



T.C.  
ULUDAĞ ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ  
BİYOİSTATİSTİK  
ANABİLİM DALI



**TESPİT EDİLEMİYEN VERİLERİN ANALİZİNDE KULLANILAN  
YÖNTEMLERİN KARŞILAŞTIRILMASI**

**GÜLSER ÇALIŞKAN**

**(YÜKSEK LİSANS TEZİ)**

**BURSA-2017**





T.C.  
ULUDAĞ ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ  
BİYOİSTATİSTİK ANABİLİM DALI



**TESPİT EDİLEMİYEN VERİLERİN ANALİZİNDE KULLANILAN  
YÖNTEMLERİN KARŞILAŞTIRILMASI**

**Gülser ÇALIŞKAN**

**(YÜKSEK LİSANS TEZİ)**

**DANIŞMAN:**

**Doç.Dr. Güven ÖZKAYA**

**Proje No- UÜ BAP KUAP(T)-2015/46**

**BURSA-2017**

**T.C.**  
**ULUDAĞ ÜNİVERSİTESİ**  
**SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**ETİK BEYANI**

Yüksek Lisans tezi olarak sunduğum “TESPİT EDİLEMİYEN VERİLERİN ANALİZİNDE KULLANILAN YÖNTEMLERİN KARŞILAŞTIRILMASI” adlı çalışmanın, proje safhasından sonuçlanmasına kadar geçen bütün süreçlerde bilimsel etik kurallarına uygun bir şekilde hazırlandığımı ve yararlandığım eserlerin kaynaklar bölümünde gösterilenlerden oluştuğunu belirtir ve beyan ederim.

**Gülser ÇALIŞKAN**  
**13/02/2017**

## SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ ' NE

Anabilim Dalı Yüksek Lisans öğrencisi Gülser ÇALIŞKAN tarafından hazırlanan  
“TESPİT EDİLEMİYEN VERİLERİN ANALİZİNDE KULLANILAN  
YÖNTEMLERİN KARŞILAŞTIRILMASI” konulu Yüksek Lisans tezi  
13/02/2017 Pazartesi günü, 10:00-12:00 saatleri arasında yapılan tez savunma  
sınavında jüri tarafından oy birliği/oy çokluğu ile kabul edilmiştir.

**Adı-Soyadı** **İmza**

**Tez Danışmanı**

**Doç.Dr.Güven ÖZKAYA**

**Üye**

**Prof.Dr.İlker ERCAN**

**Üye**

**Yrd.Doç.Dr.Ömer UYSAL**

Bu tez Enstitü Yönetim Kurulu'nun .....  
tarih ve .....  
..... sayılı toplantısında alınan ..... numaralı  
kararı ile kabul edilmiştir.

Prof Dr. Gülşah ÇEÇENER  
Enstitü Müdürü

## TEZ KONTROL ve BEYAN FORMU

13/02/2017

**Adı Soyadı:** Gülser ÇALIŞKAN

**Anabilim Dalı:** Tıp-Biyostatistik

**Tez Konusu:** Tespit Edilemeyen Verilerin Analizinde Kullanılan Yöntemlerin Karşılaştırılması

<u>ÖZELLİKLER</u>	<u>UYGUNDUR</u>	<u>UYGUN DEĞİLDİR</u>	<u>AÇIKLAMA</u>
Tezin Boyutları	<input type="checkbox"/>	<input type="checkbox"/>	
Dış Kapak Sayfası	<input type="checkbox"/>	<input type="checkbox"/>	
İç Kapak Sayfası	<input type="checkbox"/>	<input type="checkbox"/>	
Kabul Onay Sayfası	<input type="checkbox"/>	<input type="checkbox"/>	
Sayfa Düzeni	<input type="checkbox"/>	<input type="checkbox"/>	
İçindekiler Sayfası	<input type="checkbox"/>	<input type="checkbox"/>	
Yazı Karakteri	<input type="checkbox"/>	<input type="checkbox"/>	
Satır Aralıkları	<input type="checkbox"/>	<input type="checkbox"/>	
Başlıklar	<input type="checkbox"/>	<input type="checkbox"/>	
Sayfa Numaraları	<input type="checkbox"/>	<input type="checkbox"/>	
Eklerin Yerleştirilmesi	<input type="checkbox"/>	<input type="checkbox"/>	
Tabloların Yerleştirilmesi	<input type="checkbox"/>	<input type="checkbox"/>	
Kaynaklar	<input type="checkbox"/>	<input type="checkbox"/>	

**DANIŞMAN ONAYI**

**Doç.Dr.Güven ÖZKAYA**

**İmza:**

## İÇİNDEKİLER

ETİK BEYANI.....	II
KABUL ONAY .....	III
TEZ KONTROL ve BEYAN FORMU.....	IV
İÇİNDEKİLER .....	V
TABLolar DİZİNİ .....	VI
TÜRKÇE ÖZET .....	VII
İNGİLİZCE ÖZET .....	VIII
1.GİRİŞ .....	1
2.GENEL BİLGİLER .....	2
2.1. Tespit Edilemeyen Veriler.....	2
2.2. Tespit Edilemeyen Veri Analiz Yöntemleri .....	2
2.2.1. Yerine Değer Atama Yöntemi .....	2
2.2.2. Sıralı İstatistiklerin Regresyonu (Regression on order statistics-ROS):.....	3
2.2.3. Tespit Edilemeyen Veri Regresyon Analizi Yöntemi .....	6
2.2.4. Çoklu Değer Atama Yöntemi .....	8
2.2.5. Tobit Regresyon Analizi Yöntemi.....	10
3.GEREÇ VE YÖNTEM .....	18
4.BULGULAR .....	21
5.TARTIŞMA .....	52
6.KAYNAKLAR .....	54
7.TEŞEKKÜR.....	56
8. ÖZGEÇMİŞ .....	57

## TABLULAR DİZİNİ

Tablo 1: Simülasyon Tablosu .....	20
Tablo 2: $R^2=0,90$ ve $n=100$ için RMSE Değerleri .....	22
Tablo 3: $R^2=0,90$ ve $n=250$ için RMSE Değerleri .....	24
Tablo 4: $R^2=0,90$ ve $n=500$ için RMSE Değerleri .....	26
Tablo 5: $R^2=0,90$ ve $n=1000$ için RMSE Değerleri .....	28
Tablo 6: $R^2=0,90$ ve $n=5000$ için RMSE Değerleri .....	30
Tablo 7: $R^2=0,70$ ve $n=100$ için RMSE Değerleri .....	32
Tablo 8: $R^2=0,70$ ve $n=250$ için RMSE Değerleri .....	34
Tablo 9: $R^2=0,70$ ve $n=500$ için RMSE Değerleri .....	36
Tablo 10: $R^2=0,70$ ve $n=1000$ için RMSE Değerleri .....	38
Tablo 11: $R^2=0,70$ ve $n=5000$ için RMSE Değerleri .....	40
Tablo 12: $R^2=0,50$ ve $n=100$ için RMSE Değerleri .....	42
Tablo 13: $R^2=0,50$ ve $n=250$ için RMSE Değerleri .....	44
Tablo 14: $R^2=0,50$ ve $n=500$ için RMSE Değerleri .....	46
Tablo 15: $R^2=0,50$ ve $n=1000$ için RMSE Değerleri .....	48
Tablo 16: $R^2=0,50$ ve $n=5000$ için RMSE Değerleri .....	50

## TÜRKÇE ÖZET

Çalışmalarda kullanılan alet, cihaz ya da araçlar bazı durumlarda belirli bir eşik değerin altındaki sonuçları ölçemez. Bu tür verilere tespit edilemeyen veriler denir. Yapılan çalışmalarda istatistiksel analizler için veri setlerinin eksiksiz olması büyük önem taşımaktadır. Bu nedenle tespit edilemeyen verilerin tahmininde kullanılan bazı yöntemlerden bazıları yerine değer atama, sıralı istatistiklerin regresyonu, çoklu değer atama, Tobit regresyon ve tespit edilemeyen veri regresyonudur. Bu yöntemlerin karşılaştırılmasında farklı tespit edilemeyen veri oranı ve örneklem büyüklüklerine göre simülasyon senaryoları oluşturulmuştur. Simülasyon sonucunda farklı  $R^2$  değerlerine sahip her bir yöntem için örneklem büyüklüğü değiştikçe yöntemlerin hata kareler ortalaması köklerinin değişmediği görülmüştür. Karşılaştırdığımız yöntemler farklı  $R^2$  değerlerinde incelendiğinde;  $R^2$ 'nin 0,90 ve 0,70 olduğu durumlarda çoklu değer atama, alt sınır atanması ve Tobit regresyon yöntemlerinde hata kareler ortalaması köklerinin sırasıyla diğer yöntemlere göre daha düşük olduğu görülmüştür. Düşük açıklayıcılık katsayısı ve tespit edilemeyen veri oranı için tespit edilemeyen veri regresyonu alternatif bir yöntemdir.

**Anahtar kelimeler:** Tespit edilemeyen veri, Sıralı istatistiklerin regresyonu, Tobit regresyon, Tespit edilemeyen veri regresyonu



## İNGİLİZCE ÖZET

### COMPARISON OF METHODS USED IN NONDETECTS DATA ANALYSIS

Some machine, devices or tools used in some studies can not measure values below a certain threshold. Such observations are called non-detects data. It is of paramount importance that the data sets are complete for statistical analysis in the studies carried out. For this reason, some methods used for predicting non-detected data are; substitution of values, regression of on order statistics (ROS), multiple imputation, Tobit regression and non-detects data regression. For comparison of these methods, simulation scenarios were established according to the different percentage of non-detects data, sample sizes and coefficient of determination. According to simulation results, the root mean square error of methods did not change as the sample sizes changed for different levels of coefficient of determination. The methods Tobit regression, multiple imputation and substituted with detection limit have lower root mean square error than others while coefficient of determination were 0,90 and 0,70. For coefficient of determination and nondetects data proportion, nondetect data regression is an alternative method.

**Keywords:** Non-detects data, Regression on order statistics, Tobit regression, Non-detects data regression

## 1.GİRİŞ

Arařtırmalar her ne kadar iyi planlanmıř olsa da toplanan verilerde eksiklikler bulunabilir. Yapılan alıřmalarda istatistiksel analizler iin zerinde alıřılacak olan veri setlerinde tespit edilemeyen gzlem deęerlerinin olmaması yani verinin eksiksiz olması byk nem tařımaktadır.

Arařtırmalarda sre boyunca tam veri setleri elde etmeye alıřılmasına raęmen bazı alıřmalarda kullanılan alet, cihaz ya da aralar belirli bir eřik deęerin altındaki sonuları lemez. llemeyen bu gzlemler tespit edilemeyen veri olarak ifade edilmektedir. Tespit edilemeyen veriler, yapılması planlanan istatistiksel analizlerde eřitli problemlerin ortaya ıkmasına neden olmaktadır. Bu nedenle arařtırmalarda, tespit edilemeyen verileri ieren birimler bilgi yokluęunu temsil ederler ve dolayısıyla bilgi kaybına neden olurlar (Dodge, 1985; Bal ve zdamar, 2004). Tespit edilemeyen deęerler bir veri seti iin istatistiksel analizlerde yanlılık yaratmaktadır. Tespit edilemeyen verilerin stesinden gelmek nemli bir konudur ve son yıllarda bu konu iin eřitli istatistiksel yntemler geliřtirilmiřtir.

Tezin amacı; tespit edilemeyen verilerin tahmininde kullanılan yaklařımlarından tespit edilemeyen verilerin regresyonu, Tobit regresyon, oklu deęer atama yntemi gibi yntemleri inceleyerek, tretilen veri setinde oluřturulan farklı oranlarda tespit edilemeyen gzlemler zerinde bu yntemleri uygulamak ve elde edilen sonular zerinden en uygun yntemlere iliřkin bilgi edinmektir. alıřmamızın sonucunda tespit edilemeyen verilerin analizleri sonucundaki tahminlere gre hata kareler ortalamasının kare kk deęeri bakımından tespit edilemeyen veri yapılarında hangi yntemlerin daha iyi sonular verdięi, bu yntemlerin hangi durumlarda kullanılması gerektięi incelenecektir.

## 2.GENEL BİLGİLER

### 2.1. Tespit Edilemeyen Veriler

Bazı çalışmalarda kullanılan alet, cihaz ya da araçlar belirli bir eşik değerin altındaki sonuçları ölçemez. Bu tür verilere tespit edilemeyen veriler denir. Tespit edilemeyen veriler ayrıca soldan sansürlü veri olarak da ifade edilirler. Çünkü bu değerler bilinen eşik değerinin altındadırlar. Bazı durumlarda ise kullanılan cihaz belirli bir eşik değerden yukarısını ölçemeyebilir. Bu tür tespit edilemeyen veriler ise sağdan sansürlü olarak ifade edilirler. Ancak araştırmalarda sıklıkla belirli bir eşik değerin altında kalan tespit edilemeyen verilerle karşılaşmaktadır.

Bir örneklem üzerinden hesaplanan istatistikler ana kütle karakteristiklerini tahmin eder. Sansürlü veri olması durumunda hesaplamaların nasıl yapılacağı kolay anlaşılır olmayacaktır. Tespit edilemeyen gözlemlerin varlığı, tanımlayıcı istatistiklerin hesaplanmasında zorluk yaratacaktır. Tespit edilemeyen verileri incelenmesi hesaplanması istenen herhangi bir istatistiğin etkinliğini değiştirmektedir.

### 2.2. Tespit Edilemeyen Veri Analiz Yöntemleri

Temel olarak sansürlü gözlemler içeren veri setinin analizi ile ilgili bazı yöntemler vardır. Bunlar; yerine koyma, en çok olabirlik tahmini, parametrik olmayan yöntemler, sıralı istatistiklerin regresyonu (ROS, regression on order statistics), çoklu değer atama, tobit regresyon ve tespit edilemeyen veri regresyonudur (Hae ve ark., 2008; Helsel, 2005). Uygulanacak yöntem konusunda kesin ve genel bilgi vermek zordur. Çünkü seçilecek yöntemin uygunluğu örneklem büyüklüğü ve sansürleme yüzdesi dışında veri hakkında yapılan varsayımların geçerliliğine de bağlıdır.

#### 2.2.1. Yerine Değer Atama Yöntemi

Tespit edilemeyen veya kayıp gözlem değerini belirleyebilmek için oldukça sık kullanılan bir yöntemdir. Bu yöntemle tespit edilemeyen verilerin yerine tek bir değer atanır. Tespit edilemeyen verinin yerine genel olarak sıfır değeri veya eşik değerin

diğer bir ifadeyle verinin alt sınırının yarısı veya doğrudan alt sınır değeri atanmaktadır. Araştırmacılar sıklıkla daha kötü bir senaryoya karşı korunmak için alt sınırı yerine koyarak tahminlerin tutucu olması girişiminde bulunurlar. Yerine koyma yönteminin kolay bir yöntem olmasına rağmen, gerçek varyans hesaplanırken yanlışlık oluşturması, verilerin gerçek dağılımlarını bozabilmesi ve tüm tespit edilemeyen veya kayıp değerlerin aynı değeri almasından dolayı değişkenler arasındaki ilişki katsayısının etkilenmesi gibi olumsuz etkileri bulunmaktadır (Helsel, 2005).

### 2.2.2. Sıralı İstatistiklerin Regresyonu (Regression on order statistics-ROS):

Parametrik ve parametrik olmayan yöntemlere alternatif bir yaklaşım da sıra istatistikleri üzerinde robust regresyon yaklaşımı olup yarı-parametrik yöntem olarak adlandırılır. ROS yöntemleri parametrik ve nonparametrik yöntemlerin her ikisinin de fikirlerinin birleşimine dayanmaktadır (Huston ve Juarez, 2009).

ROS bir olasılık grafiğinde regresyon denklemi ile özet istatistiklerden hesaplanmaktadır. Tespit edilemeyen veriler, elde edilen veriler kullanılarak oluşturulan regresyon denkleminde yararlanarak tahmin edilmektedir. Özet istatistikler tahmin edilmiş değerler ve sansürlü gözlem değerlerinden faydalanılarak hesaplanabilmektedir.

ROS yaklaşımının altında yatan ana düşünce; eğer veri seti lognormal dağılıma ya da bilinen bir dağılıma uygunluk gösteriyorsa, sıralı verinin logaritmasının olasılık grafiğine karşı standartlaştırılmış normal çeyreklikler düz bir çizgi vermelidirler. Bu sayede verinin logaritması için ortalama ve standart sapma elde edilebilir. Grafiğin eğimi kullanılarak ortalama ve standart sapma tahmin edilebilmektedir. Daha sonra, alt sınır altındaki tespit edilemeyen değerler tahmin edilmiş olan parametrelerden faydalanarak belirlenebilir.

Çoklu belirleme limitleri olan soldan sansürlü verileri hesaplamak için eşitlik (2.1)'deki formül verilmiştir.

$pe_i$   $i$ 'inci belirleme limitini aşmanın olasılığı hesaplamak için;

$$pe_i = pe_{i+1} + \frac{A_i}{A_i+B_i} [1 - pe_{i+1}] \quad (2.1)$$

$A_i = i$  ve  $i + 1$  inci belirleme limitleri arasında belirlenen gözlemlerin sayısı

$n$   $i$  'inci belirleme limitinin altında sansürlü ve sansürsüz gözlemlerin sayısıdır.

$i$  en yüksek belirleme limiti olduğunda,  $pe_{i+1} = 0$  ve  $A_i + B_i = n$  dir.

$i$  belirleme limitinin altındaysa  $C_i$  olarak tanımlanır ve eşitlik(2.2) deki gibi hesaplanmaktadır.

$$C_i = B_i - B_{i-1} - A_{i-1} \quad i = 1, \dots, k \quad (2.2)$$

Grafğin konumunu hesaplariken gözlemlenmiş değerler için eşitlik (2.3) deki gibidir.

$$pd_{ij} = (1 - pe_i) + \left(\frac{i}{A_{i+1}}\right) [pe_i - pe_{i+1}] \quad j=1, \dots, A_i \quad i=0, \dots, k \quad (2.3)$$

Sansürlenmiş değerlerin grafik konumları eşitlik (2.4)'teki gibi hesaplanmaktadır.

$$pc_{ij} = \left(\frac{i}{C_{i+1}}\right) (1 - pe_i) \quad j=1, \dots, C_i \quad i=0, \dots, k \quad (2.4)$$

ROS yöntemi büyük örneklem büyüklüklerinde olduğu kadar küçük örneklem büyüklüklerinde de iyidir. Ancak diğer yöntemler gibi ROS tahminleri de log transformasyonundan dolayı yanlıdır. Maalesef yapılan simülasyon çalışmaları bu yöntemi kullanmanın küçük bir yanlılığa yol açtığını göstermiştir (Allison, 2001).

ROS yöntemi aşağıdaki adımları izleyerek uygulanabilir (EPA, 2009):

Adım 1: Toplam  $n$  ölçüm ile sol sansürlü bir örneklem verildiğinde;  $k$  farklı  $pe_i$  'leri tanımlayın ve sıralayın. En düşük  $pe_i$  ( $A_0$ ) değerinin altındaki tespit edilen değerlerin sayısını, en azından en yüksek  $pe_i$  ( $A_k$ ) kadar büyük olan tespit edilen değerlerin sayısını ve  $i$ 'inci ile  $(i + 1)$  arasındaki en düşük tespit edilen sayısını hesaplayın. Ayrıca  $B_0 = 0$  olsun ve toplam  $pe_i$  ( $B_i$  için  $i = 1, \dots, k$ ) altındaki tespit edilen ve tespit edilemeyenlerin toplam sayısını hesaplayın. Daha sonra,  $i$ 'inci  $pe_i$ 'nin altındaki tespit edilemeyen değerlerin ( $C_i$  için  $i = 1, \dots, k$ ) sayısını hesaplamak için eşitlik (2.2)'deki denklemi kullanın.

Adım 2:  $pe_0 = 1$  ve  $pe_{k+1} = 0$  olsun.  $i = 1, \dots, k$  için eşitlik (2.1)'deki denklem kullanılarak  $i$  numaralı farklı  $pe_i$  aşılma olasılığı hesaplanır.

Adım 3: Adım 2'den aşılma olasılıkları ile,  $A_i$  ile ilişkili tespit edilen verilerin her bir grubunu sıralayın ve ardından eşitlik (2.3) eşitliğini kullanarak bu tespit edilenler  $pd_{ij}$  için çizim konumları (yani birikimli olasılıklar) hesaplanır.

Adım 4: Normal ondalıkları (örneğin, z-skorları) tespit edilen veriler ve grafikleme konumlarıyla  $pd_{ij}$  ilişkilidir  $z_{ij}^d = \Phi^{-1}(pd_{ij})$  ile hesaplanır, burada  $\Phi^{-1}(\cdot)$  ters standart sansürlü dağılım fonksiyonudur.

Adım 5: Adım 4'teki z-skorlarını kullanarak sansürlü olasılık grafikleri oluşturun. Normal bir sansürlenmiş olasılık grafiği oluşturmak için  $z_{ij}^d$  karşı tespit edilen ölçüm değerleri  $x_{ij}^d$  'ler kullanılır. Normalleştirilmiş sansürlü olasılık grafiği oluşturmak için  $z_{ij}^d$  'lere karşı  $x_{ij}^d$  'lerin (örneğin; Log, karekök, ters, vb.) bir transformasyonu kullanılır.

Adım 6: Tek bir opsiyon olarak değiştirilmemiş gözlemleri içeren her denenmiş dönüşüm  $f(\cdot)$  için, çiftler arasındaki korelasyon katsayısı  $[f(x_{ij}^d), z_{ij}^d]$  hesaplanır.

Sansürlü olasılık çizelgesinde en yüksek korelasyon katsayısına sahip dönüşüm ve aynı zamanda lineer bir görünüm, sol sansürlenmiş örnekleme optimal olarak normalleştiren sonuçtur. Hiçbir dönüşüm yeterince doğrusal bir sansürlü olasılık grafiği ile sonuçlanmazsa, örneklemin normalleştirilemediği ve ROS'un tespit edilemeyen veriler için makul önlemler sağlamayabileceği sonucuna varır.

Adım 7: Bir normalleştirme dönüşüm tanımlanabiliyorsa, z-skorlarında ( $z_{ij}^d$ ),  $f(x_{ij}^d)$  verilerininin  $f(X) = \hat{a} + \hat{b}Z$  formunda doğrusal regresyonunu hesaplanır.

Adım 8: Eşitlik(2.4)'deki denklem kullanarak, her ayrı  $pe_i$  ile ilişkili tespit edilemeyen verilerin (diğer bir deyişle sansürlü gözlemler) grafik konumlarını ( $pc_{ij}$ ) hesaplayın. Ardından, ikinci bir z-skor kümesi oluşturun, bu sefer tespit edilemeyenlere bağlantılı olarak,  $z_{ij}^c = \Phi^{-1}(pc_{ij})$  hesaplanır.

Adım 9: Adım 7'den gelen eğim ve kesişim noktası ile Adım 8'den alınan sansürlenmiş z-skorları kullanılarak form atanan veriler için  $f(\hat{x}_{ij}^c) = \hat{a} + \hat{b} z_{ij}^c$  oluşturulur. Toplam örneklem ortalamasını ( $E(\bar{x}) = \mu$ ) ve örneklem standart sapmasını ( $E(s) = \sigma$ ) hesaplayarak anakütle ortalamasının ve standart sapmanın sansür tahminlerini elde

etmek için tespit edilemeyen veriler (dönüştürülmüş) için atanan değerleri, (dönüştürülmüş) tespit edilmiş verilerle  $f(x_{ij}^d)$  birleştirilir.

Bu sansür tahmini, tahmini ve kontrol limitleri için parametrik denklemlerde ve güven aralıkları için düzeltilmemiş örnek ortalaması ( $\bar{x}$ ) ve standart sapma (s) yerine kullanılabilir. Normalleştirici bir dönüşüme ihtiyaç duyulursa, sansürlenmiş tahminler, dönüştürülmüş ölçekte istatistiksel limitler ve aralıklar oluşturmak için kullanılmalıdır.

### 2.2.3. Tespit Edilemeyen Veri Regresyon Analizi Yöntemi

2005 yılında yaptığı çalışmasında Hersel, tespit edilemeyen veriler yerine başka veriler konulması durumundaki potansiyel veri analizi yanlılığına karşı uyarmıştır (Hersel, 2005). Tespit edilemeyen veri regresyon analizi, bir veya daha fazla verinin belirlenmiş bir eşik değerinin altında kalmasından dolayı ölçümü alınamamış verilerin analizi olup, pozitif değerli bağımlı bir değişken ile bir ya da daha fazla bağımsız değişken için incelenen bir regresyon modelidir (Hintze ve Utah, 2007).

Tespit edilemeyen veri regresyonunda tüm bağımsız değişkenler sayısal olmalıdır. Eğer kategorik değişkenler kullanılacak ise, bunlara karşılık gelen değişkenler oluşturulmalıdır. Artıkların dağılımının üstel, lojistik, log-lojistik, lognormal, normal veya Weibull dağılım gösterdiği varsayılır.

Bu regresyon modelinde, bağımlı değişkeni oluşturan veriler tespit edilen gözlem değerlerinin büyüklüğünü ve eşik değerlerini temsil eder. Bağımlı değişkeni oluşturan gözlem değeri sıfırdan büyük olmalıdır ve bu değerlerden herhangi biri eksik veya pozitif değilse tahmin aşaması sırasında kullanılmaz (Hintze ve Utah, 2007).

Bağımsız değişken ise; bağımsız değişkeni oluşturan gözlem değerinin bağımlı değişkenin tespit edilemeyen (sansürlü) bir gözlem değerini mi yoksa tespit edilmiş bir gözlem değerini mi temsil ettiğini belirtmektedir. Bağımsız değişkenler sayısal veya kategorik olabilirler.

Doğrusal regresyon eşitliği eşitlik(2.5)'daki gibidir.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + Se \quad (i=1, \dots, n) \quad (2.5)$$

Eşitlik (2.5)'de;

Y: Bağımlı veya  $\ln()$  veya  $\log()$  dönüşümü yapılmış bağımlı değişkeni

$X_1, \dots, X_i$ : Bağımsız değişkenlerin i. gözlem değerini,

$\beta_0$ : Regresyon doğrusunun Y eksenini kestiği noktanın orijine olan uzaklığını,

$\beta_1, \dots, \beta_k$ : Regresyon katsayısı olup bağımsız değişkendeki bir birimlik değişime karşılık bağımlı değişkende kendi birimi cinsinden meydana gelen ortalama değişim miktarlarını,

S: Sabit bir standart sapma değeri

e : Hata terimini,

göstermektedir.

Tespit edilemeyen veri regresyonuna ait olasılık yoğunluk fonksiyonu eşitlik (2.6)'deki gibidir. Eşitlikte M konum parametresi olarak yer almaktadır.

$$f(y|M, S) = \frac{1}{yS\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(y)-M}{S}\right)^2} \quad (2.6)$$

Daha açık bir şekilde yazmak gerekirse olasılık yoğunluk fonksiyonu eşitlik (2.7)'ye dönüşmektedir.

$$f(y|\beta_0 \dots \beta_k, S) = \frac{1}{yS\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(y)-\sum_{i=1}^k \beta_i X_i}{S}\right)^2\right\} \quad (2.7)$$

Maksimum olabilirlik tahmini, verilerin maksimum log-olabilirlik dağılım parametrelerinin değerlerinin bulunmasıyla oluşturulmaktadır. Genel olarak; bu değerler geçerli veri setinin oluşma ihtimalini en üst düzeye çıkaran parametrelerin değerleridir. Maksimum olabilirlik tahminlerinin büyük örneklerde en uygun özellikleri sahip olduğu görülmüştür (Hintze ve Utah, 2007).



#### 2.2.4. Çoklu Değer Atama Yöntemi

Alternatif olarak geliştirilen pek çok yöntem arasında 1990 başlarında teorik alt yapısı şekillendirilen ve 1990 sonlarında uygulama boyutuyla olgunlaştırılan en çok olabilirlik ve çoklu değer atama yaklaşımlarının öne çıktığı ve giderek daha yaygın bir şekilde kullanıldığı ifade edilmektedir (Allison, 2001).

Maksimum olabilirlik yaklaşımının alternatifi olarak çoklu değer atama her bir tespit edilemeyen veri yerine, olasılıkların dağılımını yansıtan, kabul edilebilir iki ya da daha fazla verinin atanmasını öngören bir yaklaşımdır (Rubin, 1978). Rubin tarafından 1978'de tavsiye edilen çoklu atama yöntemi, maksimum olabilirlik temelli tespit edilemeyen veri analizi yöntemlerindedir.

Çoklu değer atama yöntemlerinde tespit edilemeyen veri yerine  $m$  sayıda veri ataması yapılmakta ve  $m$  sayıda tamamlanmış veri seti elde edilmektedir. Her bir veri seti, atanan veriler gerçek verilermiş gibi kabul edilerek, standart eksiksiz veri süreçlerine göre analiz edilmektedir. Çoklu atama yöntemlerinde 2 ile 10 arası tamamlanmış veri setinin kullanımının mümkün olduğu belirtilmektedir (Rubin, 1978).

Bazı araştırmacıların yapmış oldukları çalışmaların sonuçlarına göre; 3,4 veya 5 tamamlanmış veri setlerinde çoklu değer atama aralığının, daha fazla değişkenlik gösterdiği gözlemlenirken, 10'dan fazla kullanılan veri setlerinde değişkenliğin bir süre sonra önemli derecede azaldığı belirtilmişlerdir (Schafer ve Graham, 2002).

Yöntem, her atama için farklı bir çözüm önerisi üretir. Eğer  $m$  tane atama sonucu birbirine benzer çıkarsa, yöntem desteklenmeli. Eğer sonuçlar önemli derecede farklı çıkarsa, bu farklılık standart hatalarla ilişkilendirilmelidir (Acock, 2005).

Çoklu değer atama yöntemleri, maksimum olabilirlik yöntemlerine göre model seçiminde daha az duyarlıdır. Çünkü basitçe, çoklu değer atama yöntemlerinde model, sadece tespit edilemeyen verilere atama yapılmasında kullanılmakta, maksimum olabilirlik yöntemlerinde olduğu gibi diğer parametrelerin kestiriminde kullanılmamaktadır (Allison, 2001). Maksimum olabilirlik kestirimlerinde olduğu gibi çoklu değer atama yöntemleri içinde en kullanışlı model çok değişkenli normal

modeldir. Bazı deęişkenler normal dağılım varsayımını karşılamasa da çok deęişkenli modelde çoklu deęer atama yöntemlerinin oldukça başarılı olduęu ifade edilmektedir (Schaffer, 1997).

En çok olabilirlik yönteminin aksine, çoklu deęer atama yöntemi her türlü veri ve modele uygulanabilmektedir. Ancak, çoklu atama her seferinde rastgele atamadan dolayı farklı tahminler üreteceęi için aynı veri setinden, aynı yöntemleri kullanarak farklı sonuçlar elde edilebilir (Allison, 2001; Elobeid ve ark., 2008). Bunun yanı sıra korelasyon deęerini olduęundan fazla gösterebilir. Örneęin X ve Y iki deęişken olsun. X deęişkeninde bazı veriler rastgele tespit edilemeyen veri olsun. X deęişkenine çoklu deęer atama yöntemi ile atanan veriler sonucunda Y deęişkeni ile aralarında 1.00 gibi mükemmel bir korelasyon deęeri çıkabilir. Bu sorun, atanan veriler arasında rastgele deęişkenlik sağlanarak çözülebilir (Allison, 2001). Yöntemin avantajı parametre tahminlerinin standart hataları; her veri setinin standart hataları ve parametre tahminlerinin veri setindeki saçılımı dikkate alınarak hesaplanmasıdır. Atama yapılmış verilerden elde edilen bu standart hatalar hipotez testi ve/veya parametre tahminlerinde güven aralığının belirlenmesinde kullanılır. Her atama sonunda rastgele deęişkenlikten dolayı farklı sonuçlar elde edildięi için standart hatalar daha doęru hesaplanır ve bu sayede elde edilen sonuçlara dair daha doęru çıkarımlar yapılır (Schlomer ve ark., 2010).

Çoklu deęer atama yönteminde tekli atamaların sonuçlarını birleştirme teknięi aşıęıdaki formüllerle açıklanabilir:

m : ataması yapılmış ve analiz edilmiş küme sayısı

$\hat{Q}_i$  : i. kümeden tahmin

$\hat{v}_i$  : i. kümenin varyans tahmini

Çoklu atamalarda elde edilen nokta tahmini her analizden elde edilenin ortalamasıdır. Bu ortalama aşıęıda belirtilen eşitlik (2.8)'daki gibidir.

$$\bar{Q} = \frac{1}{m} \sum_{i=1}^m \hat{Q}_i \quad \bar{Q} = \frac{1}{m} \sum_{i=1}^m \hat{Q}_i \quad (2.8)$$

Nokta tahmini için varyans tahmini aşağıda belirtilen eşitlik(2.9)'deki gibidir:

$$v = \frac{1}{m} \sum_{i=1}^m \hat{v}_i + \frac{m+1}{m} \left[ \frac{1}{m-1} \sum_{i=1}^m (Q_i - \bar{Q})^2 \right] \quad (2.9)$$

### 2.2.5. Tobit Regresyon Analizi Yöntemi

Tobit regresyon modeli, sınırlı bağımlı değişkenli modellerini açıklamada yaygın biçimde kullanılan ve en küçük kareler regresyonunun parametrik olmayan alternatifi olarak tanımlanan bir yöntemdir (Zorlutuna ve ark., 2016). Bazı araştırmalarda bağımlı değişkenlerin tüm değerlerine ulaşamamaktadır. Tobit modeller bir sansürlü veya budanmış sürekli değişken ile bir ya da daha fazla bağımsız değişken arasındaki ilişkiyi tanımlayan istatistiksel regresyon modelleri ailesindedir (Cunillera, 2014). Tobit modeller bağımlı değişkenin dağılım aralığı sınırlandırılmış olan sansürlü ve budanmış regresyon modelleridir. Tobit regresyon modeli, 1958 yılında Nobel ödüllü iktisatçı James Tobin tarafından geliştirilmiştir (Tobin, 1958).

Tobit modeli veya sansürlü normal regresyon modeli varsayımına göre; gizli (latent) değişken doğrusal, hata terimleri sıfır ortalama ve eşit varyanslıdır (Maddala, 1989). Tobit modeller latent değişken  $Y^*$  in doğrusal olarak bir  $\beta$  parametresi veya parametreler vektörü aracılığıyla  $X$ 'e bağlı olduğu varsayımı ile birlikte sansürlü normal dağılım fikrini regresyon modelleri içine taşır (Cunillera, 2014). Tobit modeller maksimum olabilirlik tahmini ile tahmin edilebilmektedir. Sansürlü bir eşik üzerindeki gözlemler üzerinde sıradan en küçük kareler regresyon yöntemi ve alternatif yöntemler gibi standart araçların kullanılmasının geçersiz çıkarımlar üreteceği de iyi bilinmektedir. Bu problemlerden dolayı araştırmacılar sansürlü bağımlı değişkenler ile sıklıkla Tobit Modeli kullanmaktadır (Getachew ve Huang, 2013).

#### 2.2.5.1 Sansürlü dağılım

Bir dağılım soldan sansürlü olduğunda,  $\tau'$  ya eşit ya da altındaki gözlemler  $\tau_y$  olmaktadır (Greene, 2007).

$$Y = \begin{cases} Y^* & \text{ise } Y^* > \tau \\ \tau_y & \text{ise } Y^* \leq \tau \end{cases} \quad (2.10)$$

$\tau$  ve  $\tau_y$  'nin kullanımı  $\tau$  ve  $\tau_y$  'nin sınırlanmasının bir genelleştirilmesidir. Eğer sürekli bir rastgele değişken  $Y$  bir  $f(y)$  olasılık yoğunluk fonksiyonuna sahipse ve  $\tau$  bir sabit ise, eşitlik (2.11) 'ye dönüşür (Greene, 2007).

$$f(y) = [f(y^*)]^{d_i} [F(\tau)]^{1-d_i} \quad (2.11)$$

$d$  ise  $Y > \tau$  iken 1'e;  $Y = \tau$  iken 0'a eşit olan bir gösterge değişkenidir.  $P(Y^* \leq \tau)$  ve  $P(Y^* > \tau)$  sırasıyla eşitlik(2.12) ve eşitlik(2.13)'deki gibi elde edilmektedir (Greene, 2007).

$$P(Y^* \leq \tau) = \Phi\left(\frac{\tau - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu - \tau}{\sigma}\right) \quad (2.12)$$

$$P(Y^* > \tau) = 1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right) = \Phi\left(\frac{\mu - \tau}{\sigma}\right) \quad (2.13)$$

### 2.2.5.2 Tobit Model

Tobit model  $i = 1, \dots, n$  için  $Y$  rastgele değişkeninin  $y_i$  gözlem değerlerinin eşitlik(2.14)'i yerine getirdiğini varsayar (Bierens, 2004).

$$y_i = \max(y_i^*, 0) \quad (2.14)$$

Gözlenemeyen  $y_i^*$  değerlerine karşılık gelen latent değişken  $Y^*$  eşitlik(2.15)'daki gibi yazılır (Bierens, 2004).

$$Y^* = \beta'X + \varepsilon \quad (2.15)$$

Eşitlik(2.15)'de  $X$  bağımsız rastgele değişken,  $\beta$  ise parametreler vektörüdür. Model hataları  $\varepsilon$ 'lerin bağımsız  $N(0, \sigma^2)$  dağılımlı olduğu varsayılmaktadır (Bierens, 2004).

Kukla değişken  $D$  tanımlanırsa (Bierens, 2004).

$$\begin{aligned} Y > 0 & \text{ ise } D = 1, \\ Y = 0 & \text{ ise } D = 0 \end{aligned} \quad (2.16)$$

Sonuç olarak  $Y$  eşitlik(2.17)'de olduğu gibi elde edilmektedir (Bierens, 2004).

$$Y = DY^* \quad (2.17)$$

Sansürlü bağımlı değişkeni tanımlamak için literatürde çeşitli yöntemler önerilmiştir. Sıklıkla uygulanan iki sansürlü regresyon modeli bulunmaktadır. İlk model Tobin tarafından (1958) ortaya koyulan temel Tip-1 Tobit modelidir ve yaygın olarak kullanılan modeldir (Tobin, 1958).

İkinci model ise Tip-2 Tobit modelidir. Bu modellere alternatif olarak yalnızca standart bir probit modelin ve doğrusal bir regresyon modelinin tahminini içeren iki aşamalı bir yöntem sunulmuştur. Bu yöntem de tip-3 Tobit Modeli olarak tanıtılmıştır (Heckman, 1976).

### 2.2.5.3 Tip-1 Tobit Model

Standart Tobit modelin altında yatan fikir bir ikili bağımlı değişken için Probit modelle ilişkilidir. Probit modeli gözlem dışı latent değişkeni  $Y^*$  sıfırdan küçük ya da eşit olursa ikili bağımlı değişken  $Y$ 'nin sıfır olduğunu, eğer latent değişkeni 1 olursa  $Y$ 'nin pozitif olduğunu varsaymaktadır (Franses ve Paap, 2004).

Latent değişkeni için  $X$ 'nin bir sabit de dahil  $k+1$  açıklayıcı değişken içerdiği ve  $\varepsilon \sim N(0,1)$  olduğu standart doğrusal regresyon modeli  $Y^* = X\beta + \varepsilon$  değerlendirilir. Eğer gözlem dışı latent değişkeni  $Y^*$  sıfırdan küçük ya da eşit ise sansürlü değişken  $Y = 0$  ve  $Y = Y^*$ 'dir. Eğer  $Y^*$  pozitifse  $\varepsilon \sim N(0, \sigma^2)$  iken (Franses ve Paap, 2004);

$$Y^* = X\beta + \varepsilon \leq 0 \text{ iken } Y = 0 \quad (2.18)$$

$$Y^* = X\beta + \varepsilon > 0 \text{ iken } Y = X\beta + \varepsilon \quad (2.19)$$

şeklindedir.

Sıfır olan  $y_i$  gözlemleri için yalnızca eşitlik(2.20)'i yazabiliriz.

$$\begin{aligned} P[Y = 0|X] &= P[X\beta + \varepsilon \leq 0 |X] \\ &= P[\varepsilon \leq -X\beta |X] \\ &= \Phi\left(-\frac{X\beta}{\sigma}\right) \end{aligned} \quad (2.20)$$

Bu olasılık probit modeldekiyle aynıdır. Aynı şekilde,  $Y = Y^* > 0$  probit modeldeki  $P[Y = 1|X]$  olasılığına benzemektedir. Probit modelin tersine, tobit modelde  $\sigma = 1$  kısıtlamasını getirmek zorunda değiliz. Çünkü bağımlı değişken  $Y$ 'nin pozitif gözlemleri  $\varepsilon$ 'nin varyansını belirlemektedir (Franses ve Paap, 2004).

$X$  bilindiğinde  $Y$ 'nin beklenen değeri ise eşitlik (2.21)'de gösterilmektedir.

$$E[Y|X] = \Pr[Y = 0 | X] E[Y|Y = 0, X] + \Pr[Y > 0 | X] E[Y|Y > 0, X]$$

$$\begin{aligned} &= 0 + \left(1 - \Phi\left(-\frac{X\beta}{\sigma}\right)\right) \left(-X\beta + \sigma \frac{\phi\left(-\frac{X\beta}{\sigma}\right)}{\left(1 - \Phi\left(-\frac{X\beta}{\sigma}\right)\right)}\right) \\ &= \left(1 - \Phi\left(-\frac{X\beta}{\sigma}\right)\right) X\beta + \sigma \phi\left(-\frac{X\beta}{\sigma}\right) \end{aligned}$$

(2.21)

Açıklayıcı değişkenler  $X$ 'ler bağımlı değişken  $Y$ 'nin olasılığını iki yönde etkilerler. İlk olarak pozitif bir  $\beta$  elemanı için ilgili  $X$  bileşenindeki artış  $Y$ 'nin 0'dan daha büyük olma olasılığını arttırır. İkinci olarak ise,  $X$ 'deki bir artma pozitif gözlemlerin koşullu ortalamasını da etkiler (Franses ve Paap, 2004).

$k$ 'nci açıklayıcı değişken  $X_k$ 'deki bir değişimin  $Y$ 'nin olasılığındaki toplam etkisi eşitlik(2.22)'de olduğu şekilde hesaplanmaktadır.

$$\begin{aligned} \frac{\partial E[Y|X]}{\partial X_k} &= \left(1 - \Phi\left(-\frac{X\beta}{\sigma}\right)\right) \beta_k - X\beta \phi\left(-\frac{X\beta}{\sigma}\right) \frac{\beta_k}{\sigma} \\ &\quad + \sigma \phi\left(-\frac{X\beta}{\sigma}\right) \left(-\frac{X\beta}{\sigma}\right) \left(-\frac{\beta_k}{\sigma}\right) \\ &= \left(1 - \Phi\left(-\frac{X\beta}{\sigma}\right)\right) \beta_k \end{aligned}$$

(2.22)

Yukarıda sunulan standart tobit modeli iki bilinen modelin bir kombinasyonu olarak yazılabilir. İlki  $Y$  değişkeninin sıfır mı pozitif mi olduğunu belirleyen probit modeldir (Franses ve Paap, 2004).

$$\begin{aligned} X\beta + \varepsilon \leq 0 \quad & \text{ise} \quad Y = 0 \\ X\beta + \varepsilon > 0 \quad & \text{ise} \quad Y > 0 \end{aligned} \quad (2.23)$$

İkinci model ise  $Y$ 'nin pozitif değerleri için budanmış regresyon modelidir.

$$Y = Y^* = X\beta + \varepsilon \quad Y > 0 \quad (2.24)$$

Probit modelden farkı, tobit modelde  $Y^*$  sıfırdan büyükse  $Y^*$ 'ı gözlemlerken Probit tanımlamasında hiçbir zaman  $Y^*$ 'ı gözlemlemeyiz. Bu durumda  $Y^*$ ,  $Y$ 'ye eşittir (Franses ve Paap, 2004).

#### 2.2.5.4 Tip-2 Tobit Modeli

Tip-1 Tobit modeldeki iki model de aynı hata terimi  $\varepsilon$ 'yi ve aynı  $\beta$  parametreleri ile aynı açıklayıcı değişkenler  $X$ 'leri içermektedir. Tabi ki bu varsayımı gevşetmek ve her iki modelde de farklı parametrelere ve farklı hata terimlerine yer vermek mümkündür. Bir örnek olarak;  $\alpha = (\alpha_0, \dots, \alpha_K)$  ve  $\varepsilon_1 \sim N(0,1)$  ile  $\varepsilon_2 \sim N(0, \sigma_2^2)$  iken (Franses ve Paap, 2004);

$$\begin{aligned} Y^* = X\alpha + \varepsilon_1 \leq 0 \quad & \text{ise} \quad Y = 0 \\ Y^* = X\alpha + \varepsilon_1 > 0 \quad & \text{ise} \quad Y = X\beta + \varepsilon_2 \end{aligned} \quad (2.25)$$

Her iki hata terimi de bağlantılı olabilir ve dolayısıyla da  $E[\varepsilon_1, \varepsilon_2] = \sigma_{12}$ . Bu model Tip-2 Tobit modeli olarak adlandırılmaktadır. Bu model  $Y$  sıfır veya pozitif olabildiği için bir Probit modelden ve  $Y$ 'nin pozitif değerleri için bir standart regresyon modelinden oluşmaktadır. Tip-2 Tobit Modeli, farklı  $\alpha$  ve  $\beta$  parametreleri olanak dahilinde olduğundan, Tip-1 Tobit modelinden daha esneklerdir (Franses ve Paap, 2004).

$X$  bilindiğinde  $Y = 0$  olasılığı eşitlik (2.26)'da verilmektedir.

$$\begin{aligned}
\Pr[Y = 0|X] &= \Pr[X\alpha + \varepsilon_1 \leq 0 |X] \\
&= \Pr[\varepsilon_1 \leq -X\alpha |X] \\
&= \Phi(-X\alpha)
\end{aligned} \tag{2.26}$$

$X$  bilindiğinde  $Y$ 'nin beklenen değeri ise eşitlik (2.27)'de verilmektedir.

$$\begin{aligned}
E[Y|X] &= E[Y|Y^* \leq 0, X] \Pr[Y^* \leq 0|X] \\
&= 0 + \left( X\beta + \sigma_{12} \frac{\phi(-X\alpha)}{1-\Phi(-X\alpha)} \right) (1 - \Phi(-X\alpha)) \\
&= X\beta(1 - \Phi(-X\alpha)) + \sigma_{12}\phi(-X\alpha)
\end{aligned} \tag{2.27}$$

$k$ 'inci açıklayıcı değişken  $X_k$ 'daki bir değişimin  $Y$ 'nin olasılığındaki toplam etkisi eşitlik (2.28)'da olduğu şekilde hesaplanmaktadır (Franses ve Paap, 2004).

$$\frac{\partial E[Y|X]}{\partial X_k} = (1 - \Phi(-X\alpha))\beta_k + X\beta\phi(-X\alpha)\alpha_k - \sigma_{12}(X\alpha)\phi(-X\alpha)\alpha_k \tag{2.28}$$

### 2.2.5.5. Tip-3 Tobit Model: Heckman's Model

1976 yılında yapılan çalışmalar sonucunda, yalnızca standart bir probit modelin ve lineer bir regresyon modelinin tahminini içeren iki aşamalı bir yöntem sunmuştur (Heckman, 1976). İki aşamalı yöntem koşullu ortalamayı kullanmaktadır (Schmidheiny, 2007).  $D^*$  ve  $Y^*$  olmak üzere iki latent değişkenli bir modeli değerlendiriyor (Schmidheiny, 2007).

$$D^* = Z'\gamma + \nu \tag{2.29}$$

$$Y^* = X'\beta + \varepsilon \tag{2.30}$$

$\varepsilon$  ve  $\nu$  hata terimleri gözlemlere karşı bağımsız olmakla beraber  $\rho\sigma_\varepsilon$  kovaryanslı bileşik normal dağılımlıdır (Schmidheiny, 2007).

$$(\nu, \varepsilon) \sim N\left(0, \begin{bmatrix} 1 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix}\right) \tag{2.31}$$



Latent deęişken arařtırmacı tarafından gözlenememektedir. Arařtırmacı yalnızca latent deęişkeni  $D^*$  pozitif olduęunda bir  $D$  göstergesi gözlemlemektedir.  $Y = Y^*$  deęişkeninin deęeri ise yalnızca gösterge bire eřit ise gözlenebilmektedir (Schmidheiny, 2007).

$$D = \begin{cases} D^* > 0 & \text{ise } 1 \\ \text{deęilse} & 0 \end{cases} \quad (2.32)$$

$$Y = \begin{cases} D = 1 & \text{ise } Y^* \\ \text{aksi uygulanamaz} & \end{cases} \quad (2.33)$$

Dięer bir deyişle, ilk eřitlik ( $D^*$  karar eřitlięi) bir gözlemin örnekleme olup olmadığını açıklamaktadır. İkinci eřitlik ( $Y^*$  regresyon eřitlięi)  $Y$ 'nin deęerini belirlemektedir. Standart tobit modeli bunun  $Z = X$ ,  $\gamma = \beta$ ,  $\sigma_v = \sigma_\varepsilon$  ve  $\rho = 1$  eřitliklerinin mevcut olduęu özel durumudur (Schmidheiny, 2007).

$\lambda(\alpha) = \phi(\alpha)/\Phi(\alpha)$  eřitlięi ters Mills Oranı iken  $Y$  deęişkeninin beklenen deęeri ( $D=1$  olması durumunda)  $Y^*$ 'in kořullu beklentisidir (Schmidheiny, 2007).

$$\begin{aligned} E(Y|X, Z) &= E(Y^*|D = 1, X, Z) \\ &= X'\beta + \rho\sigma_\varepsilon \frac{\phi(Z'\gamma)}{\Phi(Z'\gamma)} \\ &= X'\beta + \rho\sigma_\varepsilon \lambda(Z'\gamma) \end{aligned} \quad (2.34)$$

#### 2.2.5.6. Tobit Tahmini

Tobit tahmincileri klasik en çok olabilirlik özelliklerine sahiptir. Tobit modelin log-olabilirlik fonksiyonu küresel olarak içbükey olmasa da eřit bir maksimuma sahiptir (Getachew ve Huang, 2013). Sansürlü regresyon modellerinin parametreleri en çok olabilirlik yöntemi kullanılarak tahmin edilebilir (Franses ve Paap, 2004).

Tip-1 Tobit model için olabilirlik fonksiyonu iki kısımdan oluşmaktadır. Bir gözlemin sansürlü olma olasılıęı eřitlik (2.15) aracılıęıyla verilmiştir. Sansürsüz

gözlemlerin yoğunluğu standart normal dağılımdır.  $\theta = (\beta, \sigma)$  olduğu durumda olabilirlik fonksiyonu eşitlik(2.35)'te gösterilmektedir (Franses ve Paap, 2004).

$$L(\theta) = \prod_{i=1}^n \Phi\left(\frac{-x_i \beta}{\sigma}\right)^{I[y_i=0]} \left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i \beta)^2\right)\right)^{I[y_i>0]} \quad (2.35)$$

Tip-2 Tobit Modelinin olabilirlik fonksiyonu da iki kısımdan oluşmaktadır. Sansürlü gözlemler için olabilirlik fonksiyonu eşitlik (2.15)'inci eşitlikte verilen  $Y = 0$  veya  $Y^* \leq 0$  olasılığına eşittir. Sansürlü olmayan gözlemler için  $Y^* > 0$  bilindiğinde  $f(y|y^* > 0)$ 'in  $Y^* > 0$  olma olasılığıyla çarpılması ile ifade edilen  $Y$  rastgele değişkeninin yoğunluk fonksiyonu kullanılır. Dolayısıyla,  $\theta = (\beta, \sigma, \sigma_2^2, \sigma_{12})$  iken olabilirlik fonksiyonu eşitlik (2.36)'da gösterilmektedir (Franses ve Paap, 2004).

$$L(\theta) = \prod_{i=1}^n (P[y_i^* < 0])^{I[y_i=0]} (f(y_i|y_i^* > 0) \Pr[y_i^* > 0])^{I[y_i=1]} \quad (2.36)$$

### 3.GEREÇ VE YÖNTEM

Yöntemlerin karşılaştırılması veri türetimi aşamasında çoklu doğrusal regresyon modelinden yararlanılmıştır. Doğrusal regresyon modeline göre veri türetimi aşamasında farklı değişken sayılar için sonuçlar incelendiğinde yöntemler arasındaki ilişki yapısında büyük benzerlikler olduğu görülmüştür. Bu nedenle farklı sayıda bağımsız değişkenler için incelemeye gerek duyulmadan bağımsız değişken sayısı 2 olarak belirlenmiştir.

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\varepsilon \quad (3.1)$$

Modelde yer alan bağımsız değişkenler normal dağılımdan türetilmiştir. Modelde yer alan bağımlı değişken y'nin türetilmesi aşamasında açıklayıcılık katsayısının ( $R^2$ ) 3 farklı düzeyi 0,90; 0,70; 0,50 değerleri göz önünde bulundurulmuştur. İlgilenilen açıklayıcılık katsayısı  $R^2$  için bağımlı değişken y'nin türetilirken  $\beta_0$ ,  $\beta_1$  ve  $\beta_2$  regresyon katsayıları rastgele olarak  $\beta_0=10$ ;  $\beta_1=4$  ve  $\beta_2=2$  şeklinde belirlenmiştir.

Türetilen veri için y değişkenindeki tespit edilemeyen veri oranları %5, %10, %15, %20, %25, %30, %40, %50, %60 ve %70 şeklinde belirlenmiştir. Buna bağlı olarak örneğin; %5 tespit edilemeyen veri oranı için y değişkeninin yüzde 5'lik değeri alt sınır olarak belirlenip bu değer altındaki y değerleri tespit edilemeyen veri olarak kabul edilmiştir. Ayrıca veri türetimi için belirlenen örneklem büyüklükleri 100, 250, 500, 1000 ve 5000 şeklindedir (Tablo-1).

Veri üretim sonrasında yöntemlerin karşılaştırılmasında kullanılan istatistiksel kriter hata kareler ortalamasının kare köküdür (root mean squared error, RMSE). Veri türetimi aşamasında 1000 tekrar gerçekleştirilmiş ve RMSE değerlerinin ortalaması tablolarda verilmiştir. RMSE'yi hesaplayabilmek için eşitlik (3.2)'den yararlanılmaktadır.

$$RMSE = \frac{1}{t} \sum_{i=1}^t \left( \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-k}} \right) \quad (3.2)$$

$y_i$ : Bağımlı değişken y'nin i.gözlem değeri

$\hat{y}_i$ : Bağımlı değişken y'nin i.gözleme ilişkin tahmin değeri

n: Örneklem büyüklüğü

k: Modeldeki bağımsız değişken sayısı

Tespit edilemeyen verinin tahmin yöntemlerinden sıfır atanması, alt Sınır/2 atanması, alt sınır atanması, Tobit regresyon, ROS, tespit edilemeyen veri regresyonu, çoklu değer atama yöntemlerine ilişkin RMSE değerleri hesaplanmıştır (Tablo 2-16). Üç tip Tobit regresyon olmasına karşın simülasyonda kullanılan Tip-1 Tobit regresyondur. Veri üretim aşamasında, yöntemlerin kullanılmasında ve RMSE'nin hesaplanmasında R yazılım programı kullanılmıştır.



Tablo 1: Simülasyon tablosu

Yöntem	Örnekleme büyüklükleri	$R^2$ açıklayıcılık katsayısı	Tespit edilemeyen veri oranları
Sıfır Atanması	100 250 500 1000 5000	0,50 0,70 0,90	0,05; 0,10; 0,15;0,20;0,25; 0,30; 0,40;0,50; 0,60; 0,70
Alt Sınır/2 Atanması	100 250 500 1000 5000	0,50 0,70 0,90	0,05; 0,10; 0,15;0,20;0,25; 0,30; 0,40;0,50; 0,60; 0,70
Alt Sınır Atanması	100 250 500 1000 5000	0,50 0,70 0,90	0,05; 0,10; 0,15;0,20;0,25; 0,30; 0,40;0,50; 0,60; 0,70
Tobit Regresyon	100 250 500 1000 5000	0,50 0,70 0,90	0,05; 0,10; 0,15;0,20;0,25; 0,30; 0,40;0,50; 0,60; 0,70
ROS	100 250 500 1000 5000	0,50 0,70 0,90	0,05; 0,10; 0,15;0,20;0,25; 0,30; 0,40;0,50; 0,60; 0,70
Tespit Edilemeyen Veri Regresyonu	100 250 500 1000 5000	0,50 0,70 0,90	0,05; 0,10; 0,15;0,20;0,25; 0,30; 0,40;0,50; 0,60; 0,70
Çoklu Değer Atama	100 250 500 1000 5000	0,50 0,70 0,90	0,05; 0,10; 0,15;0,20;0,25; 0,30; 0,40;0,50; 0,60; 0,70

## 4.BULGULAR

Tespit edilemeyen verilerin tahmininde kullanılan çeşitli yöntemler içerisinde sıfır atanması, tespit edilemeyen veri için alt sınırın yarısının atanması, alt sınırın atanması, ROS, tespit edilemeyen veri regresyonu, tobit regresyon ve çoklu değer atama yöntemleri incelemeye alınmıştır. Farklı açıklayıcılık katsayısı  $R^2$ , örneklem büyüklüğü ve tespit edilemeyen veri oranları için bu yöntemler RMSE'ye göre karşılaştırılmıştır. Yöntemlerin 1000 tekrar sonucundaki ortalama RMSE değerleri Tablo 2-16 arasında yer almaktadır.

Tablo 2:  $R^2=0,90$  ve  $n=100$  için RMSE değerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Değer Atama</b>
<b>%5</b>	20,12	10,12	1,50	1,50	4,63	3,24	1,47
<b>%10</b>	26,11	13,06	1,53	1,59	4,55	6,56	1,46
<b>%15</b>	29,63	14,79	1,57	1,72	4,49	9,91	1,44
<b>%20</b>	31,86	15,92	1,61	1,89	4,42	13,27	1,45
<b>%25</b>	33,46	16,73	1,63	2,08	4,37	16,65	1,42
<b>%30</b>	34,56	17,31	1,63	2,26	4,28	20,04	1,41
<b>%40</b>	35,86	18,03	1,64	2,74	4,16	26,79	1,39
<b>%50</b>	36,55	18,46	1,59	3,31	4,00	33,50	1,37
<b>%60</b>	36,96	18,75	1,49	3,97	3,83	40,12	1,33
<b>%70</b>	36,77	18,70	1,33	4,76	3,64	46,53	1,29

Tablo 2'ye bakıldığında sıfır atanması yöntemi tespit edilemeyen veri oranının artmasıyla RMSE'de düzenli olarak artış göstermiştir. Alt sınır/2 atanması yönteminde de sıfır atanması yönteminde olduğu gibi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüş. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranlarının %5'den %40'a doğru artarken RMSE'nin de arttığı tespit edilemeyen veri oranı %50, %60 ve %70 olduğunda ise RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyon sonucunda tespit edilemeyen veri oranı arttıkça RMSE'nin de arttığı görülmüştür. ROS yöntemiyle; tespit edilemeyen veri oranı artarken RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyonu yönteminde tespit edilemeyen veri oranı artarken RMSE'nin arttığı görülmüştür. Çoklu değer atama yöntemi incelendiğinde ise ROS yöntemindeki sonuçla benzer sonuç elde edilmiş tespit edilemeyen veri oranları arttıkça RMSE'nin azaldığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır atanması, Tobit regresyon ve çoklu değer atamadan elde edilen RMSE değerlerinin diğer yöntemlere göre daha düşük olduğu ve bu üç yöntemin birbirine yakın RMSE değerlerine sahip olduğu söylenebilir.



Tablo 3:  $R^2=0,90$  ve  $n=250$  için RMSE deęerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Deęer Atama</b>
<b>%5</b>	19,62	9,84	1,50	1,51	4,63	3,10	1,47
<b>%10</b>	25,97	12,97	1,54	1,60	4,55	6,55	1,45
<b>%15</b>	29,32	14,62	1,58	1,74	4,47	9,76	1,45
<b>%20</b>	31,67	15,81	1,61	1,90	4,39	13,26	1,43
<b>%25</b>	33,21	16,60	1,63	2,09	4,34	16,50	1,43
<b>%30</b>	34,30	17,17	1,64	2,29	4,25	20,03	1,41
<b>%40</b>	35,58	17,89	1,64	2,76	4,10	26,78	1,39
<b>%50</b>	36,38	18,38	1,59	3,32	3,95	33,49	1,37
<b>%60</b>	36,70	18,62	1,50	3,99	3,78	40,11	1,34
<b>%70</b>	36,46	18,55	1,35	4,79	3,57	46,50	1,31

Tablo 3'e bakıldığında sıfır atanması yönteminde tespit edilemeyen veri oranının artmasıyla RMSE'de düzenli olarak artış göstermiştir. Alt sınır/2 atanması yönteminde de sıfır atanması yönteminde olduğu gibi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüş. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranlarının %5'den %40'a doğru artarken RMSE'nin de arttığı tespit edilemeyen veri oranı %50, %60 ve %70 olduğunda ise RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyon sonucunda tespit edilemeyen veri oranı arttıkça RMSE'nin de arttığı görülmüştür. ROS yöntemiyle; tespit edilemeyen veri oranı artarken RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyonu yönteminde tespit edilemeyen veri oranı artarken RMSE'nin arttığı görülmüştür. Çoklu değer atama yöntemi incelendiğinde ise ROS yöntemindeki sonuçla benzer sonuç elde edilmiş tespit edilemeyen veri oranları arttıkça RMSE'nin azaldığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır atanması ve çoklu değer atamadan elde edilen RMSE değerlerinin diğer yöntemlerden düşük ve birbirine benzer olduğu görülmekle birlikte Tobit regresyon yöntemi için tespit edilemeyen veri oranı %15 kadar bu yöntemlerle benzerlik gösterirken daha yüksek oranlar için RMSE artış göstermektedir.

Tablo 4  $R^2=0,90$  ve  $n=500$  için RMSE değerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Değer Atama</b>
<b>%5</b>	19,90	9,98	1,51	1,52	4,61	3,23	1,47
<b>%10</b>	25,87	12,92	1,54	1,62	4,53	6,55	1,46
<b>%15</b>	29,38	14,66	1,58	1,75	4,46	9,90	1,45
<b>%20</b>	31,66	15,81	1,61	1,91	4,38	13,26	1,43
<b>%25</b>	33,11	16,55	1,63	2,10	4,32	16,64	1,42
<b>%30</b>	34,21	17,13	1,64	2,30	4,24	20,02	1,41
<b>%40</b>	35,58	17,89	1,65	2,77	4,10	26,77	1,39
<b>%50</b>	36,32	18,35	1,60	3,32	3,93	33,48	1,37
<b>%60</b>	36,68	18,61	1,50	3,97	3,73	40,10	1,34
<b>%70</b>	36,42	18,53	1,35	4,79	3,52	46,52	1,31

Tablo 4'e bakıldığında sıfır atanması yönteminde tespit edilemeyen veri oranının artmasıyla RMSE'de düzenli olarak artış göstermiştir. Alt sınır/2 atanması yönteminde de sıfır atanması yönteminde olduğu gibi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüş. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranlarının %5'den %40'a doğru artarken RMSE'nin de arttığı tespit edilemeyen veri oranı %40 sonrası için ise RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyon sonucunda tespit edilemeyen veri oranı arttıkça RMSE'nin de arttığı görülmüştür. ROS yöntemiyle çoklu değer atama yöntemi için tespit edilemeyen veri oranı artarken RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyonu yönteminde tespit edilemeyen veri oranı artarken RMSE'nin arttığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır atanması ve çoklu değer atamadan elde edilen RMSE değerlerinin diğer yöntemlerden düşük ve birbirine benzer olduğu görülmekle birlikte Tobit regresyon yöntemi için tespit edilemeyen veri oranı %15 kadar bu yöntemlerle benzerlik gösterirken daha yüksek oranlar için RMSE artış göstermektedir.

Tablo 5:  $R^2=0,90$  ve  $n=1000$  için RMSE deęerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Deęer Atama</b>
<b>%5</b>	19,86	9,96	1,51	1,52	4,61	3,23	1,47
<b>%10</b>	25,85	12,91	1,54	1,62	4,53	6,55	1,46
<b>%15</b>	29,33	14,63	1,58	1,75	4,46	9,89	1,45
<b>%20</b>	31,59	15,77	1,61	1,91	4,38	13,26	1,43
<b>%25</b>	33,15	16,57	1,63	2,09	4,30	16,64	1,42
<b>%30</b>	34,24	17,14	1,65	2,30	4,24	20,02	1,41
<b>%40</b>	35,56	17,88	1,65	2,77	4,09	26,77	1,39
<b>%50</b>	36,32	18,35	1,60	3,32	3,91	33,49	1,37
<b>%60</b>	36,66	18,60	1,51	3,98	3,74	40,10	1,34
<b>%70</b>	36,39	18,52	1,35	4,79	3,50	46,51	1,31

Tablo 5'e bakıldığında sıfır atanması yönteminde tespit edilemeyen veri oranının artmasıyla RMSE'de düzenli olarak artış göstermiştir. Alt sınır/2 atanması yönteminde de sıfır atanması yönteminde olduğu gibi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüş. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranlarının %5'den %40'a doğru artarken RMSE'nin de arttığı tespit edilemeyen veri oranı %40 sonrası için ise RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyon sonucunda tespit edilemeyen veri oranı arttıkça RMSE'nin de arttığı görülmüştür. ROS yöntemiyle çoklu değer atama yöntemi için tespit edilemeyen veri oranı artarken RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyonu yönteminde tespit edilemeyen veri oranı artarken RMSE'nin arttığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır atanması ve çoklu değer atamadan elde edilen RMSE değerlerinin diğer yöntemlerden düşük ve birbirine benzer olduğu görülmekle birlikte Tobit regresyon yöntemi tespit edilemeyen veri oranı %5 için bu yöntemlerle benzerlik gösterirken diğer oranlar için RMSE değeri daha yüksektir.

Tablo 6:  $R^2=0,90$  ve  $n=5000$  için RMSE değerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Değer Atama</b>
<b>%5</b>	19,85	9,95	1,51	1,53	4,61	3,23	1,47
<b>%10</b>	25,82	12,89	1,54	1,62	4,53	6,55	1,46
<b>%15</b>	29,32	14,63	1,58	1,75	4,45	9,90	1,45
<b>%20</b>	31,58	15,77	1,61	1,92	4,38	13,26	1,44
<b>%25</b>	33,11	16,55	1,63	2,10	4,30	16,63	1,43
<b>%30</b>	34,15	17,10	1,65	2,31	4,22	20,01	1,41
<b>%40</b>	35,50	17,85	1,64	2,77	4,07	26,77	1,39
<b>%50</b>	36,27	18,32	1,60	3,33	3,91	33,48	1,37
<b>%60</b>	36,65	18,59	1,50	3,98	3,71	40,10	1,34
<b>%70</b>	36,37	18,51	1,35	4,79	3,48	46,51	1,31

Tablo 6'ya bakıldığında sıfır atanması yönteminde tespit edilemeyen veri oranının artmasıyla RMSE'de düzenli olarak artış göstermiştir. Alt sınır/2 atanması yönteminde de sıfır atanması yönteminde olduğu gibi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüş. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranlarının %5'den %30'a doğru artarken RMSE'nin de arttığı tespit edilemeyen veri oranı %30 sonrası için ise RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyon sonucunda tespit edilemeyen veri oranı arttıkça RMSE'nin de arttığı görülmüştür. ROS yöntemiyle çoklu değer atama yöntemi için tespit edilemeyen veri oranı artarken RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyonu yönteminde tespit edilemeyen veri oranı artarken RMSE'nin arttığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır atanması ve çoklu değer atamadan elde edilen RMSE değerlerinin diğer yöntemlerden düşük ve birbirine benzer olduğu görülmekle birlikte Tobit regresyon yöntemi tespit edilemeyen veri oranı %5 için bu yöntemlerle benzerlik gösterirken diğer oranlar için RMSE değeri daha yüksektir.



Tablo 7:  $R^2=0,70$  ve  $n=100$  için RMSE deęerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Deęer Atama</b>
<b>%5</b>	21,07	11,03	2,84	2,81	5,24	3,17	2,81
<b>%10</b>	27,83	14,36	2,77	2,78	5,16	6,41	2,75
<b>%15</b>	32,26	16,56	2,70	2,78	5,08	9,68	2,69
<b>%20</b>	35,15	18,00	2,65	2,83	5,02	12,95	2,65
<b>%25</b>	37,45	19,17	2,58	2,88	4,90	16,23	2,59
<b>%30</b>	39,10	20,00	2,50	2,98	4,84	19,49	2,54
<b>%40</b>	41,16	21,05	2,34	3,27	4,68	25,93	2,44
<b>%50</b>	42,24	21,61	2,14	3,71	4,49	32,23	2,33
<b>%60</b>	42,18	21,59	1,94	4,36	4,34	38,20	2,25
<b>%70</b>	41,13	21,04	1,66	5,15	4,08	43,66	2,13

Tablo 7'ye bakıldığında sıfır atanması yönteminde tespit edilemeyen veri oranı %5'den %50'ye doğru artarken, RMSE'nin artış gösterdiği, tespit edilemeyen veri oranı %60 ve %70 olduğunda ise RMSE'nin azaldığı görülmüştür. Alt sınır/2 atanması yönteminde ise tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüş, tespit edilemeyen veri oranı %40'dan sonra artış durağan bir hale dönüşmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı arttıkça -%5 oranı hariç- RMSE'nin arttığı görülmüştür. ROS yönteminde; tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi de Tobit regresyon yöntemiyle paralellik göstermiş ve tespit edilemeyen veri oranı arttıkça RMSE'nin de arttığı görülmüştür. Çoklu değer atama yöntemi incelendiğinde ise ROS yöntemindeki sonuçlarla benzerlik göstermiş ve tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken Tobit regresyon yöntemi için %15 tespit edilemeyen veri oranı itibariyle diğer yöntemlerden daha yüksek RMSE değerine sahiptir.

Tablo 8:  $R^2=0,70$  ve  $n=250$  için RMSE değerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Değer Atama</b>
<b>%5</b>	20,58	10,77	2,85	2,84	5,23	3,03	2,83
<b>%10</b>	27,75	14,31	2,77	2,80	5,13	6,40	2,75
<b>%15</b>	31,81	16,32	2,71	2,81	5,05	9,54	2,71
<b>%20</b>	34,95	17,90	2,65	2,84	4,98	12,94	2,66
<b>%25</b>	37,10	18,98	2,58	2,91	4,90	16,08	2,61
<b>%30</b>	38,88	19,88	2,50	3,00	4,82	19,47	2,55
<b>%40</b>	41,00	20,97	2,35	3,31	4,65	25,92	2,45
<b>%50</b>	42,01	21,50	2,17	3,75	4,46	32,20	2,36
<b>%60</b>	42,09	21,55	1,95	4,36	4,25	38,19	2,25
<b>%70</b>	40,96	20,97	1,68	5,17	4,03	43,66	2,15

Tablo 8'e bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %60'a doğru artarken, RMSE'nin artış gösterdiği, tespit edilemeyen veri oranı %70 olduğunda ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %20 ye kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS yönteminde tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi de Tobit regresyon yöntemiyle paralellik göstermiş ve tespit edilemeyen veri oranı arttıkça RMSE'nin de arttığı görülmüştür. Çoklu değer atama yöntemi incelendiğinde ise, ROS yöntemindeki sonuçlarla benzerlik göstermiş ve tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %40 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi %15 ve daha yüksek tespit edilemeyen veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.

Tablo 9:  $R^2=0,70$  ve  $n=500$  için RMSE deęerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Deęer Atama</b>
<b>%5</b>	20,88	10,91	2,84	2,84	5,22	3,16	2,82
<b>%10</b>	27,65	14,26	2,78	2,81	5,14	6,40	2,76
<b>%15</b>	31,94	16,39	2,71	2,81	5,04	9,67	2,70
<b>%20</b>	34,89	17,86	2,64	2,85	4,97	12,94	2,65
<b>%25</b>	37,14	19,00	2,57	2,91	4,87	16,21	2,60
<b>%30</b>	38,73	19,81	2,51	3,01	4,80	19,47	2,55
<b>%40</b>	40,93	20,94	2,35	3,31	4,62	25,92	2,45
<b>%50</b>	41,92	21,46	2,17	3,76	4,44	32,20	2,36
<b>%60</b>	41,98	21,49	1,95	4,38	4,22	38,18	2,26
<b>%70</b>	40,93	20,95	1,68	5,20	3,97	43,64	2,14

Tablo 9'a bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %60'a doğru artarken, RMSE'nin artış gösterdiği, tespit edilemeyen veri oranı %70 olduğunda ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %20 ye kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS yönteminde tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüştür. Çoklu değer atama yöntemi incelendiğinde ise ROS yöntemindeki sonuçlarla benzerlik göstermiş ve tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %40 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi için %15 ve daha yüksek veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.

Tablo 10:  $R^2=0,70$  ve  $n=1000$  için RMSE deęerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Deęer Atama</b>
<b>%5</b>	20,87	10,91	2,84	2,85	5,22	3,16	2,82
<b>%10</b>	27,59	14,22	2,78	2,82	5,13	6,40	2,76
<b>%15</b>	31,88	16,36	2,71	2,82	5,04	9,66	2,70
<b>%20</b>	34,87	17,86	2,65	2,85	4,96	12,93	2,65
<b>%25</b>	37,10	18,98	2,57	2,91	4,86	16,21	2,60
<b>%30</b>	38,74	19,81	2,51	3,02	4,79	19,46	2,55
<b>%40</b>	40,87	20,91	2,35	3,32	4,61	25,91	2,46
<b>%50</b>	41,91	21,45	2,17	3,77	4,43	32,19	2,37
<b>%60</b>	41,98	21,50	1,96	4,39	4,22	38,18	2,27
<b>%70</b>	40,93	20,95	1,69	5,21	3,95	43,63	2,16

Tablo 10'a bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %60'a doğru artarken, RMSE'nin artış gösterdiği, tespit edilemeyen veri oranı %70 olduğunda ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %20 ye kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS yönteminde tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüştür. Çoklu değer atama yöntemi incelendiğinde ise ROS yöntemindeki sonuçlarla benzerlik göstermiş ve tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %40 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi için %15 ve daha yüksek veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.



Tablo 11:  $R^2=0,70$  ve  $n=5000$  için RMSE değerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Değer Atama</b>
<b>%5</b>	20,84	10,89	2,84	2,85	5,22	3,16	2,82
<b>%10</b>	27,59	14,22	2,78	2,83	5,13	6,40	2,77
<b>%15</b>	31,86	16,34	2,71	2,82	5,04	9,66	2,70
<b>%20</b>	34,87	17,85	2,64	2,85	4,95	12,93	2,65
<b>%25</b>	37,07	18,96	2,58	2,92	4,87	16,20	2,60
<b>%30</b>	38,70	19,79	2,51	3,02	4,78	19,46	2,55
<b>%40</b>	40,84	20,89	2,35	3,32	4,60	25,91	2,46
<b>%50</b>	41,86	21,43	2,17	3,77	4,41	32,19	2,36
<b>%60</b>	41,94	21,48	1,96	4,39	4,20	38,18	2,27
<b>%70</b>	40,88	20,93	1,69	5,21	3,93	43,63	2,16

Tablo 11'e bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %60'a doğru artarken, RMSE'nin artış gösterdiği, tespit edilemeyen veri oranı %70 olduğunda ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %20 ye kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS yönteminde tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüştür. Çoklu değer atama yöntemi incelendiğinde ise ROS yöntemindeki sonuçlarla benzerlik göstermiş ve tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %30 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi için %15 ve daha yüksek veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.

Tablo 12:  $R^2=0,50$  ve  $n=100$  için RMSE değerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Değer Atama</b>
<b>%5</b>	22,06	12,14	4,32	4,27	6,22	3,07	4,22
<b>%10</b>	29,68	15,87	4,19	4,16	6,10	6,23	4,09
<b>%15</b>	34,64	18,31	4,01	4,04	5,99	9,41	3,94
<b>%20</b>	38,21	20,07	3,86	3,98	5,90	12,58	3,83
<b>%25</b>	41,06	21,48	3,72	3,95	5,81	15,75	3,72
<b>%30</b>	43,30	22,59	3,56	3,94	5,67	18,90	3,62
<b>%40</b>	46,05	23,92	3,25	4,09	5,51	25,08	3,43
<b>%50</b>	47,39	24,55	2,92	4,43	5,32	30,93	3,24
<b>%60</b>	47,14	24,36	2,57	5,04	5,13	36,28	3,07
<b>%70</b>	45,39	23,38	2,14	5,82	4,82	40,80	2,86

Tablo 12'ye bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %50'ye doğru artarken, RMSE'nin artış gösterdiği, tespit edilemeyen veri oranı %50'den sonra ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %30'a kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS ve çoklu değer atama yöntemleri tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %30 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi için %20 ve daha yüksek veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.

Tablo 13:  $R^2=0,50$  ve  $n=250$  için RMSE değerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Değer Atama</b>
<b>%5</b>	21,60	11,88	4,31	4,29	6,19	2,94	4,23
<b>%10</b>	29,47	15,74	4,17	4,18	6,07	6,22	4,08
<b>%15</b>	34,24	18,09	4,01	4,07	5,97	9,27	3,95
<b>%20</b>	38,08	20,00	3,87	4,01	5,88	12,57	3,85
<b>%25</b>	40,77	21,32	3,71	3,97	5,77	15,62	3,72
<b>%30</b>	42,96	22,41	3,56	3,98	5,67	18,88	3,62
<b>%40</b>	45,88	23,84	3,26	4,11	5,47	25,04	3,43
<b>%50</b>	47,17	24,44	2,93	4,44	5,26	30,91	3,25
<b>%60</b>	47,01	24,30	2,57	5,03	5,01	36,26	3,07
<b>%70</b>	45,30	23,36	2,17	5,90	4,74	40,74	2,88

Tablo 13'e bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %50'ye doğru artarken, RMSE'nin artış gösterdiği, bu orandan sonra ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %30'a kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS ve çoklu değer atama yöntemleri tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %30 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi için %20 ve daha yüksek veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.

Tablo 14:  $R^2=0,50$  ve  $n=500$  için RMSE değerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Değer Atama</b>
<b>%5</b>	21,95	12,05	4,31	4,30	6,18	3,06	4,23
<b>%10</b>	29,39	15,70	4,16	4,18	6,07	6,22	4,08
<b>%15</b>	34,36	18,15	4,00	4,08	5,95	9,39	3,94
<b>%20</b>	38,05	19,98	3,86	4,02	5,86	12,57	3,83
<b>%25</b>	40,79	21,34	3,71	3,98	5,75	15,73	3,73
<b>%30</b>	42,93	22,40	3,56	3,98	5,65	18,87	3,63
<b>%40</b>	45,80	23,80	3,27	4,13	5,45	25,03	3,44
<b>%50</b>	47,04	24,38	2,93	4,47	5,24	30,89	3,26
<b>%60</b>	46,91	24,25	2,57	5,02	4,97	36,26	3,07
<b>%70</b>	45,23	23,31	2,16	5,87	4,66	40,74	2,87

Tablo 14'e bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %50'ye doğru artarken, RMSE'nin artış gösterdiği, bu orandan sonra ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %30'a kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS ve çoklu değer atama yöntemleri tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %30 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi için %15 ve daha yüksek veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.



Tablo 15:  $R^2=0,50$  ve  $n=1000$  için RMSE deęerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Atanması</b>	<b>Alt Sınır/2 Atanması</b>	<b>Alt Sınır Atanması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Deęer Atama</b>
<b>%5</b>	21,93	12,04	4,31	4,31	6,17	3,06	4,23
<b>%10</b>	29,35	15,68	4,15	4,18	6,06	6,22	4,07
<b>%15</b>	34,35	18,14	4,00	4,08	5,95	9,39	3,94
<b>%20</b>	38,05	19,98	3,87	4,03	5,85	12,57	3,84
<b>%25</b>	40,75	21,32	3,72	4,00	5,75	15,73	3,74
<b>%30</b>	42,90	22,38	3,56	3,99	5,64	18,87	3,63
<b>%40</b>	45,76	23,78	3,26	4,13	5,44	25,03	3,44
<b>%50</b>	47,05	24,38	2,93	4,46	5,21	30,89	3,25
<b>%60</b>	46,98	24,29	2,57	5,02	4,94	36,25	3,07
<b>%70</b>	45,21	23,31	2,18	5,91	4,66	40,73	2,89

Tablo 15'e bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %50'ye doğru artarken, RMSE'nin artış gösterdiği, bu orandan sonra ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %30'a kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS ve çoklu değer atama yöntemleri tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %30 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi için %15 ve daha yüksek veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.

Tablo 16:  $R^2=0,50$  ve  $n=5000$  için RMSE deęerleri

<b>Tespit Edilemeyen Veri Oranı</b>	<b>Sıfır Ataması</b>	<b>Alt Sınır/2 Ataması</b>	<b>Alt Sınır Ataması</b>	<b>Tobit Regresyon</b>	<b>ROS</b>	<b>Tespit Edilemeyen Veri Regresyonu</b>	<b>Çoklu Deęer Atama</b>
<b>%5</b>	21,91	12,03	4,30	4,31	6,17	3,06	4,23
<b>%10</b>	29,34	15,67	4,16	4,19	6,06	6,22	4,08
<b>%15</b>	34,34	18,14	4,01	4,09	5,95	9,39	3,95
<b>%20</b>	38,00	19,96	3,86	4,02	5,84	12,57	3,83
<b>%25</b>	40,76	21,32	3,71	3,99	5,74	15,73	3,73
<b>%30</b>	42,90	22,38	3,57	4,00	5,63	18,87	3,63
<b>%40</b>	45,69	23,74	3,26	4,13	5,42	25,02	3,44
<b>%50</b>	46,99	24,35	2,94	4,47	5,20	30,88	3,26
<b>%60</b>	46,88	24,24	2,58	5,04	4,94	36,25	3,08
<b>%70</b>	45,18	23,30	2,18	5,91	4,62	40,71	2,89

Tablo 16'ya bakıldığında sıfır atanması ve alt sınır/2 atanması yöntemleri incelendiğinde tespit edilemeyen veri oranı %5'den %50'ye doğru artarken, RMSE'nin artış gösterdiği, bu orandan sonra ise RMSE'nin azaldığı görülmüştür. Alt sınır atanması yöntemi incelendiğinde; tespit edilemeyen veri oranları artarken RMSE'nin azaldığı görülmüştür. Tobit regresyon yöntemiyle yapılan simülasyonlar sonucunda tespit edilemeyen veri oranı %25'e kadar RMSE'nin kadar azaldığı bu orandan sonra arttığı görülmüştür. ROS ve çoklu değer atama yöntemleri tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülmüştür. Tespit edilemeyen veri regresyon yöntemi tespit edilemeyen veri oranı arttıkça RMSE'nin arttığı görülmüştür.

Yöntemler birbiriyle kıyaslandığında alt sınır ataması, Tobit regresyon ve çoklu atama diğer yöntemlere göre daha düşük RMSE değerlerine sahiptir. Bu üç yöntem içinde alt sınır ataması ve çoklu atama birbirlerine benzerlik gösterirken tespit edilemeyen veri oranı %30 ve üstü durumlarda alt sınır atamasının RMSE değeri daha düşük bulunmuştur. Tobit regresyon yöntemi için %15 ve daha yüksek veri oranları için diğer yöntemlerden daha yüksek RMSE değerine sahiptir.

## 5.TARTIŞMA

Araştırmalarda farklı nedenlere bağılı olarak tespit edilemeyen veriler ortaya çıkmaktadır. Bu sorun, yapılan çalışmaların veri toplanması safhasında verinin toplanmasının güç, maliyetli ve de zahmetli olduğu alanlarda çok daha önemli bir hal almaktadır. Bunların yanı sıra gözlem değerlerinin azalması, çalışmanın güvenilirliğinin azalması hatta bazı analizlerin yapılamaması gibi farklı sorunları da beraberinde getirmektedir. Bu nedenle tespit edilemeye verilerin tahmininde kullanılan bir takım yöntemler geliştirilmiş veya var olan farklı yöntemler bu alanda da uygulanmıştır. Bu konu ile ilgili yapılan çalışmalar çok fazla sayıda bulunmamakla birlikte kullanılan yöntemlerden bazıları yerine değer atama, çoklu atama, Tobit regresyon, ROS ve tespit edilemeyen veri regresyonudur (Hae, 2008; Koru-Sengul T ve ark., 2011).

Helsel (2005), yerine değer atama yönteminin yanlış sonuçlar verdiği konusunda araştırmacıları uyarmıştır. Ayrıca sıfır atanması, alt sınırın yarısının veya alt sınırın atanmasının herhangi bir temele dayanmadığı konusunda uyarmıştır. Ancak yine de bu yöntemlerin kullanıldığı araştırmalar mevcuttur (Hae, 2008; Koru-Sengul T ve ark., 2011).

Rubin (1978), çoklu değer atama yönteminin tekli atama yöntemlerine göre daha avantajlı olduğunu belirtmiştir. Çoklu atama rastgele yapıldığında tahmin değerinin etkililiğini arttırmaktadır. Yanıtsız modeller için birden fazla atama yaptığı için yapılan atamalar birleştirilerek geçerli çıkarımlar elde edilmektedir. Birden fazla model içerisinde tekrarlı atamalar yaptığı için birden fazla tam veri seti elde edilir bundan dolayı hassas çıkarımlar yapılmasını sağlamaktadır (Rubin, 1987). ROS tahminleri log transformasyonundan dolayı yanlıdır. Yapılan simülasyon çalışmaları bu yöntemi kullanmanın küçük bir yanlılığa yol açtığını göstermiştir (Elobeid ve ark., 2008).

Çalışmamıza benzer şekilde çok fazla sayıda yöntemin karşılaştırıldığı kaynak pek bulunmamakla birlikte Hae ve ark.(2008) yaptıkları çalışmalarında tespit edilemeyen veri analizi için karşılaştırdıkları yöntemlerde; tespit edilemeyen veri oranının %30 dan az olduğu durumlarda tobit regresyon analizinin iyi sonuçlar verdiğini ortaya koymuşlardır. Tespit edilemeyen veri oranı %10 dan %70 e doğru

artarken yaptıkları analizler sonucunda tüm tespit edilemeyen veri oranları göz önüne alındığında çoklu değer atama yönteminin en iyi performansı gösterdiğini belirmişlerdir.

Diğer çalışmalardan farklı olarak bu çalışmada veri türetimi belirli  $R^2$  düzeyleri için gerçekleştirilmiş ve yöntemler RMSE'ye göre karşılaştırılmıştır. Farklı  $R^2$  değerlerinde yöntemlerde benzer değişimler söz konusu olmuştur. Aynı  $R^2$  değerlerine sahip her bir yöntem için örneklem büyüklüğü değiştikçe yöntemlerin RMSE değerinin değişmediği görülmüştür.

Çoklu değer atama, ROS ve alt sınır atanması yöntemlerinde tespit edilemeyen veri oranı arttıkça RMSE'nin azaldığı görülürken; sıfır atanması, alt sınır/2 atanması, tobit regresyon ve tespit edilemeyen veri regresyonu yöntemlerinde ise tespit edilemeyen veri oranı arttıkça RMSE'nin genelde arttığı görülmüştür. Karşılaştırdığımız yöntemler farklı  $R^2$  değerlerinde incelendiğinde; çoklu değer atama, alt sınır atanması ve tobit regresyon yöntemlerinin RMSE değerinin diğer yöntemlere göre daha düşük olduğu görülmüştür. Ancak tespit edilemeyen veri oranı arttıkça tobit regresyon yönteminin RMSE değerinin diğer iki yöntemden daha yüksek olduğu bulunmuştur.

Tespit edilemeyen veri oranları açısından bakıldığında uygulamada düşük oranlar ile karşılaşılmaktadır. Tespit edilemeyen veri oranı %15 altında ve  $R^2$ 'nin 0,70'in üstünde olması durumunda çoklu değer atama, alt sınır atanması ve Tobit regresyon yöntemlerinin RMSE değerinin diğer yöntemlere göre daha düşük olduğu görülmüştür. Ancak  $R^2$ 'nin 0,50 olması durumunda ise tespit edilemeyen veri regresyonu diğer yöntemlerden daha düşük RMSE'ye sahiptir.

Sonuç olarak; farklı  $R^2$ , farklı örneklem büyüklüklerinde ve tespit edilemeyen veri oranına göre yapılan analizler sonucunda tespit edilemeyen veri analizi yöntemi olarak ön plana çıkan üç yöntem alt sınır atanması, tobit regresyon ve çoklu değer atama yöntemi birçok senaryoda en düşük RMSE'ye sahip olsa da düşük  $R^2$  sahip model için tespit edilemeyen veri regresyonu da alternatif bir yöntemdir.

## 6.KAYNAKLAR

1. Acock AC (2005) Working With Missing Values. *Journal of Marriage and Family* 67: 1012-1028.
2. Allison PD (2001) Missing Data. *Sage University Papers Series on Quantitative Applications, in the Social Sciences.* 07-136. Thousand Oaks, CA: Sage, pp:97
3. Bal C, Özdamar K (2004) Eksik gözlem sorununun türetilmiş veri setleri yardımıyla çözümlenmesi. *Osmangazi Üniversitesi Tıp Fakültesi Dergisi* 26:68.
4. Bierens, H. (2004) The Tobit model. Working Paper, Pennsylvania State University
5. Cunillera O (2014) Tobit Model, Health Services Research Unit, IMIM-Hospital del Mar, Barcelona, Spain pp:6671-6674.
6. Dodge Y (1985) Analysis of experiments with missing data. John Wiley & Sons Inc, Canada, pp: 499.
7. Elobeid MA, Padilla MA, Mctie T et al (2008) Missing data in randomized clinical trials for weight loss: Scope of the problem, State of the field, and performance of statistical methods. *PLoS ONE* 4(8): 6624.
8. EPA-Environmental Protection Agency (2009) Statistical analysis of groundwater monitoring data at RCRA facilities unified guidance. Environmental protection agency office of resource conservation and recovery press, pp:15
9. Franses PH, Paap R (2004) *Quantitative Models in Marketing Research*, Cambridge University Press, pp:137-157.
10. Getachew AD, Huang Y (2013) Bayesian semiparametric mixture Tobit models with left censoring, skewness, and covariate measurement errors. *Statistics in Medicine* 32: 3881–3898
11. Greene W (2007) *Econometric analysis. Sixth Edition*, Pearson Prentice Hall Inc, New Jersey, pp:849-850.
12. Heckman JJ (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5: 475 – 492.
13. Hae WU, Hartgers FC, Yazdanbakhsh M et al (2008) Evaluation of regression methods when immunological measurements are constrained by detection limits. *BMC Immunology* 9(59):1471-2172.
14. Helsel DR (2005) *Nondetects And Data Analysis: Statistics for censored environmental data. First Edition*, John Wiley and Sons Inc, New York, pp: 250
15. Hintze JL, Utah K (2007) *Nondetects Data Regression Users' Guide III*. Kaysville:NCSS.
16. Huston C, Juarez CE (2009) Guidelines for computing summary statistics for data-sets containing non-detects. *Simon Frase University* pp:60.
17. Koru-Sengul T, Clark JD, Fleming LE, David J Lee DJ. (2011) Toward improved statistical methods for analyzing Cotinine-Biomarker health association data. *Tob Induc Dis* Oct 3;9(1):11.
18. Maddala GS (1989) *Introduction to Econometrics. Second Edition*, Macmillan Publishing Company, New York, pp:472.
19. Rubin DB (1978) Multiple imputations in sample surveys. *American Statistical Association, Proceedings of the Survey Research Methods Section* pp: 20–34.
20. Schafer JL (1997) *Analysis of Incomplete Multivariate Data. First edition*, Chapman & Hall/CRC, Washington, pp:13.

21. Schafer J, Graham J (2002) Missing Data: Our View of the State of Art. *Psychological Methods* 7(2):147-177
22. Schlomer GL, Bauman S, Card NA (2010) Best practices for missing data management in counseling psychology. *Journal of Counselling Psychology* 57: 1-10.
23. Schmidheiny DK (2007) Limited dependent variable models. Lecture Notes in Microeconometrics. Universitat Pompeu Fabra.
24. Tobin J (1958) Estimation of relationships for limited dependent variables. *Econometrica* 26: 24-36.
25. Zorlutuna S, Erilli NA, Yücel B (2016) Lung cancer study with tobit regression analysis: sivas case. *Eurasian Econometrics, Statistics & Empirical Economics Journal* 3: 13-22.





## 7.TEŞEKKÜR

Tez çalışmam süresince, değerli yorumlarıyla, sabrıyla çalışmama yön veren, beraber çalışmaktan büyük keyif aldığım danışmanım Doç.Dr. Güven ÖZKAYA'ya,

Eğitimim boyunca bilgi, tecrübe ve sevgilerini esirgemeyen değerli anabilim dalı hocalarıma,

İtalya, Verona Üniversitesi' ndeki stajım boyunca, değerli katkılarını benimle paylaşan Erasmus danışmanım Dr. Alessandro MARCON'a çok teşekkür ederim.

Bana her zaman güvenen, hep daha iyisini yapmam için yön gösteren, hep yanımda olan, hayatımı kolaylaştıran aileme teşekkür ederim.

Ayrıca tezim, Uludağ Üniversitesi Bilimsel Araştırmalar Projesi UÜ BAP KUAP(T)-2015/46 no'lu proje kapsamında gerçekleştirilmiştir.

## 8. ÖZGEÇMİŞ

1989 yılında Bulgaristan'da doğdum. Lise öğrenimimi 2007 yılında Darıca'da Neşet Yalçın Süper Lisesi'nde tamamladım. Çukurova Üniversitesi İstatistik Bölümü'nden 2013 yılında mezun oldum. 2015 bahar döneminde Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Ana Bilim Dalı'nda yüksek lisans eğitimime başladım. Haziran-Ağustos 2016 tarihleri arasında Verona Üniversitesi Epidemiyoloji ve Medikal İstatistik Ana Bilim Dalı'nda staj yaptım.

