında düzenli olarak ölçülen CD4 ölçüm deđerlerine sahip 57 hastadan oluşan ve %31.58 durdurma oranına sahip yeni bir alt örneklem seçilmiştir. Modelde açıklayıcı deđişken olarak ilaç deđişkeni dikkate alınmıştır.

Çalışmada kullanılan nitel deđişkenlere ilişkin betimsel istatistikler Çizelge 6.10’da, sürekli deđişkenlere ilişkin betimsel istatistikler Çizelge 6.11’de verilmiştir.

Çizelge 6.10. AIDS veri setinde nitel değişkenlere ilişkin betimsel istatistikler

Değişken	Düzy	Frekans	Yüzde
Durum	0 (durdurma)	54	31.58
	1 (ölüm)	117	68.42
İlaç	0 (ddC)	87	50.88
	1 (ddI)	84	49.12

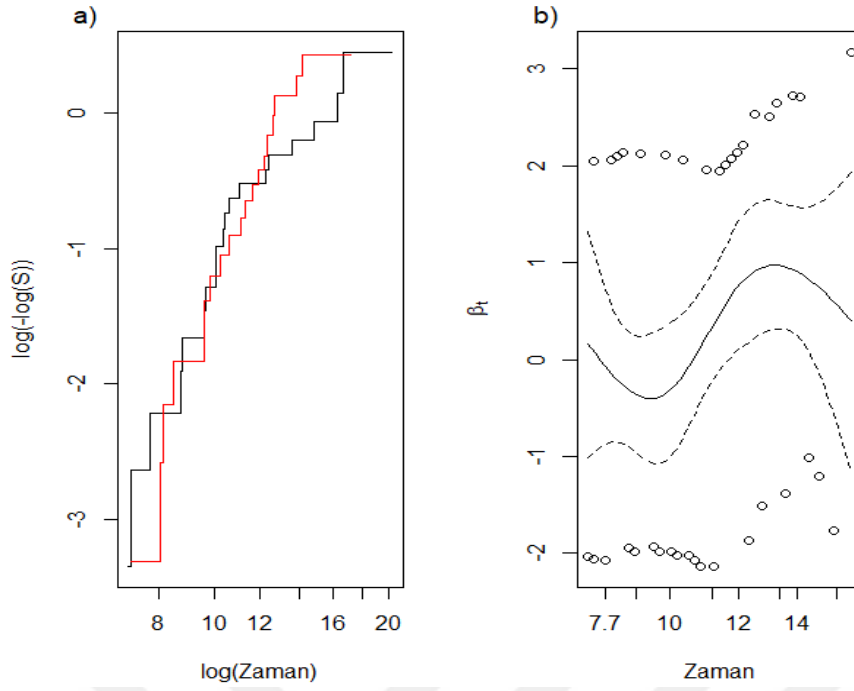
Çizelge 6.10’da verilen betimsel istatistiklere göre, hastaların %68.42’sinin öldüğü gözlemlenirken, %31.58’inin ya takipten çıktığı ya da takip süresi içerisinde olayı (ölüm) yaşamadığı söylenebilir. Ayrıca hastaların %50.88’i ddC ilacını kullanırken, %49.12’si ddI ilacını kullanmıştır.

Çizelge 6.11. AIDS veri setinde sürekli değişkenlere ilişkin betimsel istatistikler

Değişken	Min	Maks	Ortalama	Ortanca	Std. Sapma
CD4	0	17	5.1191	4.2426	3.3807
Ay (boylamsal ölçüm zamanı)	0	6	2.6667	2	2.5018
Yaşam süresi	7.1	20.27	12.1759	12.23	3.2072

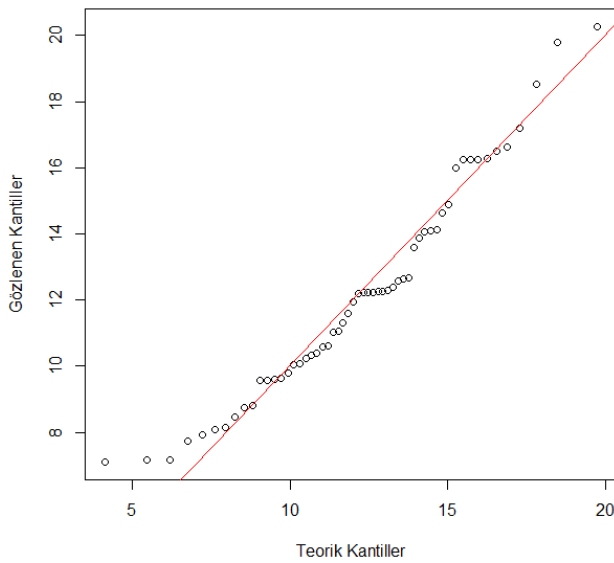
Çizelge 6.11’e göre, çalışmada yer alan hastaların bağışıklık sistemlerinin bir göstergesi olarak ölçülen CD4 hücre ölçümlerinin ortalaması  $5.1191 \pm 3.3807$ ’dir. ddI ve ddC ilaç tedavisi gören hastaların ortalama takip süresi  $2.6667 \pm 2.5018$  ve ortalama yaşam süresi  $12.1759 \pm 3.2072$ ’dir.

İlaç değişkeninin, boylamsal ölçümler ve ölüm riski üzerine etkisini araştırmak için bileşik model uygulanmadan önce orantılı tehlikeler varsayımı incelenmiştir. Ölçeklendirilmiş Schoenfeld artıkları test sonuçlarına göre, ilaç değişkeni için  $p = 0.033 < \alpha = 0.05$  olarak hesaplanmıştır. Buna göre, orantılı tehlikeler varsayımının sağlanmadığı tespit edilmiştir. Ayrıca, Şekilde 6.18a’da verilen  $\log(-\log(S))$  grafiği ve Şekil 6.18b’de verilen dönüştürülmüş zamana karşı ölçeklendirilmiş Schoenfeld artıklarının grafiği ile de varsayım test edilmiş ve orantılı tehlikeler varsayımının ilaç değişkeni için sağlanmadığı görülmüştür.



Şekil 6.18. İlaç değişkeni için a)  $\log(-\log(S))$  grafiği b) Dönüştürülmüş zamana karşı ölçeklendirilmiş Schoenfeld artıklarının grafiği

Orantılı tehlikeler varsayımı kontrol edildikten sonra yaşam sürelerinin Weibull dağılıma uygun olup olmadığını test etmek için Kolmogorov-Smirnov testi ve Şekil 6.19’da verilen Q-Q grafiği kullanılmıştır. Kolmogorov-Smirnov testi sonucuna göre,  $D = 0.130$  ve  $p = 0.299 > \alpha = 0.05$  olarak hesaplanmış ve buna göre, verilerin Weibull dağılımına uygun olduğu tespit edilmiştir.



Şekil 6.19. Yaşam süresi için Q-Q grafiği

Açıklayıcı değişken olarak dikkate alınan ilaç değişkeni orantılı tehlikeler varsayımını sağlamadığından ve yaşam süresi Weibull dağılımına uygun olduğundan, ilaç değişkeninin boylamsal ve yaşam verileri üzerindeki etkisi ve CD4 ölçümleri ile hastaların yaşam süreleri arasındaki ilişkiyi incelemek için DKEM ile Weibull HBS modellerinin birleştirilmesinden oluşan parametrik bileşik model kullanılmıştır. Üç farklı parametre tahmin yaklaşımından elde edilen sonuçlar Çizelge 6.12'de verilmiştir.

Çizelge 6.12. AIDS veri seti için Weibull HBS parametrik bileşik modelin parametre tahmin değerleri

		GH	AGH	BTS
		Katsayı	Katsayı	Katsayı
		(Std. hata)	(Std. hata)	(Std. hata)
Boylamsal alt model	ay	-0.2597 (0.0744)	-0.2326 (0.1287)	<b>-0.2665</b> <b>(0.0029)</b>
	ilaç	0.6501 (0.9426)	1.8329 (0.7052)	<b>0.9435</b> <b>(0.0589)</b>
Weibull HBS alt model	ilaç	-0.1968 (0.1129)	-0.2314 (0.1315)	<b>-0.4090</b> <b>(0.4427)</b>
	$\alpha$	0.0892 (0.0630)	0.0863 (0.0346)	<b>0.0471</b> <b>(0.0008)</b>
AIC		1029.501	1231.623	<b>507.989</b>
BIC		1053.805	1255.928	<b>512.076</b>

Çizelge 6.12 incelendiğinde, AIDS veri seti için AIC ve BIC kriterlerine göre HBS yaklaşımının, GH ve AGH yaklaşımlarına kıyasla daha iyi sonuçlar verdiği görülmüştür. BTS yaklaşımı Weibull HBS alt modelinde bileştirme katsayısına ( $\alpha$ ) göre, CD4 hücre sayısı değerleri arttığında yaşam süresinin arttığı ve ilaç değişkenine göre ddC ilacı kullananların, ddI ilacı kullananlara kıyasla yaşam süresinin  $\exp(0.4090) = 1.50$  kat daha kısa olduğu söylenebilir. AIDS veri seti için elde edilen bulgular simülasyon sonuçları ile kıyaslandığında,  $n = 100$  ve  $c = 0.30$  durumunda BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla daha az yanlı tahminler verdiği ve düşük risk durumunda GH yaklaşımının AGH yaklaşımına kıyasla daha az yanlı tahminler verdiği görülmüştür. Simülasyon sonuçlarına benzer şekilde AIDS veri seti içinde BTS yaklaşımının, GH ve

AGH yaklaşımlarına kıyasla ve GH yaklaşımının ise AGH yaklaşımına kıyasla daha iyi sonuçlar verdiği tespit edilmiştir.



## 7. SONUÇ VE TARTIŞMA

Boylamsal ve yaşam verilerinin ilişkili olduğu durumlarda daha yansız ve güvenilir tahminler elde edebilmek için bu iki verinin bileşik modellemesi üzerine çalışmalar son yıllarda giderek artmaktadır. Bileşik model temel olarak boylamsal alt model ve yaşam süresi alt modelinden oluşmakta ve bu iki alt model arasındaki ilişki, rastgele etkiler modelleri ile sağlanmaktadır.

Literatürde temel olarak kullanılan bileşik model yapısı, yaşam verileri için Cox regresyon modeli ve boylamsal veriler için DKEM'in paylaşılmış rastgele etkiler modeli ile birleştirilmesine dayanmaktadır. Bunun yanı sıra sürekli, kesikli, çoklu veya sıralı yapıya sahip boylamsal veriler ve soldan veya sağdan kesilmiş, orantılı tehlikeler varsayımını sağlamayan veya yarışan riskler durumlarından oluşan yaşam verileri ile farklı bileşik model yapıları da önerilmiştir. Ayrıca iki veri arasındaki ilişki için de paylaşılmış rastgele etkiler modeli, rastgele seçim modeli veya paylaşılmış gizli sınıflar gibi farklı modeller kullanılmaktadır.

Bileşik modelin karmaşık yapıya sahip olmasından dolayı literatürdeki bir diğer ilgi bileşik model parametre tahmin yöntemleri üzerine olmuştur. Bileşik modelde parametre tahminlerini elde etmek için iki - aşamalı yaklaşım, tam olabilirlik yaklaşımı ve Bayesci MCMC yaklaşımları gibi farklı yöntemler geliştirilmiştir. Tam olabilirlik yaklaşımında, olabilirlik fonksiyonundaki integrallerin kapalı çözümü olmadığından ve integraller karmaşık yapıya sahip olduğundan, parametre tahminleri için GH, AGH, Laplace gibi sayısal integrasyon yöntemleri kullanılmaktadır.

Bileşik model literatüründe yaşam verileri için HBS modelleri ve bu modellerden oluşan parametrik bileşik modellerde parametre tahminlerini dikkate alan az sayıda çalışma bulunmaktadır. Bu kapsamda bu tez çalışmasında, tek değişkenli boylamsal veriler için DKEM ve yaşam verilerinin Weibull HBS modelinin gizli sıfır ortalamalı ikili Gauss süreci birleştirilmesinden oluşan parametrik bileşik modelin parametre tahminlerini elde etmek için BTS yaklaşımı önerilmiştir. Önerilen yaklaşım için R programı "rstan" paketi kullanılarak bir kod geliştirilmiştir. Parametrik bileşik modellerde daha yansız parametre tahminleri elde edebilmek için alternatif olarak önerilen yaklaşımın performansı literatürde sıklıkla kullanılan GH ve AGH yaklaşımları ile farklı örneklem

büyükliklerinde, durdurma oranlarında ve regresyon parametrelerinde gerçekleştirilen simülasyon senaryoları yardımıyla kıyaslanmıştır.

Senaryo 1’de birleştirme parametresi  $\alpha$  için tüm örneklem genişliklerinde önerilen BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla daha yansız tahminler verdiği görülmüştür. Ayrıca örneklem genişliği arttıkça ve durdurma oranı küçüldükçe tüm yöntemlerde yan değerlerinin azaldığı tespit edilmiştir. GH ve AGH yaklaşımları kıyasladığında ise küçük örneklerde tüm durdurma oranlarında GH yaklaşımının AGH yaklaşımına kıyasla daha az yanlı tahminler verdiği görülmüştür. Grup parametresi  $\gamma$  içinde benzer sonuçlar elde edildiği yani BTS yaklaşımının diğer yöntemlere kıyasla daha az yanlı tahminler verdiği tespit edilmiştir. Ayrıca, grup ve birleştirme parametrelerinin tahminleri kıyaslandığında, tüm yöntemlerin grup parametresini birleştirme parametresine kıyasla daha az yanlı tahmin ettiği görülmüştür.

Senaryo 2’de birleştirme ve grup parametreleri büyük olduğundan yani riskin yüksek olmasından dolayı yöntemlerin Senaryo 1’e kıyasla iki parametreyi de daha yanlı tahmin ettiği görülmüştür. Ancak Senaryo 2 kendi içerisinde kıyaslandığında, iki parametre için tüm örneklem genişlikleri ve durdurma oranlarında BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla daha az yanlı tahminler verdiği görülmüştür. Ayrıca, BTS yaklaşımının riskin büyük olduğu bu senaryoda grup değişkenini birleştirme katsayısına kıyasla daha iyi tahmin ettiği tespit edilmiştir. Riskin büyük olduğu bu senaryo için AGH yaklaşımının birleştirme katsayısını, tüm örneklem genişlikleri ve durdurma oranlarında GH yaklaşımına kıyasla daha az yanlı tahmin ettiği görülürken, GH yaklaşımının özellikle  $n = 100$  ve  $n = 500$  örneklem genişliklerinde grup parametresini AGH yaklaşımına kıyasla daha az yanlı tahmin ettiği görülmüştür.

Riskin orta düzey olduğu Senaryo 3’te ise BTS yaklaşımının  $n = 30$  ve  $c = 0.60$  durumunda ve AGH ve GH yaklaşımlarının tüm örneklem genişlikleri ve durdurma oranlarında özellikle birleştirme parametresini, riskin yüksek olduğu Senaryo 2’ye kıyasla daha az yanlı, riskin düşük olduğu Senaryo 1’e kıyasla daha yanlı tahmin ettiği görülmüştür. Senaryo 3 kendi içerisinde kıyaslandığında ise tüm örneklem genişlikleri ve durdurma oranlarında BTS yaklaşımının iki parametreyi de GH ve AGH yaklaşımlarına kıyasla daha az yanlı tahmin ettiği görülmüştür. Ek olarak, AGH yaklaşımının birleştirme katsayısını GH yaklaşımına kıyasla daha az yanlı tahmin ettiği tespit edilmiştir.

Önerilen yöntemin uygulanabilirliğini göstermek amacıyla iki farklı veri setinde parametre tahmin yaklaşımlarının performansı incelenmiştir. Birinci uygulamada kullanılan aort kalp kapakçığı değiştirme ameliyatı veri setinde simülasyon çalışmasında kullanılan modele ve ele alınan koşullara uygun olarak kıyaslama yapabilmek için, %79.07 durdurma oranına sahip 43 hastadan oluşan bir alt örneklem seçilmiştir. Oluşturulan bu yeni veri setinde yöntemlerin performansı incelenmiş ve BTS yaklaşımının GH ve AGH yaklaşımlarına kıyasla, GH yaklaşımının ise AGH yaklaşımına kıyasla daha iyi sonuçlar verdiği görülmüştür. Elde edilen bu sonucun düşük risk ile gerçekleştirilen simülasyon çalışmasında  $n = 30$  ve  $c = 0.60$  durumu ile benzerlik gösterdiği tespit edilmiştir. İkinci uygulama olarak dikkate alınan AIDS veri setinde ise, 57 hastadan oluşan ve %31.58 durdurma oranına sahip bir alt örneklem seçilmiş ve oluşturulan bu yeni veri setinde yöntemlerin performansları karşılaştırılmıştır. Buna göre, BTS yaklaşımın GH ve AGH yaklaşımlarına kıyasla, GH yaklaşımının ise AGH yaklaşımına kıyasla daha iyi sonuçlar verdiği tespit edilmiştir. Aynı zamanda bu durumun düşük risk ile gerçekleştirilen simülasyon çalışmasında  $n = 100$  ve  $c = 0.30$  durumu ile de benzerlik gösterdiği görülmüştür.

Farklı simülasyon senaryolarından elde edilen sonuçlara göre, tüm durumlarda BTS yaklaşımının iyi sonuçlar vermesine ek olarak özellikle küçük örneklerde ve riskin yüksek olduğu durumda önerilen yaklaşımın GH ve AGH gibi klasik yaklaşımlara kıyasla daha az yanlış tahminler verdiği söylenebilir. Ayrıca risk arttıkça tüm yaklaşımların daha yanlış tahminler verdiği görülmüştür. İki parametre tahmin değerleri kıyaslandığından ise yöntemlerin grup parametresini birleştirme parametresine kıyasla daha az yanlış tahmin ettiği tespit edilmiştir.

Parametrik bileşik model üzerine yapılan çalışmalar incelendiğinde, özellikle yaşam verileri için HBS modellerini dikkate alan çalışmaların az olduğu görülmektedir. Tseng, Hsieh ve Wang (2005) yaşam süresi alt modeli için HBS model ve boylamsal veriler için DKEM'in paylaşılmış rastgele etkiler model ile birleştirilmesinden oluşan bileşik model için parametre tahminlerini Monte Carlo EM algoritması ile tahmin etmişlerdir. Çalışmalarında, rastgele etkilerin normal dağılıma sahip olduğu veya olmadığı ve %20 durdurma oranına sahip yaşam verileri için üç farklı senaryoda modelin performansını incelemişlerdir. Pericleous (2016) DKEM ile Weibull HBS modelinin paylaşılmış rastgele etkiler ile birleştirilmesine dayanan parametrik bileşik modelin Gauss Hermite

yaklaşımı ile parametre tahminlerini elde etmiş ve farklı örneklem genişliklerinde model performansını incelemiştir. Khan ve Basharat (2022) ise DKEM ile Weibull ve Log-Lojistik HBS modellerinin bileşik model tahminlerini Bayesci olabilirlik yaklaşımı ile tahmin edip, iki modelin performansını karşılaştırmıştır.

Farklı bileşik model yapılarında parametre tahminlerini ele alan çalışmalar incelendiğinde, özellikle çoklu boylamsal gözlemler, yarışan riskler ya da orantılı tehlikeler modellerinden oluşan daha karmaşık yapıya sahip bileşik modeller için iki -aşamalı yaklaşımın farklı örneklem genişliklerinde tam olabilirlik yaklaşımına kıyasla daha iyi sonuçlar verdiğini gösteren çalışmalar bulunmaktadır (Leiva-Yamaguchi ve Alvares, 2020; Mehdizadeh ve ark. 2021).

Özellikle parametrik bileşik model üzerine yapılan çalışmalar incelendiğinde, BTS yaklaşımının HBS modelleri ile DKEM modelinin bileşik model parametre tahminlerini elde etmek için ele alınmadığı görülmüştür. Aynı zamanda parametrik bileşik modellerde GH ve AGH gibi tam olabilirlik yaklaşımlarının, farklı örneklem genişlikleri ve durdurma oranlarında karşılaştıran bir çalışma da bulunmamaktadır. Bu kapsamda bu tez çalışmasında öncelikle Weibull HBS model ile DKEM'in paylaşılmış rastgele etkilerin farklı bir versiyonu olan gizli sıfır ortalamalı ikili Gauss süreci ile birleştirilmesinden oluşan parametrik bileşik model için BTS yaklaşımı önerilmiştir. Önerilen yaklaşımın performansı, farklı senaryolarda oluşturulan simülasyon çalışmalarında GH ve AGH yaklaşımları ile kıyaslanmış ve farklı örneklem genişlikleri ve durdurma oranlarında yöntemlerin performansları detaylıca incelenmiştir.

## 8. KAYNAKLAR

- Alam, K., Maity, A., Sinha, S. K., Rizopoulos, D. and Sattar, A., Joint Modeling of Longitudinal Continuous, Longitudinal Ordinal, and Time-To-Event Outcomes, *Lifetime Data Anal.*, 27(1) (2021) 64-90.
- 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.
- Alkhatami, A., Joint Modeling of Longitudinal and Survival Data, Doctoral Thesis, Probability and Statistics, Carleton University, Ontario, 2021.
- Atlı, Z., Boylamsal ve Sağ Kalım Verilerinin Birleşik Modellenmesi Üzerine Bir Çalışma, Doktora Tezi, Mimar Sinan Güzel Sanatlar Üniversitesi, İstatistik Anabilim Dalı, İstanbul, 2020.
- Basharat, N., Joint Analysis of Longitudinal and Time to Event Data using Accelerated Failure Time Models: A Bayesian Approach, Master thesis, University of Saskatchewan, Collaborative Biostatistics Program of School of Public Health, Saskatoon, 2019.
- Bennett, S., Loglogistic Regression Models for Survival Data, *Journal of the Royal Statistical Society-Series C (Applied Statistics)*, 32 (1983) 165-171.
- Bewick, V., Cheek, L. and Ball, J., *Statistics Review 12: Survival Analysis*, Crit Care, 8 (5) (2004) 389-394.
- Bolstad, W. M. and Curran, J. M., *Introduction to Bayesian Statistics*. John Wiley&Sons, 2016.
- Brown, E. R. and Ibrahim, J. G., A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data, *Biometrics*, 59(2) (2003) 221–228.
- Brown, E. R., Ibrahim, J. G. and DeGruttola V., A Flexible B-spline Model for Multiple Longitudinal Biomarkers and Survival, *Biometrics*, 61(1) (2005) 64-73.
- Chi, Y. Y. and Ibrahim, J. G., Joint Models for Multivariate Longitudinal and Multivariate Survival Data, *Biometrics*, 62(2) (2006) 432-445.
- Collett, D., *Modelling Survival Data in Medical Research*, Chapman and Hall/CRC, 2015.
- Cowles, M. K. and Carlin, B. P., Markov Chain Monte Carlo Convergence Diagnostics: a Comparative Review, *Journal of the American Statistical Association*, 91(434) (1996) 883-904.
- Cox, D.R., Regression Models and Life - Tables, *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2) (1972) 187-202.

- Cox, D. R. and Oakes, D. Analysis of Survival Data. Chapman & Hall, CRC Press London. Volume 21, chapter 5, **1984**.
- Çetin, M. S., Çetin, E. H. Ö., Ocak, K., Özcan Çelebi, Ö., Aras, D., Temizhan, A., ve Aydoğdu, S. The Association Between Left Ventricular Mass Index and Coronary Collateral Circulation in Patients with Chronic Total Occlusion. *Türk Kardiyoloji Derneği Arşivi*, 47(1) (**2019**) 21-28.
- DeGruttola, V. and Tu, X. M., Modelling Progression of CD4-Lymphocyte Count and its Relationship to Survival Time, *Biometrics*, 50 (**1994**) 1003-1014.
- Deslandes, E. and Chevret, S., Joint Modeling of Multivariate Longitudinal Data and The Dropout Process in a Competing Risk Setting: Application to ICU Data, *BMC Med Res Methodol*, 10 (69) (**2010**).
- Diggle, P. J., Sousa, I. and Chetwynd, A. G., Joint Modelling of Repeated Measurements and Time-To-Event Outcomes: The Fourth Armitage Lecture. *Stat Med.*, 27(16) (**2008**) 2981-2998.
- Diggle, P., Heagerty, P., Liang, K. Y., and Zeger, S. L., Analysis of Longitudinal Data. Oxford University Press. Vol. 2, New York, **2002**.
- Doğanay, B., Uzunlmasına Çalışmaların Analizinde Karma Etki Modelleri, Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara, **2007**.
- 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.
- Ekman, A., Variable Selection for The Cox Proportional Hazards Model: A Simulation Study Comparing The Stepwise, Lasso and Bootstrap Approach, Master thesis, Department of Mathematics and Mathematical Statistics, Umeå University, **2017**.
- 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.
- 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.
- Fine, J. and Gray, R., A Proportional Hazards Model for The Subdistribution of a Competing Risk, *JASA*, 94(446) (**1999**) 496-509.
- Fitzmaurice, G. M., Laird, N. M. and Ware, J. H., Applied Longitudinal Analysis, John Wiley and Sons, Vol. 2, New York, **2011**.
- 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**).

- Furgal, A. K. C., Sen, A. and Taylor, J. M. G., Review and Comparison of Computational Approaches for Joint Longitudinal and Time-to-Event Models, *International statistical review = Revue internationale de statistique*, 87(2) (**2019**) 393–418.
- Gelman, A. and Rubin, D. B., Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 7(4) (**1992**) 457-472.
- George, B., Seals, S. and Aban, I., Survival Analysis and Regression Models, *Journal of Nuclear Cardiology*, 21(4) (**2014**) 686-694.
- Geyer, C. J., Practical Markov Chain Monte Carlo, *Statistical Science*, 7(4) (**1992**) 473-483.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J., *Introducing Markov Chain Monte Carlo*, Chapman & Hall, London, **1995**.
- Givens, G. H. and Hoeting, J. A., *Computational Statistics*, Wiley Series in Computation Statistics, **2005**.
- Goldman, A., Carlin, B., Crane, L., Launer, C., Korvick, J., Deyton, L., Abrams, K., Response of CD4+ and clinical consequences to treatment using ddI or ddC in patients with advanced HIV infection, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 11 (**1996**) 161–169.
- Guler, I., Faes, C., Cadarso-Suárez, C., Teixeira, L., Rodrigues, A. and Mendonça, D., Two-Stage Model for Multivariate Longitudinal and Survival Data with Application to Nephrology Research, *Biom. J.*, 59 (**2017**) 1204-1220.
- 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.
- Henderson, R., Diggle, P. and Dobson, A., Joint Modeling of Longitudinal Measurements and Event Time Data, *Biostatistics*, 1 (**2000**) 465-480.
- Hu, B., Li, L. and Greene, T., Joint Multiple Imputation for Longitudinal Outcomes and Clinical Events that Truncate Longitudinal Follow-Up, *Statistics in Medicine*, 35(17) (**2016**) 2991-3006.
- 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., Chen, M. H. and Sinha, D., Bayesian Methods for Joint Modeling of Longitudinal and Survival Data with Applications to Cancer Vaccine Trials, *Statistica Sinica*, 14 (**2004**) 863–83.
- James, M. and Philipson, P., A Fast Approximate EM Algorithm for Joint Models of Survival and Multivariate Longitudinal Data, *Computational Statistics and Data Analysis*, 170 (**2022**) 107438.

- Jennrich, R. I. and Schluchter, M. D., Unbalanced Repeated-Measures Models with Structured Covariance Matrices, *Biometrics*, 42(4) (**1986**) 805-820.
- Kalbfleisch, J. D. and Prentice, R. L., *The statistical analysis of failure time data*, Vol. 1, **2002**.
- Karasoy, D. ve Tutkun, N. A., *Yaşam Çözümlemesi*, 1. Baskı, Nobel Akademik Yayıncılık, **2016**.
- Khan, S. A. and Basharat, N., Accelerated Failure Time Models for Recurrent Event Data Analysis and Joint Modeling, *Comput Stat*, 37 (**2022**) 1569–1597.
- Klein, J. P. and Moeschberger, M. L., *Survival Analysis Techniques for Censored and Truncated Data*, Springer-Verlag, Vol. 2, New York, **2003**.
- 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.
- Lawless, J. F., *Statistical models and methods for lifetime data* (Vol. 362). John Wiley & Sons, **2011**.
- Lee, E. T. and Wang, J., *Statistical Methods for Survival Data Analysis*. John Wiley & Sons. **2003**.
- Leiva-Yamaguchi, V. and Alvares, D., A Two-Stage Approach for Bayesian Joint Models of Longitudinal and Survival Data: Correcting Bias with Informative Prior, *Entropy*, 23(1) (**2020**) 50.
- 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. Y. and Zeger, S. L., Longitudinal Data Analysis using Generalized Linear Models, *Biometrika*, 73(1) (**1986**) 13-22.
- Lin, J. and Luo, S., Deep Learning for The Dynamic Prediction of Multivariate Longitudinal and Survival Data, *Statistics in medicine*, 41(15) (**2022**) 2894–2907.
- Lunn, M. and McNeil D., Applying Cox Regression to Competing Risks, *Biometrics*, 51(2) (**1995**) 524-32.
- Mao, M. J., *The Effects of Misspecification of Submodels in Joint Modeling of Repeated Measures and Time-to-Event Outcomes*, Master thesis, University of California, Los Angeles, **2016**.

- Mauff, K., Steyerberg, E., Kardys, I., Boersma, E. and Rizopoulos, D., Joint Models with Multiple Longitudinal Outcomes and a Time-to-Event Outcome: A Corrected Two-Stage Approach, *Stat. Comput.*, 30 (2020) 999–1014.
- 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.
- Mehdizadeh, P., BAGHfalaki, T., Esmailian, M. and Ganjali, M. A two-stage approach for joint modeling of longitudinal measurements and competing risks data, *Journal of biopharmaceutical statistics*, 31 (2) (2021) 1-21.
- Meyn, S. P. and Tweedie, R. L., *Markov Chains and Stochastic Stability*, Springer Science & Business Media, 2012.
- Molenberghs, G. and Verbeke, G., *Linear Mixed Models for Longitudinal Data*, Springer, 2000.
- Mondal, P. K., *Joint Modeling of Longitudinal Measurements and Survival Data with Competing Risks: Application to Hiv/Aids Study*, Doctora Thesis, University of Saskatchewan, Saskatoon, 2017.
- Murawska, M., Rizopoulos, D. and Lesaffre, E., A Two-Stage Joint Model for Nonlinear Longitudinal Response and a Time-to-Event with Application in Transplantation Studies, *Journal of Probability and Statistics*, (2012) 1-18.
- Naylor, J. C. and Smith, A. F. M., Applications of a Method for the Efficient Computation of Posterior Distributions, *Applied Statistics*, 31(3) (1982) 214-225.
- Öztürk, Z. and Cengiz, M. A., Bayesci Genelleştirilmiş Lineer Karma Modellerde Önsel Seçimleri ve Karşılaştırılması. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 21(3) (2017) 999-1010.
- Pericleous, P., *Parametric Joint Modelling for Longitudinal and Survival Data*, Doctoral Thesis, School of Computing Sciences University of East Anglia, UK, 2016.
- Philipson, P., Sousa, I., Diggle, P., Williamson, P., Kolamunnage-Dona, R., Henderson, R., and Hickey, G. *joineR: Joint Modelling of Repeated Measurements and Time-to-Event Data*. R Package Version 1.2.8, (2018).
- 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.
- Powney, A. M., *Joint Modelling of Longitudinal and Time-to-Event Data*, Doctoral Thesis, University of Liverpool, 2015.
- Press, W., Teukolsky, S., Vetterling, W. and Flannery, B., *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge University Press, New York, 2007.

- Proust-Lima, C., Sène M, Taylor, J. M., and Jacqmin-Gadda, H., Joint Latent Class Models for Longitudinal and Time-To-Event Data: A Review, *Stat Methods Med Res*, 23(1) (2014) 74-90.
- Putter, H., Fiocco, M. and Geskus, R. B., Tutorial in Biostatistics: Competing Risks and Multi-State Models, *Stat Med*, 26(11) (2007) 2389-430.
- Rizopoulos, D. and Ghosh, P., A Bayesian Semiparametric Multivariate Joint Model for Multiple Longitudinal Outcomes and a Time-To-Event, *Statistics in medicine*, 30(12) (2011) 1366–1380.
- 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.
- 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., Verbeke, G. and Molenberghs, G., Shared Parameter Models under Random Effects Misspecification, *Biometrika* 95 (2008) 63-74.
- Schoenfeld, D., Partial Residuals for The Proportional Hazards Regression Model, *Biometrika*, 69(1) (1982) 239–241.
- 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, *Perspect Clin Res.*, 2(4) (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.
- Sousa, I., A Review on Joint Modelling of Longitudinal Measurements and Time-to-Event, *REVSTAT-Statistical Journal*, 9 (2011) 57-81.
- 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.
- 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.

- Teixeira, L., Sousa, I., Rodrigues, A. and Mendonça, D., Joint Modelling of Longitudinal and Competing Risks Data in Clinical Research, *REVSTAT-Statistical Journal*, 17(2) (2019) 245–264.
- Thomadakis, C., Meligkotsidou, L., Yiannoutsos, C. T. and Touloumi, G., Joint Modeling of Longitudinal and Competing-Risk Data using Cumulative Incidence Functions for The Failure Submodels Accounting for Potential Failure Cause Misclassification Through Double Sampling, *Biostatistics*, 25(1) (2023) 80-97.
- 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. and Davidian, M. Joint Modeling of Longitudinal and Time-to-Event data: An Overview, *Statistica Sinica*, 14 (2004) 809-834.
- Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J. and Van Aken, M. A., A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child development*, 85(3) (2014) 842-860.
- 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., Bayesian Joint Modeling of Longitudinal and Time-to-Event Data with Application to Alzheimer's Disease, Doctoral Thesis, Public Health Sciences, University of Illinois at Chicago, Chicago, 2021.
- Wang, Y. and Taylor, J. M. G., Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome, *Journal of the American Statistical Association*, 455(96) (2001) 895-905.
- Wei, L. J., The Accelerated Failure Time Model: A Useful Alternative to The Cox Regression Model in Survival Analysis, *Statistics in medicine*, 11(14-15) (1992) 1871-1879.
- Wienke, A., *Frailty Models in Survival Analysis*, Chapman and Hall/CRC, 2010.
- 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., *Mixed Effects Models for Complex Data*, Chapman and Hall/CRC, 2009.

- Wu, L., Liu, W. and Hu, X. J., Joint Inference on HIV Viral Dynamics and Immune Suppression in Presence of Measurement Errors, *Biometrics*, 66(2) (2010) 327-335.
- 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) 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.
- Xu, J. and Zeger, S. L., The Evaluation of Multiple Surrogate Endpoints, *Biometrics*, 57(1) (2001) 81-87.
- Ye, W., Lin, X. and Taylor, J. M. G., Semiparametric Modeling of Longitudinal Measurements and Time-to-Event Data Two-Stage Regression Calibration Approach, *Biometrics*, 64(4) (2008) 1238-1246.
- Yildirim, E. and Karasoy, D., gsem: A Stata Command for Parametric Joint Modelling of Longitudinal and Accelerated Failure Time Models, *Computer methods and programs in biomedicine*, 196 (2020) 105612.
- Yildirim, E. and Karasoy, D., A Two-Stage Approach for Bayesian Parametric Joint Modelling of Longitudinal and Weibull Accelerated Failure Time Data, *Filomat*, 38(23) (2024) 8265–8281.
- Yu, M., Law, J. N., Taylor, M. G. J. and Sandler, M. H., 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.
- Zhang, H. and Wu, L., Joint Model of Accelerated Failure Time and Mechanistic Nonlinear Model for Censored Covariates, With Application in HIV/AIDS. *Ann Appl Stat*, 13(4) (2019) 2140–2157.
- Zhao, X., Yu, C., and Tong, H., A Bayesian Approach to Weibull Survival Model for Clinical Randomized Censoring Trial based on MCMC Simulation, 2nd International Conference on Bioinformatics and Biomedical Engineering, Shanghai, 16-18 May, 2008, 1181-1184.

## **EKLER**

### **EK 1 - Tezden Türetilmiş Yayınlar**

Yildirim, E. and Karasoy, D., A Two-Stage Approach for Bayesian Parametric Joint Modelling of Longitudinal and Weibull Accelerated Failure Time Data, *Filomat*, 38(23) (2024) 8265–8281.



## **EK 2 - Tezden Türetilmiş Bildiriler**

Yıldırım, E. ve Karasoy, D., Boylamsal ve Yaşam Verilerinin Parametrik Bileşik Modellemesinde Parametre Tahmin Yaklaşımlarının Karşılaştırılması, 23. Ulusal ve 6. Uluslararası Biyoistatistik Kongresi, Ankara, 2022.

Yıldırım, E. ve Karasoy, D., Boylamsal ve Weibull Hızlandırılmış Başarısızlık Süresi Verilerinin Parametrik Bileşik Model Tahminlerinde Bayesci İki Aşamalı Yaklaşım, IRSYSC, Adana, 2024.

Yıldırım, E. ve Karasoy, D., Boylamsal ve Weibull Hızlandırılmış Başarısızlık Süresi Verilerinin Parametrik Bileşik Model Tahminlerinde Düzeltilmiş Bayesci İki - Aşamalı Yaklaşım, 25. Ulusal ve 8. Uluslararası Biyoistatistik Kongresi, Sakarya, 2024.