

Çoklu İç İlişki Durumunda Kısmi En Küçük Kareler Regresyonu ve Alternatif  
Yöntemlerle Karşılaştırılması

Yakup Murat Bulut

**YÜKSEK LİSANS TEZİ**

İstatistik Anabilim Dalı

Aralık-2011

Comparison of Partial Least Squares Regression and it's Alternative Methods When  
Multicollinearity Exists

Yakup Murat Bulut

**MASTER OF SCIENCE THESIS**

Department of Statistics

December-2011

Çoklu İç İlişki Durumunda Kısmi En Küçük Kareler Regresyonu ve Alternatif  
Yöntemlerle Karşılaştırılması

Yakup Murat Bulut

Eskişehir Osmangazi Üniversitesi

Fen Bilimleri Enstitüsü

Lisansüstü Yönetmeliği Uyarınca

Uygulamalı İstatistik Bilim Dalında

YÜKSEK LİSANS TEZİ

Olarak Hazırlanmıştır

Danışman: Doç. Dr. Zeki Yıldız

Aralık-2011

## ONAY

İstatistik Anabilim Dalı Yüksek Lisans öğrencisi Yakup Murat Bulut'un YÜKSEK LİSANS tezi olarak hazırladığı "Çoklu İç İlişki Durumunda Kısmi En Küçük Kareler Regresyonu ve Alternatif Yöntemlerle Karşılaştırılması" başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

**Danışman** : Doç. Dr. Zeki Yıldız

**İkinci Danışman** : -----

### **Yüksek Lisans Tez Savunma Jürisi:**

**Üye** : Doç. Dr. Zeki Yıldız

**Üye** : Prof. Dr. Veysel Yılmaz

**Üye** : Doç. Dr. Süleyman Dünder

**Üye** : Yrd. Doç. Dr. Arzu Altın Yavuz

**Üye** : Yrd. Doç. Dr. Fatih Çemrek

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ..... tarih ve ..... sayılı kararıyla onaylanmıştır.

Prof. Dr. Nimetullah BURNAK

Enstitü Müdürü

## ÖZET

Kısmi en küçük kareler regresyonu çoklu iç ilişki durumunda çokça kullanılan istatistiksel yöntemlerden biridir. Çoklu iç ilişki problemi regresyon analizinde sıkça karşılaşılan bir problemdir. Açıklayıcı değişkenler arasında çoklu iç ilişki problemi varsa yapılan çoklu lineer regresyon analizi sonucunda elde edilen regresyon katsayıları durağan olmayan sonuçlar verecektir. Bu durumda ise modelin tahmin gücü oldukça düşük olacaktır. Literatürde çoklu iç ilişki problemi ile başa çıkabilmek için birçok yöntem geliştirilmiştir. Bu yöntemlerden birisi ise klasik en küçük kareler tahmin edicisini kullanmak yerine yanlı regresyon tekniklerini kullanmaktır.

Bu çalışmada çoklu iç ilişki problemi ile karşılaşıldığı durumlarda kullanılan yanlı regresyon yöntemlerinden birisi olan kısmi en küçük kareler regresyon yöntemi tanıtılmıştır. Ayrıca kısmi en küçük kareler regresyonu (PLSR) yine aynı amaca hizmet eden temel bileşenler regresyonu (PCR) ve ridge regresyon (RR) ile karşılaştırılmıştır. Yapılan karşılaştırma sonucunda tahmin gücü bakımından Ridge regresyon yönteminin en iyi sonucu verdiği gözlemlenmiştir. PLSR ve PCR yöntemleri birbirleriyle karşılaştırıldığında PLSR yöntemi PCR yöntemine göre daha az bileşenle model kurmuştur. Bu yüzden de PLSR yöntemi PCR yöntemine göre daha başarılı bulunmuştur.

**Anahtar Kelimeler:** Kısmi En Küçük Kareler Regresyonu, Temel Bileşenler Regresyonu, Ridge Regresyon, Çoklu İç İlişki

## SUMMARY

Partial least square regression is one of the statistical method which is used greatly in multicollinearity. Multicollinearity problem occurs frequently in regression analysis. If there is a multicollinearity problem in data set, regression coefficients which were found with linear regression analysis give non-stable results. In this state model's estimation power will be very low. In literature lots of methods are developed to cope with multicollinearity problem. One of that methods is using biased regression techniques rather than least square estimators.

In this study partial least square regression method is introduced which is one of the biased regression methods that used when multicollinearity problem occurs. Also partial least square regression compared with principal components regression and ridge regression. As a result of this compare it is found that ridge regression gives best result about estimation power. When PLSR and PCR methods are compared with each other PLSR method obtains a model with less components than PCR method. So that PLSR method is found more successful rather than PCR method.

**Keywords:** Partial Least Squares Regression, Principal Component Regression, Ridge Regression, Multicollinearity

## TEŞEKKÜR

Bu tez çalışmasının her aşamasında değerli destek ve yardımlarını esirgemeyen danışman hocam Doç. Dr. Zeki Yıldız'a teşekkürü borç bilirim.

İstatistik alt yapımın oluşmasında bana yardımcı olan ihtiyaç duyduğum her anda sorularıma yılmadan cevap veren değerli hocalarım Prof. Dr. Olcay Arslan'a, Doç. Dr. Güzin Yüksel'e ve Yrd. Doç. Dr. Arzu Altın Yavuz'a sonsuz teşekkürlerimi sunarım.

Her zaman yanımda olduklarını bildiğim değerli arkadaşlarım Yonca Ücal, Yusuf Kuvvetli ve Talha Arslan'a verdikleri desteklerden dolayı teşekkürlerimi sunarım.

Ayrıca eğitim hayatım boyunca maddi ve manevi olarak bana destek olan aldığım her kararı sorgulamadan arkamda duran babam Ahmet Bulut'a ve annem Nevriye Bulut'a sonsuz teşekkürler.

## İÇİNDEKİLER

	<u>Sayfa</u>
<b>ÖZET .....</b>	<b>v</b>
<b>SUMMARY .....</b>	<b>vi</b>
<b>TEŞEKKÜR .....</b>	<b>vii</b>
<b>İÇİNDEKİLER .....</b>	<b>viii</b>
<b>ŞEKİLLER DİZİNİ .....</b>	<b>xi</b>
<b>ÇİZELGELER DİZİNİ .....</b>	<b>xii</b>
<b>SİMGELER VE KISALTMALAR DİZİNİ .....</b>	<b>xiii</b>
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. ÇOKLU LİNEER REGRESYON .....</b>	<b>4</b>
2.1. Çoklu Lineer Regresyon Modeli.....	4
2.2. En Küçük Kareler ve Normal Denklemler ile Elde Edilişi.....	6
2.2.1. En Küçük Kareler Tahmin Edicisinin Özellikleri.....	7
2.2.2. . En Küçük Kareler Tahmin Edicisinin Varsayımları.....	8
2.2.2.1. Normal dağılım varsayımı .....	9
2.2.2.2. Ortak varyans varsayımı .....	9
2.2.2.3. Hataların ilişkisiz olması varsayımı.....	10
<b>3. ÇOKLU İÇ İLİŞKİ PROBLEMİ.....</b>	<b>11</b>
3.1. Çoklu İç İlişkinin Kaynakları .....	11
3.2. Çoklu İç İlişkinin Derecesi .....	12
3.2.1. Tam çoklu iç ilişki .....	13
3.2.2. Tama yakın çoklu iç ilişki.....	13
3.3. Çoklu İç İlişkinin Belirlenmesi.....	14
3.3.1. Korelasyon matrisinin incelenmesi.....	14



## İÇİNDEKİLER (devam)

	<u>Sayfa</u>
3.3.2. Varyans şişirme faktörü (VIF).....	14
3.3.3. $X'X$ matrisinin özdeğerlerinin incelenmesi.....	15
3.4. Çoklu İç İlişkinin Giderilmesi .....	16
<b>4. YANLI REGRESYON .....</b>	<b>17</b>
4.1. Ridge Regresyon.....	20
4.1.1. Ridge tahmin edicisinin en küçük kareler tahmin edici ile karşılaştırılması .....	21
4.1.2. Ridge tahmin edicisinin yanlılık değerinin hesaplanması. ....	22
4.1.3. Ridge tahmin edicisinin hata kareler ortalamasının hesaplanması .....	24
4.1.4. Ridge tahmin edicisinde yanlılık parametresi $k$ 'nın seçimi.....	26
4.1.4.1. Ridge izi yöntemi.....	26
4.1.4.2. Analitik yöntemler .....	27
4.2. Temel Bileşenler Regresyonu.....	28
4.2.1. Temel bileşenler regresyonunda bileşenlerin seçimi için yaklaşımlar .....	34
4.3. Kısmi En Küçük Kareler.....	37
4.3.1. Kısmi en küçük karelerin elde edilişi .....	37
4.3.2. Kısmi En Küçük Kareler Regresyonu .....	41
4.3.3. Kısmi En Küçük Kareler Regresyonunda Kullanılan Algoritmalar .....	45
4.3.3.1. Lineer olmayan yinelemeli en küçük kareler algoritması (NIPALS) .....	46
4.3.3.2. PLS yönteminin istatistiksel olarak esinlenilmiş değişikliğinin basit bir uygulaması (SIMPLS) .....	51
4.3.3.3. Standart PLS algoritması ile SIMPLS algoritmasının karşılaştırılması.....	57
4.3.3.4. Evrensel kısmi en küçük kareler algoritması (UNIPALS) .....	57
4.3.3.5. Çekirdek algoritması.....	58

**İÇİNDEKİLER (devam)****Sayfa**

4.3.3.6. Örnek-uzaklık kısmi en küçük kareler algoritması (SAMPLS) .....	59
4.3.3.7. Özvektör algoritması.....	60
4.3.4. Kısmi en küçük kareler regresyonunda kullanılan ideal bileşen sayısının belirlenmesi.....	61
4.3.5. Kısmi en küçük kareler regresyonunun temel bileşenler regresyonun ve ridge regresyon ile ilişkisi .....	64
<b>5. UYGULAMA .....</b>	<b>65</b>
5.1. Kısmi En Küçük Kareler Regresyonu .....	66
5.2. Temel Bileşenler Regresyonu.....	68
5.3. Ridge Regresyon.....	69
5.4. PLSR, PCR ve RR Yöntemlerinin Karşılaştırılması .....	70
<b>6 SONUÇ VE TARTIŞMA .....</b>	<b>72</b>
<b>KAYNAKLAR DİZİNİ .....</b>	<b>73</b>

## ŞEKİLLER DİZİNİ

<u>Şekil</u>		<u>Sayfa</u>
Şekil 4.1	Yanlı tahmin edicinin yansız tahmin ediciden daha küçük MSE değerine sahip olduğunun gösterilmesi. ....	18
Şekil 4.2	Hata kareler ortalaması fonksiyonları .....	25
Şekil 4.3	Çoklu lineer regresyon yöntemi olarak PLS .....	42
Şekil 4.4	PLS'nin matris gösterimi .....	45
Şekil 4.5	$\ F_H\ $ değerine karşı bileşen sayısının grafiği .....	62
Şekil 4.6	PRESS değerine karşı bileşen sayısının grafiği .....	62
Şekil 5.1	RMSECV değerine karşı gizil değişken sayısının grafiği .....	67
Şekil 5.2	RMSECV değerine karşı temel bileşenlerin sayısının grafiği .....	68
Şekil 5.3	Ridge izi grafiği.....	70

## ÇİZELGELER DİZİNİ

<u>Çizelge</u>		<u>Sayfa</u>
Çizelge 4.1	Düzeltilmiş PLS algoritmasının temel konsepti .....	54
Çizelge 5.1	Hesaplanan özdeğerler ve VIF değerleri .....	65
Çizelge 5.2	PLSR sonucunda açıklanan varyans değerleri .....	67
Çizelge 5.3	PCR sonucunda açıklanan varyans değerleri .....	69
Çizelge 5.4	PLSR, PCR ve RR yöntemleri için RMSECV değerleri.....	70

## SİMGELER ve KISALTMALAR DİZİNİ

BLUE	En iyi yansız tahmin edici
CV	Çapraz geçerlilik
EKK	En küçük kareler
EÇO	En çok olabilirlik
MSE	Hata kareler ortalaması
NIPALS	Lineer olmayan kısmi en küçük kareler algoritması
PC	Temel bileşen
PCR	Temel bileşenler regresyonu
PLS	Kısmi en küçük kareler
PLS1	Tek değişkenli kısmi en küçük kareler algoritması
PLSR	Kısmi en küçük kareler regresyonu
PRESS	Tahmin edilen hata kareler toplamı
RMSECV	Hata kareler ortalamasının karekökü
RR	Ridge regresyon
SAMPLS	Örnek-uzaklık kısmi en küçük kareler algoritması
SIMPLS	PLS yönteminin istatistiksel olarak esinlenilmiş değişikliğinin basit bir uygulaması
SVD	Tekil değer ayrışımı
UNIPALS	Evrensel kısmi en küçük kareler algoritması
VIF	Varyans şişirme faktörü

## BÖLÜM 1

### GİRİŞ

Birçok lineer regresyon ve tahmin probleminde açıklayıcı değişkenler çok yüksek derecede ilişkili olabilir. Bu durum genelde çoklu iç ilişki olarak bilinir. Çoklu iç ilişki problemi yapılan analizler sonucunda elde edilen en küçük kareler tahmin edicisinin varyans değerlerinin olması gerekenden büyük olmasına ve regresyon katsayılarının gerçek değerlerinden uzaklaşmasına neden olmaktadır. Çoklu iç ilişki durumunda en küçük kareler tahmin edicisi regresyon katsayıları için güvenilir olmayan sonuçlar vermektedir (Gunst and Mason, 1980). Bu problemle başa çıkabilmek için birçok yöntem geliştirilmiştir. Bu yöntemler arasında en çok kullanılanlar Temel Bileşenler Analizi, Ridge Regresyon ve Gizil Kök Regresyonudur. Bu tahmin yöntemleri birçok durumda başarılı sonuçlar vermişlerdir. Ayrıca bu tahmin yöntemlerinin performanslarını karşılaştırmak için bir çok çalışma yapılmıştır. Son birkaç yıl içinde kemometri literatüründe çoklu iç ilişki problemi ile başa çıkmak için yeni bir tahmin yöntemi görülmeye başlanmıştır. Bu tahmin yöntemi kısmi en küçük kareler tahmin yöntemidir. Bu yöntem 1960'lı yılların sonlarına doğru Herman Wold tarafından geliştirilmiştir. Kısmi en küçük kareler yöntemi temel bileşenler analizi ve çoklu lineer regresyonun birleştirilmesi ve genelleştirilmesiyle elde edilen bir yöntemdir. Bu yöntemin amacı bağımlı değişkenleri açıklayıcı değişkenler yardımıyla tahmin etmek ya da analiz etmektir. Bu analizde açıklayıcı değişkenlerden elde edilen ve gizil değişkenler olarak adlandırılan ortogonal faktörler yardımıyla regresyon yapılır. Elde edilen bu ortogonal faktörlerin en iyi tahmin yeteneğine sahip oldukları belirlenmiştir (Abdi, 2007). Bu özelliği sayesinde kısmi en küçük kareler yöntemi kemometride kimyasal veri problemlerinin çözümünde standart olarak kullanılan bir yöntem haline gelmiştir. Kısmi en küçük karelerin kemometrideki bu başarılı sonuçları nedeniyle yöntemin kökeni sosyal bilimler özellikle de ekonomi olmasına karşın gıda araştırmaları, tıp, farmakoloji ve psikolojinin de dahil olduğu bir çok bilimsel alanda başarılı bir şekilde kullanılmaya başlanmıştır (Frank and Kowalski, 1984; Haaland and Thomas, 1988; Toscas et. al., 1999; McIntosh and Lobaugh, 2004; Nadler and Coifman, 2005; Carrascal et. al., 2009; Alın et. al., 2009;). Kısmi en küçük kareler yönteminin kimyasal uygulamalarda kullanılmasının öncüsü ise S. Wold ve H. Martens'in araştırma

grubudur. Bu grup yöntemi kimyasal problemlere uygulamaya ilk olarak 1970'li yılların sonunda başlamışlardır. Daha sonraları ise Kowalski bu uygulamalara devam etmiştir (Geladi and Kowalski, 1986). Wold ve Otto yaptıkları çalışmaların sonunda kısmi en küçük karelerin klasik çoklu lineer regresyon yöntemlerine iyi bir alternatif olduğunu göstermişlerdir. Kısmi en küçük karelerin diğer çoklu lineer regresyon yöntemlerine iyi bir alternatif olmasının en önemli nedeni daha sağlam bir tahmin edici olmasıdır. Sağlam olmanın anlamı ise ana kitleden alınan yeni bir örneklem için model parametrelerinin çok fazla değişmemesidir. Bağımlı değişken sayısının tek olduğu durumlarda kısmi en küçük kareler algoritması regresyon problemi için Paige ve Saunders tarafından geliştirilen tekil değer ayrışımına (singular value decomposition) eşittir (Wold, 1993).

Kısmi en küçük kareler regresyonu açıklayıcı ve bağımlı değişkenler arasındaki lineer ilişkiyi modellemek için kullanılan yanlı bir yöntemdir. Bu yüzden kısmi en küçük kareler doğal olarak regresyon problemlerine genişletilebilir. Açıklayıcı ve bağımlı değişkenlerin her birisi değişkenler bloğu olarak düşünülebilir. Kısmi en küçük kareler yeni açıklayıcı değişkenler olarak kullanılacak skor vektörlerini hesaplar ve daha sonra bu skor vektörlerini kullanarak bağımlı değişkene ilişkin regresyon denklemini kurar. Skor vektörleri hesaplanırken açıklayıcı değişkenler ile bağımlı değişkenler arasındaki doğal asimetri yansıtılır. Kısmi en küçük kareler regresyonu istatistikçiler tarafından göz ardı edilmiş ve istatistiksel bir model olmasından ziyade bir algoritma olarak düşünülmüştür. Son yıllarda kısmi en küçük karelerin istatistiksel özellikleriyle ilgilenenlerin sayısı artmaya başlamıştır (Lorber et. al.,1987). Kısmi en küçük karelerin değişkenlik ve daraltıcı özellikleri açısından etkinliği teorik olarak çalışılmıştır (Lingjærde and Christophersen, 2000; Goutis, 1996; Butler and Denham, 2000). Ayrıca kısmi en küçük karelerin performansı bir çok simülasyon çalışmasında ortaya konulmuştur (Naes and Martens, 1986; Geladi and Kowalski, 1986a; Frank, 1987; Frank and Friedman, 1993).

Bu tez çalışmasında amaç kısmi en küçük kareler regresyonunu ve bu yöntemde kullanılan algoritmaları tanıtmaktır. Ayrıca bu yöntem ile aynı amaca hizmet eden yanlı regresyon yöntemleri olan Temel bileşenler regresyonu ve Ridge regresyonu ile arasındaki ilişkiden bahsedilecektir.

Tez çalışmasının birinci bölümünde kısmi en küçük kareler yönteminin tarihsel gelişiminden bahsedilmiştir. İkinci bölümünde ise çoklu lineer regresyon modeli, en küçük kareler yöntemi ve bu yöntemin varsayımlarından bahsedilmiştir. Üçüncü bölümde regresyon analizinde en sık karşılaşılan problemlerden birisi olan çoklu iç ilişki probleminin ne olduğu, kaynakları, belirlenmesi ve bu problem ile başa çıkma yolları açıklanmıştır. Dördüncü bölümde çoklu iç ilişki problemi ile başa çıkmak için en küçük kareler yöntemine alternatif olarak geliştirilen yanlı regresyon yöntemlerinden kısmi en küçük kareler regresyonu, temel bileşenler regresyonu ve ridge regresyon yöntemleri ele alınmıştır. Beşinci bölümde ise, kısmi en küçük kareler regresyonu, temel bileşenler regresyonu ve ridge regresyon yöntemleri literatürden alınmış bir veri seti üzerinde tahmin güçleri bakımından karşılaştırılmıştır.



## BÖLÜM 2

### ÇOKLU LİNEER REGRESYON

Regresyon analizi değişkenler arasındaki ilişkiyi inceleyen ve modelleyen istatistiksel bir tekniktir. İlişki, bir veya daha fazla açıklayıcı değişken ve bağımlı değişkenle bir denklem formunda ifade edilir. Regresyon denklemi bir bağımlı değişken ve birden fazla açıklayıcı değişken içeriyor ise çoklu regresyon olarak adlandırılır.

Bu bölümde çoklu lineer regresyon modeli açıklanacak ve normal denklemler elde edilecektir, daha sonra herhangi sayıda açıklayıcı değişken içeren lineer modeller için normal denklemleri çözülecektir.

#### 2.1. Çoklu Lineer Regresyon Modeli

Bir tane bağımlı değişken ve  $p$  tane açıklayıcı değişken içeren lineer regresyon modeli genel olarak aşağıdaki gibidir.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i \quad (1)$$

$n$  örneklem büyüklüğünü,  $p$  açıklayıcı değişken sayısını göstermektedir. Modeldeki  $\beta_0$  sabit terimi ile birlikte tahmin edilecek toplam  $(p + 1)$  tane parametre vardır. Gösterimde sadelik olması açısından  $(p + 1) = p'$  olarak alınacaktır.

Burada,  $\mathbf{Y}$ ;  $(n \times 1)$  boyutlu bağımlı değişken vektörü,  $\mathbf{X}$ ; 1'inci sütunu 1'lerden oluşan  $(n \times p')$  boyutlu açıklayıcı değişkenler matrisi,  $\boldsymbol{\beta}$ ;  $(p' \times 1)$  tipinde tahmin edilecek parametrelerin vektörü,  $\boldsymbol{\varepsilon}$ ;  $(n \times 1)$  tipinde rassal hataların vektörüdür.

(1)'deki modelin matris gösterimi (2) şeklinde

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

veya

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & X_{23} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \dots & X_{np} \end{bmatrix}_{n \times p'} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p' \times 1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n \times 1}$$

şeklinde yazılabilir.

Burada  $\mathbf{X}$  matrisinin her bir sütunu açıklayıcı değişkenlerin belirli değerlerini içerir.  $\mathbf{X}$  matrisinin ilk sütununu  $\beta$  ile çarpıp sonra  $\varepsilon$  hata vektörünün ilk terimi eklenirse ilk gözlem için model aşağıdaki şekilde elde edilir.

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \dots + \beta_p X_{1p} + \varepsilon_1$$

Burada  $\mathbf{Y}$  ve  $\varepsilon$  vektörleri rassal vektörlerdir.  $\mathbf{X}$  matrisi bilinen sabitlerin matrisi olarak düşünülebilir. Burada  $\mathbf{X}$  matrisinin tam ranklı olduğu düşünüldüğünde modele tam ranklı model denir. Ayrıca buradaki  $\beta$  verilerden tahmin edilmeye çalışılan bilinmeyen sabitlerin vektörüdür.  $\beta_j$  parametresi  $X_i$  ( $i \neq j$ ) açıklayıcı değişkeni sabit iken  $X_j$ 'deki 1 birimlik değişimin  $y$ 'de ne kadarlık bir değişime neden olduğunu gösterir.

Burada aşağıdaki 3 durum gözlemlenebilir.

- (1)  **$p > n$  durumu:** Bu durumda açıklayıcı değişken sayısı gözlem sayısından büyüktür. Bu durumla karşılaşıldığında regresyon katsayıları için sonsuz sayıda çözüm vardır.
- (2)  **$p = n$  durumu:** Bu durumda gözlem sayısı açıklayıcı değişken sayısına eşittir. Uygulamada genellikle bu durumla karşılaşılmaz. Eğer  $\mathbf{X}$  tam ranklı bir matris ise regresyon katsayıları için tek bir çözüm vardır. Bu durum  $\varepsilon = \mathbf{y} - \mathbf{X}\beta = \mathbf{0}$  yazmamıza olanak sağlar.
- (3)  **$p < n$  durumu:** Bu durumda açıklayıcı değişken sayısından daha çok gözlem değeri vardır. Bu regresyon katsayıları için tam bir çözüme izin vermez. Fakat  $\varepsilon$  hata vektörünün uzunluğu minimize edilerek bir çözüm elde edilebilir.

Genellikle regresyon analizinde  $\varepsilon_i$  hata terimlerinin birbirinden bağımsız, 0 ortalamalı ve  $\sigma^2$  varyanslı normal dağılıma uydukları varsayılır. Varsayımda olduğu gibi  $\varepsilon_i$  'lerin her biri diğerleri ile bağımsız ise her  $i \neq j$  için  $\varepsilon_i$  ile  $\varepsilon_j$  arasındaki kovaryans sıfırdır.  $\varepsilon_i$  'ler varsayıldığı gibi bağımsız ve  $N(0, \sigma^2)$  dağılımına sahip rassal değişkenler olduğundan;

- 1)  $Y_i$  'ler  $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$  ortalamalı ve  $\sigma^2$  varyanslı normal rassal değişkenlerdir.
- 2)  $Y_i$  'ler birbirlerinden bağımsızdırlar.

Her  $i \neq j$  için  $Y_i$  ile  $Y_j$  arasındaki kovaryans sıfırdır.  $Y_1, Y_2, \dots, Y_n$  'in ortak olasılık yoğunluk fonksiyonu;

$$(2\pi)^{-n/2} \sigma^{-n} e^{-\sum [Y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})]^2 / 2\sigma^2} \quad (3)$$

şeklinde dir. Hipotez testleri ve parametrelerin güven aralığı tahminleri normal dağılım varsayımına dayanmaktadır. Bundan dolayı,  $\varepsilon_i$  'lerin normallik varsayımı kritik öneme sahiptir. Varsayımları sağlandığı durumlarda en küçük kareler tahmin edicileri bütün lineer yansız tahmin ediciler arasında minimum varyansa sahip en iyi tahmin edicilerdir. Normallik varsayımı sağlandığında EKK tahmin edicisi ile en çok olabilirlik (EÇO) tahmin edicisi aynı sonucu vermektedir. Bunun nedeni EÇO yönteminde  $L(\beta; x_1, x_2, \dots, x_n)$  şeklinde gösterilen olabilirlik fonksiyonunu maksimize eden değer EKK'daki hata kareler toplamını minimize eden değere eşit çıkmasıdır (Graybill and Iyer, 1961).

## 2.2. En Küçük Kareler ve Normal Denklemler ile Elde Edilişi

En küçük kareler tahmin yönteminde  $y$  bağımlı rassal değişkeni ile tahmini olan  $\hat{y}$  değerleri arasındaki farkların kareleri toplamı minimum yapılarak  $\beta$  parametre vektörü aşağıdaki şekilde bulunur.

$$y = X\hat{\beta} + e \Rightarrow e = y - X\hat{\beta}$$

$$e'e = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

olur.  $\widehat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y}$ ,  $1 \times 1$  boyutlu bir sabittir.

$$\widehat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} = (\widehat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y})' = \mathbf{y}' \mathbf{X} \widehat{\boldsymbol{\beta}}$$

olduğundan,

$$\mathbf{e}' \mathbf{e} = \mathbf{y}' \mathbf{y} - 2\mathbf{y}' \mathbf{X} \widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}' (\mathbf{X}' \mathbf{X}) \widehat{\boldsymbol{\beta}} = \mathbf{y}' \mathbf{y} - 2(\mathbf{X}' \mathbf{y})' \widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}' (\mathbf{X}' \mathbf{X}) \widehat{\boldsymbol{\beta}}$$

olarak elde edilir. Buradan;

$$\frac{\partial \mathbf{e}' \mathbf{e}}{\partial \widehat{\boldsymbol{\beta}}} = -2\mathbf{X}' \mathbf{y} + (\mathbf{X}' \mathbf{X}) \widehat{\boldsymbol{\beta}} + (\mathbf{X}' \mathbf{X})' \widehat{\boldsymbol{\beta}} = -2\mathbf{X}' \mathbf{y} + 2(\mathbf{X}' \mathbf{X}) \widehat{\boldsymbol{\beta}}$$

eşitliği elde edilir. Bu türevin sıfıra eşitlenmesi ile,

$$-2\mathbf{X}' \mathbf{y} + 2(\mathbf{X}' \mathbf{X}) \widehat{\boldsymbol{\beta}} = \mathbf{0} \quad (4)$$

eşitliği elde edilir. (4) nolu eşitlik yardımıyla;

$$\mathbf{X}' \mathbf{X} \widehat{\boldsymbol{\beta}} = \mathbf{X}' \mathbf{y}$$

denklemini bulunur. Bu denkleme en küçük kareler normal denklemini denir (Montgomery and Peck, 2001). Bu denklemin her iki tarafını soldan  $(\mathbf{X}' \mathbf{X})^{-1}$  ile çarparsak  $\boldsymbol{\beta}$ 'nin en küçük kareler tahmin edicisi

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$$

olarak bulunur.

### 2.2.1. En küçük kareler tahmin edicisinin özellikleri

1) En küçük kareler tahmin edicisi normallik varsayımı altında yansız bir tahmin edicidir. Yani;  $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$  dir.

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}) &= E[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}] \\ &= E[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\ &= E[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \boldsymbol{\beta} + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varepsilon}] \end{aligned}$$

$$\begin{aligned}
&= E(\boldsymbol{\beta}) + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}] \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta}
\end{aligned}$$

2) Gauss Markov teoremine göre en iyi yansız tahmin edicidir (BLUE). Yani diğer lineer tahmin ediciler arasında minimum varyansa sahiptir.

$$\text{cov}(\widehat{\boldsymbol{\beta}}) = \text{cov}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{cov}(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

$\text{cov}(\widehat{\boldsymbol{\beta}})$ ,  $(k+1) \times (k+1)$  tipinde simetrik bir matris olup bu matrisin  $j$ 'inci köşegen elemanı  $\widehat{\boldsymbol{\beta}}_j$ 'nin varyansını,  $ij$ 'inci elemanı  $\widehat{\boldsymbol{\beta}}_i$  ile  $\widehat{\boldsymbol{\beta}}_j$  arasındaki kovaryansı verir.

$\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$  alınırsa;

$$\text{Var}(\widehat{\boldsymbol{\beta}}_j) = \sigma^2 c_{jj} \text{ ve } \text{cov}(\widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\beta}}_j) = \sigma^2 c_{ij}$$

olarak elde edilir.

Yansızlık ve minimum varyansa sahip olmanın yanı sıra etkinlik, tutarlılık, yeterlilik ve dayanıklılıkta iyi bir tahmin edicide aranan özelliklerdir.

### 2.2.2. En küçük kareler tahmin edicisinin varsayımları

En küçük kareler tahmin yönteminin kullanılabilmesi için aşağıdaki varsayımların sağlanması gerekmektedir. Bu varsayımlar;

- 1) Model parametreleri lineerdir.
- 2) Hatalar istatistiksel olarak birbirinden bağımsızdır.
- 3) Hataların beklenen değeri sıfırdır.
- 4) Bağımsız değişkenler arasında ilişki olmamalıdır. Yani çoklu iç ilişki problemi olmamalıdır.
- 5) Bağımsız değişkenler ölçüm değişkenleridir.
- 6) Hatalar sabit varyanslıdır.
- 7) Hatalar normal dağılıma sahiptir.

Bu varsayımlardan ilk 5'i sağlanıyorsa en küçük kareler tahmin edicisi yansızdır. Eğer bunlara ek olarak 6'ncı varsayımda sağlanıyor ise en küçük kareler tahmin edicisi yansız tahmin ediciler arasında minimum varyansa sahiptir.

#### **2.2.2.1.Normal dağılım varsayımı**

Normallik varsayımı hataların normal dağıldığını varsayar. Fakat bu varsayım parametrelerin tahmininde ve toplam değişimin bölümlendirilmesinde çok önemli bir varsayım değildir. Normallik varsayımı, sadece parametrelerin güven aralıklarının oluşturulmasında ve anlamlılığının test edilmesinde kullanılan bir varsayımdır. t testi, F testi ve ki-kare testlerinde rasgele değişkenlerin normal dağıldığı varsayımına ihtiyaç vardır. Ayrıca, student-t dağılımından dolayı geleneksel güven aralığı tahminleri normal dağılım varsayımına bağlıdır. Normallik varsayımının sağlanmamasının en küçük karelere etkisi normallikten ayrılışın derecesine bağlıdır. Normallik varsayımı sağlanmadığı durumlarda eğer diğer varsayımlar sağlanıyor ise en küçük kareler tahmin edicisi yine en iyi yansız tahmin edicidir (Rawlings et. al., 1998).

#### **2.2.2.2.Ortak varyans varsayımı**

Ortak varyans varsayımı klasik en küçük kareler yönteminde önemli bir rol oynamaktadır. Ortak varyans varsayımında bağımlı değişkendeki her bir gözlem aynı miktarda bilgi içermektedir. Sonuç olarak, klasik en küçük karelerde yer alan bütün gözlemler aynı ağırlığa sahiptir. Diğer yandan ortak varyansın sağlanmaması durumunda bazı gözlemler diğer gözlemlere göre daha fazla bilgi içermektedir. Verilerin akla yakın kullanımında bilginin çoğunluğunu içeren bu gözlemlere daha fazla ağırlık verilebilir.

En küçük kareler tahmin edicisinin minimum varyans özelliği direkt olarak bu varsayımın sağlanmasına bağlıdır. Eğer varyanslar eşit değilse ve gözlemlere eşit ağırlık verilirse EKK yöntemi parametrelerin minimum varyans tahminlerini vermez (Rawlings et. al., 1998).

### **2.2.2.3.Hataların ilişkisiz olması varsayımı**

Hatalar arasındaki ilişki bir çok farklı durumdan kaynaklanabilmektedir. İlişkili hatalar genellikle zaman serilerinde yaygındır. Hemen hemen sürekli olan her fiziksel süreçte seri korelasyon görülecektir. Biyolojik çalışmalardaki tekrarlı ölçümler aynı birey üzerinde farklı zamanlarda yapılır. Örneğin bitki ve hayvanların büyümesi ile ilgili çalışmalarda ya da klinik deneylerde genellikle ilişkili hatalar olacaktır.

Çoklu lineer regresyon analizinde en sık karşılaşılan problemlerin başında otokorelasyon ve çoklu iç ilişki gelmektedir. Bu tez çalışmasında çoklu iç ilişki problemi ile karşılaşılması durumunda kullanılan parametre tahmin yöntemlerinden bazıları anlatılacak ve bu yöntemlerin karşılaştırılması için uygulama çalışması yapılacaktır.

## BÖLÜM 3

### ÇOKLU İÇ İLİŞKİ PROBLEMİ

Çoklu iç ilişki kavramı ilk kez Ragnar Frisch tarafından 1934 yılında tanımlanmıştır. Çoklu iç ilişki problemi son yıllarda çoklu regresyon analizinin ciddi sorunlarından birisidir (Frisch, 1934).

Çoklu lineer regresyon modelinin varsayımlarından birisi sütunları açıklayıcı değişkenlerden oluşan ( $n \times p$ ) tipindeki  $\mathbf{X}$  veri matrisinin rankının  $p$ 'ye eşit olması ve bunun gözlem sayısı olan  $n$ 'den küçük olmasıdır.  $\mathbf{X}$  matrisinin rankının  $p$  olmasının anlamı bu matrisin sütunlarının bağımsız olmasıdır. Yani diğer bir ifade ile açıklayıcı değişkenlerin birbirleriyle ilişkisiz olduğudur. Eğer  $\mathbf{X}$  matrisinin rankı  $p$  ise ( $p \times p$ ) tipindeki  $\mathbf{X}'\mathbf{X}$  matrisinde rankı  $p$ 'ye eşittir. Yani bu matrisin tüm sütunları birbirinden bağımsızdır. Bu durumda  $\mathbf{X}'\mathbf{X}$  matrisinin tersi alınabilir ve parametreleri en küçük kareler (EKK) yöntemi ile tahmin edilebilir.

Çoklu iç ilişki probleminin ortaya çıkması durumunda EKK tahmin edicisinin varyansı yüksek ve parametre tahminlerinin işaretleri yanlış hesaplanabilir. Bu durumda elde edilen EKK tahminleri güvenilmez olmaktadır. Çoklu iç ilişki probleminin ortaya çıktığı durumlarda EKK tahmin edicisine alternatif olarak bazı yanlı tahmin ediciler önerilmiştir. Bu tahmin edicilerden bazıları ilerleyen bölümlerde ayrıntılı olarak incelenecektir.

Bu bölümde çoklu iç ilişkinin kaynakları, çoklu iç ilişkinin derecesi, çoklu iç ilişkinin belirlenmesi, çoklu iç ilişkinin etkileri ve çoklu iç ilişkinin giderilmesi konuları incelenecektir.

#### 3.1. Çoklu İç İlişkinin Kaynakları

Çoklu iç ilişkinin birçok kaynağı olabilir. Bu kaynaklardan bazıları uygulanan veri toplama metodu, modeldeki veya kitledeki zorluklar, modelin belirlenmesi ve aşırı tanımlanmış modeldir (Montgomery and Peck, 2001).



### **Uygulanan Veri Toplama Metodu**

Analiz yapmak için örneklem alınan bölgedeki verilerin birbirleri ile ilişkili olması durumunda çoklu iç ilişki problemi ortaya çıkar. Örneğin, elektrik tüketimi ile ilgili olarak ailenin gelirinin ve yaşanılan evin büyüklüğünün elektrik tüketimi üzerinde etkisi araştırılmak istendiğinde ailenin geliri ile evin büyüklüğü açıklayıcı değişkenleri arasında ilişki olduğu görülür. Çoklu iç ilişkinin nedeni geliri yüksek olan ailelerin daha büyük evlerde oturuyor olmasıdır (Montgomery and Peck, 2001).

### **Modeldeki Veya Kitledeki Zorluklar**

Kullanılan veri toplama tekniğinden bağımsız olarak, model veya açıklayıcı değişkenler üzerindeki kısıtlamalar çoklu iç ilişkiye neden olabilir.

### **Modelin Belirlenmesi**

İki yada daha fazla açıklayıcı değişkenin birbirleriyle yüksek derecede ilişkili olduğu durumlarda ortaya çıkar. Bu gibi durumlarda açıklayıcı değişkenlerin tamamının modelde yer alması iç ilişkiyi güçlendirir. Ayrıca  $X$  değişkenlerinin değişim aralığı küçük iken bir regresyon modeline bu değişkenlerin fonksiyonlarının eklenmesi de çoklu iç ilişkiye neden olabilir (Özkale, 2007).

### **Aşırı Tanımlanmış Model**

Aşırı tanımlanmış modeller gözlem sayısından daha çok açıklayıcı değişkenin olduğu modellerdir. Bu modellerde çoklu iç ilişkiye neden olur. Genellikle ekonometrik modeller ve tıbbi modeller aşırı tanımlanmış modellerdir. Bu gibi modellerde bazı değişkenlerin modelden çıkarılarak daha az değişken ile çalışılması gerekmektedir (Montgomery and Peck, 2001).

### **3.2. Çoklu İç İlişkinin Derecesi**

Çoklu iç ilişki problemi açıklayıcı değişkenlerin birbirleriyle ilişkisiz olması varsayımının sağlanmadığı durumlarda ortaya çıkmaktadır. Bu varsayımdan ne kadarlık bir sapma olduğu ve bu sapmanın ne gibi sonuçlar ortaya çıkaracağı çoklu iç ilişkinin derecesine bağlıdır. Çoklu iç ilişkinin varlığının ya da yokluğunun

araştırılması yerine derecesinin saptanması, problemle karşılaşıldığında sonuçları ortaya çıkarmak açısından daha önemlidir.

### 3.2.1. Tam çoklu iç ilişki

Tam çoklu iç ilişki problemi açıklayıcı değişkenlerin biri yada birden fazlasının diğer açıklayıcı değişkenlerin lineer bir fonksiyonu olması durumunda ortaya çıkmaktadır. Bunun anlamı;  $X_i$ ,  $X$  matrisini  $i$ 'inci sütunu ve  $t_i$ 'ler sabit sayılar ( $i = 1, 2, \dots, p$ ) olmak üzere;

$$\sum_{i=1}^p t_i X_i = \mathbf{0} \quad (5)$$

(5) eşitliğinde hepsi sıfır olmayan  $t_i$  değerleri bulunabiliyorsa  $X_1, X_2, \dots, X_p$  vektörleri lineer bağımlıdır.

Tam çoklu iç ilişki  $X'X$  matrisinin singüler olması durumudur. Yani bu durumda  $X'X$  matrisinin determinanı sıfıra eşit olmaktadır. Dolayısıyla matrisin tersi hesaplanamaz ve EKK tahmin yöntemiyle parametreler hesaplanamaz.

### 3.2.2. Tama yakın çoklu iç ilişki

Tama yakın çoklu iç ilişki açıklayıcı değişkenler arasında tama yakın şekilde yüksek bir ilişkinin olması durumudur.

$$\sum_{i=1}^p t_i X_i \cong \mathbf{0} \quad (6)$$

Bir başka ifade ile bu durum (6) eşitliği ile verilen ifadenin sağlanması durumudur. Bu durumda  $X'X$  matrisi singüler olmayan bir matristir. Bu matrisin determinanı sıfırdan farklıdır. Fakat sıfıra oldukça yakın bir değer alır. Yani  $X'X$  matrisinin tersi hesaplanabilir ancak determinantın sıfıra çok yakın bir değer olmasından dolayı matrisin elemanları çok büyük değerler olur. Bu durumda EKK yönteminde parametre tahminlerinin standart hataları gerçek değerlerinden oldukça büyük olarak elde edilecektir.

### 3.3. Çoklu İç İlişkinin Belirlenmesi

Çoklu iç ilişkinin varlığının belirlenmesi ve derecesinin ölçülmesi amacıyla bir çok farklı yöntem geliştirilmiştir. Geliştirilen bu yöntemlerin ortak özelliği çoklu iç ilişkinin varlığının belirlenmesinden çok derecesinin ölçülmesi ile ilgilenmeleridir. Çoklu iç ilişki probleminin belirlenmesinde ve derecesinin ölçülmesinde yaygın olarak kullanılan bazı yöntemler aşağıda verilmiştir.

#### 3.3.1. Korelasyon matrisinin incelenmesi

Açıklayıcı değişkenler arasındaki ilişkiyi gösteren basit korelasyon katsayılarının yüksek olması çoklu iç ilişkinin bir göstergesidir.  $\mathbf{X}'\mathbf{X}$  matrisinin köşegen elemanları dışındaki elemanları açıklayıcı değişkenler arasındaki basit korelasyon katsayılarını göstermektedir. Eğer  $\mathbf{X}_i$  ve  $\mathbf{X}_j$  değişkenleri arasında lineer bir ilişki var ise aralarındaki basit korelasyon katsayısının mutlak değeri 1'e çok yakındır. Fakat korelasyon katsayılarının incelenmesi sadece iki açıklayıcı değişken arasında var olan ilişkiyi ortaya çıkarmaktadır. Eğer açıklayıcı değişken sayısı ikiden fazla ise bu değişkenler arasındaki ilişkiyi açıklamakta korelasyon katsayıları yeterli olmayabilir (Farrar and Glauber, 1967).

#### 3.3.2. Varyans şişirme faktörü (VIF)

$(\mathbf{X}'\mathbf{X})^{-1}$  matrisinin köşegen elemanlarının incelenmesi çoklu iç ilişkinin belirlenmesinde yaygın olarak kullanılan bir yöntemdir.  $(\mathbf{X}'\mathbf{X})^{-1}$  matrisinin  $j$ 'inci köşegen elemanı  $j$ 'inci varyans şişirme faktörü olarak adlandırılır. Varyans şişirme faktörü çoklu iç ilişkiyi belirlemek amacıyla ilk olarak 1967 yılında Farrar ve Glauber tarafından kullanılmıştır. Fakat  $j$ 'inci regresyon katsayısının varyansı  $\sigma^2 \mathbf{C}_{jj}$  olduğundan buradaki  $\mathbf{C}_{jj}$  açıklayıcı değişkenler arasındaki ilişkiden dolayı  $\hat{\beta}_j$ 'nin varyansını arttıran bir faktör olarak değerlendirilebilir. Bu ifade 1970 yılında Marquardt tarafından VIF olarak isimlendirilmiştir (Marquardt, 1970).  $R_j^2$ ,  $x_j$  açıklayıcı değişkeninin diğer  $p - 1$  açıklayıcı değişken ile regresyonundan elde edilen çoklu belirlilik katsayısı olmak üzere  $VIF_j = \frac{1}{(1-R_j^2)}$  şeklinde ifade edilir. Eğer  $x_j$  açıklayıcı değişkeni ile diğer açıklayıcı değişkenler arasında çoklu iç ilişki yoksa  $R_j^2$  değeri çok

küçük olacağından  $VIF_j$  değeri 1'e yaklaşacaktır. Eğer  $x_j$  açıklayıcı değişkeni ile diğer açıklayıcı değişkenler arasında çoklu iç ilişki varsa  $R_j^2$  değeri 1'e yaklaşacağından  $VIF_j$  değeri çok büyük bir değer alacaktır. VIF değerinin 10'dan büyük olması çoklu iç ilişkinin varlığını göstermektedir (Vinod and Ullah, 1981).

### 3.3.3. $X'X$ matrisinin özdeğerlerinin incelenmesi

$X'X$  matrisinin özdeğerlerinin incelenmesi çoklu iç ilişkinin tespit edilmesinde yaygın olarak kullanılan yöntemlerden birisidir.  $\lambda_1, \lambda_2, \dots, \lambda_p$  'ler ve  $v_1, v_2, \dots, v_p$  'ler sırasıyla  $X'X$  matrisinin özdeğerleri ve özvektörleri olmak üzere, veride çoklu iç ilişki olması durumunda en az bir özdeğer sıfıra yakın olacaktır ve özvektörlerden en az birisi (7) nolu eşitliği sağlayacaktır.

$$\sum_{i=1}^p v_{ij} X_i = \mathbf{0} \quad (7)$$

Burada  $v_{ij}$ , j'inci özvektörün i'inci elemanını,  $X_i$ ,  $X$  matrisinin i'inci sütununu göstermektedir.

Genellikle özdeğerlerin tek tek incelenmesi hiçbir şey ifade etmemektedir. Fakat özdeğerlerin birbirleriyle kıyaslanması çoklu iç ilişkinin varlığı ve şiddeti hakkında bize bilgi vermektedir. En büyük özdeğerin en küçük özdeğere bölünmesiyle  $K = \frac{\lambda_{max}}{\lambda_{min}}$  olarak tanımlanan koşul sayısı çoklu iç ilişkinin şiddetinin belirlenmesinde çok iyi bir göstergedir (Vinod and Ullah, 1981). Koşul sayısı olarak tanımlanan K değeri 100'den küçük olduğu durumlarda çok zayıf şiddette çoklu iç ilişki olduğunu, 100'den büyük olduğu durumlarda ise şiddetli bir çoklu iç ilişkinin olduğunu söyleyebiliriz.

Yukarıda kısaca açıklanan yöntemlerin dışında çoklu iç ilişkiyi belirlemek için başka yöntemlerde vardır. Bunlardan birisi  $X'X$  matrisinin determinantının incelenmesidir.  $X'X$  matrisi standartlaştırıldığında alacağı değerler  $0 \leq |X'X| \leq 1$  aralığında olacaktır.  $|X'X| = 0$  olduğunda tam çoklu iç ilişki vardır denir.  $|X'X| = 1$  olduğunda ise açıklayıcı değişkenler birbirine diktir yani çoklu iç ilişki yoktur denir. Bu durumda  $|X'X|$  değeri sıfıra yaklaştıkça çoklu iç ilişkinin şiddeti artacaktır (Farrar and Glauber, 1967).

Regresyon katsayılarının işaretlerinin ve büyüklüklerinin beklenenden farklı olması ve açıklayıcı değişkenlerin eklenmesi veya çıkarılmasıyla regresyon katsayılarının tahminlerinde büyük değişiklikler meydana gelmesi de çoklu iç ilişkinin varlığını gösteren bir durumdur.

### **3.4. Çoklu İç İlişkinin Giderilmesi**

Çoklu iç ilişki probleminin giderilmesi için birçok yöntem önerilmiştir. Bu yöntemlerden bazıları ek verinin toplanması, modelin yeniden belirlenmesi ve yanlı tahmin yöntemlerinin kullanılmasıdır. Bu yöntemler kısaca aşağıdaki şekilde özetlenebilir.

#### **Ek verinin toplanması**

Farrar ve Glauber (1967) ve Silvey (1969) çoklu iç ilişki probleminin giderilmesi için ek verinin toplanmasını önermişlerdir. Fakat ek verinin toplanması modeldeki veya kitledeki kısıtlamalardan dolayı her zaman mümkün olmayabilir.

#### **Modelin yeniden belirlenmesi**

Çoklu iç ilişkinin nedeni model seçimi de olabilir. Bu yüzden çoklu iç ilişki problemi ile başa çıkabilmek için modelin yeniden belirlenmesi bir yol olarak izlenebilir. Modeli yeniden belirlemenin yolları açıklayıcı değişkenlerin yeniden tanımlanması ya da ilişkili açıklayıcı değişkenlerden birisinin modelden çıkarılması olabilir.

#### **Yanlı Tahmin Tekniklerinin Kullanılması**

Regresyon analizinde parametrelerin tahmini için en çok kullanılan yöntem EKK yöntemidir. EKK yönteminin uygulanabilmesi için bazı koşulların sağlanması gerekmektedir. Bu koşullardan birisi açıklayıcı değişkenlerin lineer bağımsız olması koşuludur. Çoklu iç ilişki problemi ile karşılaşıldığında bu koşul sağlanmamaktadır. Bu yüzden çoklu iç ilişki problemini ortadan kaldırmak için EKK tahmin ediciye alternatif olan yanlı tahmin tekniklerini kullanmak gerekmektedir (Özkale, 2007).

## BÖLÜM 4

### YANLI REGRESYON

Regresyon katsayılarının en küçük kareler tahmin edicileri varsayımları sağlandığı takdirde en iyi yansız tahmin edicilerdir. Yani, en küçük kareler tahmin edici bütün yansız tahmin ediciler arasında en küçük varyansa sahip tahmin edicidir. Çoklu iç ilişkinin varlığında ise minimum varyanslı olma özelliği sağlanmamaktadır. Bu durumda EKK tahmin edicilerinin bazı özelliklerinin esnetilmesiyle regresyon modeli kurulabilir. Bunlardan birisi yansızlık özelliğinin esnetilmesidir. Yanlı regresyon yöntemleri çoklu iç ilişki durumunda yaygın olarak kullanılan yöntemlerdir. Yanlı regresyon yansızlığın çok önemli olmadığı bu tip regresyon yöntemlerini ifade eder. Bu tip yöntemler çoklu iç ilişki probleminin mümkün çözümleri için önerilmektedir. Yanlı regresyonda elde edilen tahmin ediciler parametrelerin tahmini için en küçük karelere göre gerçek değerine daha yakın sonuçlar vermektedir.

Bir tahmin edicinin tahmin edilen parametreye yakınlığının ölçüsü tahmin edicinin hata kareler ortalamasıdır (MSE). Eğer  $\tilde{\theta}$ ,  $\theta$ 'nın bir tahmin edicisi ise,  $\tilde{\theta}$ 'nin hata kareler ortalaması aşağıda tanımlandığı gibidir.

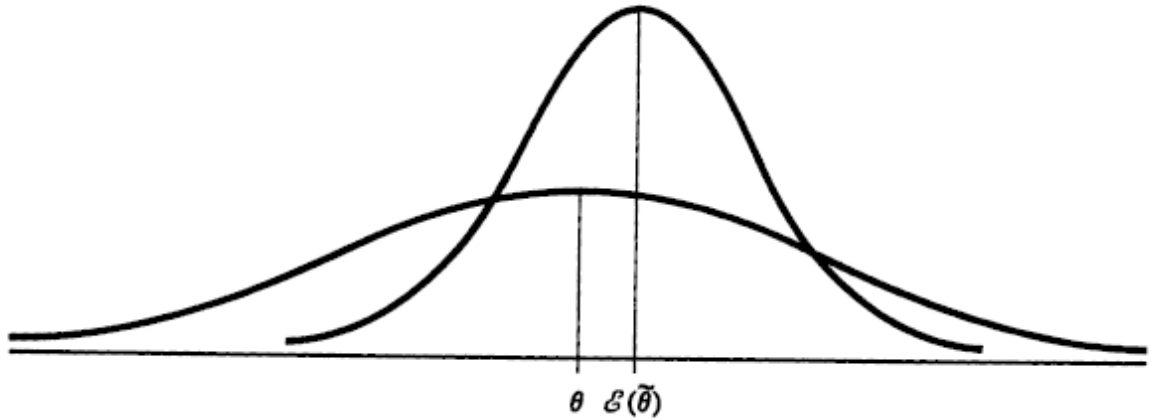
$$MSE(\tilde{\theta}) = E(\tilde{\theta} - \theta)^2 \quad (8)$$

$\tilde{\theta}$  tahmin edicisinin varyansı ise aşağıdaki gibi tanımlanır.

$$Var(\tilde{\theta}) = E[\tilde{\theta} - E(\tilde{\theta})]^2 \quad (9)$$

Eğer kullanılan tahmin edici yansız bir tahmin edici ise  $E(\tilde{\theta}) = \theta$  olur, bundan dolayı da  $MSE(\tilde{\theta}) = \sigma^2(\tilde{\theta})$  olur. Diğer yandan, MSE; tahmin edicinin varyansı ile yanlılığının karesinin toplamıdır. Burada yanlılık  $Bias(\tilde{\theta}) = E(\tilde{\theta}) - \theta$  şeklindedir. Yanlı tahmin edicilerin varyansının yansız tahmin edicilerin varyansından daha küçük olması mümkündür. Böyle bir durumda yanlı tahmin edici tahmin edilen parametrenin ortalamasına yansız bir tahmin ediciden daha yakındır (Rawlings et. al., 1961).

Yanlı tahmin edicilerin avantajı Şekil 4.1'de gösterilmektedir.



**Şekil 4.1. Yanlı tahmin edicinin yansız tahmin ediciden daha küçük MSE değerine sahip olduğunu gösterilmesi (Rawlings et. al., 1998):**

Yukarıdaki şekilde  $E(\tilde{\theta})$  merkezli normal eğri  $\theta$ 'nın yansız tahmin edicisi olan  $\tilde{\theta}$ 'nin olasılık dağılımını göstermektedir. Burada yanlılık parametresi Bias  $E(\tilde{\theta})$  ile  $\theta$  arasındaki farka eşittir. Dağılımdaki daha küçük yayılmalar daha küçük varyansın olacağını yansıtmaktadır. Biraz yanlılığa izin verildiğinde yansız tahmin ediciye göre daha küçük MSE değerine sahip bir yanlı tahmin edici bulmak mümkündür (Rawlings et. al., 1998).

Yanlı regresyon yönteminin konseptini tanımlamak için aşağıda verilen lineer modeli düşünelim.

$$Y_i = \beta_0 + Z_{i1}\beta_1 + Z_{i2}\beta_2 + \varepsilon_i \quad i = 1, \dots, n \quad (10)$$

Burada  $Z_{i1}$  ve  $Z_{i2}$  merkezleştirilmiş ve ölçeklendirilmiş değişkenlerdir ayrıca  $\varepsilon_i \sim NID(0, \sigma^2)$  dir. Merkezleştirilme ve ölçeklendirmenin anlamı  $\sum_{i=1}^n Z_{i1} = \sum_{i=1}^n Z_{i2} = 0$  ve  $\sum_{i=1}^n Z_{i1}^2 = \sum_{i=1}^n Z_{i2}^2 = 1$  olmasıdır.  $\rho$  'yi  $\sum_{i=1}^n Z_{i1}Z_{i2}$  olarak tanımlayalım. Bundan dolayı

$$\begin{bmatrix} n & 0 & 0 \\ 0 & \sum_{i=1}^n Z_{i1}^2 & \sum_{i=1}^n Z_{i1}Z_{i2} \\ 0 & \sum_{i=1}^n Z_{i1}Z_{i2} & \sum_{i=1}^n Z_{i2}^2 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{n} & 0 & 0 \\ 0 & (1-\rho^2)^{-1} & -\rho(1-\rho^2)^{-1} \\ 0 & -\rho(1-\rho^2)^{-1} & (1-\rho^2)^{-1} \end{bmatrix}$$

olmaktadır.  $\beta_0, \beta_1$  ve  $\beta_2$ 'nin en küçük kareler tahmin edicileri olan  $\hat{\beta}_0, \hat{\beta}_1$  ve  $\hat{\beta}_2$ 'nin varyansları sırası ile  $\sigma^2/n, \sigma^2/(1 - \rho^2)$  ve  $\sigma^2/(1 - \rho^2)$  olmaktadır. Eğer  $\rho$  değeri 1'e yakın ise  $Z_{i1}$  ve  $Z_{i2}$  değişkenleri yüksek derecede ilişkilidir ve bundan dolayı da çoklu iç ilişki problemi ile karşılaşılır.  $\rho$  değeri 1'e eşit olduğunda  $\hat{\beta}_1$  ve  $\hat{\beta}_2$ 'nin varyansları  $\sigma^2/(1 - \rho^2)$  olacaktır. Bu değer ise oldukça büyüktür.  $\hat{\beta}_1, \beta_1$ 'in en iyi lineer yansız tahmin edicisi olmasına rağmen daha küçük hata kareler ortalamasına (MSE) sahip yanlı bir tahmin edici bulunabilir. Örneğin  $\beta_1$ 'in tahmin edicisinin aşağıdaki şekilde verildiğini düşünelim.

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n Z_{i1} Y_i}{\sum_{i=1}^n Z_{i1}^2} \quad (11)$$

Burada verilen  $\tilde{\beta}_1$  tahmin edicisi  $\beta_1$ 'in en küçük kareler tahmin edicisidir. Eşitlik (10)'da verilen modelde  $\beta_2 = 0$  olduğunu varsayalım.  $E(\tilde{\beta}_1) = \beta_1 + \rho\beta_2$  olduğu için  $\tilde{\beta}_1$  tahmin edicisi  $\beta_1$  için yansız değildir.  $\tilde{\beta}_1$ 'nin yanlılık değeri  $E(\tilde{\beta}_1) - \beta_1 = \rho\beta_2$  ayrıca  $\tilde{\beta}_1$ 'nin varyansı  $\sigma^2$ 'dir. Bundan dolayı  $\tilde{\beta}_1$ 'nin MSE değeri

$$MSE(\tilde{\beta}_1) = var(\tilde{\beta}_1) + [Bias(\tilde{\beta}_1)]^2 = \sigma^2 + \rho^2\beta_2^2 \quad (12)$$

şeklinindedir.  $\beta_2$ 'nin küçük değerleri için  $MSE(\tilde{\beta}_1)$  değeri,  $MSE(\hat{\beta}_1)$  değerinden daha küçük olabilir (Rawlings et. al., 1998).

Çoklu iç ilişki problemini çözmek için literatürde bir çok yanlı regresyon yöntemi önerilmiştir. Bunlardan bazıları; Stein Regresyon (Stein, 1960), Ridge Regresyon (Hoerl and Kennard, 1970), Temel Bileşenler Regresyonu (Hawkins, 1973; Marquardt, 1970; Webster et. al., 1974) ve Kısmi en küçük kareler regresyonudur (Wold, 1985).

Çoklu iç ilişki problemi X uzayında gözlemlenen değerlerin, tahmin edilen bağımlı değişkenin hassasiyetini etkilememesine rağmen tahmin edilen bağımlı değişkenlerin varyanslarının şişmesine neden olmaktadır. Park (1981), temel bileşenler regresyonunda tam olarak tahmin edilen parametreler üzerine kısıtlar koymanın bağımlı değişkenlerin tahmini için optimal MSE değerini verdiğini göstermiştir. Bu da yanlı



regresyon yöntemlerinin bazı durumlarda bağımlı değişkenlerin tahmin edilmesinde yararlı olacağını ifade etmektedir (Rawlings et. al., 1998).

Çalışmanın bu bölümünde yanlı regresyon yöntemlerinden olan Kısmi En Küçük Kareler Regresyonu, Temel Bileşenler Regresyonunu ve Ridge Regresyonu incelenecektir.

#### 4.1. Ridge Regresyon

Ridge regresyon yöntemi çoklu iç ilişki problemi ile karşılaşıldığında kullanılan yanlı regresyon yöntemlerinden birisidir. Ridge tahmin edicisi yanlı regresyon tahmin edicileri arasında en geniş uygulama alanına sahip tahmin edicidir. Bu tahmin edici ilk olarak 1970 yılında Hoerl ve Kennard tarafından önerilmiştir.

Çoklu iç ilişki problemi yapılan analizler sonucunda elde edilen en küçük kareler tahmin edicisinin varyans değerlerinin olması gerekenden büyük olmasına ve regresyon katsayılarının gerçek değerlerinden uzaklaşmasına neden olmaktadır. Bu problemle karşılaşıldığı durumlarda en küçük kareler tahmin edicisi regresyon katsayıları için güvenilir olmayan sonuçlar vermektedir (Gunst and Mason, 1980). Bu durumda kullanılmak üzere Hoerl ve Kennard  $X'X$  matrisinin köşegen elemanlarına küçük bir sabit ekleyerek Ridge tahmin edicisini önermişlerdir. Ridge tahmin edicisi parametreleri yanlı olarak tahmin eder. Fakat elde edilen bu tahminler EKK tahmin edicisine göre daha küçük varyansa sahiptirler. Hoerl ve Kennard (1970),  $X'X$  matrisinin köşegen elemanlarında küçük değişiklik yaparak Ridge tahmin edicisini (13) nolu eşitlikte gösterildiği gibi elde etmişlerdir.

$$\hat{\beta}_R = (X'X + kI)^{-1}X'y \quad (13)$$

Burada  $k$  ( $0 < k < 1$ ) yanlılık sabiti veya daraltma parametresi olarak adlandırılır.  $k$  sabitinin seçilmesi Ridge tahmin edicisinin performansını etkilemektedir.  $k = 0$  olması durumunda Ridge tahmin edicisi EKK tahmin edicisine dönüşmektedir (Hoerl and Kennard, 1970).

#### 4.1.1. Ridge tahmin edicisinin en küçük kareler tahmin edicisi ile karşılaştırılması

Ridge tahmin edicisinin verildiği (13) nolu eşitlikte  $X'y$  yerine  $X'X\hat{\beta}$  yazılırsa (14) nolu eşitlik elde edilir.

$$\hat{\beta}_R = (X'X + kI)^{-1}X'X\hat{\beta} \quad (14)$$

(14) nolu eşitlikte  $(X'X + kI)^{-1}X'X = Z$  olarak alınırsa

$$\hat{\beta}_R = Z\hat{\beta} \quad (15)$$

eşitliği elde edilir. (15) nolu eşitlikten Ridge tahmin edicisinin EKK tahmin edicisinin doğrusal bir fonksiyonu şeklinde yazılabileceği görülmektedir (Hoerl and Kennard, 1970).

$Z = (X'X + kI)^{-1}X'X$  matrisinin tersi alınarak (15) nolu eşitlikte yerine yazılırsa Ridge tahmin edicinin farklı bir gösterimi olan (16) nolu eşitlik elde edilebilir.

$$\hat{\beta}_R = [I + k(X'X)^{-1}]^{-1}\hat{\beta} \quad (16)$$

Ayrıca (13) nolu eşitlikteki  $(X'X + kI)^{-1}$  matrisi  $W$  matrisi olarak tanımlanır ise Ridge tahmin edicisinin farklı bir gösterimi (17) nolu eşitlikteki gibi yazılabilir.

$$\hat{\beta}_R = WX'y \quad (17)$$

Hoerl ve Kennard (1970), yayınladıkları makalelerinde  $W$  ve  $Z$  matrislerini tanımladıktan sonra  $W$  ve  $Z$  matrisleri ile EKK tahmin edicisinin bazı özelliklerini kullanarak Ridge tahmin edicisi ile EKK tahmin edicisi arasındaki ilişkiyi aşağıdaki şekilde tanımlamışlardır.

- i.*  $\lambda_i, \xi_i(W)$  ve  $\xi_i(Z)$  sırasıyla  $X'X, W$  ve  $Z$  matrislerinin özdeğerlerini göstermek üzere;

$$\xi_i(W) = \frac{1}{\lambda_i + k} \quad (18)$$

$$\xi_i(Z) = \frac{\lambda_i}{\lambda_i + k} \quad (19)$$

eşitlikleri yazılabilir. Elde edilen bu sonuçlar  $|\mathbf{W} - \xi\mathbf{I}|$  ve  $|\mathbf{Z} - \xi\mathbf{I}|$  karakteristik denklemlerinin çözülmesiyle bulunmuştur (Hoerl and Kennard, 1970).

ii.  $\mathbf{Z} = \mathbf{I} - k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X})$  matrisi alternatif formu olan  $\mathbf{Z} = \mathbf{W}\mathbf{X}'\mathbf{X}$  şeklinde yazılır ve eşitliğin her iki tarafı soldan  $\mathbf{W}^{-1}$  matrisi ile çarpılırsa  $\mathbf{Z}$  matrisi (20) nolu eşitlikte olduğu gibi  $\mathbf{W}$  matrisinin bir fonksiyonu olarak yazılabilir.

$$\mathbf{Z} = \mathbf{I} - k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} = \mathbf{I} - k\mathbf{W} \quad (20)$$

iii.  $k \neq 0$  olduğu zaman  $\widehat{\boldsymbol{\beta}}_R$  vektörü  $\widehat{\boldsymbol{\beta}}$  vektöründen küçüktür. Dolayısıyla  $\widehat{\boldsymbol{\beta}}_R' \widehat{\boldsymbol{\beta}}_R < \widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{\beta}}$  olarak yazılabilir.

#### 4.1.2. Ridge tahmin edicisinin yanlılık değerinin hesaplanması

Ridge tahmin edicisinin yanlılığını gösterebilmek için (15) nolu eşitlik kullanılmalıdır. (15) nolu eşitliğin beklenen değeri alınır,

$$E(\widehat{\boldsymbol{\beta}}_R) = E(\mathbf{Z}\widehat{\boldsymbol{\beta}}) = \mathbf{Z}\boldsymbol{\beta} \quad (21)$$

elde edilir. (21) nolu eşitlikten  $\widehat{\boldsymbol{\beta}}_R$  tahmin edicisinin beklenen değerinin  $\boldsymbol{\beta}$ 'ya eşit olmadığı görülebilir. Bir başka ifade ile  $\widehat{\boldsymbol{\beta}}_R$  tahmin edicisi  $\boldsymbol{\beta}$ 'nin yansız bir tahmin edicisi değildir. Ridge tahmin edicisinin yanlılık miktarı Ridge tahmin edicisinin beklenen değeri alınarak hesaplanabilir (Hoerl and Kennard, 1970). Ridge tahmin edici,

$$\widehat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$$

olmak üzere, beklenen değeri alınır ise,

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}_R) &= E[(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}] \quad (22) \\ &= E[(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\ &= E[(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}] \\ &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) \\ &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

eşitliği elde edilir. Elde edilen bu eşitliğin her iki tarafını soldan  $(\mathbf{X}'\mathbf{X} + k\mathbf{I})$  ile çarparsak,

$$(\mathbf{X}'\mathbf{X} + k\mathbf{I})E(\widehat{\boldsymbol{\beta}}_R) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

$$(\mathbf{X}'\mathbf{X} + k\mathbf{I})E(\widehat{\boldsymbol{\beta}}_R) = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

$$(\mathbf{X}'\mathbf{X} + k\mathbf{I})E(\widehat{\boldsymbol{\beta}}_R) = \mathbf{X}'\mathbf{X}\boldsymbol{\beta} + k\mathbf{I}\boldsymbol{\beta} - k\mathbf{I}\boldsymbol{\beta}$$

$$(\mathbf{X}'\mathbf{X} + k\mathbf{I})E(\widehat{\boldsymbol{\beta}}_R) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})\boldsymbol{\beta} - k\mathbf{I}\boldsymbol{\beta}$$

$$(\mathbf{X}'\mathbf{X} + k\mathbf{I})E(\widehat{\boldsymbol{\beta}}_R) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})\boldsymbol{\beta} - k[(\mathbf{X}'\mathbf{X} + k\mathbf{I})(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}]\boldsymbol{\beta}$$

Yukarıdaki son eşitliği soldan  $(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}$  ile çarparsak,

$$E(\widehat{\boldsymbol{\beta}}_R) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}[(\mathbf{X}'\mathbf{X} + k\mathbf{I})\boldsymbol{\beta} - k(\mathbf{X}'\mathbf{X} + k\mathbf{I})(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}]\boldsymbol{\beta}$$

$$E(\widehat{\boldsymbol{\beta}}_R) = \boldsymbol{\beta} - k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\boldsymbol{\beta} \quad (23)$$

eşitliği elde edilir. Ridge tahmin edicisinin yanlılığını gösterebilmek için (23) nolu eşitlikte verilen Ridge tahmin edicisinin beklenen değerinden yararlanılacaktır.

$$\text{Bias}(\widehat{\boldsymbol{\beta}}_R) = E(\widehat{\boldsymbol{\beta}}_R) - \boldsymbol{\beta}$$

olduğundan  $E(\widehat{\boldsymbol{\beta}}_R)$  yerine (23) nolu eşitlikten elde edilen sonuç yazılırsa,

$$\begin{aligned} \text{Bias}(\widehat{\boldsymbol{\beta}}_R) &= \boldsymbol{\beta} - k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\boldsymbol{\beta} - \boldsymbol{\beta} \\ &= -k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\boldsymbol{\beta} \end{aligned} \quad (24)$$

eşitliği elde edilir. (24) nolu eşitlik Ridge tahmin edicisinin yanlılık değerini göstermektedir (Hoerl and Kennard, 1970).

#### 4.1.3. Ridge tahmin edicisinin hata kareler ortalamasının hesaplanması

Tahmin edicilerin hata kareler ortalaması değeri, tahmin vektörünün gerçek parametre vektörüne olan uzaklığının karesinin beklenen değeri olarak tanımlanmaktadır. Bu uzaklık (25) nolu eşitlikte verildiği gibi hesaplanmaktadır (Hoerl and Kennard, 1970).

$$L^2 = (\hat{\beta}_R - \beta)'(\hat{\beta}_R - \beta) \quad (25)$$

(25) nolu eşitlikte verilen bu uzaklığın beklenen değeri alınır ise;

$$\begin{aligned} E(L^2) &= E[(\hat{\beta}_R - \beta)'(\hat{\beta}_R - \beta)] \\ &= E[(\hat{\beta} - \beta)'Z'Z(\hat{\beta} - \beta)] + (Z\beta - \beta)'(Z\beta - \beta) \\ &= \sigma^2 \text{trace}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z} + \beta'(Z - I)'(Z - I)\beta \\ &= \sigma^2 [\text{trace}(\mathbf{X}'\mathbf{X} + kI)^{-1} - k\text{trace}(\mathbf{X}'\mathbf{X} + kI)^{-2}] \\ &\quad + k^2 \beta'(\mathbf{X}'\mathbf{X} + kI)^{-2}\beta \\ &= \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \beta'(\mathbf{X}'\mathbf{X} + kI)^{-2}\beta \\ &= \gamma_1(k) + \gamma_2(k) \end{aligned} \quad (26)$$

eşitliği elde edilir.

(26) nolu eşitlikteki ikinci bileşen  $\gamma_2(k)$ ,  $Z\beta$ 'dan  $\beta$ 'ya olan uzaklığın karesini göstermektedir.  $k = 0$  olduğunda  $Z = I$  olacağından dolayı  $\gamma_2(k)$  sifıra eşit olacaktır. (26) nolu eşitliğin birinci bileşeni  $\gamma_1(k)$  ise parametre tahmin edicilerinin varyanslarının toplamını göstermektedir.

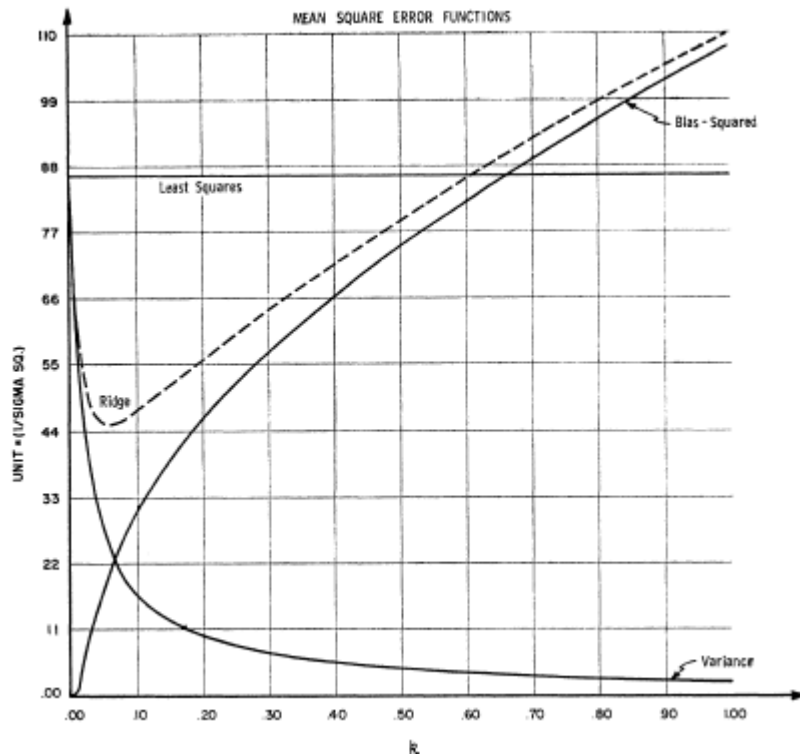
$$\hat{\beta}_R = Z\hat{\beta} = Z(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

olduğundan, Ridge tahmin edicisinin varyansı

$$\text{Var}(\hat{\beta}_R) = \text{Var}(Z(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})$$

$$\begin{aligned}
&= \mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}' \\
&= \sigma^2\mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}' \\
&= \sigma^2\mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}' \tag{27}
\end{aligned}$$

şeklinde elde edilir. Ridge tahmin edicisinin varyansı, yanlılığı ve  $k$  parametresi arasındaki ilişki Şekil 4.2'de verilmiştir.



**Şekil 4.2. Hata kareler ortalaması fonksiyonları (Hoerl and Kennard, 1970).**

Şekil 4.2'de yatay eksen  $k$  değerlerini gösterirken dikey eksen ise varyans değerlerini göstermektedir. Yatay eksene paralel olan çizgi ise en küçük kareler tahmin edicisinin hata kareler ortalaması değerini göstermektedir. En küçük kareler tahmin edicisi yansız bir tahmin edici olduğu için hata kareler ortalaması değeri varyans değerine eşit olmaktadır. Ayrıca en küçük kareler tahmin edicisinin varyans değeri  $k$ 'nın bir fonksiyonu olmadığı için değişen  $k$  değerlerine karşılık varyansı sabit kalmaktadır.  $k$ 'nın artan fonksiyonu olarak ifade edilen eğri Ridge tahmin edicisinin yanlılık değerini göstermektedir. Ayrıca azalan bir fonksiyon gösteren eğri ise Ridge tahmin edicisinin varyans değerini göstermektedir.

Hoerl ve Kennard (1970) yaptıkları çalışmalarında ortogonal olan  $\mathbf{X}'\mathbf{X}$  matrisi için yanlılık parametresi olan  $k$  değeri sıfıra yaklaştıkça  $\gamma_1(k)$  fonksiyonunun  $-2k\sigma^2$  gibi negatif bir türeve sahip olduğu ve  $\mathbf{X}'\mathbf{X}$  matrisi ortogonallikten uzaklaştıkça ise türev değerinin  $-\infty$  değerine yaklaştığını göstermişlerdir. Diğer yandan  $k$  değeri sağdan sıfıra yaklaştıkça  $\gamma_2(k)$  fonksiyonu orjin etrafında sıfırdır.

Bu özellikler  $k > 0$  olduğunda yani çok az bir yanlılık karşılığında varyans değerinde önemli ölçüde azalma sağlayacak ve en küçük kareler tahmin edicisinden daha küçük hata kareler ortalaması değerine sahip Ridge tahmin edicileri bulunabileceği sonucunu desteklemektedir (Hoerl and Kennard, 1970).

#### **4.1.4. Ridge tahmin edicisinde yanlılık parametresi $k$ 'nın seçimi**

Yanlılık parametresi olan  $k$ 'nın seçimi elde edilen Ridge tahmin edicisinin çoklu iç ilişki problemi ile ne derecede başa çıkacağını belirlediğinden dolayı önemli bir konudur. Ayrıca  $k$  parametresi tahmin edicinin ne kadar yanlı olacağını belirlediği için bu parametrenin sıfıra yakın bir değer olarak seçilmesi yaygın olarak kullanılmaktadır. Yanlılık parametresi olan  $k$ 'nın seçimi için bir çok yöntem önerilmiştir. Bu yöntemler grafiksel yöntemler ve analitik yöntemler olmak üzere iki gruba ayrılabilir. Bu çalışmada grafiksel yöntemler arasında yaygın olarak kullanılan Ridge izi yöntemi tartışılacaktır.

##### **4.1.4.1. Ridge izi yöntemi**

Ridge izi yöntemi yanlılık parametresinin seçimi için kullanılan grafiksel bir yöntemdir. Ridge iz her bir regresyon katsayısının  $k$ 'nın farklı değerlerine karşılık gelen değerlerinin grafiksel bir gösterimidir. Marquart ve Snee (1975)'e göre Ridge izi yönteminde dikkat edilmesi gereken önemli bir nokta, bir karışıklığın meydana gelmemesi için bu yöntemde 10'dan fazla regresyon katsayısının grafikte gösterilmemesi gerektiğidir.

Ridge izi yönteminin çalışma prensibi grafikte gösterilen regresyon katsayılarının sabitleşmeye başladığı en küçük  $k$  değerinin seçilmesi şeklindedir. Hoerl ve Kennard (1970),  $k$  değerinin  $0 \leq k \leq 1$  aralığında olacağını önermişlerdir. Ridge izi

yönteminde  $k$  yanlılık parametresinin seçimi araştırmacıdan araştırmacıya değiştiği için bu durum bir dezavantaj yaratmaktadır. Bu sorunu aşmak için daha sonra analitik yöntemler geliştirilmiştir.

#### **4.1.4.2. Analitik yöntemler**

Ridge izi yönteminde yanlılık parametresi olan  $k$ 'nın seçimi araştırmacıdan araştırmacıya değiştiği için daha objektif olan analitik yöntemler önerilmiştir. Önerilen bu analitik yöntemler MSE değerinin minimum yapılması esasına dayanmaktadır.

- Hoerl, Kennard (1970), yanlılık parametresinin seçimi için aşağıdaki istatistiği önermişlerdir. Burada amaç Ridge tahmin edicisinin EKK tahmin edicisinden daha küçük MSE değerine sahip olacağı  $k$  değerinin belirlenmesidir.

$$\hat{k}_{HK} = \frac{s^2}{\hat{\alpha}_{max}^2}$$

- Mallows (1973), aşağıda verilen eşitliği minimum yapacak şekilde seçilen  $k$  değerinin uygun  $k$  değeri olarak seçilebileceğini önermiştir.

$$C_k = \frac{SSE(k)}{s^2} - n + 2 + 2trace[\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}']$$

- Theobald (1974), her  $S$  için  $E[(\hat{\beta}_R - \beta)' S(\hat{\beta}_R - \beta)]$  beklenen değerinin,  $E[(\hat{\beta} - \beta)' S(\hat{\beta} - \beta)]$  beklenen değerinden daha küçük olacak şekilde yanlılık parametresinin seçimini

$$k_T = \frac{2s^2}{\hat{\beta}'\hat{\beta}}$$

olarak önermişlerdir.

- Hoerl, Kennard ve Baldwin (1975),  $k$ 'nın uygun değeri için Ridge tahmin edicisinin MSE değerini minimum yapan  $k_i = \frac{\sigma^2}{\alpha_i^2}$  değerlerinin harmonik ortalaması olan

$$\hat{k}_{HKB} = \frac{ps^2}{\hat{\beta}'\hat{\beta}}$$

değerini önermişlerdir.



- Mc Donald ve Galarneau (1975),

$$\hat{\beta}'_R \hat{\beta}_R = \hat{\beta}' \hat{\beta} - \sigma^2 \text{trace}(\mathbf{X}' \mathbf{X})^{-1}$$

olacak şekilde  $k$  yanlılık parametresini seçmeyi önermişlerdir.

- Lawless ve Wang (1976), bayesian yaklaşımla yanlılık parametresinin seçimini

$$k_{LW} = \frac{ps^2}{\hat{\beta}' \mathbf{X}' \mathbf{X} \hat{\beta}}$$

olarak önermişlerdir.

#### 4.2. Temel Bileşenler Regresyonu

Temel bileşenler regresyonu çoklu iç ilişki problemiyle başa çıkmak için kullanılan yanlı regresyon yöntemlerinden birisidir. Temel bileşenler regresyonu iyi yorumlanabilir regresyon modelleri kurabilir fakat bunun için özellikle açıklayıcı değişkenler çok fazla olduğu durumlarda çok fazla hesaplama gerektirmektedir. Temel bileşenler regresyonu çoklu iç ilişki problemini çözebilir ve açıklayıcı değişkenlerin sayısını azaltabilir. Temel bileşenler regresyonu açıklayıcı değişkenlerin lineer kombinasyonlarını alarak çözüm yaptığı için bu yöntem temel bileşenler analizi ile çoklu lineer regresyon yönteminin bir kombinasyonudur (Varmuza and Filzmoser, 2009).

Temel bileşenler regresyonu açıklayıcı değişkenler yerine açıklayıcı değişkenlerden elde edilen temel bileşenleri kullanmaktadır. Elde edilen temel bileşenler ilişkisiz olduklarında bu temel bileşenler arasında çoklu iç ilişki problemi görülmez. Eğer regresyon analizinde bütün temel bileşenler kullanılırsa elde edilen regresyon denklemi en küçük karelerde elde edilen denkleme eşit olur. Fakat temel bileşenler regresyonu yardımıyla hesaplanan en küçük kareler tahmin edicisi orijinal veri ile hesaplanan en küçük kareler tahmin edicisine göre daha durağan olabilir (Flury and Riedwyl, 1988).

Bu bölümde temel bileşenler regresyonu, regresyon denkleminde silinecek temel bileşenlerin belirlenme yöntemleri ve diğer yanlı regresyon yöntemleriyle ilişkisi incelenecektir.

(28) nolu eşitlikte verilen standart regresyon denklemini düşünelim.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (28)$$

Burada  $\mathbf{y}$ ;  $n$  gözlemlili bağımlı değişkenler vektörü,  $\mathbf{X}$ ;  $n \times p$  tipinde açıklayıcı değişkenler vektörü,  $\boldsymbol{\beta}$ ; regresyon katsayılarının vektörü ve  $\boldsymbol{\varepsilon}$ ; hata terimlerinin vektörüdür. Literatürde temel bileşenler regresyonunda açıklayıcı değişkenlerin standartlaştırıldığı varsayılmaktadır. Bu yüzden  $\mathbf{X}'\mathbf{X}$  açıklayıcı değişkenlerin korelasyon matrisi formundadır.

Her bir gözlem için temel bileşenlerin değerleri (29) nolu eşitlikte verildiği gibidir.

$$\mathbf{Z} = \mathbf{X}\mathbf{A} \quad (29)$$

(29) nolu eşitlikte  $\mathbf{Z}$  matrisinin  $(i, k)$ 'inci elemanı  $i$ 'inci bileşen için  $k$ 'inci temel bileşenin değeridir ve  $\mathbf{A}$ ;  $p \times p$  tipinde bir matristir.  $\mathbf{A}$  matrisinin  $k$ 'inci sütunu  $\mathbf{X}'\mathbf{X}$  matrisinin  $k$ 'inci özvektörünü ifade etmektedir.  $\mathbf{A}$  matrisi ortogonal olduğundan dolayı,  $\mathbf{X}\boldsymbol{\beta}$  vektörü  $\mathbf{X}\mathbf{A}\mathbf{A}'\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma}$  olarak yeniden yazılabilir. Burada  $\boldsymbol{\gamma} = \mathbf{A}'\boldsymbol{\beta}$  şeklindedir. (28) nolu eşitlikte  $\mathbf{X}\boldsymbol{\beta}$  yerine  $\mathbf{Z}\boldsymbol{\gamma}$  yazılırsa temel bileşenler regresyonu (30) nolu eşitlikteki gibi elde edilir.

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (30)$$

İndirgenmiş temel bileşenler regresyon modeli (31) nolu eşitlikte verildiği gibidir.

$$\mathbf{y} = \mathbf{Z}_m\boldsymbol{\gamma}_m + \boldsymbol{\varepsilon}_m \quad (31)$$

(31) nolu eşitlikte verilen  $\boldsymbol{\gamma}_m$ ;  $\boldsymbol{\gamma}$  vektörünün  $m$  tane elemanının oluşturduğu alt kümesinin vektörüdür.  $\mathbf{Z}_m$  ise  $n \times m$  tipinde  $\mathbf{Z}$  matrisinin sütunlarının alt kümesinin oluşturduğu matristir. (30) nolu eşitlikteki  $\boldsymbol{\gamma}$ 'nın tahmini için ilk olarak en küçük kareler tahminleri elde edilir daha sonra ise bu tahminler (32) nolu eşitlikte yerine konularak  $\boldsymbol{\beta}$  regresyon katsayıları vektörü hesaplanır. Bu durumda elde edilen  $\widehat{\boldsymbol{\beta}}$  tahminleri (28) nolu denklemlerle direkt olarak elde edilen en küçük kareler tahminine eşittir (Jolliffe, 1982, 2002).

$$\hat{\beta} = A\hat{\gamma} \quad (32)$$

Orijinal açıklayıcı değişkenler yerine temel bileşenlerin kullanılması yeni bir fikir değildir (Hotelling, 1957; Kendall, 1957). Temel bileşenlerin kullanılmasının birçok avantajı vardır. İlk olarak,  $\mathbf{Z}$  matrisinin sütunları ortogonal olduğundan (30) nolu denklemden  $\hat{\gamma}$ 'nın hesaplanması (28) nolu denklemden  $\hat{\beta}$ 'nin hesaplanmasından daha kolaydır. (33) nolu eşitlikte  $\hat{\gamma}$ 'nin hesaplanması verilmiştir.

$$\hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \mathbf{L}^{-2}\mathbf{Z}'\mathbf{y} \quad (33)$$

Burada  $\mathbf{L}$  matrisi  $l_k^{1/2}$  elemanlarından oluşan diagonal matristir. Burada tanımlanan  $l_k$ 'lar ise  $\mathbf{X}'\mathbf{X}$  matrisinin en büyük  $k$ 'ıncı özdeğeridir. Ayrıca, regresyon denklemi açıklayıcı değişkenler yerine temel bileşenler kullanılarak hesaplanırsa temel bileşenlerin regresyon denklemine katkısı orijinal açıklayıcı değişkenlerin katkısına göre daha kolay yorumlanabilir. Temel bileşenlerin ilişkisiz olmasından dolayı, bir temel bileşenin tahmin edilen katsayısı ve katkısı regresyon denklemindeki diğer temel bileşenlerden etkilenmez. Fakat orijinal açıklayıcı değişkenler için katsayılar ve katkılar regresyon denklemine yeni açıklayıcı değişkenler eklendiğinde ya da denklemden açıklayıcı değişken silindiğinde önemli ölçüde değişebilir. Çoklu iç ilişki meydana geldiğinde genellikle bu durumla karşılaşmaktadır. Çoklu iç ilişki bir problem olarak görülmediğinde orijinal açıklayıcı değişkenler yerine temel bileşenlerle yapılan regresyon, hesaplama ve yorumlama açısından avantajlara sahiptir. Çoklu iç ilişki meydana geldiğinde temel bileşenler regresyonun en büyük avantajı; özellikle küçük varyansa sahip temel bileşenlerin alt kümesi silindiği zaman  $\beta$ 'nin daha durağan tahminleri elde edilebilir (Jolliffe, 2002). Bunu görebilmek için (32) nolu eşitlikte (33) nolu eşitlik yerine yazılırsa;

$$\hat{\beta} = A(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \quad (34)$$

$$= A\mathbf{L}^{-2}\mathbf{Z}'\mathbf{y}$$

$$= A\mathbf{L}^{-2}\mathbf{A}'\mathbf{X}'\mathbf{y}$$

$$= \sum_{k=1}^p l_k^{-1} \mathbf{a}_k \mathbf{a}_k' \mathbf{X}'\mathbf{y} \quad (35)$$

Burada  $l_k$ ,  $L^2$ 'nin  $k$ 'inci diagonal elemanıdır ve  $a_k$  ise  $\mathbf{A}$  matrisinin  $k$ 'inci sütunudur.  $(\mathbf{X}'\mathbf{X})^{-1}$  matrisinin spektral ayrışımı kullanılarak,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

yardımıyla (35) nolu eşitlik direkt olarak türetilir. Regresyon analizinin klasik varsayımlarından biri  $\mathbf{y}$  vektörünün elemanlarının ilişkisiz olması ve aynı varyanslı olmasıdır.  $\hat{\boldsymbol{\beta}}$ 'nin varyans-kovaryans matrisi (34) nolu eşitlik yardımıyla (36) nolu eşitlikteki gibi elde edilir.

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}) &= \text{Var}(\mathbf{A}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}) \\ &= \mathbf{A}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\text{Var}(\mathbf{y})\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{A}' \\ &= \sigma^2 \mathbf{A}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{A}' \\ &= \sigma^2 \mathbf{A}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{A}' \\ &= \sigma^2 \mathbf{A}\mathbf{L}^{-2}\mathbf{A}' \\ &= \sigma^2 \sum_{k=1}^p l_k^{-1} \mathbf{a}_k \mathbf{a}_k' \end{aligned} \quad (36)$$

Bu anlatım çoklu iç ilişkinin  $\hat{\boldsymbol{\beta}}$  vektörünün elemanları için nasıl büyük varyans ürettiğini açık bir şekilde ifade etmektedir. Çoklu iç ilişki problemi ortaya çıkarsa, temel bileşenlerin daha küçük varyans değerlerine sahip oldukları görülür. Diğer bir ifade ile sırasıyla temel bileşenlerin  $l_k$  değerlerinin monoton bir şekilde azalmasıdır. Dolayısıyla da  $l_k^{-1}$  çok büyük değerler alır. Bundan dolayı da (36) nolu eşitlikten çok küçük özdeğerlerle ilişkili olan temel bileşenlerin varyans değerleri çok büyük olur (Jolliffe, 2002).

Bu etkiyi azaltmanın bir yolu (35) nolu eşitlikte çok küçük  $l_k$  değerine karşılık gelen terimlerin silinmesidir. Bu yolla elde edilen bir tahmin edici (37) nolu eşitlikte verilmiştir.

$$\tilde{\boldsymbol{\beta}} = \sum_{k=1}^m l_k^{-1} \mathbf{a}_k \mathbf{a}_k' \mathbf{X}'\mathbf{y} \quad (37)$$

Burada  $l_{m+1}, l_{m+2}, \dots, l_p$  çok küçük özdeğerlerdir. Bu  $\boldsymbol{\gamma}$ 'nın son  $(p - m)$  elemanının sıfıra eşit olmasına denktir.

$\tilde{\boldsymbol{\beta}}$  tahmin edicisinin varyansı,

$$\sigma^2 \sum_{j=1}^m l_j^{-1} \mathbf{a}_j \mathbf{a}_j' \mathbf{X}' \mathbf{X} \sum_{k=1}^m l_k^{-1} \mathbf{a}_k \mathbf{a}_k'$$

şeklinde hesaplanmaktadır.

$\mathbf{X}' \mathbf{X}$  matrisinin spektral ayrışımını ( $\mathbf{X}' \mathbf{X} = \sum_{h=1}^p l_h \mathbf{a}_h \mathbf{a}_h'$ ) varyans ifadesinde yerine yazarsak;

$$\text{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2 \sum_{h=1}^p \sum_{j=1}^m \sum_{k=1}^m l_h l_j^{-1} l_k^{-1} \mathbf{a}_j \mathbf{a}_j' \mathbf{a}_h \mathbf{a}_h' \mathbf{a}_k \mathbf{a}_k'$$

$h = 1, 2, \dots, p$  için  $\mathbf{a}_h$  vektörleri ortonormal olduğundan dolayı  $h = j = k$  olduğunda varyans (38) nolu eşitlikteki gibi hesaplanır.

$$\text{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2 \sum_{k=1}^m l_k^{-1} \mathbf{a}_k \mathbf{a}_k' \quad (38)$$

İlk  $m$  tane özdeğerden hiçbirisi çok küçük değilse, (35) nolu eşitliğin diagonal elemanları ile gösterilen varyansların hiçbirisi büyük olmayacaktır. (37) nolu eşitlikte verilen  $\tilde{\boldsymbol{\beta}}$  tahmin edicisinin varyansındaki azalış  $\hat{\boldsymbol{\beta}}$ 'nin varyansı ile kıyaslandığında  $\tilde{\boldsymbol{\beta}}$  tahmin edicisinin yanlılık değerinin varyanstaki azalışı karşıladığı görülmektedir. Eğer çoklu iç ilişki ciddi bir problem ise varyanstaki azalış önemli olabilir, halbuki yanlılık nispeten küçük olabilir. Aslında, silinen elemanlara karşılık gelen  $\boldsymbol{\gamma}$ 'nın elemanları genellikle sıfırsa, yanlılık ortaya çıkmayacaktır. Bu durum aşağıdaki gibi özetlenebilir.

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \sum_{k=m+1}^p l_k^{-1} \mathbf{a}_k \mathbf{a}_k' \mathbf{X}' \mathbf{y}, \quad E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

ve

$$\begin{aligned}
E \left[ \sum_{k=m+1}^p l_k^{-1} \mathbf{a}_k \mathbf{a}_k' \mathbf{X}' \mathbf{y} \right] &= \sum_{k=m+1}^p l_k^{-1} \mathbf{a}_k \mathbf{a}_k' \mathbf{X}' \mathbf{X} \boldsymbol{\beta} \\
&= \sum_{k=m+1}^p \mathbf{a}_k \mathbf{a}_k' \boldsymbol{\beta}
\end{aligned}$$

En son verilen terim  $E(\tilde{\boldsymbol{\beta}}) \neq \boldsymbol{\beta}$  olduğundan dolayı sıfırdan farklıdır. Diğer bir ifade ile, küçük özdeğerlere karşılık (35) nolu denklemden silinen terimler veya  $\boldsymbol{\gamma}$ 'nın anlamsız olan elemanları çıkarılarak da yanlışlık ortadan kaldırılabilir. Burada önemli olan hangi elemanların sıfırdan farklı olduğunun bulunmasıdır. Bu durumda değişken seçme problemi karşımıza çıkar. Burada değişken olarak bahsedilen temel bileşenlerden çok orijinal açıklayıcı değişkenlerdir. Bunun için iyi bilinen değişken seçme yöntemleri kullanılabilir (Draper and Smith, 1998; Jolliffe, 2002). Temel bileşen regresyonuna ilişkin (30) ve (31) nolu eşitlikler (28) nolu doğrusal modelde ve (39) nolu denklemde tahmin edilen  $\boldsymbol{\beta}$  değerine eşittir.

$$\tilde{\boldsymbol{\beta}} = \sum_M l_k^{-1} \mathbf{a}_k \mathbf{a}_k' \mathbf{X}' \mathbf{y} \quad (39)$$

Burada M değeri açıklayıcı değişken sayısı  $p$ 'nin herhangi bir alt kümesidir. Temel bileşenler regresyonunun genel tanımında M ilk  $p$  sayının herhangi bir alt kümesi olabilir.

$n \geq p$  olduğu ve  $\mathbf{X}$  matrisini  $p$  ranklı olduğu varsayımı yapılırsa,  $\mathbf{X}$  matrisinin tekil değer ayrışımı aşağıdaki şekilde gösterilebilir.

$$\mathbf{X} = \mathbf{U} \mathbf{L} \mathbf{A}'$$

Burada, kullanılan  $\mathbf{A}$  ve  $\mathbf{L}$  matrisleri (29) ve (33) nolu eşitliklerde tanımlanmıştır. Ayrıca  $\mathbf{U}$  matrisinin sütunları  $\mathbf{X}'\mathbf{X}$  matrisinin sıfırdan farklı özdeğerlerine karşılık gelen özvektörlerinden oluşmaktadır.

Bu yüzden  $\mathbf{X}\boldsymbol{\beta}$ ,  $\mathbf{U} \mathbf{L} \mathbf{A}' \boldsymbol{\beta} = \mathbf{U} \boldsymbol{\delta}$  şeklinde yeniden yazılabilir. Burada,  $\boldsymbol{\delta} = \mathbf{L} \mathbf{A}' \boldsymbol{\beta}$  şeklindedir, bu yüzden  $\boldsymbol{\beta} = \mathbf{A} \mathbf{L}^{-1} \boldsymbol{\delta}$  olur.  $\hat{\boldsymbol{\beta}} = \mathbf{A} \mathbf{L}^{-1} \hat{\boldsymbol{\delta}}$  olmasına neden olan  $\boldsymbol{\delta}$ 'nin en küçük kareler tahmin edicisi,

$$\hat{\boldsymbol{\delta}} = (\mathbf{U}' \mathbf{U})^{-1} \mathbf{U}' \mathbf{y} = \mathbf{U}' \mathbf{y}$$

şeklindedir.

$\gamma$  ve  $\delta$  arasındaki ilişki,

$$\gamma = A' \beta = A' (AL^{-1} \delta) = (A' A) L^{-1} \delta = L^{-1} \delta$$

eşitliğinde kolaylıkla görülmektedir.

#### 4.2.1. Temel bileşenler regresyonunda bileşenlerin seçimi için yaklaşımlar

Çoklu iç ilişkinin neden olduğu büyük varyans problemini ortadan kaldırmak için temel olarak varyans değeri küçük olan bileşenler silinebilir. Fakat aynı zamanda, bağımlı değişken  $y$  ile yüksek korelasyona sahip bileşenlerin silinmesi istenmeyen bir durumdur. (39) nolu eşitlikte verilen  $M$  değerinin seçilmesindeki bir yaklaşım varyansı belirli bir değerden küçük olan bütün bileşenlerin silinmesidir. Burada bahsedilen değerlerin seçimi keyfidir. Fakat korelasyon matrisiyle ilgilenildiğinde özdeğerlerin ortalaması 1 ise bu değerlerin 0.01 ile 0.1 aralığında seçilmesi pratikte uygun görülmektedir.

Bu değerlerin seçimi için daha karmaşık bir yol ise  $p$  tane açıklayıcı değişken için varyans şişirme faktörlerine (VIF) bakmaktır. Standartlaştırılmış değişkenler kullanıldığında  $j$ 'inci değişken için VIF değeri  $c_{jj} / \sigma_2$  olarak tanımlanmaktadır. Aynı zamanda  $j$ 'inci değişkenin VIF değeri  $(X'X)^{-1}$  matrisinin  $j$ 'inci köşegen elemanına eşittir (Marquardt, 1970). Buradaki  $c_{jj}$  değeri  $\beta'$ 'nin en küçük kareler tahmin edicisinin  $j$ 'inci elemanının varyansıdır. Eğer bütün değişkenler ilişkisiz ise VIF değerleri 1'e eşittir. Eğer çoklu iç ilişki meydana gelirse, çoklu iç ilişkiye neden olan değişkenlerin VIF değerleri çok büyük olacaktır. (35) nolu eşitlikteki son birkaç terimin silinmesiyle elde edilen yanlış tahmin edicilerin VIF değerleri azalacaktır. VIF değerleri istenilen seviyede olana kadar silme işlemine devam edilir. Bir değişkenin VIF değeri o değişken ile diğer  $(p - 1)$  açıklayıcı değişken arasındaki  $R^2$  değeri ile ilişkilidir. Bundan dolayı da VIF değerinin ne kadar küçük seçileceği de  $R^2$  değeri ile ilişkilendirilebilir. Örneğin, VIF değerleri 10'dan büyük ise  $R^2$  değeri 0.90'dan büyüktür (Hill et. al.,1977).

Bazı arařtırmacılar (Hocking, 1976; Mosteller and Tukey, 1977; Gunst and Mason, 1980) regresyon denkleminde yer alacak temel bileřenlerin seřimini varyans indirgeme temelinde tartıřmıřlardır. Fakat, Jolliffe (1982) tarafından verilen örneęe bakıldıęında düşük varyanslı bileřenler tahmin deęerine sahip ise bu gibi prosedürler tehlikeli olabilir. Aynı zamanda Jolliffe bu tip örneklerin pratikte yaygın olmadıęını not etmiřtir. Berk's 1984 yılında 6 veri seti ile yaptıęı çalıřmasında tam ters bir sonuç bulmuřtur. Fakat kullandıęı veriler açıklayıcı deęiřkenler ile baęımlı deęiřken arasında güçlü pozitif korelasyonun olduęu özel verilerdir. Bu gibi durumlarda, ilk temel bileřen açıklayıcı deęiřkenlerin aęırlıklandırılmıř ortalamasıdır ve aęırlıklar pozitifdir, ayrıca baęımlı deęiřken açıklayıcı deęiřkenler ile pozitif yönde iliřkilidir.

Temel bileřenlerin varyansların büyüklüęüne göre seřilmesinin aksine diđer bir yöntem regresyon denklemindeki her bir bileřenin baęımsız katkısını ölçen t istatistięinin kullanılmasıdır. Mason ve Gunst (1985) yaptıkları çalıřmalarında düşük varyansa sahip temel bileřenlerin t istatistiklerinin yüksek varyansa sahip temel bileřenlerle karřılařtırıldıęında indirgeyici güce sahip olduklarını göstermiřlerdir. Varyansa dayalı seřim ile t testine dayalı seřim arasındaki ortak nokta ise en küçük varyansa sahip temel bileřenden başlayarak sırasıyla en küçük varyansa sahip bileřenlerin silinmesidir (Mason and Gunst, 1985).

Hill ve arkadaşları (1977) regresyon denkleminde silinecek temel bileřenlere karar vermek için çeřitli yöntemler önermiřlerdir. Önerdikleri önemli iki kriter vardır. Birinci kriter  $\hat{\beta}$ 'yi  $\beta$ 'ya yaklařtırmak, ikinci kriter ise,  $y$ 'nin tahminini  $y$  deęerine yada  $E(y)$  deęerine yaklařtırmaktır. Birinci durumda  $\beta$ 'nin tahmini ile ilgilenilirken ikinci durumda ise  $y$ 'nin tahmini ile ilgilenilmektedir.

Özel olarak, birinci kriterin saęlanması için iki yol vardır. Bunlar literatürde zayıf ve güçlü olarak sınıflandırılmaktadır. Zayıf olan kriter Wallace (1972) tarafından önerilmiřtir. Wallace göre,  $tr[MSE(\tilde{\beta})] \leq tr[MSE(\hat{\beta})]$  olduęunda  $\tilde{\beta}$ ,  $\hat{\beta}$ 'ya tercih edilir. Bunun anlamı  $\tilde{\beta}$  ile  $\beta$  arasındaki beklenen öklidyen uzaklık  $\hat{\beta}$  ile  $\beta$  arasındakine göre daha küçük olduęunda  $\tilde{\beta}$  tercih edilir. Güçlü kriter ise sıfırdan farklı  $p$  elemanlı her  $c$  vektörü için  $MSE(c'\tilde{\beta}) \leq MSE(c'\hat{\beta})$  kořuluna dayanmaktadır.



İkinci kriterin sağlanması içinde yine iki farklı yol önerilmiştir. Bunlardan ilki Wallace (1972) tarafından önerilmiştir. Wallace aşağıdaki koşul sağlanırsa  $\tilde{\beta}$ 'yi,  $\hat{\beta}$ 'a tercih etmektedir.

$$E[(\mathbf{X}\tilde{\beta} - \mathbf{X}\beta)'(\mathbf{X}\tilde{\beta} - \mathbf{X}\beta)] \leq E[(\mathbf{X}\hat{\beta} - \mathbf{X}\beta)'(\mathbf{X}\hat{\beta} - \mathbf{X}\beta)]$$

Alternatif bir MSE kriteri olarak tahmin edilen  $\hat{y}$  değeri ile gerçek  $y$  değeri arasındaki uzaklığa bakılır. Eğer,

$$E[(\mathbf{X}\tilde{\beta} - y)'(\mathbf{X}\tilde{\beta} - y)] \leq E[(\mathbf{X}\hat{\beta} - y)'(\mathbf{X}\hat{\beta} - y)]$$

koşulu sağlanıyorsa  $\tilde{\beta}$ ,  $\hat{\beta}$ 'ya tercih edilir.  $y = \mathbf{X}\beta + \varepsilon$  ifadesi yerine yazılırsa,

$$E[(\mathbf{X}\tilde{\beta} - y)'(\mathbf{X}\tilde{\beta} - y)] = E[(\mathbf{X}\tilde{\beta} - \mathbf{X}\beta)'(\mathbf{X}\tilde{\beta} - \mathbf{X}\beta)] + n\sigma^2$$

elde edilir. İlk bakışta, ikinci kriterin birinci kriterle eşit olduğu görülür. Fakat,  $\sigma^2$  bilinmediğinde denklemler sırasıyla  $\tilde{\beta}$ ,  $\hat{\beta}$  kullanılarak oluşturulduğunda farklı tahminler elde edilebilir.

Gunst ve Mason (1979), farklı regresyon denklemlerini karşılaştırmak için tahminlerin birleştirilmiş MSE değerlerini ele almışlardır. Friedman ve Montgomery (1985), Gunst ve Mason (1979) tarafından önerilen potansiyel gözlemlerin tahmin yeteneklerinin ortalamasını kullanmak yerine her bir gözlemin tahmin yeteneklerinin kullanılmasını önermişlerdir.

$y$ 'nin gözlenen ve beklenen değerlerini karşılaştırmanın diğer bir yolu çapraz geçerlilik yöntemidir. Martens (1995) temel bileşenler regresyonunda kaç tane bileşenin bulunduğuna karar vermek için bu kriteri kullanmıştır. Bu kriter,

$$\sum_{i=1}^n (y_i - \hat{y}_{M(i)})^2$$

şeklinde dir. Burada  $\hat{y}_{M(i)}$ ,  $\mathbf{X}_i$  matrisinin ( $\mathbf{X}$  matrisinin  $i$ 'inci sütunu silinerek elde edilen matris) kullanılmasıyla oluşturulan temel bileşenler regresyonundaki  $y_i$  değerlerinin tahminini ifade etmektedir. Tüm temel bileşenlerin hesaplanması için

gözlem değerlerinin sırasıyla silindiği bir algoritma vardır. Bu algorithmada korelasyon matrisi yerine kovaryans matrisi kullanılmaktadır.

Temel bileşenlerin silinmesindeki diğer bir yaklaşım Lott (1973) tarafından önerilmiştir. Bu yaklaşıma göre, varyans ve yanlılık aynı anda hesaba katılmaktadır. Bu yaklaşımda temel olarak düzeltilmiş çoklu belirlilik katsayısı hesaplanmaktadır.

$$\bar{R}^2 = 1 - \frac{(n-1)}{(n-p-1)}(1-R^2)$$

Bu yaklaşıma göre seçilecek en iyi alt küme düzeltilmiş çoklu belirlilik katsayısını maksimize eden kümedir. Lott yaptığı kısıtlı simülasyon çalışmalarında bu basit prosedürün çok iyi çalıştığını göstermiştir.

Temel bileşen regresyonunda olması gereken açıklayıcı değişken sayısı  $M$ 'nin belirlenmesine yönelik karar vermek oldukça zordur. Varyansın büyüklüğüne göre  $M$ 'nin belirlenmesi tavsiye edilmeyen bir yöntemdir. Küçük varyansa sahip temel bileşenlerin kullanılması da tehlikeli olabilir. Çünkü bunlar tahmin edilen regresyon denkleminde durağan olmayan sonuçlar verebilir.

Temel bileşenler regresyonu en küçük kareler regresyonunun başaramadığı çoklu iç ilişki problemi ile başa çıkabilmektedir. Bunun sonucunda regresyon katsayılarının durağan olmaması durumu ortadan kalkar.

### 4.3. Kısmi En Küçük Kareler

#### 4.3.1. Kısmi en küçük karelerin elde edilişi

Kısmi en küçük kareler yöntemindeki temel fikir ortogonal gizil değişkenler yardımıyla  $\mathbf{X}$  matrisi ile  $\mathbf{Y}$  matrisi arasındaki ilişkiyi bulmaktır. Kısmi en küçük kareler yönteminin ilk adımında  $\mathbf{X}$  ve  $\mathbf{Y}$  matrisleri arasındaki kovaryansı maksimum yapacak şekilde ortogonal gizil değişkenler hesaplanır (Garthwaite, 1994).

İki veri seti (değişkenler bloğu) arasındaki ilişkiyi modellemek için lineer olmayan kısmi en küçük kareler algoritmasının genel yapısı şu şekilde verilmektedir.  $\mathbf{X}$  matrisi  $n$  boyutlu değişkenler uzayında ilk veri setini (bloğu) benzer şekilde  $\mathbf{Y}$  matrisi

$m$  boyutlu değişkenler uzayında ikinci veri setini (bloğunu) gösterdiği varsayılır. Kısmi en küçük kareler skor vektörleri yardımıyla bu iki veri seti (bloğu) arasındaki ilişkiyi modellemeye çalışır. Her bir değişken bloğunda  $n$  gözlem elde edildikten sonra, kısmi en küçük kareler sıfır ortalamalı  $\mathbf{X}$  matrisini ve sıfır ortalamalı  $\mathbf{Y}$  matrisini aşağıdaki şekilde ayrıştırır.

$$\begin{aligned} \mathbf{X} &= \mathbf{TP}' + \mathbf{E} \\ \mathbf{Y} &= \mathbf{UQ}' + \mathbf{F} \end{aligned} \quad (40)$$

Burada,  $\mathbf{T}$  ve  $\mathbf{U}$  matrisleri  $n \times p$  tipinde hesaplanan skorların matrisidir.  $n \times p$  tipindeki  $\mathbf{P}$  matrisi ve  $m \times p$  tipindeki  $\mathbf{Q}$  matrisi de yüklerin matrisini göstermektedir. Ayrıca  $\mathbf{E}$  ve  $\mathbf{F}$  matrisleri sırasıyla  $n \times N$  ve  $n \times M$  tipinde rezidü matrislerini göstermektedir. (40) nolu eşitlikte gösterilen ayrıştırma 1989 yılında Martens ve Naes tarafından derinlemesine incelenmiştir. Kısmi en küçük kareler yönteminin klasik formu lineer olmayan yinelemeli kısmi en küçük kareler (NIPALS) algoritmasına dayanmaktadır. Kısmi en küçük kareler yöntemi aşağıdaki eşitliği sağlayacak şekilde  $\mathbf{w}$  ve  $\mathbf{c}$  yük matrislerini bulmaktadır.

$$[\text{cov}(\mathbf{t}, \mathbf{u})]^2 = \text{cov}[(\mathbf{X}\mathbf{w}, \mathbf{Y}\mathbf{c})]^2 = \max_{|r|=|s|=1} [\text{cov}(\mathbf{X}\mathbf{r}, \mathbf{Y}\mathbf{s})]^2 \quad (41)$$

Yukarıda verilen eşitlikte  $\text{cov}(\mathbf{t}, \mathbf{u}) = \mathbf{t}'\mathbf{u}/n$  skor vektörleri olan  $\mathbf{t}$  ve  $\mathbf{u}$  arasındaki kovaryansı tanımlamaktadır. NIPALS algoritması  $y$  uzayındaki herhangi bir  $\mathbf{u}$  skor vektörü ile başlar ve yakınsama sağlanana kadar aşağıdaki adımlar dizisini tekrar eder.

- 1)  $\mathbf{w} = \mathbf{X}'\mathbf{u}/(\mathbf{u}'\mathbf{u})$
- 2)  $\|\mathbf{w}\| \rightarrow 1$
- 3)  $\mathbf{t} = \mathbf{X}\mathbf{w}$
- 4)  $\mathbf{c} = \mathbf{Y}'\mathbf{t}/(\mathbf{t}'\mathbf{t})$
- 5)  $\|\mathbf{c}\| \rightarrow 1$
- 6)  $\mathbf{u} = \mathbf{Y}\mathbf{c}$

Eğer sadece tek bir tane bağımlı değişken var ise  $\mathbf{u} = \mathbf{y}$  olur. Bu durumla karşılaşıldığında ise NIPALS algoritmasında yakınsama tek iterasyonda sağlanır.  $\mathbf{w}$

ağırlık vektörünün aşağıda verilen özdeğer probleminin ilk özvektörüne karşılık geldiği gösterilebilir.

$$\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{w} = \lambda\mathbf{w} \quad (42)$$

Ayrıca benzer olarak,  $\mathbf{t}, \mathbf{u}$  ve  $\mathbf{c}$  vektörlerinin hesaplanması için özdeğer problemleri türetilir. Bu özdeğer problemlerinden birisini çözen bir araştırmacı NIPALS algoritmasında tanımlanan ilişkileri kullanarak skor ve ağırlık vektörlerini hesaplayabilir (Rosipal and Kramer, 2006).

Kısmi en küçük kareler modeli klasik NIPALS algoritmasını ya da çekirdek algoritmayı kullanarak hesaplanabilir. Fakat bunun farklı yolları da vardır. Bunlardan birisi Höskuldsson (1988) tarafından gösterilmiştir. Bu algoritmalar ardışık algoritmalarlardır. Bu algoritmalarda bir gizil vektör bir seferde hesaplanır ve sıradaki gizil vektörün hesaplanması için  $\mathbf{X}$  ve  $\mathbf{Y}$  matrisleri indirgenir.

Genellikle PLS algoritmasında her bir gizil vektör hesaplandıktan sonra  $\mathbf{X}$  ve  $\mathbf{Y}$  matrislerinin her ikisi birden indirgenir. Fakat, Höskuldsson (1988)  $\mathbf{Y}$  matrisini indirgemenin isteğe bağlı olduğunu göstermiştir. Lindgren ve arkadaşları bu avantajı kullanarak sadece  $\mathbf{X}$  matrisini indirgeyerek daha hızlı çekirdek algoritmayı tanımlamışlardır (Dayal and MacGregor, 1997). Bu algoritmalar ilerleyen bölümde daha ayrıntılı bir şekilde incelenecektir.

Fakat kısmi en küçük kareler regresyonunda kullanılan algoritmalara geçmeden önce  $\mathbf{X}$  ve  $\mathbf{Y}$  matrislerinin her ikisini birden indirgemek yerine bunlardan sadece birisinin indirgenmesinin yeterli olduğunu ispatını verelim (Dayal and MacGregor, 1997).

İspatı yapabilmek için öncelikle indirgeme işleminin nasıl yapıldığını gösteren aşağıdaki eşitliği bilmemiz gerekmektedir.

$$\mathbf{X}_{a+1} = \mathbf{X}_a - \mathbf{t}_a\mathbf{p}'_a \quad (43)$$

$$\mathbf{Y}_{a+1} = \mathbf{Y}_a - \mathbf{t}_a\mathbf{q}'_a \quad (44)$$

İndirgeme işlemini nasıl yapılacağını tanımladıktan sonra aşağıda verilen eşitliğin sağlandığını göstermeliyiz.

$$\mathbf{X}'_{a+1}\mathbf{Y}_{a+1} = \mathbf{X}'_a\mathbf{Y}_{a+1} = \mathbf{X}'_{a+1}\mathbf{Y}_a \quad (45)$$

(43) ve (44) nolu eşitlikler kullanılarak  $\mathbf{X}'_{a+1}\mathbf{Y}_{a+1}$  aşağıdaki şekilde genişletilebilir.

$$\begin{aligned} \mathbf{X}'_{a+1}\mathbf{Y}_{a+1} &= (\mathbf{X}_a - \mathbf{t}_a\mathbf{p}'_a)'(\mathbf{Y}_a - \mathbf{t}_a\mathbf{q}'_a) \\ &= \mathbf{X}'_a\mathbf{Y}_a - \mathbf{X}'_a\mathbf{t}_a\mathbf{q}'_a - \mathbf{p}_a\mathbf{t}'_a\mathbf{Y}_a + \mathbf{p}_a\mathbf{t}'_a\mathbf{t}_a\mathbf{q}'_a \quad (46) \\ &= \mathbf{X}'_a\mathbf{Y}_a - \mathbf{X}'_a\mathbf{t}_a\mathbf{q}'_a - \mathbf{p}_a\mathbf{t}'_a\mathbf{Y}_a + \mathbf{p}_a\mathbf{q}'_a\mathbf{t}'_a\mathbf{t}_a \end{aligned}$$

Kısmi en küçük karelerde kullanılan  $\mathbf{q}_a$  ve  $\mathbf{p}_a$  vektörlerine ilişkin birimler sırasıyla (47) ve (48) da verildiği gibi hesaplanmaktadır.

$$\mathbf{q}_a = \frac{\mathbf{Y}'_a\mathbf{t}_a}{\mathbf{t}'_a\mathbf{t}_a} \quad (47)$$

$$\mathbf{p}_a = \frac{\mathbf{X}'_a\mathbf{t}_a}{\mathbf{t}'_a\mathbf{t}_a} \quad (48)$$

(47) ve (48) nolu eşitlikleri yeniden düzenlersek

$$\mathbf{t}'_a\mathbf{Y}_a = \mathbf{q}'_a(\mathbf{t}'_a\mathbf{t}_a) \quad (49)$$

$$\mathbf{X}'_a\mathbf{t}_a = (\mathbf{t}'_a\mathbf{t}_a)\mathbf{p}_a \quad (50)$$

(49) ve (50) nolu yeni eşitlikleri elde ederiz.

- I. İsteğe bağlı olan  $\mathbf{Y}$  matrisinin indirgenmesinin ispatı (50) nolu eşitliğin (46) nolu eşitlikte yerine konulmasıyla elde edilir.

$$\begin{aligned} \mathbf{X}'_{a+1}\mathbf{Y}_{a+1} &= \mathbf{X}'_a\mathbf{Y}_a - \mathbf{p}_a\mathbf{q}'_a(\mathbf{t}'_a\mathbf{t}_a) - \mathbf{p}_a\mathbf{t}'_a\mathbf{Y}_a + \mathbf{p}_a\mathbf{q}'_a(\mathbf{t}'_a\mathbf{t}_a) \quad (51) \\ &= \mathbf{X}'_a\mathbf{Y}_a - \mathbf{p}_a\mathbf{t}'_a\mathbf{Y}_a \\ &= (\mathbf{X}'_a - \mathbf{p}_a\mathbf{t}'_a)\mathbf{Y}_a \\ &= (\mathbf{X}_a - \mathbf{t}_a\mathbf{p}'_a)'\mathbf{Y}_a \\ &= \mathbf{X}'_{a+1}\mathbf{Y}_a \end{aligned}$$

- II. İsteğe bağlı olan  $\mathbf{X}$  matrisinin indirgenmesinin ispatı (49) nolu eşitliğin (46) nolu eşitlikte yerine konulmasıyla elde edilir.

$$\begin{aligned}
X'_{a+1}Y_{a+1} &= X'_aY_a - X'_at_aq'_a - p_aq'_a(t'_at_a) + p_aq'_a(t'_at_a) \quad (52) \\
&= X'_aY_a - X'_at_aq'_a \\
&= X'_a(Y_a - t_aq'_a) \\
&= X'_aY_{a+1}
\end{aligned}$$

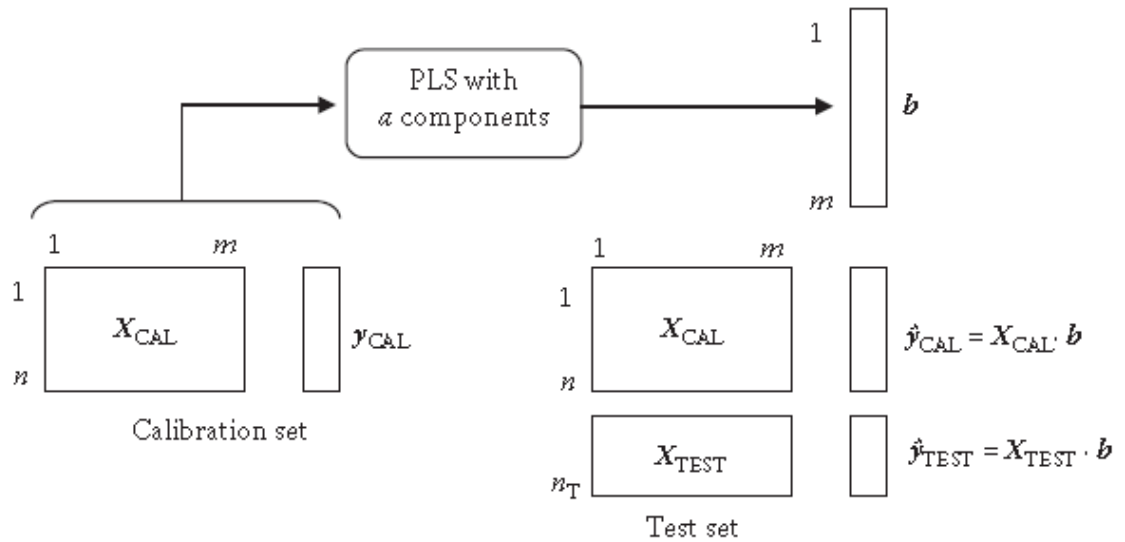
### 4.3.2. Kısmi En Küçük Kareler Regresyonu

Regresyon analizi istatistiksel analizin çok önemli bir alanıdır. Çoklu lineer regresyon analizinde katsayılar vektörünün en küçük kareler tahmin edicisi yansız tahmin ediciler arasında minimum varyansa sahiptir. Çoklu iç ilişki problemi ile karşılaşıldığı zaman katsayılar vektörünün varyansı aşırı derecede büyümektedir. Bundan dolayı en küçük kareler yöntemine alternatif olarak birçok yanlı regresyon yöntemi önerilmiştir. Yanlı regresyon yöntemlerinin amacı biraz yansızlıktan vazgeçerek tahminlerin hata kareler ortalamasını düşürmektir. Bu amaçla geliştirilen regresyon yöntemlerinden birisi de kısmi en küçük kareler regresyonudur. Kısmi en küçük kareler regresyonu da Temel Bileşenler Regresyonu ve Ridge Regresyon gibi yanlı bir regresyon yöntemidir. Bu yöntem regresyon analizi içerisinde kendisine birçok uygulama alanı bulmuştur. Kısmi en küçük kareler yönteminin regresyon analizindeki gelişimi aşağıdaki gibi özetlenebilir (Kondylis, 2006).

1. Wold (1966,1975): Lineer olmayan yinelemeli kısmi en küçük kareler algoritması (NIPALS)
2. Wold et al. (1983): Tek değişkenli kısmi en küçük kareler regresyonu (PLS1)
3. Höskuldsson (1988): Kısmi en küçük kareler regresyon algoritmaları
4. Helland (1988): Kısmi en küçük kareler regresyonu ve Krylov uzayları
5. Stone ve Brooks (1990): Sürekli Regresyon
6. Phatak (1993): Kısmi en küçük kareler regresyonu ve Lanczos-Conjugate yöntemi
7. Rosipal ve Trejo (2001): Çoğaltılmış çekirdek Hilbert uzaylarında çekirdek kısmi en küçük kareler
8. Bastien et al. (2005): Kısmi en küçük kareler geliştirilmiş regresyon

Kısmi en küçük kareler regresyonu ilk olarak bir çeşit kalibrasyon yöntemi olarak Wold (1975) tarafından incelenmiştir. Kısmi en küçük kareler regresyonu kendisine kemometri literatüründe önemli bir yer bulmuştur. Bu yöntemin istatistiksel özellikleri Wold (1984), Manne (1987), Naes ve Martens (1986), Helland (1988) ve Stone ve Brook (1990) gibi bir çok istatistikçi ve matematikçinin destekleriyle netlik kazanmaya başlamıştır.

Genel olarak PCR ile PLSR benzer yapıdadırlar. Her iki yöntemde de açıklayıcı değişkenler yeni bileşenlere dönüştürülerek elde edilen bu yeni bileşenlerle bağımlı değişken vektörüne EKK uygulanmaktadır. Fakat iki yöntem arasındaki temel fark; PCR bileşenleri elde ederken sadece açıklayıcı değişkenlerden yararlanmaktadır. Bunun aksine PLS bileşenleri elde ederken açıklayıcı değişkenlerden yararlanmanın yanı sıra bağımlı değişken ile maksimum kovaryansı sağlayacak şekilde bileşenler üretmektedir (Naes and Martens, 1986). PLS sürekli regresyonun özel bir halidir (Stone and Brooks, 1990). PLS’de kullanılan algoritmaya bağlı olarak, yük vektörleri ortogondur yada skor vektörleri ilişkisizdir. PLSR Şekil 4.3’teki gibi özetlenebilir.



**Şekil4.3. Çoklu lineer regresyon yöntemi olarak PLS (Varmuza and Filzmoser, 2008).**

- PLS güçlü bir lineer regresyon yöntemidir. Ayrıca ilişkili verilere karşı duyarsızdır ve çok fazla sayıda değişken içinde işlem yapabilir.
- PLS'de model oluşturulurken daha az sayıda bileşen hesaplanır.
- PLS'deki bileşenlerin sayısı model yeterliliğine karar verir ve yüksek tahmin performansına sahiptir.

Standart PLS'de bileşen skorları ilişkisizdir ve yük vektörlerine karşı ortogondur. Modele bileşen eklenmesi modelin tahmin yeteneğini artırır. Ayrıca modelde bulunması gereken değişken sayısı genellikle çapraz geçerlilik (CV) yardımıyla hesaplanır. PLS'de bileşenler elde edilirken  $y$  bağımlı değişkeni ile yüksek derecede ilişkili bileşenler oluşturulduğundan PLS'deki bileşen sayısı PCR'a göre daha azdır. Diğer bir yandan PLS modelleri PCR modellerine göre daha az duranıdır. Bunun nedeni ise PLS modelinde daha az varyans açıklanmasıdır. PLS modellerinde daha çok bileşen kullanılırsa PCR ile benzer sonuçlar elde edilebilir. PLS algoritmalarının karmaşık olan kısmı bileşenler hesaplanırken adım adım hesaplanmalarıdır. Bir sonraki bileşen hesaplanırken  $X$  matrisinin artık matrisleri bulunur ve bileşen bu artık matrisi yardımıyla hesaplanır. Bileşen artık matrisinden hesaplandığından dolayı da  $X$  matrisi ile ilişkili değildir. Ancak  $x$  verisine ait PLS- $x$ -yüklerine ve PLS- $x$ -skorlarına ilişkin var olan eşitlik de orijinal  $x$  verisine ait son modeldeki regresyon katsayılarını sağlar.

Literatürde PLS'den genel olarak belirli kısıtlar altında amaç fonksiyonunu maksimize eden nümerik bir algoritma olarak bahsedilmektedir. Algoritmadaki amaç fonksiyonu  $x$  ve  $y$  skorları arasındaki kovaryanstır ve kısıt genellikle skorların ortogonal olmasıdır.

PLS2 regresyonunda bağımlı ve açıklayıcı değişkenlerin çok değişkenli oldukları varsayılmaktadır. Burada kullanılan  $X$  matrisi  $n \times m$  boyutludur,  $Y$  matrisi ise  $n \times q$  boyutludur. PLS1 regresyonunda ise bağımlı değişken bir vektördür. PLS2 regresyonunda amaç  $m \times q$  boyutlu  $B$  regresyon katsayıları vektörü yardımıyla (53) nolu eşitlikte gösterildiği gibi  $X$  ve  $Y$  matrisleri arasında lineer bir ilişki bulmaktır.

$$Y = XB + E \quad (53)$$



Bu ilişki direk olarak  $\mathbf{X}$  ve  $\mathbf{Y}$  matrislerini kullanmak yerine  $\mathbf{X}$  ve  $\mathbf{Y}$  matrislerinin gizil değişkenler yardımıyla (54) ve (55) nolu eşitliklerde görüldüğü gibi modellenmesi ile bulunmaktadır.

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}_X \quad (54)$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{E}_Y \quad (55)$$

Burada  $\mathbf{T}$  ve  $\mathbf{U}$  skor matrislerini,  $\mathbf{P}$  ve  $\mathbf{Q}$  yük matrislerini,  $\mathbf{E}_X$  ve  $\mathbf{E}_Y$  ise artık matrislerini göstermektedir.  $\mathbf{T}$  matrisindeki x skorları açıklayıcı değişkenlerin lineer kombinasyonlarıdır. Benzer şekilde  $\mathbf{U}$  matrisindeki y skorları bağımlı değişkenlerin lineer kombinasyonlarıdır. Bu skorlar açıklayıcı değişkenleri ve bağımlı değişkenleri iyi bir şekilde özetlemektedir.  $\mathbf{t}_j$ ,  $\mathbf{u}_j$ ,  $\mathbf{p}_j$  ve  $\mathbf{q}_j$  sırasıyla  $\mathbf{T}$ ,  $\mathbf{U}$ ,  $\mathbf{P}$  ve  $\mathbf{Q}$  matrislerinin j'inci sütunlarıdır ( $j = 1, 2, \dots, a$ ). Ek olarak x ve y skorları lineer içsel bir ilişki ile birbirlerine bağlıdır. Bu içsel ilişki (56) nolu eşitlikte gösterildiği gibidir.

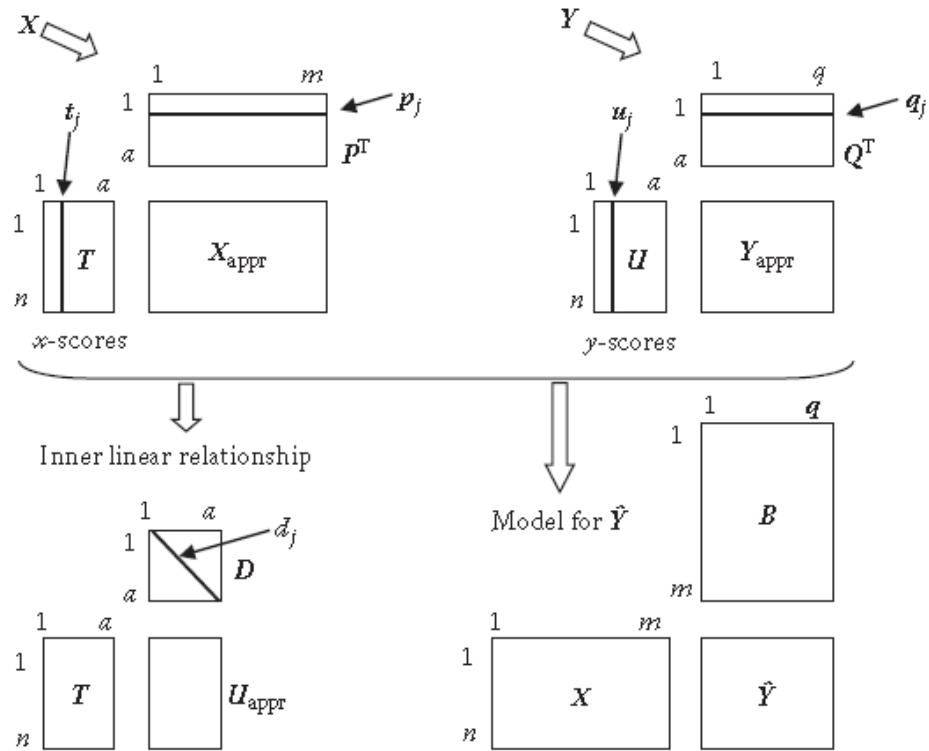
$$\mathbf{u}_j = d_j \mathbf{t}_j + \mathbf{h}_j \quad (56)$$

Burada  $d_j$  regresyon parametresi ve  $\mathbf{h}_j$  artıklardır. Eğer  $\mathbf{u}_1$  ile  $\mathbf{t}_1$  arasındaki kuvvetli bir ilişki var ise, birinci PLS bileşeninin skorları y skorları için iyi bir tahmin edicidir, dolayısıyla bağımlı değişkenleri iyi bir şekilde tahmin eder. Genellikle  $\mathbf{X}$  matrisi yardımıyla  $\mathbf{Y}$  matrisinin modellenmesi için birden çok PLS bileşeni kullanılmaktadır. İdeal bileşen sayısının belirlenmesi için birçok yöntem vardır. Bu yöntemler ilerleyen bölümlerde açıklanacaktır. Skorlar arasındaki ilişki (57) nolu eşitlikte gösterildiği gibidir.

$$\mathbf{U} = \mathbf{TD} + \mathbf{H} \quad (57)$$

Burada  $\mathbf{D}$  elemanları  $d_1, d_2, \dots, d_a$  olan diagonal matristir ve  $\mathbf{H}$  sütunları  $h_j$  olan artık matrisidir. PLS'nin matris gösterimi Şekil4'te verilmiştir. Kısmi en küçük kareler denmesinin sebebi ise regresyon için x ve y verilerindeki kısmi bilginin kullanılmasıdır. PLS1'de y skorları yoktur, bundan dolayı (57) nolu eşitlik (58) nolu eşitlik gibi yazılabilir.

$$\mathbf{y} = \mathbf{Td} + \mathbf{h} \quad (58)$$



Şekil4.4. PLS'nin matris gösterimi (Varmuza and Filzmoser, 2008).

#### 4.3.3. Kısmi En Küçük Kareler Regresyonunda Kullanılan Algoritmalar

Wold tarafından geliştirilen kısmi en küçük kareler algoritmasının birçok çeşidi bulunmaktadır. Bu algoritmalarının amacı hesaplama hızını arttırmaktır. Genellikle algoritmalarda kullanılan yöntem daha kullanışlı varyans-kovaryans matrislerini ve ilişki matrislerini hesaplamaktır (Lindgren and Rännar, 1998). Bu algoritmalarından hangisinin kullanılacağı konusu veri matrisinin şekliyle ilişkilidir. Literatürde NIPALS algoritmasıyla başlayan kısmi en küçük kareler regresyonunda birçok algoritma önerilmiştir. Bu algoritmaların tarihsel gelişimi aşağıda verilmiştir (Kondylis, 2006).

- Ortogonal skorlar kısmi en küçük kareler regresyonu (Wold et al., 1983,1984)
- Ortogonal yükler kısmi en küçük kareler regresyonu (Martens, 1985)
- Özvektör algoritması (Höskuldsson, 1988)
- Helland algoritması (Helland, 1988)
- SIMPLS algoritması (De Jong, 1993)
- SAMPLS algoritması (Bush ve Nachbar, 1993)

- Çekirdek algoritması (Rännar, 1994)
- UNIPALS algoritması (Glen, 1998)

Bu bölümde amaç geliştirilen bazı algoritmaları tanıtmak ve standart algoritma olarak bilinen NIPALS (Lineer olmayan yinelemeli kısmi en küçük kareler algoritması) algoritması ile karşılaştırmaktır.

#### **4.3.3.1. Lineer olmayan yinelemeli kısmi en küçük kareler algoritması (NIPALS)**

Literatürde klasik kısmi en küçük kareler algoritması olarak bilinen bu algoritma 1966 yılında Herman Wold tarafından geliştirilmiştir. Daha sonra Wold (1975) algoritmada bağımlı değişkenin ya da değişkenlerinde hesaba katılması için algoritmayı geliştirmiştir. NIPALS algoritması modelin boyutlarını hesaplamak için iteratif bir yöntem kullanmaktadır. Bu algoritmanın temel amacı kovaryans matrisini en çoklayan bileşenleri elde etmektir. Algoritmada bileşenlerin aynı anda elde edilmeleri olanaksızdır. Bu yüzden bileşenlerin elde edilebilmesi için yineleme yöntemi uygulanmaktadır. Fakat bağımlı değişkenimiz bir tane olduğu durumlarda algoritma yinelemeli değildir. Algoritmanın yakınsama kriteri istenilen sayıda bileşen elde edilmesi ya da  $\mathbf{X}$  açıklayıcı değişkenler matrisinin sıfır olmasıdır. NIPALS algoritması aşağıdaki adımlardan oluşmaktadır. Burada öncelikle  $\mathbf{X}$  ve  $\mathbf{Y}$  matrislerinin standartlaştırıldığı varsayılmaktadır.

- (1) Bağımlı değişken sayısı birden çok olduğu durumlarda bağımlı değişkenler matrisinin herhangi bir sütunu başlangıç vektörü olarak alınır. Eğer bağımlı değişken sayısı tek ise direkt o vektör alınır. ( $\mathbf{u}=\mathbf{y}$ )
- (2)  $\mathbf{X}$ 'in  $\mathbf{u}=\mathbf{y}$  üzerine regresyonundan  $\mathbf{X}$  ve  $\mathbf{y}$  arasındaki kovaryansı en çok açıklayan  $\mathbf{w}$  ağırlık vektörü elde edilir. ( $\mathbf{w} = \mathbf{X}'\mathbf{u}/\mathbf{u}'\mathbf{u}$ )
- (3)  $\mathbf{w}$  ağırlık vektörü birim uzunluğa sahip olacak şekilde ölçeklendirilir. ( $\mathbf{w}_y = \mathbf{w}/\|\mathbf{w}\|$ )
- (4)  $\mathbf{X}$  matrisinin skorları olan  $\mathbf{t}$  vektörü hesaplanır. ( $\mathbf{t} = \mathbf{X}\mathbf{w}_y$ )
- (5)  $\mathbf{t}$  vektörün  $\mathbf{y}$  bağımlı değişkenini modellemedeki katkısını açıklayacak  $\mathbf{c}$  ağırlık vektörü bulunur. ( $\mathbf{c} = \mathbf{u}'\mathbf{t}/\mathbf{t}'\mathbf{t}$ )

- (6)  $\mathbf{c}$  ağırlık vektörü birim uzunluğa sahip olacak şekilde ölçeklendirilir. ( $\mathbf{c}_y = \mathbf{c}/\|\mathbf{c}\|$ )
- (7)  $\mathbf{u}$  vektörü  $\mathbf{c}$  ağırlık vektörü ile  $\mathbf{Y}$  nin doğrusal bir kombinasyonu olacak şekilde hesaplanır. ( $\mathbf{u}_y = \mathbf{Y}\mathbf{c}_y/\mathbf{c}_y'\mathbf{c}_y$ )
- (8) Algoritmada yeniden hesaplanan  $\mathbf{u}_y$  vektörü ile  $\mathbf{u}$  vektörü arasında bir yakınsama olup olmadığına bakılır. Bu yakınsamayı test etmek için  $\|\mathbf{u}_y - \mathbf{u}\|/\|\mathbf{u}_y\| < 10^{-6}$  değerinden yararlanılır. Eğer bu değer sağlanıyor ise algoritmanın sonraki adımlarına geçilir, sağlanmıyor ise 7. adımda hesaplanan  $\mathbf{u}_y$  değeri 2. adımda yerine konularak algoritma yeniden işletilir.
- (9)  $\mathbf{X}$  matrisinin  $\mathbf{t}$  vektörü üzerine regresyonundan  $\mathbf{p}$  yük vektörü hesaplanır. Hesaplanan  $\mathbf{p}$  yük vektörü,  $\mathbf{t}$  vektörünün açıklayıcı değişken üzerindeki etkisini ifade etmektedir. ( $\mathbf{p} = \mathbf{X}'\mathbf{t}/(\mathbf{t}'\mathbf{t})$ )
- (10)  $\mathbf{Y}$  'nin  $\mathbf{u}$  vektörü üzerine regresyonundan  $\mathbf{q}$  yük vektörü hesaplanır. Hesaplanan  $\mathbf{q}$  yük vektörü,  $\mathbf{u}$  vektörünün bağımlı değişken üzerindeki etkisini ifade etmektedir. ( $\mathbf{q} = \mathbf{Y}'\mathbf{u}_y/(\mathbf{u}_y'\mathbf{u}_y)$ )
- (11)  $\mathbf{X}$  ve  $\mathbf{Y}$  matrisleri için bileşenler ayrı ayrı hesaplandığından, hesaplanan bu bileşenler arasında zayıf bir ilişki meydana gelmektedir. Bu durumu önlemek için  $\mathbf{Y}$ 'nin bileşeni  $\mathbf{u}$ 'nun  $\mathbf{X}$ 'in bileşeni  $\mathbf{t}$  üzerine regresyonundan elde edilecek olan  $\mathbf{b}$  katsayısı hesaplanır. Hesaplanan bu  $\mathbf{b}$  katsayısı içsel ilişkiyi tanımlamaktadır.
- (12) Algoritmada elde edilen bileşenler ve yükler açıklayıcı ve bağımlı değişkeni modelleme de kullanılmaktadır. Bu modelleme için aşağıdaki ifadeler kullanılmaktadır.

$$\mathbf{X} = \mathbf{TP}' \quad \mathbf{Y} = \mathbf{BTC}'$$

Ayrıca algoritmanın bu adımında bir sonraki bileşeni elde etmek için kullanılacak olan artık matrisleride hesaplanmaktadır. Bu hesaplama aşağıda verilen şekilde gerçekleşmektedir.

$$\mathbf{X} \rightarrow \mathbf{X} - \mathbf{tp}' \quad \mathbf{Y} \rightarrow \mathbf{Y} - \mathbf{btc}'$$

Bağımlı değişkendeki değişimin büyük bir kısmı açıklandığında algoritma sonlandırılır.

İlk olarak Höskuldsson NIPALS algoritmasını bir özdeğer/özvektör problemi olarak ele almıştır. Höskuldsson NIPALS algoritmasında hesaplanan skor ve ağırlık vektörlerinin ( $\mathbf{w}$ ,  $\mathbf{c}$ ,  $\mathbf{t}$ ,  $\mathbf{u}$ ), karesel varyans-kovaryans matrislerinin kümesi için özvektör olarak hesaplanabileceğini aşağıdaki eşitlikler yardımıyla göstermiştir.

$$\mathbf{w}a_1 = (\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X})\mathbf{w} \quad (59)$$

$$\mathbf{c}a_2 = (\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y})\mathbf{c} \quad (60)$$

$$\mathbf{t}a_3 = (\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}')\mathbf{t} \quad (61)$$

$$\mathbf{u}a_4 = (\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}')\mathbf{u} \quad (62)$$

Eşitliklerde kullanılan  $a_1, a_2, a_3$  ve  $a_4$  değerlerinin hepsi özdeğerlerdir. Ayrıca  $\mathbf{w}$ ,  $\mathbf{c}$ ,  $\mathbf{t}$  ve  $\mathbf{u}$  matrislerinin her birinin normu bire eşittir. (59)-(62) eşitliklerinde kullanılan matrislerin avantajı bu matrislerin boyutları ile ilgilidir. (59) numaralı eşitlikteki matrisin boyutu  $p \times p$ , (60) numaralı eşitlikte kullanılan matrisin boyutu ise  $m \times m$  dir. Burada  $m$  bağımlı değişkenler matrisindeki değişken sayısını göstermektedir. (59) ve (60) eşitliklerinde kullanılan matrislerin boyutları gözlem sayılarından bağımsızdır, bu matrislerin boyutları sadece değişken sayılarına bağlıdır. Bu durumun aksine (61) ve (62) eşitliklerinde kullanılan matrislerin boyutları gözlem sayılarına bağlıdır. Bu durumda da  $\mathbf{X}$  ve  $\mathbf{Y}$  matrislerindeki değişken sayıları matrislerin boyutlarını etkilememektedir. Sonuç olarak, çok fazla gözlemlili matrisler ya da çok fazla değişkene sahip matrisler algoritmanın çalışması için önemli olan tüm bilgiyi içeren küçük boyutlu matrislerle özetlenebilir (Lindgren and Rännar, 1998).

Bu algoritmada ortogonallik  $\mathbf{t}_i$  vektörlerinin  $\mathbf{E}_i$  artık matrislerinin lineer kombinasyonları olarak hesaplanmasıyla sağlanır. Yani bir başka deyişle;

$$\mathbf{t}_i = \mathbf{E}_{i-1}\mathbf{w}_i, \quad \mathbf{E}_i = \mathbf{X} - \sum_{j=1}^i \mathbf{t}_j\mathbf{p}_j', \quad \mathbf{E}_0 = \mathbf{X} \quad (63)$$

şeklinde dir. Burada  $\mathbf{w}_i$  yük vektörleri ortonormaldir. Çok değişkenli ve tek değişkenli bir çok kısmi en küçük kareler algoritmasında ilk adım  $\mathbf{t}_i$  vektörlerinin lineer kombinasyonlarını hesaplayabilmek için  $\mathbf{w}_i$  ağırlık vektörlerinin hesaplanmasıdır. Daha sonra  $\mathbf{X}'$ 'in  $\mathbf{t}_i$  üzerine regresyonundan  $\mathbf{p}_i$  yükleri hesaplanır.  $m$  tane faktör hesaplandıktan sonra aşağıdaki ilişkiler yazılabilir.

$$\mathbf{T}_m = \mathbf{X}\mathbf{R}_m \quad (64)$$

$$\mathbf{P}_m = \mathbf{X}'\mathbf{T}_m(\mathbf{T}_m'\mathbf{T}_m)^{-1} \quad (65)$$

$$\mathbf{R}_m = \mathbf{W}_m(\mathbf{P}_m'\mathbf{W}_m)^{-1} \quad (66)$$

Buradaki ilk  $m$  faktör  $\mathbf{X}$  matrisindeki varyansın büyük çoğunluğunu açıklamaktadır. (66) nolu eşitlik lineer dönüşümler yardımıyla ağırlık vektörlerinin iki seti arasında bağlantı kurmaktadır. (64) ve (65) nolu eşitliklerden  $\mathbf{R}_m'\mathbf{P}_m'$ ' nin  $\mathbf{I}_m$  'ye eşit olacağı kolayca görülmektedir.

$$\mathbf{R}_m'\mathbf{P}_m = \mathbf{R}_m'\mathbf{X}'\mathbf{T}_m(\mathbf{T}_m'\mathbf{T}_m)^{-1} = \mathbf{T}_m'\mathbf{T}_m(\mathbf{T}_m'\mathbf{T}_m)^{-1} = \mathbf{I}_m \quad (67)$$

$m$  boyut hesaplandıktan sonra, kısmi en küçük kareler ile elde edilen değerler vektörü  $\mathbf{T}_m$  ile gösterilebilir. Buradan (68) eşitliği elde edilir.

$$\hat{\mathbf{y}}_{PLS}^m = \mathbf{T}_m(\mathbf{T}_m'\mathbf{T}_m)^{-1}\mathbf{T}_m'\mathbf{y} \quad (68)$$

(68) eşitliğinde  $\mathbf{T}_m$  yerine  $\mathbf{X}\mathbf{R}_m$  ve  $\mathbf{y}$  yerine  $\mathbf{X}\hat{\boldsymbol{\beta}}_{OLS}$  yazılırsa

$$\hat{\mathbf{y}}_{PLS}^m = \mathbf{X}\mathbf{R}_m(\mathbf{R}_m'\mathbf{X}'\mathbf{X}\mathbf{R}_m)^{-1}\mathbf{R}_m'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{OLS} \quad (69)$$

eşitliği elde edilir. (69) eşitliğine bakılırsa;

$$\hat{\boldsymbol{\beta}}_{PLS}^m = \mathbf{R}_m(\mathbf{R}_m'\mathbf{X}'\mathbf{X}\mathbf{R}_m)^{-1}\mathbf{R}_m'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{OLS} \quad (70)$$

olarak yazılabileceği açıkça görülmektedir. (64) nolu eşitlik (65) nolu eşitlikte yerine konular ve elde edilen sonuç (70) nolu eşitlikte kullanılırsa  $\hat{\boldsymbol{\beta}}_{PLS}$  için daha basit bir ifade aşağıdaki gibi elde edilir (Yeniay ve Göktaş, 2002).

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{PLS}^m &= \mathbf{R}_m(\mathbf{R}_m'\mathbf{X}'\mathbf{X}\mathbf{R}_m)^{-1}\mathbf{R}_m'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{OLS} \\ &= \mathbf{R}_m\mathbf{P}_m'\hat{\boldsymbol{\beta}}_{OLS} \\ &= \mathbf{W}_m(\mathbf{P}_m'\mathbf{W}_m)^{-1}\mathbf{P}_m'\hat{\boldsymbol{\beta}}_{OLS} \end{aligned} \quad (71)$$

NIPALS algoritmasında bir başka yaklaşım orijinal değişken olarak PLS faktörlerinin kullanılmasıdır. Bu durum kısaca aşağıdaki gibi açıklanabilir.

$w_a$  ( $a = 2, 3, \dots, A$ ) ağırlık matrislerinin her birisi faktör skorlarını tanımlamak için kullanıldı. Bu ifade,

$$t_a = X_{a-1} w_a, a=1, \dots, A \quad (72)$$

şeklindedir. Burada  $X_0$  orijinal merkezleştirilmiş veri değildir. İndirgenmiş  $X_a$  matrisinde indirgeme işlemi devam ettiğinde bilgi kaybı artmaktadır bundan dolayı da faktörlerin yorumlanması da zorlaşmaktadır. İlk faktörde bazı  $X$  değişkenleri kullanılıyorken diğerlerinde daha başkaları kullanılmaktadır. Faktörler ile değişkenler arasındaki ilişki  $p_a$  yük vektörleri tarafınan daha iyi bir şekilde gösterilebilir.  $W$  ağırlık matrisinin kısmi en küçük kareler regresyon modelinde yük vektörlerine göre daha az kullanıldığı söylenebilir. Bundan dolayı da orijinal merkezleştirilmiş veri matrisi  $X_0$  yerine  $t_a$  yük vektörlerinin ifade edilmesi avantajlıdır.

$$t_a = X_0 r_a, a=1, \dots, A \quad (73)$$

yada ağırlık vektörlerinin matrisi  $R \equiv [r_1, r_2, \dots, r_A]$  olmak üzere;

$$T = X_0 R \quad (74)$$

şeklindedir. İndirgenmiş  $X$  matrisleri yardımıyla hesaplanan  $T$  faktör skorları, standartlaştırılmış  $X$  matrislerinin lineer kombinasyonu olarak ifade edilebilir. İndirgenmiş  $X_a$  matrislerinin tamamı ve  $t_a$  faktör skorları  $X_0$ 'ın sütun uzayında yer almaktadır. Bundan dolayı  $R$  matrisi  $T$ 'nin  $X_0$  üzerine regresyonundan aşağıdaki şekilde elde edilebilir.

$$R = X_0^+ T = (X_0' X_0)^- X_0' T = (X_0' X_0)^- P (T' T)^{-1} \quad (75)$$

Burada  $P = [p_1, p_2, \dots, p_A]$  faktör yükleri matrisini, (-) genelleştirilmiş inversi ve (+) ise tek Moore-Penrose inversini göstermektedir.  $r_b' p_a = \frac{r_b' X_0' t_a}{t_a' t_a} = t_b' t_a / (t_a' t_a) = \delta_{ab}$  olduğundan dolayı ;

$$P' R = P' (X_0' X_0)^- P (T' T)^- = I_A \quad (76)$$

ilişkisi elde edilir.

Burada  $I_A$ ,  $A \times A$  tipinde birim matris,  $\delta_{ab}$  ise Kronecker deltasıdır. (76) eşitliğinden dolayı  $R$  matrisi  $P'$  matrisinin genelleştirilmiş inversidir.  $R$  matrisi aşağıdaki formda yazılabilir.

$$R = W(P'W)^{-1} \quad (77)$$

(75) ve (77) eşitliklerindeki inverse matrislerin tam olarak hesaplanması PLS-NIPALS algoritması ile bir şekilde azaltılmaktadır. Höskuldsson aşağıdaki tekrarlayan ilişkiyi göstermiştir (Höskuldsson, 1988).

$$r_a = w_a - (p'_{a-1}w_a)r_{a-1}, \quad a > 1 \quad (78)$$

Bu eşitlik  $r_1 = w_1$  ile başlamaktadır. Fakat bu eşitlik  $P'P$  matrisinin tridiagonal yapısına bağlıdır ve sadece bağımlı değişken sayısı bir olduğu durumda doğrudur. (79) ve (80) nolu eşitliklerde verilen güncellenmiş formüller genel formüller olarak kabul edilebilir.

$$r_a = G_a w_a, \quad a=1,2,\dots,A \quad (79)$$

$$G_{a+1} = G_a - r_a p'_a, \quad a=1,2,\dots,A-1 \quad (80)$$

Bu eşitliklerde  $G_1 = I_p$  alınarak işlemlere başlanır. Burada  $w_a$  ağırlık vektörünün aksine  $r_a$  vektörü standartlaştırılmamıştır. Bundan dolayı (73) nolu eşitlikteki  $t_a$  ve  $r_a$  vektörleri de normalleştirilmemiştir (De Jong, 1993).

#### **4.3.3.2. PLS yönteminin istatistiksel olarak esinlenilmiş değişikliğinin basit bir uygulaması (SIMPLS)**

Bu algoritma 1993 yılında De Jong tarafından önerilmiştir. Bu algoritmanın diğer algoritmalarından farkı burada PLS bileşenlerinin ortogonalizasyonudur. SIMPLS algoritması skorları indirgenmiş  $X$  matrisleri yerine orijinal  $X$  matrislerinin kombinasyonlarını kullanarak hesaplamayı amaçlamaktadır. Bu yaklaşımdan dolayı algoritma klasik algoritmada elde edilen sonuçlardan farklı sonuçlar vermektedir. Fakat bu fark çok küçüktür ve bir çok durumda dikkate alınmaz. Bu küçük farklılığa  $X'Y$



matrisinin klasik algoritmadaki yada çekirdek algoritmadaki gibi indirgenmemesi neden olmaktadır. SIMPLS algoritmasında önceki bileşenlerden elde edilen yük vektörleri üzerine orjinal  $\mathbf{X}'\mathbf{Y}$  matrisi yansıtılarak özvektör analizi yapılmaktadır. İndirgemedeki bu farklılık SIMPLS algoritması ile diğer algoritmalar arasında farklılığa neden olmaktadır (Lindgren and Rännar, 1998).

De Jong tarafından alternatif olarak önerilen bu yaklaşım klasik PLS algoritmasının aksine önce amacı belirler, amaç belirlendikten sonra optimizasyon kriteri oluşturulur ve bu kriter optimize edilmeye çalışılır. Son adımda ise algoritma kurulur. Bu yaklaşım çok değişkenli istatistikte kullanılan klasik yaklaşımdır. Algoritmanın amacı yanlı regresyon metotlarının aksine parametre tahminlerini durağanlaştıran  $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{B}$  şeklinde bir tahmin modeli bulmaktır. Standart PLS metoduyla aynı çizgide kalmak amacıyla  $\mathbf{X}$ 'in ardışık ortogonal faktörlerini hesaplamak için yük vektörleri  $\mathbf{t}_a = \mathbf{X}_0\mathbf{r}_a$  şeklinde seçilir. Bu yük vektörleri  $\mathbf{Y}$ 'nin  $\mathbf{u}_a = \mathbf{Y}_0\mathbf{q}_a$  şeklindeki faktörleri ile arasındaki kovaryans maksimum olacak şekilde hesaplanır. Önerilen bu düzeltmede  $\mathbf{R}$  ağırlık matrisi direk olarak hesaplanmaktadır. Bu yolla indirgenmiş veri matrislerinin ve  $\mathbf{W}$  ağırlık matrisinin hesaplanmasından kaçınılmaktadır. Eşitlik (65) ve (66)'da verilen ters matrislerin kesin hesaplanmasından kaçınılmaktadır. Burada yeni tanımlanan  $\mathbf{R}$  matrisi eşitlik (74) de verilen  $\mathbf{R}$  matrisi ile benzerdir. Hesaplanan bu yeni  $\mathbf{R}$  matrisi standart PLS algoritmasındaki  $\mathbf{W}$  gibi normalleştirilmiş vektörlerini içermektedir.  $\mathbf{r}_a$  ve  $\mathbf{q}_a$  ağırlık vektörleri merkezleştirilmiş veriden direk olarak aşağıdaki şekilde hesaplanabilir (De Jong,1993).

$$\mathbf{t}_a = \mathbf{X}_0\mathbf{r}_a \quad a = 1,2, \dots A \quad (81)$$

$$\mathbf{u}_a = \mathbf{Y}_0\mathbf{q}_a \quad a = 1,2, \dots A \quad (82)$$

Ağırlık vektörleri belirlenirken  $\mathbf{t}_a$  ve  $\mathbf{u}_a$  skor vektörleri arasındaki kovaryans maksimum olacak şekilde bazı kısıtlar altında karar verilmelidir. Özellikle aşağıda verilen 4 koşul çözümü kontrol etmektedir:

- i. Kovaryans maksimize edilmelidir:  $\mathbf{u}'_a\mathbf{t}_a = \mathbf{q}'_a(\mathbf{Y}'_0\mathbf{X}_0)\mathbf{r}_a = \max!$
- ii.  $\mathbf{r}_a$  ağırlık vektörü normalleştirilmelidir:  $\mathbf{r}'_a\mathbf{r}_a = 1$
- iii.  $\mathbf{q}_a$  ağırlık vektörü normalleştirilmelidir:  $\mathbf{q}'_a\mathbf{q}_a = 1$

iv.  $\mathbf{t}$  skorları ortogonalleştirilmelidir:  $\mathbf{t}'_b \mathbf{t}_a = 0$ ,  $a > b$  için

Son kısıt olmadığı zaman sadece tek bir çözüm vardır.  $\mathbf{r}_1$  ve  $\mathbf{q}_1$   $p \times m$  boyutlu  $\mathbf{S} \equiv \mathbf{X}'_0 \mathbf{Y}_0$  çapraz çarpımlar matrisinin ilk sol ve sağ tekil vektörleridir. Fakat, birden fazla çözüm elde etmek ve  $\mathbf{X}$ 'in ortogonal faktörlerinin bir alt kümesini oluşturmak için son kısıt olan ortogonalite kısıtı eklenmek zorundadır. Bundan dolayı  $a > b$  için aşağıdaki eşitliğin kullanılması gerekir.

$$\mathbf{t}'_b \mathbf{t}_a = \mathbf{t}'_b \mathbf{X}_0 \mathbf{r}_a = (\mathbf{t}'_b \mathbf{t}_b) \mathbf{p}'_b \mathbf{r}_a = 0 \quad (83)$$

Burada  $\mathbf{p}_b$ , orijinal  $\mathbf{X}$  değişkenleri ile  $b$ 'inci PLS faktörü arasındaki ilişkiyi açıklayan yük vektörüdür. Eşitlik (83)'deki herhangi yeni  $\mathbf{r}_a$  ağırlık vektörü önceki bütün yük vektörlerine dik olmak zorundadır. Örneğin;  $\mathbf{P}_{a-1} \equiv [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{a-1}]$ 'in sütunlarına dik olmalıdır.  $\mathbf{P}_{a-1}^\perp$  gerekli ortogonal yansıtıcı olmak üzere aşağıdaki (17) nolu eşitlikte gösterildiği gibidir.

$$\mathbf{P}_{a-1}^\perp = \mathbf{I}_p - \mathbf{P}_{a-1} (\mathbf{P}'_{a-1} \mathbf{P}_{a-1})^{-1} \mathbf{P}'_{a-1} \quad (84)$$

Bu eşitlikten dolayı;

$$\mathbf{r}_a = \mathbf{P}_{a-1}^\perp \mathbf{r}_a \quad (85)$$

eşitliğini yazabiliriz. (84) ve (85) nolu eşitlikler dördüncü ortogonalite kısıtını sağlamaktadır.  $\mathbf{q}_a$  ve  $\mathbf{r}_a$  için çözüm  $\mathbf{P}_{a-1}$  üzerine yansıtılan  $\mathbf{S}_0$ 'ın tekil değer ayrışımının (SVD) ilk çifti tarafından verilmektedir. (Örneğin;  $\mathbf{P}_{a-1}^\perp \mathbf{S}_0$  matrisinin tekil değer ayrışımı gibi.) Genellikle  $a$  yük vektörleri  $\mathbf{S}_a$  gibi dışarı yansıtıldıktan sonra çapraz çarpım belirtilecektir. Burada  $\mathbf{S}_a$ ;

$$\mathbf{S}_a \equiv \mathbf{P}_a^\perp (\mathbf{X}'_0 \mathbf{Y}_0) = \mathbf{P}_a^\perp \mathbf{S}_0 \quad (86)$$

şeklindedir. Bu fikir doğrultusunda PLS çözümlerinin nasıl elde edileceğinin basit adımları Çizelge 4.1'de özetlenmiştir. Fakat bir algoritma tanımlanamamaktadır (De Jong, 1993).

#### Çizelge.4.1. Düzeltilmiş PLS algoritmasının temel konsepti

- 
- Çapraz çarpımı hesapla:  $S = X_0'Y_0$   
a=1,2,...,A için;
  - a=1 için S matrisinin tekil değer ayrışımını (SVD) hesapla.
  - a>1 için  $S - P(P'P)^{-1}P'S$  matrisinin tekil değer ayrışımını(SVD) hesapla.
  - Ağırlıkları bul:  $r$ =ilk sol singüler vektör.
  - Skorları hesapla:  $t = X_0r$
  - Yükleri hesapla:  $p = X_0't/(t't)$
  - $r, t$  ve  $p$ 'yi sırası ile  $R, T, P$  içine depola.
  - Sonlandır.
  - Regresyon katsayılarını hesapla:  $B_{PLS} = RT^{-1}Y_0$
- 

$S_{a+1}$  'in kendinden önce gelen  $S_a$ 'dan hesaplanması uygundur.  $P_a$ 'nın sütun uzayı üzerine yansımaları başarmak için ortogonal yansımaların bir dizisi olarak uygulanacaktır. Bunun için  $V_a \equiv [v_1, v_2, \dots, v_a]$  şeklinde  $P_a$  'nın ortonormal bir temeline ihtiyaç vardır.  $V_a, P_a$ 'nın aşağıdaki eşitlikte verildiği gibi Gram-Schmidt ortonormalleştirilmesiyle elde edilebilir.

$$v_a \propto p_a - V_{a-1}(V_{a-1}'p_a), a=2,3,\dots, A \quad (87)$$

Bu eşitlikte  $V_1 = v_1 \propto p_1$  alınarak başlanmaktadır. Bağımlı değişken tek olduğu durumda ilave olarak bir basitleştirme mümkündür. Bu durumda,  $b \leq a - 2$  için  $p_b'p_a = 0$  şeklindeki  $P$  nin ortogonalite özelliklerinden yararlanılabilir. Bu özellikler ortonormal  $V$  yüklerine taşınabilir. Bundan dolayı,  $V_{a-2}$ 'ye göre  $p_a$  nin ortogonalitesi otomatik olarak dikkate alınır ve (87) nolu eşitlik aşağıdaki gibi basitleştirilebilir.

$$v_a \propto p_a - v_{a-1}(v_{a-1}'p_a), a > 1, m = 1 \quad (88)$$

İlk  $\alpha$  tane yük vektörü tarafından gerilen alt uzaya yansıtılan,  $P_a(P_a'P_a)^{-1}P_a', V_aV_a'$  ve ortogonal tamamlayıcı  $P_a^\perp$  üzerine yansıtılan  $I_p - V_aV_a' = \prod_1^a(I_p - v_bv_b')$  yer

değiştirebilir. Bundan dolayı,  $V$ 'nin ortogonallığı kullanılarak  $S_a$  ( $a=1,2,\dots$ ) çarpım matrisleri dikey yönde yansıtılan  $v_a$  yardımıyla sürekli olarak indirgenebilir (De Jong, 1993). Bu durum eşitlik (89) de verilmiştir (De Jong, 1993).

$$S_a = S_{a-1} - v_a(v_a' S_{a-1}), \quad a > 1 \quad (89)$$

SIMPLS algoritması ile standart PLS algoritması arasındaki temel fark indirgeme işleminden kaynaklanmaktadır. SIMPLS yönteminde indirgeme işlemi daha büyük boyutlu  $X_0$  ve  $Y_0$  matrisleri yerine çapraz çarpımlar matrisi  $S_0$  'a uygulanmaktadır. Her  $S_a$  'nın ilk singüler vektör çiftleri iteratif bir yöntemle hesaplanmaktadır. Bu işlem sadece ilgilenilen singüler vektör çiftlerini hesaplama avantajını sağlamaktadır. Genellikle  $Y$  değişkenlerinin sayısı  $X$  değişkenlerinin sayısından daha küçüktür ( $m < p$ ). Örneğin, çok değişkenli kimya uygulamalarında  $p$  yüzlerle ifade edilirken  $m$  oldukça az sayıda olabilir (5'ten daha küçük olabileceğini söyleyebiliriz).  $m \times m$  tipindeki küçük boyutlu simetrik matris  $S_{a-1}' S_{a-1}$  in baskın özvektörü olarak  $q_a$  'nın hesaplanması etkili olacaktır. Bundan dolayı da  $r_a$  aşağıdaki şekilde bulunacaktır.

$$r_a \propto S_{a-1} q_a \quad (90)$$

Tek değişkenli  $Y(=y)$  için,  $S_{a-1}(=s_{a-1})$  vektörü,  $X$  değişkenleri ile  $y$  değişkeni arasındaki kovaryans vektörü olacaktır. Burada,  $S_{a-1}' S_{a-1} = s_{a-1}' s_{a-1}$  skaler,  $q_a = \mathbf{1}$  ve  $r_a \propto s_{a-1}$  olacaktır. Bundan dolayı, çözüm tek değişkenli standart PLS algoritmasına benzer şekilde iteratif olmayacaktır.

SIMPLS algoritmasının standart PLS algoritmasından daha basit bir şekilde uyguladığı diğer bir durum  $X$  ve  $Y$  matrislerinin merkezleştirilmesidir. Standart indirgeme işlemindeki ilk adım olan merkezleştirme işlemi burada uygulamak önemli değildir.

$S_0, S_0 = X'Y - n\bar{x}'\bar{y}$  şeklinde hesaplanabilir. Alternatif olarak,  $S_0 = X_0'Y_0 = X'Y_0$  olduğundan dolayı sadece  $Y$  matrisi merkezleştirilebilir. Genel olarak  $Y$  matrisi iki veri setinden daha küçük olandır. Sıklıkla da  $Y$  matrisi tek sütundan oluşur, yani  $y$  vektörüdür.  $X$  matrisi  $X_0$  matrisini verecek şekilde merkezleştirildiğinde  $t_a = X_0 r_a$

skor vektörleri otomatik olarak merkezleştirilmiş olur. Eğer  $\mathbf{X}$  matrisi merkezleştirilmezse, skor vektörleri ilk olarak  $\mathbf{X}\mathbf{r}_a$  şeklinde hesaplanacaktır. Bundan dolayı da skor vektörlerinin merkezleştirilmesi gerekir. Skor vektörleri  $\mathbf{u}_a = \mathbf{Y}\mathbf{q}_a$  şeklinde merkezleştirilebilir. Önceki  $\mathbf{t}$  skorları için  $(\mathbf{t}_1, \dots, \mathbf{t}_{a-1})$   $\mathbf{u}_a$  skorlarının ortogonalleştirilmesi yararlı olabilir. Bu gerekli olmamasına karşın  $\mathbf{u}$  skorları arasındaki iç ilişkiyi azaltır ve daha iyi bir yorum için standart PLS algoritması ile benzerlikler sağlar. Tablo1'de verilen basit konsepte bağlı kalınarak standart PLS algoritmasından esinlenilerek bazı değişiklikler yapıldığı için bu algoritmaya SIMPLS denilmesi uygun bulunmuştur (De Jong, 1993).

SIMPLS algoritmasının adımları aşağıda verildiği gibidir.

$$(1) \mathbf{Y}_0 = \mathbf{Y} - \text{ort}(\mathbf{Y})$$

$$(2) \mathbf{S} = \mathbf{X}'\mathbf{Y}_0$$

a=1,...,A için

$$(3) \mathbf{q} = \mathbf{S}'\mathbf{S} \text{ matrisinin baskın özvektörü}$$

$$(4) \mathbf{r} = \mathbf{S}\mathbf{q}$$

$$(5) \mathbf{t} = \mathbf{X}\mathbf{r}$$

$$(6) \mathbf{t} = \mathbf{t} - \text{ort}(\mathbf{t})$$

$$(7) \text{norm}t = \sqrt{\mathbf{t}'\mathbf{t}}$$

$$(8) \mathbf{t} = \mathbf{t}/\text{norm}t$$

$$(9) \mathbf{r} = \mathbf{r}/\text{norm}t$$

$$(10) \mathbf{p} = \mathbf{X}'\mathbf{t}$$

$$(11) \mathbf{q} = \mathbf{Y}_0'\mathbf{t}$$

$$(12) \mathbf{u} = \mathbf{Y}_0\mathbf{q}$$

$$(13) \mathbf{v} = \mathbf{p}$$

Eğer a>1 ise

$$(14) \mathbf{v} = \mathbf{v} - \mathbf{V}(\mathbf{V}'\mathbf{p})$$

$$(15) \mathbf{u} = \mathbf{u} - \mathbf{T}(\mathbf{T}'\mathbf{u})$$

sonlandır.

$$(16) \mathbf{v} = \mathbf{v}/\sqrt{\mathbf{v}'\mathbf{v}}$$

$$(17) \mathbf{S} = \mathbf{S} - \mathbf{v}(\mathbf{v}'\mathbf{S})$$

(18)  $r, t, p, q, u$  ve  $v$  yi  $R, T, P, Q, U$  ve  $V$  içine kaydet.

sonlandır.

$$B = RQ'$$

$$h = \text{diag}(TT') + 1/n$$

$$\text{var}(X) = \text{diag}(P'P)/(n-1)$$

$$\text{var}(Y) = \text{diag}(Q'Q)/(n-1)$$

#### 4.3.3.3. Standart PLS algoritması ile SIMPLS algoritmasının karşılaştırılması

SIMPLS ve standart PLS algoritmalarının her ikisinde de kovaryans kriteri maksimize edilir. Burada  $t_a$  ardışık vektörleri ortogonal olmak zorundadır. İlk bakışta, iki algortmada da eşit koşulların sağlandığı garanti edilebilir gibi görünmektedir. Fakat A faktörden sonra veri matrisi NIPALS-PLS ile açıkça  $X_A = (I_n - TT')X_0$  ve SIMPLS algoritması ile dolaylı olarak  $X_0(I_p - VV')$  matrislerine indirgenmektedir. Bu artık matrisleri farklı oldukları için iki algoritma arasındaki esas fark görülebilecektir. Ayrıca SIMPLS algoritmasında faktörler direkt olarak orijinal veri matrisinden hesaplanmaktadır. SIMPLS algoritmasında hesaplanan  $R$  ağırlıkları PLS algoritmasında hesaplanan  $W$  ağırlıklarından daha basit bir yoruma sahiptir. Buna ek olarak  $R$  ağırlıkları hesaplanırken yukarıda da değinildiği gibi ters matrisler hesaplanmak zorunda değildir. SIMPLS algoritmasında  $X$  matrisinin indirgenme işlemi yapılmadığından standart PLS algoritmasına göre daha hızlıdır (De Jong, 1993).

Bağımlı değişkenler matrisi tek boyutlu olduğunda ( $m=1$ ) her iki algoritmanın benzer sonuçlar vereceği bunun aksine çok değişkenli durumda ( $m>1$ ) ise sonuçların oldukça farklı olacağı görülmektedir (De Jong, 1993).

#### 4.3.3.4. Evrensel kısmi en küçük kareler algoritması (UNIPALS)

1998 yılında Glen tarafından gösterilen bu algoritma kısmi en küçük karelerin hesaplanmasında daha küçük varyans-kovaryans matrislerini kullanan ilk algortmadır. Bu algoritma UNIPALS olarak bilinmektedir. Bu algoritma  $m \times m$  boyutlu  $Y'XX'Y$

matrisine dayanmaktadır.  $Y'XX'Y$  matrisinin en büyük özdeğerine karşılık gelen özvektör  $Y$  bağımlı değişkenler matrisinin ilk ağırlık vektörü  $c$ 'ye karşılık gelmektedir. Bu ağırlık vektörüne ve orjinal  $X$  ve  $Y$  vektörlerine bağlı olarak iterasyon olmaksızın diğer vektörler hesaplanabilir. Fakat klasik kısmi en küçük kareler algoritmasına benzer olarak boyutlar arasındaki güncelleştirme işlemi orijinal  $X$  ve  $Y$  matrisleri üzerinden gerçekleştirilmektedir. Bundan dolayı her yeni boyut için  $Y'XX'Y$  matrisi indirgenmiş  $X$  ve  $Y$  matrislerinden yeniden türetilmelidir. Bu algorithmada da orijinal veri matrisi klasik algorithmadaki gibi indirgendiğinden sonuçlar her iki algoritma için de benzerdir (Lindgren and Rännar,1998).

#### 4.3.3.5. Çekirdek algoritması

Değişken sayısının gözlem sayısından ya da gözlem sayısının değişken sayısından çok olduğu veri setleri için NIPALS algoritması hesaplamalar için çok fazla zaman harcamaktadır. Bu tip veri setlerinde kullanılmak üzere Lindgren tarafından bir algoritma geliştirilmiştir. Bu algoritmanın adı Çekirdek Algoritmasıdır. Çekirdek algoritması UNIPALS algoritmasındaki  $Y'XX'Y$  matrisi ile çalışmak yerine  $pxp$  boyutlu  $X'YY'X$  matrisinden  $X$  matrisi için ağırlık vektörü olan  $w$  vektörünü hesaplar.  $w$  ağırlık vektörü,  $X'Y$  ve  $X'X$  alt matrisleri yardımıyla kısmi en küçük kareler algoritmasındaki vektörler hesaplanabilir. İlk çekirdek algoritmasındaki düşünce orjinal  $X$  ve  $Y$  matrislerine müdahale etmeden varyans-kovaryans matrislerinin nasıl güncelleneceğidir.  $X'YY'X$  matrisinin  $pxp$  boyutlu  $I - wp'$  güncelleştirme matrisi ile çarpılmasıyla algoritmanın hesaplama hızında önemli gelişmeler olmaktadır.

$$E'YY'E = (I - wp')'X'YY'X(I - wp')$$

Bu gelişmenin nedeni ise varyans-kovaryans matrisini oluştururken zaman alan adımların sadece bir kez tekrarlanmasıdır. Burada sadece  $X$  matrisinin indirgenmesi önerilebilir,  $X$  matrisi yerine  $Y$  matrisi indirgenerek işlem yapılırsa bu sonuçları etkilemez.

İkinci Çekirdek algoritması Rännar (1994) tarafından önerilmiştir. Önerilen bu algoritma ilk algorithmaya oldukça benzemektedir. Fakat aralarındaki en önemli fark ikinci algoritmanın açıklayıcı değişken sayısı gözlem sayısından büyük olduğu

durumlarda daha iyi sonuç vermesidir. Açıklayıcı değişken sayısının gözlem sayısından büyük olduğu veri setlerine genelde endüstriyel süreçlerde rastlanmaktadır. Algoritmada kullanılan  $XX'YY'$  matrisi açıklayıcı değişken sayısından bağımsızdır. Bundan dolayı da çekirdek algoritmasının bu versiyonu için  $XX'YY'$  matrisinin özvektörleri ile çözüme başlamak daha iyidir. Bu özvektörler  $X$  matrisi için  $t$  skor vektörlerini vermektedir. Buradan elde edilen  $t$  skor vektörü ve  $YY'$  matrisi ile  $Y$  matrisinin skor vektörü olan  $t$  vektörü hesaplanır. Çekirdek algoritmasının bu versiyonunda indirgeme küçük varyans-kovaryans matrisi elde etme üzerinde doğrudan etkilidir. Algoritmanın son adımında orijinal  $X$  ve  $Y$  matrisleri kullanılarak kısmi en küçük kareler ağırlıkları ( $w$  ve  $c$ ) ve yük vektörleri ( $p$ ) hesaplanır. Bu vektörler regresyon katsayılar matrisi  $B$ 'nin elde edilmesinde gereklidir.  $B$  matrisi (91) nolu eşitlikte gösterildiği gibi hesaplanmaktadır.

$$B = W(P'W)^{-1}C' \quad (91)$$

Çekirdek algoritmasının her iki versiyonunda da en önemli özellik her iki algoritmanın çoklu bağımlı değişken olması durumunda klasik algoritmadaki gibi iyi sonuçlar vermesidir.

Çekirdek algoritması son zamanlarda De Jong (1997) tarafından da düzeltilmiştir. Bu düzeltme ile elde edilen algoritma daha hızlı ve basit sonuçlar vermektedir. Çekirdek algoritması üzerinde bir düzeltmede Dayal ve arkadaşları (1997) tarafından yapılmıştır. Önerdikleri versiyonda  $X$  veya  $Y$  matrislerinin sadece bir tanesini indirgemişlerdir.  $Y$  matrisinde daha az sayıda değişken olduğundan dolayı  $Y$  matrisinin indirgenmesi  $X$  matrisine göre daha hızlı gerçekleşmektedir. Dolayısıyla  $Y$  matrisinin indirgenmesi zaman bakımından avantaj sağlamaktadır (Lindgren and Rännar,1998).

#### **4.3.3.6. Örnek-uzaklık kısmi en küçük kareler algoritması (SAMPLS)**

Örnek-uzaklık kısmi en küçük kareler algoritması (SAMPLS) Bush ve Nachbar (1993) tarafından önerilmiştir. Algoritma özel olarak çok fazla açıklayıcı değişkenin ve az sayıda gözlemin olduğu durumlar için önerilmiştir. Ayrıca bu algoritmada bağımlı



değişken sayısı bir tanedir. Algoritmada sadece bir tane bağımlı değişken olması sınırlayıcı bir durumdur. Hesaplama zamanı açısından bakıldığında bu algoritmanın performansı klasik algoritmadan ve çekirdek algoritmalarından daha üstündür.

SAMPLS algoritması  $\mathbf{t}$  skor vektörünü hesaplamak için  $\mathbf{XX}'$  matrisi ve bağımlı değişken vektörü  $\mathbf{y}'$ 'yi kullanmaktadır. Ayrıca bu algoritmada iterasyon olmaksızın klasik matris-vektör çarpımı kullanılmaktadır. SAMPLS algoritması diğer algoritmalarındaki gibi ağırlık ve yük matrislerini vermemesine rağmen tahmin için kullanılabilir. Değişkenler arasındaki korelasyon bilgisi kaybolacağından dolayı ağırlık ve yük matrislerinin olmaması ciddi bir dezavantaj olabilir. Bush ve Nachbar (1993) bu algoritmada indirgeme işlemi için  $\mathbf{X}$  veya  $\mathbf{Y}$  matrislerinin birisini seçme avantajını sağlamaktadırlar. Ayrıca bu algoritmada bağımlı değişken sadece bir tane olduğundan dolayı bu vektörü indirgemek daha hızlıdır. Bu yapı güncelleştirme sürecini klasik algoritmadakine benzer kılmaktadır. Bundan dolayı da sonuçları klasik algoritmanın sonuçlarına benzer olacaktır. Ancak ortogonal yapıyı korumak için algoritmayı kullanmaksızın yeni skor vektörleri bir önceki vektörlerle ortogonalleştirilmelidir (Lindgren and Rännar,1998).

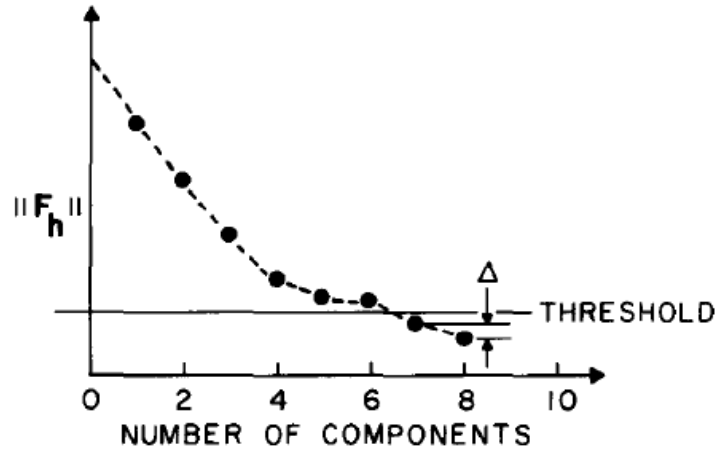
#### **4.3.3.7. Özvektör algoritması**

Höskuldsson (1988) tarafından tanımlanan bu algoritma çekirdek algoritma ile benzer bir yaklaşımdır, fakat bu algoritmanın çalışması çekirdek algoritmaya göre daha kolaydır. Bu algoritmadaki fikir  $\mathbf{X'YY'X}$  ve  $\mathbf{Y'XX'Y}$  matrislerinin en büyük özdeğerine karşılık gelen özvektörü hesaplamak yerine PLS bileşenlerinin sayısı kadar olan en büyük özdeğerlere karşılık gelen özvektörleri hesaplamaktır. Bu yaklaşımdan dolayı  $\mathbf{X'YY'X}$  matrisinin en büyük  $a$  özdeğerlerine karşılık gelen özvektörler olan  $\mathbf{p}_1, \dots, \mathbf{p}_a$  vektörleri PLS yük vektörlerine ortogonal olacaktır. Burada PLS yük vektörleri olan  $\mathbf{q}_1, \dots, \mathbf{q}_a$  vektörleri  $\mathbf{Y'XX'Y}$  matrisinin en büyük  $a$  özdeğerlerine karşılık gelen özvektörlerdir.  $x$  ve  $y$  skorları verilerin yük vektörleri üzerine yansıtılması ile bulunmaktadır. Bu algoritmada indirgeme işlemi uygulanmadığından, skor vektörleri ilişkisiz değildirler. Bundan dolayı bu yaklaşım maksimizasyon problemi için bir çözüm üretemez. Fakat özvektörlerin tanımından dolayı yük vektörleri ortogonal olmaktadır. Böylece bu yaklaşım özellikle yol göstermesi açısından tercih edilebilir.

Skor vektörleri kullanılarak x ve y verileri daha düşük boyutlu ortogonal koordinat sistemlerine taşınabilir, bu işlemde dolayı da ilginç veri yapıları meydana çıkabilir (Varmuza and Filzmoser, 2008).

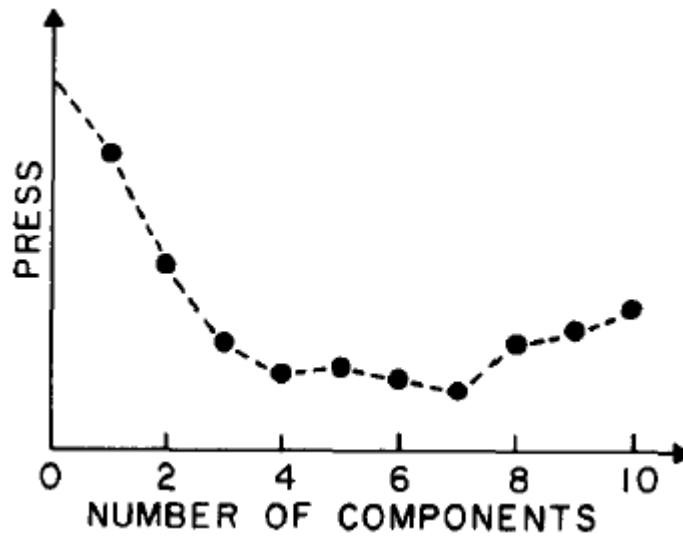
#### **4.3.4. Kısmi en küçük kareler regresyonunda kullanılan ideal bileşen sayısının belirlenmesi**

Modelde bulunan bileşen sayısı PLS modelinin en önemli özelliklerinden birisidir. Bağımlı değişken ile açıklayıcı değişkenler arasındaki ilişkiyi gösteren model lineer bir model ise, bu modelde bulunan bileşen sayısı modelin boyutuna eşittir. Ayrıca lineer olmayan modellerde lineer olmamayı tanımlamak için daha fazla bileşene ihtiyaç duyulmaktadır. Bir model oluşturulurken açıklayıcı değişkenler matrisinin rankı kadar bileşen hesaplamak mümkündür fakat bu bileşenlerin hepsi modelde kullanılmayabilir. Bunun temel nedenleri ölçülen değişkenlerin hiçbir zaman gürültüden bağımsız olmaması ve diğer bileşenlere göre daha önemsiz bileşenlerin gürültüyü tanımlamasıdır. Çoklu iç ilişki problemine neden olduklarından dolayı genellikle önemsiz bileşenler ihmal edilir. Burada önemli olan hangi bileşenlerin ihmal edileceği konusudur ve bu konuda karar vermek için birçok yöntem önerilmiştir. Bu yöntemlerden birisi  $F_h = F_{h-1} - b_h t_h q_h'$  denklemindeki  $F_h$ 'ın normunun küçük olmasıdır. Bunu belirlemek için Şekil4'teki gibi bileşen sayısına karşı  $\|F_H\|$ 'ın grafiği çizilmelidir. Çizilen bu grafikte bir eşik değer belirlenir ve  $\|F_H\|$  değeri bu eşik değerinin altına düştüğü zaman durulur. Diğer bir yol ise  $\|F_H\|$ 'ın gerçek değeri ile önceki değeri arasındaki farkın alınması ve bu fark daha önce belirlenmiş bir ölçüm hatasından daha küçük olduğunda durulmasıdır. Bu iki yöntemin birleştirilerek bileşen sayısına karar verilmesi de tercih edilen bir yöntemdir.



Şekil.4.5.  $\|F_H\|$  değerine karşı bileşen sayısının grafiği (Geladi and Kowalski, 1986).

Eğer tahmin söz konusu ise kullanılacak bileşen sayısına karar vermek için farklı yöntemler bulunmaktadır. Bu yöntemler çapraz geçerlilik (CV) olarak adlandırılmaktadır. Bu yöntem için PRESS (tahmin edilen hata kareler toplamı) değeri hesaplanabilir. Hesaplanan bu PRESS değerine karşı bileşen sayısının grafiği Şekil5'te görüldüğü gibi çizilirse modelde bulunması gereken bileşen sayısına karar vermek için bu grafik yardımcı olabilir. PRESS değeri modelin tahmin gücü hakkında bilgi vermek için kullanılan bir istatistiktir. En iyi tahmini veren model PRESS değeri minimum olan modeldir. Dolayısıyla bu modeldeki bileşen sayısı modelde olması gereken ideal bileşen sayısıdır (Geladi and Kowalski, 1986).



Şekil.4.6. PRESS değerine karşı bileşen sayısının grafiği (Geladi and Kowalski, 1986).

Çapraz geçerlilik yöntemi (CV) tahmin edicinin öneminin test edilmesi için kullanılan pratik ve güvenilir bir yöntemdir. Bu yöntem PLS analizinde standart olarak kullanılmaktadır ve bu yöntemin çeşitli versiyonları PLS yazılımlarına karşımıza çıkmaktadır. Genel olarak CV yöntemi veriyi belirli sayıda (G) gruplara bölerek (genellikle 5 ile 9 arasında olmaktadır) her seferinde bu gruplardan birisini veriden atarak farklı modeller oluşturmaktadır. Model oluşturulduktan sonra,  $y$  değerlerinin gerçek değerleri ile tahmin edilen değerleri arasındaki fark hesaplanmaktadır. Daha sonra hesaplanan bu farkların kareler toplamı her bir model için ayrı ayrı hesaplanmaktadır ve elde edilen bu değerler modelin tahmin yeteneğini açıklamak için kullanılmaktadır. CV yöntemi ardışık olarak kullanıldığında, bir bileşenden sonra diğer bileşen üzerinde uygulanmaktadır fakat bir sonraki bileşenin CV değeri için artık matrisleri **E** ve **F** gruplara bölündükten sonra tam veri matrisleri üzerinde indirgeme işlemi bir kez uygulanır. Her bir bileşenden sonra  $PRESS_a/SS_{a-1}$  oranı hesaplanır, hesaplanan bu oran en az bir  $y$  değişkeni için 0,90 değerinden küçük ise o bileşen önemli bir bileşen olarak kabul edilir. Burada  $SS_{a-1}$ ,  $a$  bileşeninden önceki bileşen için uydurulan artık kareler toplamını ifade etmektedir. Bu hesaplamalar bir bileşenin önemsiz olduğuna karar verene kadar devam etmektedir.

Alternatif olarak “toplam CV” yöntemi ilk olarak veri setini gruplara böler, daha sonra her bir CV grubuna ait veri matrislerini indirgeyerek her bir bileşen için PRESS değeri hesaplar. Bu hesaplamalardan sonra en küçük  $PRESS/(n - A - 1)$  oranını veren model kullanılır. Hesaplamalar bakımından “toplam CV” yöntemi daha yorucu olmasına karşın ardışık yöntem ile benzer sonuçları vermektedir.

Her iki yöntemde de tahmin edilen önemli bileşenler ile kurulan son model için PRESS değerleri hesaplanmaktadır. Hesaplanan bu PRESS değeri yardımıyla  $Q^2$  istatistiği  $1 - PRESS/SS$  şeklinde hesaplanmaktadır.  $Q^2$  istatistiği  $R^2$  istatistiği ile karşılaştırılabilir. Birden fazla  $y$  değerine sahip modellerde her bir  $y$  değeri için  $R_m^2$  ve  $Q_m^2$  değerleri hesaplanabilir.

#### 4.3.5. Kısmi En Küçük Kareler Regresyonunun Temel Bileşenler Regresyonu ve Ridge Regresyon İle İlişkisi

PCR, PLSR ve RR için tanımlanan algoritmalar model tahmininde küçük farklılıklar göstermektedir. Bunun sebebi olarak üç yöntemde birbirinden farklı prosedürler izliyor olması gösterilebilir.

Kısmi en küçük kareler regresyonu, temel bileşenler regresyonu ve ridge regresyon genellikle düzenleyici regresyon yöntemleri olarak adlandırılmaktadır. Klasik en küçük kareler tahmin edicisi yansız bir tahmin edicidir fakat sözü edilen yanlı tahmin ediciler kullanılarak daha küçük hata kareler ortalaması hesaplanabilir. PLSR, PCR ve RR yöntemlerinin temel amacı açıklayıcı değişkenler uzayındaki çözümleri daraltmak olduğundan bu yöntemlere genel olarak daraltıcı tahmin ediciler denilmektedir (Frank and Friedman, 1993).

Stone ve Brooks (1990) yaptıkları çalışmalarında EKK, PCR ve PLS yöntemlerini içeren özel bir regresyon yöntemi önermişlerdir. Bu yöntem adımsal bir yöntemdir ve her bir adımda genelleştirilmiş kriter maksimum yapılır. Bu kriter  $[0,1]$  aralığında değişmekte olan  $\alpha$  değerine bağlıdır. Eğer  $\alpha = 0$  ise EKK,  $\alpha = 1/2$  ise PLSR ve son olarak  $\alpha = 1$  ise PCR elde edilmektedir. Bu yöntem bazı örnekler için iyi sonuçlar vermiştir fakat yöntemin zayıf noktası adım sayısına ve  $\alpha$  değerine karar vermek için çapraz geçerlilik yöntemini kullanmasıdır.

Stone ve Brooks (1990) tarafından kullanılan kriter üç metodun özelleştirilmiş hali olarak bilinmektedir. EKK;  $\mathbf{y}$  ve  $\hat{\beta}'\mathbf{x}$  arasındaki gözlenen korelasyonun maksimizasyonuna dayanmaktadır. PCR;  $\mathbf{X}$  'in lineer kombinasyonlarının varyanslarının her adımda maksimizasyonuna dayanır. PLSR ise  $\mathbf{y}$  ve diğer lineer kombinasyonlar arasındaki varyansı her adımda maksimize etmeyi amaçlamaktadır (Höskuldsson, 1988). Bahsedilen bu üç metot ve Ridge regresyonun bazı özellikleri temel alınarak kapsamlı bir karşılaştırılması Frank ve Friedman'ın 1993 yılında yaptıkları çalışmalarında bulunabilir (Helland, 2001).

## BÖLÜM 5

### UYGULAMA

Bu bölümde kısmi en küçük kareler, temel bileşenler ve ridge regresyon yöntemleriyle elde edilen modellerin tahmindeki başarılarının kıyaslanması için Robin H. Lock tarafından 1993 yılında derlenen/kullanılan veri seti kullanılmıştır. Analizler için Matlab ve Matlab PLS\_Toolbox programlarından yararlanılmıştır.

Literatürden alınan veri setinde 93 farklı marka otomobilin 11 açıklayıcı değişken yardımıyla şehir içi yakıt tüketimi araştırılmak istenmektedir. Öncelikle veri setinde çoklu iç ilişki bulunup bulunmadığı araştırılmak istenmektedir. Bu yüzden özdeğerlerin ve VIF değerlerinin hesaplanması gerekmektedir. Matlab programında hesaplanan özdeğer ve VIF değerleri Çizelge 5.1’de verilmiştir.

**Çizelge 5.1. Hesaplanan özdeğerler ve VIF değerleri.**

Değişkenler	Özdeğerler	VIF
Motor Hacmi ( $x_1$ )	6,5873	<b>11,9126</b>
Beygir Gücü ( $x_2$ )	1,6171	<b>12,6516</b>
Yakıt Deposu Hacmi ( $x_3$ )	1,0140	5,7228
Uzunluk ( $x_4$ )	0,8264	7,0289
Dingil Açıklığı ( $x_5$ )	0,2877	7,8262
Genişlik ( $x_6$ )	0,2242	7,9316
Ağırlık ( $x_7$ )	0,1452	<b>20,5015</b>
Arka Koltuk Genişliği ( $x_8$ )	0,1101	3,1220
Bagaj Hacmi ( $x_9$ )	0,0928	2,3936
Dakikadaki Devir Sayısı ( $x_{10}$ )	0,0666	4,8627
Kilometredeki Devir Sayısı ( $x_{11}$ )	0,0286	3,6163

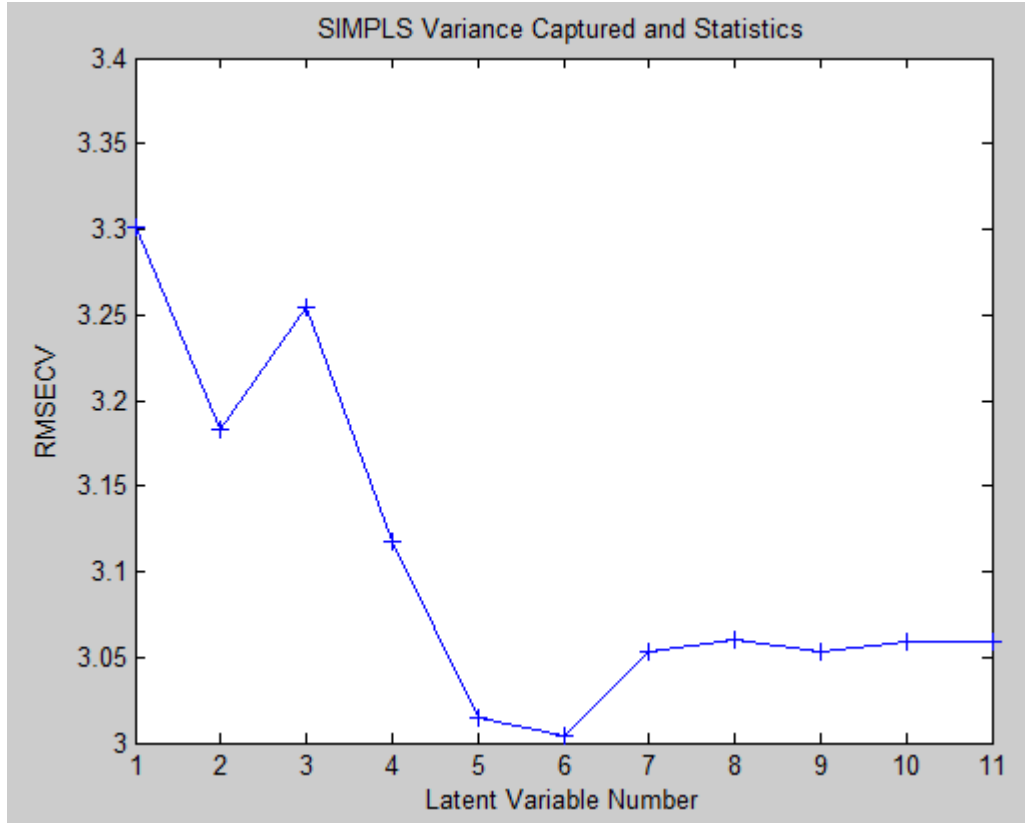
Çizelge 5.1 incelendiğinde VIF değerlerinden üç tanesinin (11,9126; 12,6516; 20,5015) kritik değer olan 10 değerinden daha büyük oldukları gözlemlenmektedir. Bu durum çoklu iç ilişkinin varlığını göstermektedir. Ayrıca en büyük özdeğerin en küçük özdeğere oranlanması ile elde edilen koşul sayısı incelendiğinde;  $K = \frac{6,5873}{0,0286} = 230,325$  değeri 100'den büyük olduğundan dolayı şiddetli çoklu iç ilişkinin varlığından söz edilebilir.

Veri setinde çoklu iç ilişki problemi olduğundan dolayı çoklu lineer regresyon yöntemi ile elde edilen regresyon denkleminin tahmin gücü iyi olmayabilir. Bu yüzden çoklu iç ilişki problemi ile karşılaşıldığında kullanılan yanlı regresyon yöntemleri ile analizi yapmak uygun olacaktır.

Yanlı regresyon yöntemlerini uygulamadan önce bağımlı ve açıklayıcı değişkenler standartlaştırılmıştır. Daha sonra standartlaştırılmış veri setine PLSR, PCR ve RR yöntemleri uygulanmıştır.

### **5.1. Kısmi En Küçük Kareler Regresyonu**

Matlab PLS\_Toolbox yardımıyla kısmi en küçük kareler regresyonu yapıldığında SIMPLS algoritması kullanılmaktadır. Kısmi en küçük kareler regresyonu yaparken ilk olarak modelde bulunması gereken ideal bileşen sayısına karar verilmelidir. Bunun için RMSECV değerine karşı gizil değişkenlerin sayısının grafiği çizdirilmelidir. Bu grafik Şekil 5.1'de verilmiştir.



**Şekil 5.1. RMSECV değerine karşı gizil değişken sayısının grafiği.**

Şekil 5.1 incelendiğinde PLSR analizinde modelde bulunması gereken ideal bileşen sayısının 6 olduğu açıkça görülmektedir. PLSR analizinde modelde bulunması gereken gizil değişken sayısına karar verdikten sonra 6 gizil değişken için yakalanan varyans oranlarını görmek için Çizelge 5.2 incelenmelidir.

**Çizelge 5.2. PLSR sonucunda açıklanan varyans değerleri.**

Gizil değişken (LV) Sayısı	X Bloğu		Y Bloğu	
	LV	Toplam	LV	Toplam
1	59,68	59,68	67,38	67,38
2	14,49	74,17	4,17	71,55
3	3,50	77,67	5,19	76,74
4	6,28	83,96	1,49	78,23
5	6,56	90,52	0,23	78,45
6	3,19	<b>93,71</b>	0,29	<b>78,75</b>
7	1,89	95,60	0,06	78,80
8	1,56	97,16	0,01	78,82
9	1,36	98,52	0,01	78,83
10	0,54	99,06	0,00	78,83
11	0,00	99,06	0,00	78,83

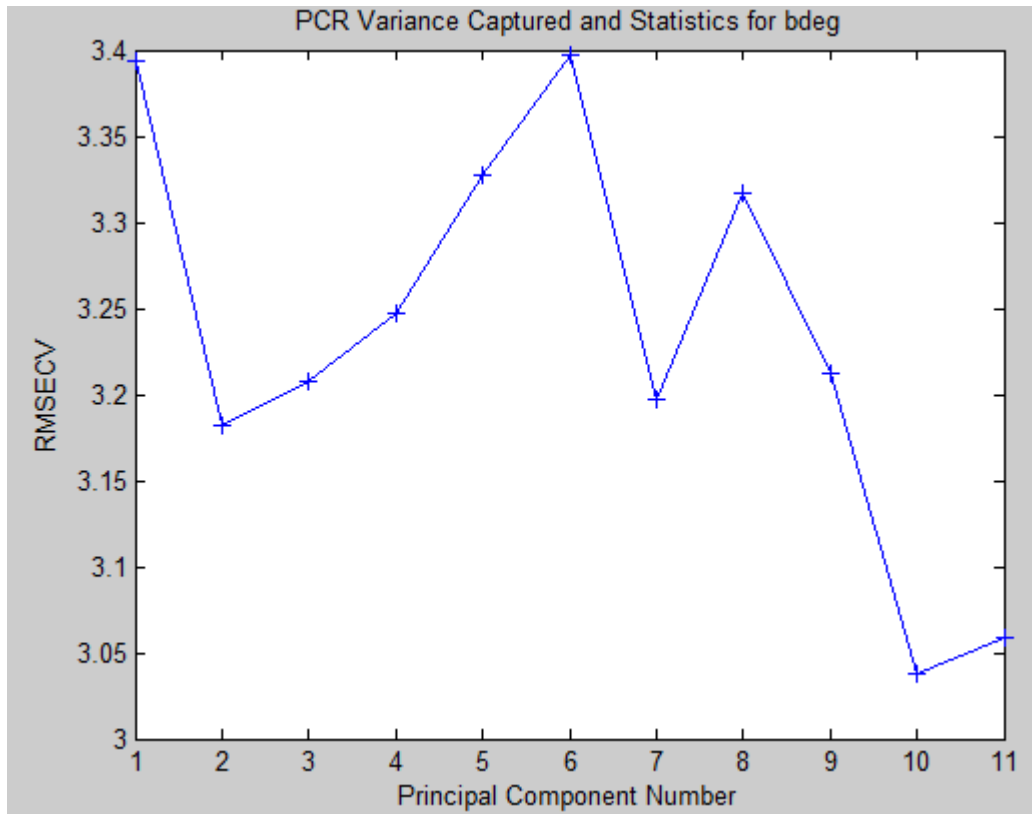


Çizelge 5.2 incelendiğinde 6 gizil değişken ile kurulan modelin bağımlı değişkeni % 78,75 oranında açıkladığı görülmektedir. Eğer modelde 11 bileşenin hepsi kullanılırsa bağımlı değişken %78,83 oranında açıklanabilmektedir. Ayrıca buna ek olarak bu 6 gizil değişken açıklayıcı değişkenleri % 93,71 oranında açıklamaktadır. 11 bileşen ile çalışmak yerine 6 bileşen ile modelin neredeyse tamamen açıklanabildiği görülmektedir.

Şekil 5.1 yardımıyla modelde 6 gizil değişken bulunacağına karar verildikten sonra bu gizil değişkenler ile kurulan modelin RMSECV değeri 3,00354 olarak hesaplanmıştır.

## 5.2. Temel Bileşenler Regresyonu

Matlab PLS\_Toolbox yardımıyla temel bileşenler regresyonunda ilk olarak modelde bulunması gereken ideal bileşen sayısına karar vermek için RMSECV değerine karşı temel bileşenlerin sayısının grafiği çizdirilmelidir. Bu grafik şekil 5.2'de görüldüğü gibidir.



### Şekil5.2. RMSECV değerine karşı Temel Bileşenlerin sayısının grafiği.

Şekil 5.2 incelendiğinde PCR modelinde kalması gereken ideal temel bileşen sayısının 10 olduğu görülmektedir. Modelde bulunması gereken temel bileşen sayısına karar verdikten sonra modelin varyans açıklama oranlarını incelenirse modelin bağımlı değişkeni ne kadar açıkladığı görülebilir. Bunun için oluşturulan Çizelge 5.3 aşağıdaki gibidir.

**Çizelge 5.3. PCR sonucunda açıklanan varyans değerleri.**

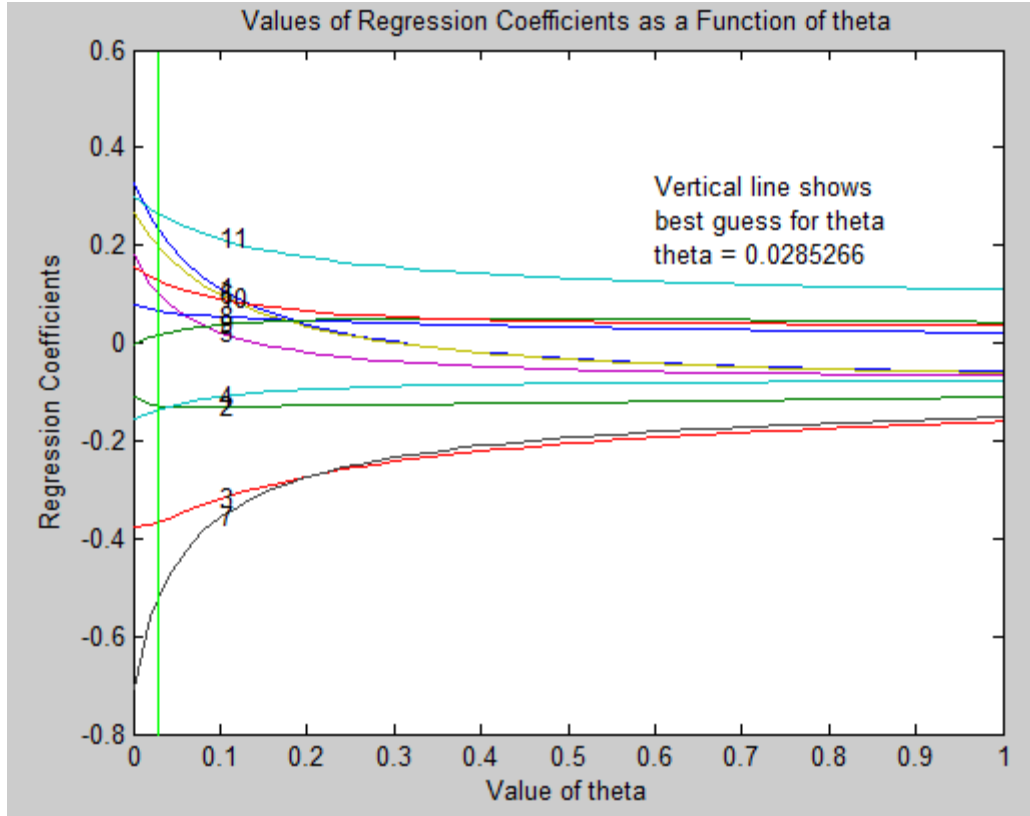
Temel Bileşen (PC) Sayısı	X Bloğu		Y Bloğu	
	PC	Toplam	PC	Toplam
1	59,88	59,88	64,95	64,95
2	14,70	74,59	4,70	69,64
3	9,22	83,80	0,11	69,75
4	7,51	91,32	0,24	69,99
5	2,62	93,93	0,02	70,01
6	2,04	95,97	0,14	70,15
7	1,32	97,29	4,46	74,61
8	1,00	98,29	0,29	74,90
9	0,84	99,13	1,85	76,75
10	0,61	<b>99,74</b>	1,96	<b>78,71</b>
11	0,26	100,00	0,12	78,83

Çizelge 5.3 incelendiğinde 10 temel bileşen ile kurulan modelin bağımlı değişkeni % 78,71 oranında açıkladığı görülmektedir. Ayrıca buna ek olarak bu 10 temel bileşen açıklayıcı değişkenleri % 99,74 oranında açıklamaktadır.

Şekil 5.2 yardımıyla modelde 10 temel bileşen bulunacağına karar verildikten sonra bu temel bileşenler ile kurulan modelin RMSECV değeri 3,03834 olarak hesaplanmıştır.

### 5.3. Ridge Regresyon

Ridge regresyonda önemli olan nokta ridge tahmin edicisini hesaplarken k değerinin seçimidir. Daha önceki bölümlerde açıklandığı üzere k değerinin seçimi için bir çok yöntem geliştirilmiştir. Burada k değeri ridge izi yöntemiyle seçilmektedir. Şekil 5.3 incelendiğinde uygun k değerinin 0,0285266 olacağı görülmektedir.



**Şekil 5.3. Ridge izi grafiği.**

Ridge izi yöntemiyle uygun  $k$  değerine karar verdikten sonra bu değer ile kurulan modelin bağımlı değişkeni %76,57 oranında açıkladığı görülmektedir. Ayrıca ridge regresyon yöntemi ile hesaplanan RMSECV 0,0513 olarak hesaplanmıştır.

#### 5.4. PLSR, PCR ve RR Yöntemlerinin Karşılaştırılması

Uygulama bölümünde ele alınan veri seti için kısım 5.1, 5.2 ve 5.3'te ayrı ayrı PLSR, PCR ve RR yöntemleri sırasıyla uygulanmıştır. Bu uygulamalar sonucunda hangi modelin tahmin gücünün daha iyi olacağına ilişkin yorum yapmamızı sağlayacak RMSECV değerleri hesaplanmıştır. Bu kısımda hesaplanan RMSECV değerleri yardımıyla yöntemler birbirleri ile karşılaştırılacaktır. Çizelge 5.4'te her bir yöntem için hesaplanan RMSECV değerleri verilmiştir.

**Çizelge 5.4. PLSR, PCR ve RR yöntemleri için RMSECV değerleri.**

	PLSR	PCR	RR
Boyut Sayısı	6	10	11
RMSECV	3,00354	3,03834	0,0513

Çizelge 5.4 incelendiğinde en küçük RMSECV değerine sahip regresyon modelinin Ridge regresyon olduğu açıkça görülmektedir. Bunun anlamı Ridge regresyon yöntemi tahmin başarısı bakımından PLSR ve PCR yöntemlerinden daha başarılıdır. Literatüre bakıldığında ridge regresyon yöntemi boyut indirgeme ile ilgilenmediğinden dolayı daha iyi tahmin yeteneğine sahip olması beklenen bir durumdur. PLSR ve PCR yöntemleri kıyaslandığında RMSECV değerlerinin birbirlerine oldukça yakın oldukları görülmektedir. Fakat PLSR yönteminde PCR yöntemine oranla daha az sayıda bileşen kullanıldığından yorumlama açısından kolaylık sağlamaktadır. Bu yüzden PLSR yöntemiyle elde edilen model PCR yöntemiyle elde edilen modele göre tahmin gücü bakımından daha iyi sonuçlar vermektedir.

## 6. SONUÇLAR VE TARTIŞMA

Bu tez çalışmasında lineer regresyon analizinde çok sık karşılaşılan bir problem olan çoklu iç ilişki problemi incelenmiştir. Ayrıca bu problemle başa çıkabilmek için geliştirilen yöntemlerden birisi olan yanlı regresyon yöntemi açıklanmış ve yanlı regresyon yöntemlerinden olan PLSR, PCR ve RR yöntemleri ayrıntılı olarak incelenmiştir.

Tez çalışmasının birinci bölümünde kısmi en küçük kareler yönteminin tarihçesinden ve gelişiminden bahsedilmiş ve kullanıldığı alanlar açıklanmıştır. İkinci bölümde çoklu lineer regresyon yöntemi ve burada sıklıkla kullanılan en küçük kareler tahmin edicisi ayrıntılı olarak incelenmiştir. Üçüncü bölümde lineer regresyon analizinde sıklıkla karşılaşılan çoklu iç ilişki probleminin ne olduğu neden kaynaklandığı ve nasıl giderileceği hakkında bilgiler verilmiştir. Dördüncü bölümde yanlı regresyon yöntemlerinden bahsedilmiştir. Yanlı regresyon yöntemlerinden olan kısmi en küçük kareler regresyonunda kullanılan algoritmalar tanıtılmış ve bu yöntem diğer yanlı yöntemlerle karşılaştırılmaya çalışılmıştır. Beşinci bölümde, anlatılan yanlı regresyon yöntemleri ile ilgili literatürden alınan bir veri seti üzerinde uygulama yapılmıştır. Yapılan uygulama çalışmasının sonucunda Ridge regresyon yönteminin tahmin gücünün beklendiği gibi diğerlerinden yüksek olduğu bulunmuştur. PLSR ve PCR karşılaştırıldığında ise PLSR yönteminin tahmin gücünün PCR ile çok yakın olduğu görülmüştür. Fakat PLSR yöntemi PCR yöntemine göre daha az bileşen ile model kurmaktadır. Bundan dolayı modelin yorumlanması daha kolay olacağından PLSR yönteminin PCR yöntemine göre daha üstün olacağı görülmüştür.

Uygulamanın sonucuna bakıldığında tahmin gücü bakımından modellerin birbirine yakın sonuçlar vermesine karşın PLSR yöntemi ile elde edilen modelin daha az bileşenli olduğu bulunmuştur. Modelde daha az bileşen bulunması yorumlamayı kolaylaştıracağından dolayı daha az bileşen ile model kuran PLSR yönteminin uygulanmasının daha iyi olacağı görülmüştür.

## KAYNAKLAR DİZİNİ

- Abdi, H. (2007). Partial Least Square Regression PLS-Regression. *Encyclopedia of Measurement and Statistics* .
- Alin, A., Kurt, S., McIntosh, A. R., Öniz, A., & Özgören, M. (2009). Partial Least Squares Analysis in Electrical Brain Activity. *Journal of Data Science* , 99-110.
- Bastien, P., Vinzi, V., & Tenenhaus, M. (2005). PLS generalized linear regression. *Computational Statistics & Data Analysis*, 17-46.
- Belsley, D. A., Kuh, E., & Welsch, R. E. (2004). *Regression Diagnostics*. New Jersey: John Wiley & Sons.
- Bush, B. L. & Nachbar, J. B. (1993). Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput. Aided Mol. Des.* , 587-619.
- Butler, N. A., & Denham, M. C. (2000). The peculiar shrinkage properties of partial least squares regression. *Journal of Royal Statistical Society* , 585-593.
- Carrascal, L. M., Galvan, I., & Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* , 681-690.
- D.M.Hawkins. (1973). On the investigation of alternative regressions by principal component analysis. *Applied Statistics* , 275-286.
- Dayal, B. S., & MacGregor, J. F. (1997). Improved PLS Algorithms. *Journal of Chemometrics* , 73-85.
- De Jong, S. (1993). SIMPLS: An Alternative Approach to Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems* , 251-263.
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. New York: Wiley.
- Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics* , 92-107.

**KAYNAKLAR DİZİNİ (devam)**

- Flury, B. & Riedwyl, H. (1988). *Multivariate Statistics. A practical approach.* London: Chapman and Hall.
- Frank, I. E. (1987). Intermediate least squares regression methods. *Chemometrics and Intelligent Laboratory Systems* , 233-242.
- Frank, I. E., & Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics* , 109-147.
- Frank, I. E., & Kowalski, B. R. (1984). Prediction of Wine Quality and Geographic Origin from Chemical Measurements by Partial Least Squares Regression Modeling. *Analytica Chimica Acta* , 241-251.
- Frank, I. E., Feikema, J., Constantine, N., & Kowalski, B. R. (1984). Prediction of Product Quality from Spectral Data Using the Partial Least Squares Method. *J. Chem. Inf. Comput. Sci.* , 20-24.
- Friedman, D. J., & Montgomery, D. C. (1985). Evaluation of the Predictive Performance of Biased Regression Estimators. *Journal of Forecasting*, 153-163.
- Frisch, R. (1934). *Statistical Confluence Analysis by Means of Complete Regression Systems.* Universitets Okonomiske Institut, Oslo.
- Garthwaite, P. H. (1994). An Interpretation of Partial Least Squares. *Journal of the American Statistical Association* , 122-127.
- Geladi, P., & Kowalski, B. R. (1986a). An example of 2-Block predictive partial least squares regression with simulated data. *Analytica Chimica Acta* , 19-32.
- Geladi, P., & Kowalski, B. R. (1986b). Partial Least Squares Regression: A Tutorial. *Analytica Chimica Acta* , 1-17.
- Goutis, C. (1996). Partial Least Squares Algorithm Yields Shrinkage Estimator. *The Annals of Statistics* , 816-824.
- Graybill, F. A., & Iyer, H. K. (1961). *Regression Analysis: Concepts and Applications.*

**KAYNAKLAR DİZİNİ (devam)**

Gunst, R. F., & Mason, R. L. (1980). *Regression Analysis and its Application: A Data-oriented Approach*. New York: Marcel Dekker.

Gunst, R. F., & Mason, R. L. (1979). Some considerations in the evaluation of alternate prediction equations. *Technometrics* , 55-63.

Haaland, D. M., & Thomas, E. V. (1988). Partial least squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* , 1193-1202.

Hawkins, D. M. (1973). On the investigation of alternative regressions by principal component analysis. *Appl. Statist.* , 275-286.

Helland, I. S. (1988). On the structure of partial least squares regression. *Commun. Statist.-Simula* , 581-607.

Helland, I. S. (1990). Partial Least Squares Regression and Statistical Models. *Scandinavian Journal of Statistics* , 97-114.

Helland, I. S. (2001). Some theoretical aspects of partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* , 97-107.

Hill, R. C., Fomby, T. B., & Johnson, S. R. (1977). Component selection norms for principal component regression. *Commun. Statist.-Theor. Method* , 309-334.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* , 1-49.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Non-orthogonal Problems. *Technometrics* , 55-67.

Hoerl, A.E., Kennard, R.W., & Baldwin, K. F. (1975). Ridge Regression: Some Simulation. *Communication in Statistics*, 105-123.

Hotelling, H. (1957). The relations of the newer multicariate statistical methods to factor analysis. *Brit. J. Statist. Psychol*, 69-79.



**KAYNAKLAR DİZİNİ (devam)**

- Höskuldsson, A. (1988). PLS Regression Methods. *Journal of Chemometrics* , 211-228.
- İnal, C., & Günay, S. (1999). *Olasılık ve Matematiksel İstatistik*.
- Jolliffe, I. T. (1982). A Note on the Use of Principal Components in Regression. *Applied Statistics* , 300-303.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. New York: Springer.
- Kendall, M. G. (1957). A course in multivariate analysis. London: Griffin.
- Kondylis, A. (2006). PLS methods in regression.
- Lawless, J. F. & Wang, P., 1976. A Simulation Study of Ridge and Other Regression Estimators. *Communication in Statistics*,139-164.
- Lindgren, F., & Rännar, S. (1998). Alternative Partial Least squares algorithms. *Perspective in Drug and Discovery and Design* , 105-113.
- Lingjærde, O. C., & Christophersen, N. (2000). Shrinkage Structure of Partial Least Squares. *Scandinavian Journal of Statistics* , 459-473.
- Lock, R. H. (1993). 1993 New Car Data. *Journal of Statistics Education*.
- Lorber, A., Wangen, L. E., & Kowalski, B. R. (1987). A theoretical foundation for the PLS algorithm. *Journal of Chemometrics* , 19-31.
- Lott, W. F. (1973). The optimal set of principal component restrictions on least squares regression. *Commun. Statist*, 449-464.
- Mallows, C. L. (1973). Some Comments on  $p$  . *C Technometrics*, 15, 661-675.
- Manne, R. (1987). Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometr. & Int. Lab. Syst.* , 187-197.
- Marquardt, D. W. (1970). Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation. *Technometrics* , 591-612.

### KAYNAKLAR DİZİNİ (devam)

- Marquart, D. W., & Snee, R. D. (1975). Ridge regression in practise. *The American Statistician*, 3-20.
- Martens, H. (1985). *Multivariate Calibration*. PhD thesis, Technical University of Norway, Trondheim, Norway (unpublished).
- Martens, H. & Naes, T. (1989). *Multivariate Calibration*. New York: Wiley.
- Mason, R. L., & Gunst, R. F. (1985). Selecting principal components in regression . *Statistics&Probability Letters* , 299-301.
- Mcdonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association*, 407-416.
- McIntosh, A. R., & Lobaugh, N. J. (2004). Partial Least Squares Analysis of Neuroimaging Data: Applications and Advances. *Neuroimage* , 250-263.
- Miller, A. J. (1984). Selection of subsets of regression variables(with discussion). *Journal of Royal Statistical Society* , 389-425.
- Miller, A. J. (1990). *Subset selection in regression*. Lodon: Chapman and Hall.
- Montgomery, D. C., & Peck, E. A. (2001). *Introduction to Linear Regression Analysis*. New York: John Wiley and Sons.
- Mosteller, F., & Tukey, J. W. (1977). *Data Analysis and Regression*. Addison-Wesley.
- Nadler, B., & Coifman, R. R. (2005). Partial least squares, Beer's law and the net analyte signal: statistical modeling and analysis. *Journal of Chemometrics* , 45-54.
- Naes, T., & Martens, H. (1985). Comparison of prediction methods for multicollinear data. *Communication Statistics-Simulation and Computation* , 545-576.
- Naes, T., Irgens, C., & Martens, H. (1986). Comparison of Linear Statistical Methods for Calibration of NIR Instruments. *Journal of the Royal Statistical Society* , 195-206.
- Özkale, M. R. (2007). *Çoklu İç İlişki İle İlgili Problemler, Doktora Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, 107 s. (yayınlanmamış).*

### KAYNAKLAR DİZİNİ (devam)

- Phatak, A. (1993). *Evaluation of some Multivariate Methods and their Applications in Chemical Engineering*. PhD thesis, University of Waterloo, Ontario, Canada (unpublished).
- Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied Regression Analysis: A Research Tool*.
- Rosipal, R. & Trejo, L. (2001). Kernel PLS in Reproducing Kernel Hilbert Spaces. *Journal of Machine Learning Research*, 97-123
- Rosipal, R., & Kramer, N. (2006). Overview and Recent Advances in Partial Least Squares. *Lecture Notes in Computer Science* , 34-51.
- Silvey, S. D. (1969). Multicollinearity and Imprecise Estimation. *Journal of the Royal Statistical Society, Series B* , 539-552.
- Stein, C. M. (1960). *Multiple Regression*. Stanford, California: Stanford University Press.
- Stone, M., & Brooks, R. J. (1990). Continuum Regression: Cross-Validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal component regression. *J. R. Statist. Soc.* , 237–269.
- Theobald, C. M. (1974). Generalizations of mean square error applied to ridge regression. *Journal of the Royal Statistical Society. Series B (methodological)*, 103-06
- Toscas, P. J., Shaw, F., & Beilken, B. L. (1999). Partial least squares (PLS) regression for the analysis of instrument measurements and sensory meat quality data. *Meat Science* , 173-178.
- Varmuza, K., & Filzmoser, P. (2008). *Introduction to Multivariate Statistical Analysis in Chemometrics*. New York: CRC Press, Taylor & Francis Group.
- Vinod, H. D., & Ullah, A. (1981). *Recent Advances in Regression Methods*. Marcel Dekker, New York
- Wallace, T. D. (1972). Weaker criteria and tests for linear restrictions in regression. *Econometrica* , 689-698.

**KAYNAKLAR DİZİNİ (devam)**

Webster, J. T., Gunst, R. F., & Mason, R. L. (1974). Latent Root Regression Analysis. *Technometrics* , 513-522.

Wold, H. (1966). Estimation of principal components and related methods by iterative least squares. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (pp. 391–420). New York: Academic Press.

Wold, H. (1975). Soft modelling by latent variables: the nonlinear iterative partial least squares approach. In Gani, J., editor, *Perspectives in Probability and Statistics, Papers in Honour of M.S. Bartlett, London, Academic Press*.

Wold, S., Martens, H., & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. *Lecture notes in mathematics. Proc. Conf. Matrix Pencils*, In A. Ruhe, and B. Kgstrm (ed.).

Wold, H. (1982). Soft Modelling: the basic design and some extensions. *Systems Under Indirect Observation* , 1-53.

Wold, H. (1984). Partial least squares. In *Encyclopedia o Statistical Scince*, Vol 6, eds. N. L. Johnson and S. Kotz, New York: Wiley, 581-591.

Wold, H. (1985). Partial Least Squares. *Encyclopedia of the Statistical Sciences* , 581-591.

Wold, S. (1993). Discussion: PLS in Chemical Practice. *Technometrics* , 136-139.

Yeniay, Ö., & Göktaş, A. (2002). A Comparison of Partial Least Squares Regression with Other Prediction Methods. *Hacettepe Journal of Mathematics and Statistics* , 99-111.