

**ÇUKUROVA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ**

**Yasemin CAN**

**SÜREKLİ REGRESYON VE İLİŞKİLİ REGRESYON MODELLERİNİN  
İNCELENMESİ**

**İSTATİSTİK ANABİLİM DALI**

**ADANA, 2013**

**ÇUKUROVA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**SÜREKLİ REGRESYON VE İLİŞKİLİ REGRESYON MODELLERİNİN  
İNCELENMESİ**

**Yasemin CAN**

**YÜKSEK LİSANS TEZİ**

**İSTATİSTİK ANABİLİM DALI**

Bu Tez 29/04/2013 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından  
Oybirliği/Oyçokluğu ile Kabul Edilmiştir.

.....  
Doç. Dr. Mahmude Revan ÖZKALE  
DANIŞMAN

.....  
Prof. Dr. Selahattin KAÇIRANLAR  
ÜYE

.....  
Yrd. Doç. Dr. Hüseyin GÜLER  
ÜYE

Bu Tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.

**Kod No:**

**Prof. Dr. Mustafa GÖK  
Enstitü Müdürü**

**Not:** Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

**ÖZ**

**YÜKSEK LİSANS TEZİ**

**SÜREKLİ REGRESYON VE İLİŞKİLİ REGRESYON MODELLERİNİN  
İNCELENMESİ**

**Yasemin CAN**

**ÇUKUROVA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
İSTATİSTİK ANABİLİM DALI**

Danışman : Doç. Dr. Mahmude Revan ÖZKALE  
Yıl: 2013, Sayfa:96

Jüri : Doç. Dr. Mahmude Revan ÖZKALE  
: Prof. Dr. Selahattin KAÇIRANLAR  
: Yrd. Doç. Dr. Hüseyin GÜLER

Regresyon analizi değişkenler arasındaki ilişkiyi araştırmak ve modellemek amacıyla çeşitli bilim dallarında yaygın olarak kullanılan istatistiksel bir yöntemdir. Bu amaç için ilk olarak regresyon katsayılarının tahminleri ile ilgilenilir ve bilinen en eski yöntem en küçük kareler yöntemidir. Fakat en küçük kareler tahmin edici, açıklayıcı değişkenlerin arasında lineer bağımlılık olması durumunda ortaya çıkan çoklu iç ilişki probleminden etkilenmektedir. Bu durumda alternatif tahmin yöntemleri kullanılmaktadır. Bu çalışmada çoklu iç ilişki olması durumunda kullanılan tahmin edicilerden ridge regresyon, temel bileşenler regresyon, kısmi en küçük kareler regresyon ve sürekli regresyon yöntemleri söz konusu yöntemler arasındaki ilişki incelenmiştir. Bu amaç için Fearn'nın (1983) verisi kullanılarak ilgili yöntemler örneklendirilmiştir.

**Anahtar Kelimeler:** Sürekli Regresyon, En Küçük Kareler Tahmin Edici, Temel Bileşenler Regresyon, Kısmi En Küçük Kareler Regresyon, Ridge Regresyon.

## ABSTRACT

### MSc THESIS

<b>CONTINUUM REGRESSION AND EXAMINING RELATED REGRESSION MODELS</b>
---

**Yasemin CAN**

**ÇUKUROVA UNIVERSITY  
INSTITUTE OF NATURAL AND APPLIED SCIENCES  
DEPARTMENT OF STATISTICS**

Supervisor : Assoc. Prof. Dr. Mahmude Revan ÖZKALE  
Year: 2013, Pages: 96  
Jury : Assoc. Prof. Dr. Mahmude Revan ÖZKALE  
: Prof. Dr. Selahattin KAÇIRANLAR  
: Asst. Prof. Dr. Hüseyin GÜLER

Regression analysis is a common statistical method used in various scientific disciplines in order to search and model the relationships among variables. For this purpose, firstly estimation of regression coefficients are dealt with and the oldest known method is the least squares method. However, linear least squares estimated coefficients are affected from the multicollinearity problem arises when there is a linear relationship between the explanatory variables. In this case, alternative estimation methods are used. In this study, ridge regression, principle component regression, partial least squares regression and continuum regression methods among the estimation methods are used in case of multicollinearity problem and the relationships between these methods have been examined. For this purpose, the data set in Fearn (1983) is exemplified.

**Key Words:** Continuum Regression, Ridge Regression, Partial Least Squares Regression, Principal Components Regression, Least Squares Regression

## **TEŐEKKÜR**

Tez alıřmamın her ařamasında ilminden faydalandığım, insani ve ahlaki deęerleri ile örnek edindiğim, kendisiyle alıřmaktan onur duyduğum ve ayrıca tecrübelerinden yararlanırken göstermiş olduęu hořgörü ve sabırdan dolayı deęerli hocam Sayın Do. Dr. M. Revan ÖZKALE'ye teőekkürlerimi sunarım.

Beni her zaman destekleyerek bu günlere gelmemi saęlayan aileme sonsuz teőekkürler.

## İÇİNDEKİLER

## SAYFA

ÖZ.....	I
ABSTRACT .....	II
TEŞEKKÜR .....	III
İÇİNDEKİLER.....	IV
TABLolar DİZİNİ.....	VI
ŞEKİLLER DİZİNİ .....	VIII
SİMGELER VE KISALTMALAR .....	X
1. GİRİŞ .....	1
2. EN KÜÇÜK KARELER YÖNTEMİ .....	3
3. ÇOKLU İÇ İLİŞKİ .....	7
3.1. Çoklu İç İlişkinin Tanımlanması.....	7
3.2. Çoklu İç İlişkinin Belirlenmesi .....	8
3.2.1. Korelasyon Matrisinin İncelenmesi .....	8
3.2.2. Varyans Şişirme Faktörü.....	8
3.2.3. $X'X$ Matrisinin Özdeğer Analizi .....	8
3.2.4. Diğer Tanımlama Yöntemleri.....	9
3.3. Çoklu İç İlişkinin İstatistiksel Sonuçları.....	9
3.4. Çoklu İç İlişkinin Ortadan Kaldırılması .....	10
4. ÇOKLU İÇ İLİŞKİ OLMASI DURUMUNDA KULLANILAN YÖNTEMLER	11
4.1. Ridge Regresyon .....	11
4.2. Temel Bileşenler Regresyon .....	14
4.3. Kısmi En Küçük Kareler Regresyon .....	20
5. SÜREKLİ REGRESYON .....	25
5.1. Sürekli Regresyon ile En Küçük Kareler Yöntemi Arasındaki İlişki.....	32
5.2. Sürekli Regresyon ile Temel Bileşenler Regresyon Arasındaki İlişki .....	34
5.3. Sürekli Regresyon ile Kısmi En Küçük Kareler Regresyon Arasındaki İlişki	35
5.4. Sürekli Regresyon ile Ridge Regresyon Arasındaki İlişki .....	36
5.5. EKK, PCR, PLSR, RR ve CR Arasındaki İlişki .....	47
6. SÜRELİ REGRESYONDA SÜREKSİZLİKLER .....	51

7. MODEL SEÇME KRİTERLERİ.....	59
7.1. Hata Kareler Ortalamasının Karekökü .....	59
7.2. Ayarlama Kümesine Dayalı Geçerlilik.....	60
7.3. Öntahmin Testi.....	60
7.4. Çapraz Geçerlilik.....	61
8. UYGULAMA.....	65
9. SONUÇLAR VE ÖNERİLER .....	85
KAYNAKLAR.....	87
ÖZGEÇMİŞ.....	91
EKLER.....	92

**TABLULAR DİZİNİ****SAYFA**

Tablo 5.1. EKK, PLSR ve PCR yöntemlerinin sürekli regresyon ve ridge regresyona karşılık gelen parametre değerleri .....	45
Tablo 7.1. RMSE ve MSE ölçülerinin amaçlarına göre sınıflandırılması .....	59
Tablo 8.1. EKK yöntemi ile bulunan katsayılar, VIF değerleri ve özdeğerler.....	67
Tablo 8.2. 24 gözlem PLSR modeline ait RMSECV değerleri.....	68
Tablo 8.3. 24 gözlem için PLSR yöntemi kullanarak oluşturulan regresyon modeli tarafından açıklanan varyans yüzdeleri.....	68
Tablo 8.4. 24 gözlem için PCR modeline ait RMSECV değerleri .....	69
Tablo 8.5. 24 gözlem için PCR yöntemi kullanılarak oluşturulan regresyon modeli tarafından açıklanan varyans yüzdeleri.....	70
Tablo 8.6. $k$ 'nın çeşitli değerlerine karşı gelen $PRESS(k)$ ve RMSECV değerleri .....	70
Tablo 8.7. EKK, PCR, PLSR ve RR modelleri için RMSEC, RMSEP ve RMSECV değerleri .....	74
Tablo 8.8. EKK yöntemi ile bulunan katsayılar, VIF değerleri ve özdeğerler.....	76
Tablo 8.9. 12 gözlem için PLSR modeline ait RMSECV değerleri .....	77
Tablo 8.10. 12 gözlem için PLSR yöntemi kullanarak oluşturulan regresyon modeli tarafından açıklanan varyans yüzdeleri.....	77
Tablo 8.11. 12 gözlem için PCR modeline ait RMSECV değerleri .....	78
Tablo 8.12. 12 gözlem için PCR yöntemi kullanılarak oluşturulan regresyon modeli tarafından açıklanan varyans yüzdeleri.....	79
Tablo 8.13. $k$ 'nın çeşitli değerlerine karşı gelen $PRESS(k)$ ve RMSECV değerleri .....	79
Tablo 8.14. 12 gözlem için EKK, PCR, PLSR ve RR modelleri için RMSEC, RMSEP ve RMSECV değerleri .....	83





## ŞEKİLLER DİZİNİ

## SAYFA

Şekil 2.1. En küçük kareler uyumu.....	4
Şekil 4.1. Temel bileşenler fikri .....	15
Şekil 4.2. PCA ve PCR arasındaki ilişki.....	16
Şekil 4.3. Ridge regresyon ve temel bileşen regresyon yöntemlerinde büzülme.....	18
Şekil 5.1. CR parametresinin çeşitli değerlerine karşı gelen regresyon yöntemleri ..	25
Şekil 5.2. Mümkün olan tüm ( $K^2(c), V(c)$ ) değerlerinin gösterilmesi.....	44
Şekil 5.3. Gerçek regresyon katsayısı $\beta = (4 \quad 2)$ ve aralarındaki korelasyon $\mp 0.5$ olan iki açıklayıcı değişkenli regresyon için farklı yöntemlerin katsayı profilleri.....	50
Şekil 6.1. $X'y = (3,1)$ ve $XX = diag(100,0.01)$ olmak üzere $\gamma = 0.6$ (düz çizgi) ve $\gamma = 0.7$ (kesikli çizgi) değerleri ile $c$ değerleri kullanılarak $T(\gamma, c)$ fonksiyonunun maksimum yapılması .....	51
Şekil 6.2. Ordinatu $\gamma(k)$ ve apsisi $k$ olmak üzere (6.2) eşitliğinin gösterimi. ....	54
Şekil 8.1. 24 gözlem için PLSR yöntemine ilişkin RMSECV grafiği.....	67
Şekil 8.2. 24 gözlem için PCR yöntemine ilişkin RMSECV grafiği.....	69
Şekil 8.3. Ridge regresyon için $RMSECV$ grafiği.....	71
Şekil 8.4. RR, PLSR, PCR ve EKK modellerindeki regresyon katsayıları .....	72
Şekil 8.5. 12 gözlem için PLSR yöntemine ilişkin RMSECV grafiği.....	76
Şekil 8.6. 12 gözlem için PCR yöntemine ilişkin RMSECV grafiği.....	78
Şekil 8.7. Ridge regresyon için $RMSECV$ grafiği.....	80
Şekil 8.8. RR, PLSR, PCR ve EKK modellerindeki regresyon katsayıları .....	81



## SİMGELER VE KISALTMALAR

- BLUE : En İyi Lineer Yansız Tahmin Edici (Best Linear Unbiased Estimator)
- CR : Sürekli Regresyon (Continuum Regression)
- CV : Çapraz Geçerlilik (Cross-Validation)
- EKK : En Küçük Kareler (Ordinary Least Squares)
- MSE : Hata Kareler Ortalaması (Mean Squared Error)
- NIPALS : Doğrusal Olmayan Yinelemeli En Küçük Kareler Algoritması  
(Nonlinear Iterative Partial Least Squares)
- PCR : Temel Bileşenler Regresyon (Principal Components Regression)
- PLSR : Kısmi En Küçük Kareler Regresyon (Partial Least Squares Regression)
- RMSE : Hata Kareler Ortalamasının Karekökü (Root Mean Squared)
- RMSEC : Ayarlama Hata Kareler Ortalamasının Karekökü (Root Mean Squared  
Error of Calibration)
- RMSECV : Çapraz Geçerlilik Hata Kareler Ortalamasının Karekökü (Root Mean  
Squared Error of Cross-Validation)
- RMSEP : Öntahmin Hata Kareler Ortalamasının Karekökü (Root Mean Squared  
Error of Prediction)
- RR : Ridge Regresyon (Ridge Regression)
- SIMPLS : PLSR yöntemine alternatif bir yaklaşım (An Alternative Approach to  
Partial Least Squares Regression)
- SMSE : Skaler Hata Kareler Ortalaması (Scaler Mean Squared Error)
- VIF : Varyans Şişirme Faktörü (Variance Inflation Factor)





## 1. GİRİŞ

Regresyon (bağlanım) sözlük anlamıyla bir şeyi başka bir şeye bağlama işi ve biçimidir. Bilimsel olarak regresyon terimi bir değişkenle başka bir (ya da birden çok) değişken arasında ilişki kurma işini ve ilişkinin biçimini anlatır.

İstatistiksel anlamda biri yanıt (bağımlı) diğeri açıklayıcı (bağımsız) değişken olmak üzere en az iki değişken arasındaki fonksiyonel ilişkiyi belirlemek için kullanılan bir yöntemdir.

Yanıt değişken, açıklayıcı değişkenlerin etkisi altında bulunan değişken olarak tanımlanır. Ele aldığımız bir konuda, yanıt değişken sonuç değişken olarak da adlandırılabilir. Değeri rasgele koşullara göre belirlenen ve başka değişkenlerin değişimi üzerinde etkide bulunan değişkenlere ise açıklayıcı veya bağımsız değişken denir.

Regresyon yönteminin temelini en küçük kareler (EKK) prensibi oluşturur. EKK prensibi ilk olarak Adrian Marie Legendre tarafından “Moindres Carrés” olarak 1805 yılında ortaya atılmıştır. C.F. Gauss EKK prensibi üzerinde çalışarak 1821 yılında yayınladığı bir makalede kendi adını verdiği Gauss-Markov teoremini bulmuştur. Udny Yüle ve Karl Pearson ise daha geniş istatistiksel alanlara uygulayarak geliştirmiştir.

EKK yöntemi parametre tahmini yaparak istatistiksel sonuç çıkarmaya yardımcı olur. Regresyon analizinde yaygın kullanıma sahip EKK yönteminde parametre tahmininde kullanılacak varsayımlar, bu yöntemle elde edilen tahmin edicilerin özelliklerini belirlemede yardımcı olmaktadır.

Bu çalışma teorik anlatım ve uygulama olmak üzere iki kısımdan oluşmaktadır. Teorik kısımda ilk olarak EKK regresyon ele alınmıştır. Ancak çoklu iç ilişki olması durumunda EKK yöntemi sağlıklı sonuçlar vermeyeceğinden yanlı regresyon yöntemleri olan ridge regresyon (RR), temel bileşenler regresyon (PCR) ve kısmi en küçük kareler regresyon (PLSR) yöntemleri EKK yöntemine alternatif olarak incelenmiştir. Ayrıca Stone ve Brooks’un (1990) EKK, PLSR ve PCR ile aralarındaki ilişkiyi daha iyi anlayabilmek için önerdiği sürekli regresyon (CR)

yöntemi incelenerek yukarıda belirtilen regresyon yöntemleri ile ilişkisi eşitlikler yardımıyla gösterilmiştir.

Uygulama kısmında ise Matlab programı kullanılarak Fearn'nın (1983) verisi üzerinde EKK, PCR, PLSR, RR ve CR yöntemlerinin uygulaması yapılmıştır.



## 2. EN KÜÇÜK KARELER YÖNTEMİ

$X$ ,  $n$  gözlem sayısı olmak üzere  $p$  açıklayıcı değişkenin  $n \times p$  boyutlu matrisi,  $y$   $n \times 1$  boyutlu yanıt değişken vektörü,  $\beta$   $p \times 1$  boyutlu regresyon katsayı vektörü ve  $\varepsilon$   $n \times 1$  boyutlu rasgele hatalar vektörü olmak üzere verilerin merkezileştirildiği varsayılarak çoklu doğrusal regresyon modeli

$$y = X\beta + \varepsilon \quad (2.1)$$

şeklinde gösterilsin. (2.1) modeli tüm gözlemler dikkate alındığında

$$\begin{aligned} y_i &= f(X) + \varepsilon_i \\ &= \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \\ &= \sum_{j=1}^p x_{ij} \beta_j + \varepsilon_i \quad i = 1, \dots, n \end{aligned}$$

olarak da yazılabilir. İstatistiksel çalışmalarda tahmin edicilerin tutarlı olması için, yanıt değişkenin tahmini  $f(X)$  olmak üzere, kayıp fonksiyonunun beklenen değeri  $E[L(y, f(X))]$ 'in minimum olması gerekir. Buna göre kayıp fonksiyonu

$$L(y, f(X)) = (y - f(X))^2 \quad (2.2)$$

ve  $E(y/X)$  regresyon fonksiyonu

$$\begin{aligned} f^*(X) &= \min_{f(X)} E(y - f(X))^2 \\ &= E(y/X) \end{aligned}$$

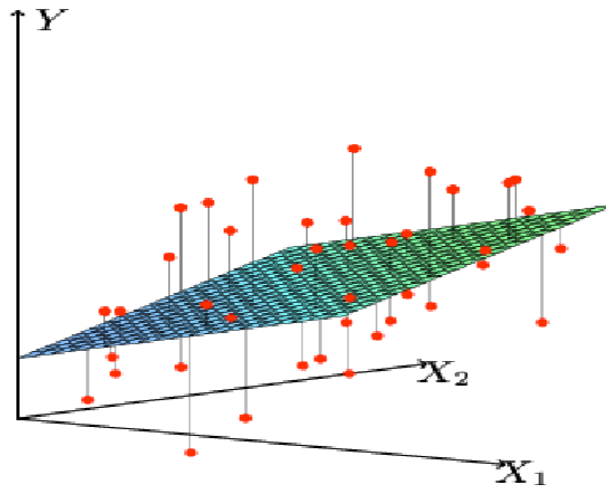
olarak gösterilir (<http://sites.stat.psu.edu/~jiali/course/stat597e/notes2/lreg.pdf>).

Regresyon modeli uydurulduktan sonra modelin doğru modele ne derece yaklaştığını ve regresyon analizinin tüm varsayımlarını sağlayıp sağlamadığını kontrol etmek gerekir.

Doğrusal regresyon yöntemini kullanmak için temelde şu varsayımların sağlandığı kabul edilmektedir:

1. Hata terimi  $\varepsilon_i$  ortalaması sıfıra eşit ve varyansı  $\sigma^2$  (sabit yani değişen varyanslılık yok) olmak üzere normal dağılıma sahip rasgele değişkendir  $\varepsilon_i \sim N(0, \sigma^2)$ .
2. Hatalar birbiri ile ilişkisizdir:  $Cov(\varepsilon_i, \varepsilon_j) = 0$  ( $i \neq j$ ).
3. Doğrusal regresyon analizinde açıklayıcı değişkenler sabit (deterministik) olmasına rağmen, yanıt değişken rasgeledir.
4. Değişkenler hatasız ölçülmüştür.
5. Yanıt ve açıklayıcı değişkenler arasındaki ilişki doğrusaldır.
6. Gözlem sayısı ( $n$ ), değişken sayısından ( $p$ ) büyük olmalıdır:  $n > p$

EKK, Şekil 2.1'de görüldüğü gibi yanıt değişkene karşı açıklayıcı değişkenlerin grafiği çizildiğinde gözlem değerlerinin düzlemden sapmalarının kareleri toplamının minimum olmasını amaçlar.



Şekil 2.1. En küçük kareler uyumu (Hastie ve ark., 2008)

$rank(X) = p$  ( $X$  değişkeni tam kolon ranklı) iken  $\beta$  parametresinin EKK tahmin edicisi  $\hat{\beta}$ ,  $\beta$ 'nin en iyi lineer yansız tahmin edicisidir (BLUE).

$$\begin{aligned} RSS(\hat{\beta}) &= \sum_{i=1}^n \hat{\varepsilon}_i^2 = \hat{\varepsilon}'\hat{\varepsilon} = (y - X\hat{\beta})'(y - X\hat{\beta}) \\ &= (y' - \hat{\beta}'X')(y - X\hat{\beta}) \\ &= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

olarak ifade edilen rezidü kareler toplamı,  $(y'X\hat{\beta})' = \hat{\beta}'X'y$  olduğundan  $\hat{\varepsilon}'\hat{\varepsilon} = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$  olarak yazılır.  $\hat{\varepsilon}'\hat{\varepsilon}$  ifadesinin  $\hat{\beta}$ 'ya göre türevi alınıp sıfıra eşitlendiğinde normal denklemler

$$\frac{\partial(\hat{\varepsilon}'\hat{\varepsilon})}{\partial\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

$$\Rightarrow X'X\hat{\beta} = X'y$$

olarak elde edilir. Buradan EKK tahmin edicisi,  $S = X'X$  ve  $s = X'y$  gösterimlerinin kullanılması ile

$$\hat{\beta} = (X'X)^{-1} X'y = S^{-1}s \quad (2.3)$$

olur.  $\hat{\beta}$ 'nin beklenen değeri

$$\begin{aligned} E(\hat{\beta}) &= E[(X'X)^{-1} X'y] \\ &= (X'X)^{-1} X'E(y) \\ &= \underbrace{(X'X)^{-1} X'X}_I \beta \\ &= \beta \end{aligned}$$

ve kovaryans matrisi

$$\begin{aligned}\text{cov}(\hat{\beta}) &= \text{cov}[(X'X)^{-1}X'y] \\ &= [(X'X)^{-1}X'] \underbrace{\text{Cov}(y)}_{\sigma^2 I} [(X'X)^{-1}X'] \\ &= \sigma^2 \underbrace{(X'X)^{-1}X'X}_{I} (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1}\end{aligned}$$

olarak bulunur.

### 3. ÇOKLU İÇ İLİŞKİ

#### 3. 1. Çoklu İç İlişkinin Tanımlanması

Çoklu doğrusal regresyon modeli oluşturulurken açıklayıcı değişkenlerin birbirinden bağımsız olduğu varsayılır.  $X_1, X_2, \dots, X_p$ ,  $X$  matrisinin kolonlarını göstermek üzere

$$\sum_{j=1}^p t_j X_j = 0 \quad (3.1)$$

eşitliğinde hepsi sıfır olmayan  $t_1, t_2, \dots, t_p$  sabitleri varsa  $X_1, X_2, \dots, X_p$  vektörleri lineer bağımlıdır. (3.1) eşitliği  $X$ 'in kolonlarının bir kümesi için tam olarak sağlanıyorsa "tam çoklu iç ilişki" vardır denir (Montgomery ve ark., 2001; Özkale, 2007).Çoklu iç ilişkinin pek çok nedeni olabilir (Judge ve ark., 1985; Montgomery ve ark., 2001).

1. Uygulanan Veri Toplama Yöntemi: Uygulama yapacak kişinin (3.1) ile tanımlı bir bölgenin alt uzayından örneklem alması çoklu iç ilişki problemine neden olur.
2. Modeldeki veya Kitledeki Zorunluluklar: Kitledeki zorunluluklar daha çok kimyasal ve üretim proseslerinde ortaya çıkar. Örneğin, bir kimyasal reaksiyonun gerçekleşmesi için belli içeriklerin sabit oranlarda alınması.
3. Modelin Belirlenmesi:  $X$  değişkenlerinin değişim aralığı küçük iken bir regresyon modeline polinom terimin eklenmesi çoklu iç ilişkiye neden olabilir.
4. Aşırı Tanımlı Modeller: Gözlemlerden çok açıklayıcı değişkenlerin olduğu modellere "aşırı tanımlanmış model" denir. Daha çok tıbbi ve ekonomik araştırmalarda ortaya çıkar. Bu gibi durumlarda açıklayıcı değişkenlerin bazılarını atmak gerekmektedir.

## 3.2. Çoklu İç İlişkinin Belirlenmesi

### 3.2.1. Korelasyon Matrisinin İncelenmesi

Açıklayıcı değişkenler standartlaştırılmış formda olmak üzere  $X'X$  matrisinin köşegen dışı elemanları  $r_{ij}$  olsun.  $|r_{ij}| \rightarrow 1$  iken  $x_i$  ve  $x_j$  açıklayıcı değişkenleri lineer olarak bağımlıdır.  $r_{ij}$ 'nin incelenmesi sadece ikili açıklayıcı değişkenler arasındaki lineer bağımlılığı incelemeye yararlıdır. Fakat ikiden fazla açıklayıcı değişken arasında lineer bağımlılık varsa  $r_{ij}$ 'ler uygun ölçü olmayabilir (Montgomery ve ark, 2001).

### 3.2.2. Varyans Şişirme Faktörü

Varyans şişirme faktörü (VIF) ilk kez Marquardt (1970) tarafından önerilmiştir.  $(X'X)^{-1}$  matrisinin  $j$ . köşegen elemanları,  $(VIF)_j$  değerini verir.  $R_j^2$  ise  $j$ . açıklayıcı değişkenin diğer  $p-1$  açıklayıcı değişken ile oluşturulan regresyon modelinden elde edilen çoklu belirleyicilik katsayısı olmak üzere,  $VIF_j = c_{jj} = (1 - R_j^2)^{-1}$  şeklinde tanımlanır.  $R_j^2 > 0,99$  ile  $(VIF)_j > 100$  aynı anlama gelmektedir.  $R_j^2$  1'e yakınsadıkça yani tolerans değeri  $TV_j = 1 - R_j^2$  küçük iken  $VIF_j$  değeri de artar. Bu durum açıklayıcı değişkenler arasında çoklu iç ilişki olduğunu gösterir. Analiz sonucunun güvenilir olması için VIF değerlerinin 10'dan küçük olması beklenir.

### 3.2.3. $X'X$ Matrisinin Özdeğer Analizi

Çoklu iç ilişkiyi belirlemenin diğer bir yolu da  $X'X$  matrisinin özdeğerlerini  $(\lambda_1, \lambda_2, \dots, \lambda_p)$  incelemektir. Bir veya daha fazla özdeğerin sıfıra eşit olması çoklu iç ilişkinin varlığını gösterir.  $\lambda_{\min}$  ve  $\lambda_{\max}$ ,  $X'X$  matrisinin en küçük ve en büyük

özdeğerleri olmak üzere,  $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$  olarak tanımlanan koşul sayısı çoklu iç ilişkiyi belirlemek için kullanılır. Genel olarak koşul sayısı 100'den küçük ise çoklu iç ilişki probleminin ciddi boyutlarda olmadığı, 100 ile 1000 arasında ise güçlü bir iç ilişki problemi olduğu, 1000'den büyük ise çok ciddi boyutlarda iç ilişki problemi olduğu söylenebilir (Özkale, 2007).

### 3.2.4. Diğer Tanımlama Yöntemleri

$X'X$  matrisinin determinantı alınarak da açıklayıcı değişkenler arasında çoklu iç ilişki olup olmadığı anlaşılabilir. Standartlaştırılmış formda  $0 \leq |X'X| \leq 1$ 'dir.  $|X'X|=1$  olması durumunda açıklayıcı değişkenler birbirine ortogondur,  $|X'X|=0$  olduğunda ise tam çoklu iç ilişki vardır.  $|X'X|$  sifıra yaklaştıkça çoklu iç ilişkinin şiddeti artacaktır (Farrar ve Glauber, 1967; Özkale, 2007).

### 3.3. Çoklu İç İlişkinin İstatistiksel Sonuçları

Açıklayıcı değişkenler arasında tam çoklu iç ilişki problemi söz konusu iken,  $X'X$  matrisi tekil olacağı için tersi hesaplanamaz. Yaklaşık çoklu iç ilişki ((3.1) eşitliği  $X$ 'in kolonlarının bazı alt kümeleri için yaklaşık olarak sağlanıyorsa) durumunda  $(X'X)^{-1}$  hesaplanırsa dahi  $\hat{\beta}_j$ 'lerin varyansı kötü koşulluluk nedeni ile şişirileceğinden  $\hat{\beta}_j$ 'ler kararlı olmaz. Çoklu iç ilişki katsayıların tahminlerini işaret bakımından da etkiler. Yani araştırmacının pozitif çıkmasını beklediği katsayı tahmininin işareti negatif, negatif çıkmasını beklediği katsayı tahmininin işareti pozitif çıkabilir. Ayrıca çoklu iç ilişki katsayıların tahminlerini değerce ve işaretçe etkilediğinden yanıt değişkenin öntahminine ait standart hata da büyük olur.

Çoklu iç ilişkinin güven aralıkları üzerinde de olumsuz etkileri vardır. EKK tahmin edicisinin varyansını arttırarak regresyon katsayıları için EKK tahminlerinin güven aralıklarının genişlemesine neden olur (Özkale, 2007).

#### 3.4. Çoklu İç İlişkinin Ortadan Kaldırılması

Çoklu iç ilişki probleminin ortadan kaldırılması veya etkisinin azaltılması konusunda çeşitli düşünceler önerilmiştir.

Ek verinin toplanması: Farrar ve Glauber (1967) ve Silvey (1969) çoklu iç ilişkiyi ortadan kaldıracak şekilde ek verinin toplanmasını önermişlerdir. Bazen yeni gözlem değerlerinin elde edilmesiyle çoklu iç ilişki problemi ortadan kaldırılabilir. Ancak modeldeki veya kitledeki kısıtlamalardan dolayı bu her zaman mümkün olmayabilir.

Modelin yeniden belirlenmesi: Çoklu iç ilişki problemi modelin seçiminden de kaynaklanabilir (Gujarati, 1995; Netter ve ark., 1990).

- Açıklayıcı değişkenlerden bazılarının modelden çıkarılması çoklu iç ilişkiyi azaltabilir. Ancak bu modelin etkinliğini de azaltabilir.
- Farklar alınarak değişkenler dönüştürülebilir. Fakat böyle bir dönüşüm hatalar arasında otokorelasyon problemine yol açabilir. Ayrıca böyle bir dönüşüm sadece zaman serilerine uygulanabilirken, kesit verilerine uygulanamamaktadır.
- Birbirleriyle ilişkili olan iki değişken yerine bu iki değişkeni temsil edebilecek tek bir değişken alınabilir.

Alternatif regresyon analizlerinin kullanılması: Çoklu iç ilişkinin etkilerini azaltmaya yönelik geliştirilmiş tahmin edicilerin kullanılması da en etkin yöntemlerden birisidir. Bu tahmin ediciler genellikle yanlış tahmin ediciler olarak isimlendirilir.



#### 4. ÇOKLU İÇ İLİŞKİ OLMASI DURUMUNDA KULLANILAN YÖNTEMLER

##### 4. 1. Ridge Regresyon

Ridge regresyon (RR) ilk kez 1962 yılında Hoerl tarafından ele alınmıştır.

Regresyon katsayılarının  $k \sum_{j=1}^p \beta_j^2$  “ceza” terimi ile  $\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2$  ifadesinin  $\beta$ ’ya göre minimum yapılmasıyla

$$\begin{aligned} \hat{\beta}(k) &= \arg \min_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + k \sum_{j=1}^p \beta_j^2 \right\} \\ &= \min_{\beta} \{ (y - X\beta)'(y - X\beta) + k\beta'\beta \} \end{aligned} \quad (4.1)$$

ifadesi elde edilir. Ridge sabiti  $k > 0$  olmak üzere büzülme miktarını kontrol eder. (4.1) denkleminin çözümü yapıldığında Hoerl ve Kennard’ın (1970) verdiği ridge tahmin edici

$$\begin{aligned} \hat{\beta}(k) &= \left[ \sum_{i=1}^n x_i x_i' + kI_p \right]^{-1} \left( \sum_{i=1}^n x_i y_i \right) \\ &= (X'X + kI)^{-1} X'y \\ &= (S + kI)^{-1} s \end{aligned}$$

olur. Böylece  $X'X$  matrisinin köşegen elemanlarına pozitif  $k$  sabitinin eklenmesiyle tam ranklı olan, dolayısıyla tekil olmayan  $(X'X + kI)$  matrisi tanımlanır.

$\hat{\beta}(k)$ ’nın yanlı bir tahmin edici olduğunu göstermek amacıyla  $Q = [I + k(X'X)^{-1}]^{-1}$  ve  $W_k = (X'X + kI)^{-1}$  ifadeleri verilsin.  $Q$  ile  $W_k$  arasında

$$Q = I - k(X'X + kI)^{-1} = I - kW_k$$

şeklinde bir ilişki olduğu varsayalım. Bu durumda  $\hat{\beta}(k)$  tahmin edicisi,

$$\begin{aligned}\hat{\beta}(k) &= (X'X + kI)^{-1} X'y \\ &= (X'X + kI)^{-1} X'X\hat{\beta} \\ &= (X'X + kI)^{-1} [(X'X)^{-1}]^{-1} \hat{\beta} \\ &= [(X'X)^{-1} (X'X + kI)^{-1}]^{-1} \hat{\beta} \\ &= [I + k(X'X)^{-1}]^{-1} \hat{\beta} \\ &= Q\hat{\beta}\end{aligned}$$

olarak ifade edilir. Böylece ridge tahmin edici ile EKK tahmin edici arasında bir ilişki olduğu görülür. Ortonormal girdiler (yani  $X'X = I$  iken) olması durumunda

ridge tahmin edicisi  $\hat{\beta}(k) = \frac{\hat{\beta}}{(1+k)}$  olarak bulunur.  $k \neq 0$  olmak üzere ridge tahmin

edici, EKK tahmin edici üzerine büzülme uyguladığından  $\hat{\beta}(k)$ 'nin boyu  $\hat{\beta}$ 'den daha kısadır:  $\hat{\beta}(k)' \hat{\beta}(k) < \hat{\beta}' \hat{\beta}$ .

$X$  matrisinin tekil değer ayrışımı RR hakkında daha ayrıntılı bilgi elde edilmesini sağlar.  $n \times p$  boyutlu ( $n > p$ )  $X$  matrisinin tekil değer ayrışımı

$$X = UDV' \tag{4.2}$$

olarak ifade edilir. (4.2) eşitliğinde  $U_{n \times p}$ ,  $X$  matrisinin kolon uzayını ve  $V_{n \times p}$ ,  $X$  matrisinin satır uzayını gösteren ortogonal matrisler ( $U'U = I$  ve  $V'V = I$ ),  $D_{p \times p}$  ( $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$ ) ise  $X$  matrisinin özdeğer matrisidir.  $X'X$  matrisinin spektral ayrışımı ise  $X'X = VDU'UDV' = VD\underbrace{U'U}_I D V' = VD^2 V' = V\Lambda V'$  olarak elde edilir. Burada  $\lambda_i = d_i^2$   $i = 1, \dots, p$ ,  $X'X$  matrisinin özdeğerlerini,  $V = [V_1, \dots, V_p]$  bu özdeğerlere karşılık gelen ortonormal özvektörleri gösterir.

Tekil değer ayrışımı kullanılarak EKK tahmini

$$\begin{aligned}
X\hat{\beta} &= X(X'X)^{-1} X'y \\
&= UU'y \\
&= \sum_{j=1}^p u_j u_j' y
\end{aligned} \tag{4.3}$$

şeklinde yazılır.

$u_j$  değerleri  $U$  'nun kolonları olmak üzere ridge çözümleri

$$\begin{aligned}
X\hat{\beta}(k) &= X(X'X + kI)^{-1} X'y \\
&= UD(D^2 + kI)^{-1} DU'y \\
&= \sum_{j=1}^p u_j \frac{d_j^2}{d_j^2 + k} u_j' y \\
&= \sum_{j=1}^p u_j \frac{\lambda_j}{\lambda_j + k} u_j' y
\end{aligned} \tag{4.4}$$

olarak yazılabilir.  $k \geq 0$  olduğu için  $\frac{d_j^2}{d_j^2 + k} \leq 1$  olur. (4.3) ve (4.4) eşitlikleri

karşılaştırıldığında EKK regresyonda olduğu gibi RR'de de tahmin değerinin

koordinatları  $U$  matrisine göre hesaplanır. Sonra RR tarafından verilen  $\frac{d_j^2}{d_j^2 + k} \leq 1$

büzülme faktörü ile söz konusu hesaplanan koordinatlar büzdürülür (Hastie ve ark., 2008).

Ridge tahmin edicisinin beklenen değeri  $E(\hat{\beta}(k)) = E[(Q\hat{\beta})] = Q\beta$  şeklindedir. Yanlılığı

$$\begin{aligned}
Bias(\hat{\beta}(k)) &= E(\hat{\beta}(k)) - \beta \\
&= [(X'X + kI)^{-1} X'X - I]\beta \\
&= (Q - I)\beta = -k(X'X + kI)^{-1}\beta,
\end{aligned}$$

kovaryans matrisi

$$\begin{aligned}
Cov(\hat{\beta}(k)) &= Cov[(X'X + kI)^{-1} X'y] \\
&= (X'X + kI)^{-1} X' Cov(y) X (X'X + kI)^{-1} \\
&= \sigma^2 [I + k(X'X)^{-1}]^{-1} (X'X)^{-1} [I + k(X'X)^{-1}]^{-1} \\
&= \sigma^2 Q (X'X)^{-1} Q'
\end{aligned}$$

ve matris hata kareler ortalaması (MSE)

$$\begin{aligned}
MSE(\hat{\beta}(k)) &= E\{[\hat{\beta}(k) - \beta][\hat{\beta}(k) - \beta]'\} \\
&= Cov(\hat{\beta}(k)) + Bias(\hat{\beta}(k))Bias(\hat{\beta}(k))' \\
&= \sigma^2 Q (X'X)^{-1} Q' + (Q - I)\beta\beta'(Q - I)' \\
&= \sigma^2 Q (X'X)^{-1} Q' + k^2 (X'X + kI)^{-1} \beta\beta'(X'X + kI)^{-1}
\end{aligned}$$

olur. Skaler MSE ise

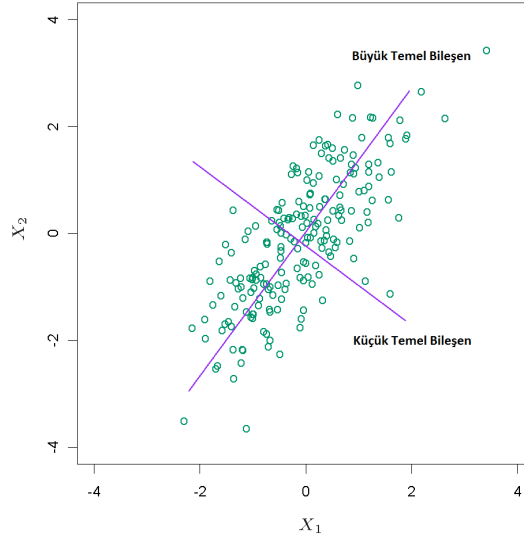
$$\begin{aligned}
sMSE(\hat{\beta}(k)) &= trace[MSE(\hat{\beta}(k))] \\
&= \sum_{i=1}^p \frac{\sigma^2 \lambda_i}{(\lambda_i + k)^2} + k^2 \beta'(X'X + kI)^{-2} \beta
\end{aligned}$$

olarak ifade edilir. İlk terim varyanslar toplamı ve ikinci terim yanlılığın karesini göstermek üzere sırasıyla  $k$ 'nın azalan ve artan fonksiyonlarıdır.

## 4.2. Temel Bileşenler Regresyon

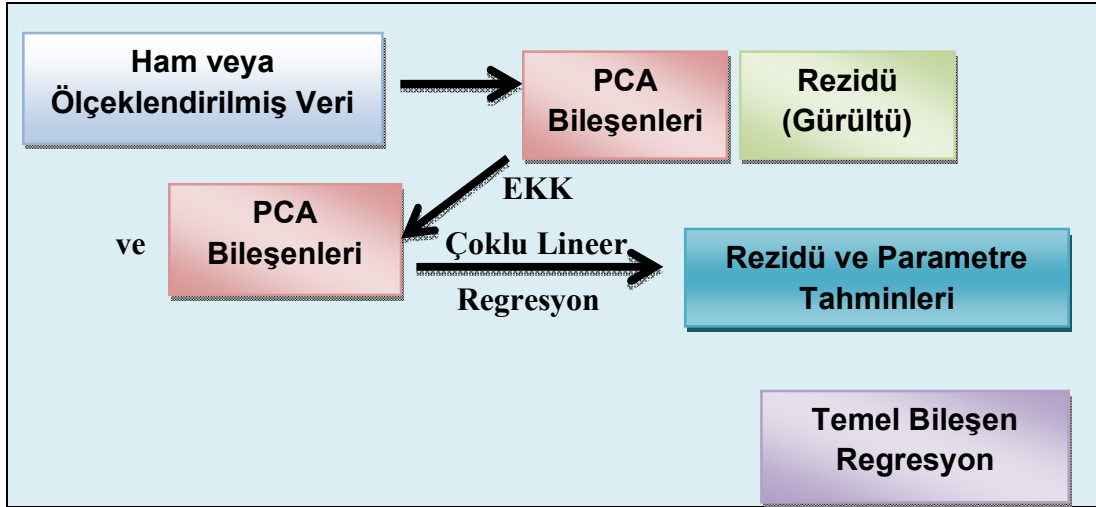
Bazı durumlarda birbirleriyle ilişkili çok sayıda girdi olur. Temel bileşenler analizi (PCA) değişkenlerin varyans-kovaryans yapısını bu değişkenlerin doğrusal bileşimleri yardımıyla açıklayarak boyut indirgemesi ve yorumlamasını sağlayan çok değişkenli istatistiksel bir yöntemdir. 1901 yılında Karl Pearson'nın başlattığı PCA çalışmaları 1933 yılında Hotelling tarafından geliştirilmiştir. PCA eldeki veriyi daha az sayıda değişken ile ifade edilebilecek en iyi dönüşümü belirlemeyi amaçlar. Dönüşüm sonrasında elde edilen değişkenler temel bileşenler (PC) olarak

adlandırılır. İlk temel bileşen varyans değeri en büyük olandır ve diğer temel bileşenler varyans değeri azalacak şekilde sıralanır. Şekil 4.1 iki boyutlu veri için temel bileşenleri örneklendirmektedir. Büyük temel bileşen verinin varyansını maksimum yapan yönde ve küçük temel bileşen geriye kalan kısmın varyansını maksimum yapan yöndedir.



Şekil 4. 1. Temel bileşenler fikri (Hastie ve ark., 2008)

Temel bileşenler analizi ve EKK regresyonun birleşimi olarak ifade edilen temel bileşenler regresyon (PCR) ilk kez Massy (1965) tarafından önerilmiştir. PCR yönteminde amaç varyansın maksimum yapılmasıdır. PCA ve PCR arasındaki ilişki Şekil 4.2’de açıklanmıştır. Şekil 4.2’de olduğu gibi ilk olarak veri için PCA modeli oluşturulur. Modeldeki PCA bileşenleri (skorları) kullanılarak çoklu lineer regresyon elde edilir.



Şekil 4. 2. PCA ve PCR arasındaki ilişki

$W = XV$  ve  $\alpha = V'\beta$  olmak üzere (2.1) modeli

$$\begin{aligned}
 y &= X\beta + \varepsilon \\
 &= XVV'\beta + \varepsilon \\
 &= W\alpha + \varepsilon
 \end{aligned} \tag{4.5}$$

şeklinde kanonik formda yazılabilir. Burada  $p$  tam modeldeki açıklayıcı değişken sayısını göstermek üzere  $W$ ,  $n \times p$  boyutlu ortogonal yeni açıklayıcı değişkenlerin matrisi ve  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_p]'$  yeni katsayıların  $p \times 1$  boyutlu vektörüdür.  $W = [W_1 \quad \dots \quad W_p]$  matrisinin kolonları temel bileşenler olarak adlandırılır:

$$\begin{aligned}
 W'W &= (XV)'(XV) \\
 &= V'X'XV \\
 &= \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p).
 \end{aligned}$$

$W = [W_s, W_r]$  matrisinin parçalanmasıyla (4.5) modeli  $y = W_s\alpha_s + W_r\alpha_r + \varepsilon$  şeklinde yazılır. Burada  $s$  indirgenmiş modeldeki açıklayıcı değişken sayısını ve  $r$  tam modelden atılan açıklayıcı değişken sayısını gösterir. Temel bileşenler tahmin edicisi

$$\begin{aligned}
\hat{\beta}_{PC} &= V_s \hat{\alpha}_s \\
&= V_s (V_s' X' X V_s)^{-1} V_s' X' y \\
&= V_s \Lambda_s^{-1} V_s' X' y
\end{aligned}$$

beklenen değeri

$$E(\hat{\beta}_{PC}) = V_s (V_s' X' X V_s)^{-1} V_s' X' E(y) = V_s \alpha_s = V_s V_s' \beta$$

yanlılığı

$$\begin{aligned}
Bias(\hat{\beta}_{PC}) &= E(\hat{\beta}_{PC}) - \beta \\
&= (V_s V_s' \beta) - \beta \\
&= (V_s V_s' - I_p) \beta \\
&= -V_r V_r' \beta
\end{aligned}$$

ve kovaryans matrisi

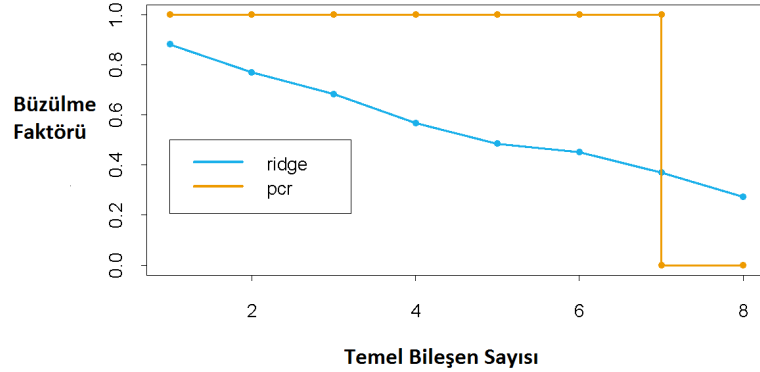
$$\begin{aligned}
Cov(\hat{\beta}_{PC}) &= Cov(V_s \Lambda_s^{-1} V_s' X' y) \\
&= V_s \Lambda_s^{-1} V_s' X' Cov(y) X V_s \Lambda_s^{-1} V_s' \\
&= \sigma^2 V_s \Lambda_s^{-1} V_s' X' X V_s \Lambda_s^{-1} V_s' \\
&= \sigma^2 V_s \Lambda_s^{-1} V_s'
\end{aligned}$$

şeklinde yazılır. Tam model için MSE değeri  $V_s V_s' - I_p = -V_r V_r'$  olmak üzere

$$\begin{aligned}
MSE(\hat{\beta}_{PC}) &= Cov(\hat{\beta}_{PC}) + Bias(\hat{\beta}_{PC}) Bias(\hat{\beta}_{PC})' \\
&= \sigma^2 V_s \Lambda_s^{-1} V_s' + (V_s V_s' - I_p) \beta \beta' (V_s V_s' - I_p) \\
&= \sigma^2 V_s \Lambda_s^{-1} V_s' + V_r V_r' \beta \beta' V_r V_r'
\end{aligned}$$

olarak ifade edilir.

RR ve PCR yöntemlerinin her ikisi de katsayılar üzerinde büzülme uygulaması bakımından birbirleriyle benzerlik gösterir. Fakat Şekil 4.3'ten görüldüğü üzere RR özdeğerlere karşılık gelecek şekilde bir büzülme uygularken PCR küçük belirli sayıdaki özdeğerleri göz ardı ederek büzülme uygular.



Şekil 4. 3. Ridge regresyon ve temel bileşen regresyon yöntemlerinde büzülme (Hastie ve ark., 2008)

Kısmi en küçük kareler regresyonun (PLSR) daha anlaşılır olması ve PCR ile ilişkisinin görülebilmesi için bu bölümde bahsedilen PCR yöntemi doğrusal olmayan yinelemeli en küçük kareler (NIPALS) algoritması kullanılarak aşağıdaki şekilde de açıklanabilir.

Çok değişkenli analizde genellikle değişkenlerin ölçü birimleri birbirinden farklı olmaktadır. Bu nedenle standartlaştırma yapılır. PCA sadece açıklayıcı değişkenlerle ilgilendiğinden  $X$  matrisi PCA kullanılarak

$$\begin{aligned} X &= \hat{T}\hat{P}' + E \\ &= \hat{t}_1\hat{p}'_1 + \hat{t}_2\hat{p}'_2 + \dots + \hat{t}_s\hat{p}'_s + \hat{E} \end{aligned} \quad (4.6)$$

şeklinde ayrıştırılır.  $\hat{T}$ ,  $n \times s$  boyutlu bileşen (skor) matrisi;  $\hat{P}$ ,  $p \times s$  boyutlu yük matrisi ( $X'X$  matrisinin özvektörleri) ve  $\hat{E}$ ,  $n \times p$  boyutlu artık matrisidir (yani  $X$ 'in  $s$  tane temel bileşen tarafından açıklanamayan kısmını gösterir). Bu modelde  $\hat{P}$ ,  $\hat{P}' = (\hat{T}'\hat{T})^{-1}\hat{T}'X$  şeklinde  $X$ 'in  $\hat{T}$ 'ya EKK uyumu olarak yazılabilir. Benzer



olarak  $\hat{T}$ ,  $\hat{P}'\hat{P}=I$  olduğundan  $\hat{T}=X\hat{P}(\hat{P}'\hat{P})^{-1}=X\hat{P}$  şeklinde  $X$ 'in  $\hat{P}$ 'ya EKK uyumu olarak bulunabilir.

PCA'da kullanılan algoritmalarından biri NIPALS algoritmasıdır. NIPALS algoritmasına göre her bir bileşen için önceki  $a-1$  bileşenin tahmininden sonra elde edilen  $\hat{E}$  artık matrisinden yararlanılarak  $\hat{p}_a$  yük ve  $\hat{t}_a$  bileşen vektörleri elde edilir. PCA algoritması için  $\hat{E}$  yerine kullanılan  $X_{a-1}$ ,  $X_{a-1}=X-1\bar{x}'-\hat{t}_1\hat{p}_1-\dots-\hat{t}_{a-1}\hat{p}'_{a-1}$  şeklinde gösterilir. Burada "1",  $n \times 1$  boyutlu 1'lerden oluşan vektörü göstermektedir. PCA'ya ilişkin NIPALS algoritmasının adımları aşağıdaki gibidir (Polat, 2009):

**Adım 1:**  $X$ 'den bir  $x_j$  vektörü alınır ve  $t_a$  olarak adlandırılır.

**Adım 2:**  $p'_a$  hesaplanır:  $p'_a = t'_a X / t'_a t_a$

**Adım 3:**  $p'_a$  birim uzunluğa sahip olacak şekilde normalleştirilir:

$$p'_{ayeni} = p'_{aeski} / \|p'_{aeski}\|$$

**Adım 4:**  $t_a$  hesaplanır:  $t_a = X p_a / p'_a p_a$

**Adım 5:** 2. adımda kullanılan  $t_a$ , 4. adımda elde edilen  $t_a$  ile karşılaştırılır. Bu iki  $t_a$  değeri yakınsadığında durulur. Farklı ise 2.adıma gidilir.

Burada  $t_a$ ,  $n \times 1$  boyutlu  $X$  matrisi için bileşenlerin bir kolon vektörü ve  $p'_a$ ,  $1 \times p$  boyutlu  $X$  matrisi için yüklerin bir satır vektörüdür.  $a$  ( $a=1, \dots, s$ ) ise bileşenlerin sayısını ifade eden yapay bir göstergedir. İlk bileşen elde edildikten sonra 2. ve 4. adımdaki  $X$  yerine artıkları kullanarak daha sonraki bileşenler elde edilir (Naes ve Martens, 1989; Naes ve ark., 2002).

Algoritma sonucunda PCR katsayı tahmini  $\hat{\beta}_{PC} = P_s(T'_s T_s)^{-1} T'_s y$  olarak bulunur (Wise ve Ricker, 1990). Burada  $P_s$ ,  $V_s$ 'ye ve  $T_s$ ,  $X V_s$ 'ye denk olarak elde edilir.

### 4.3. Kısmi En Küçük Kareler Regresyon

Kısmi En Küçük Kareler Regresyon (PLSR) ilk olarak sosyal bilimlerden ekonomi alanında Herman Wold tarafından 1966 yılında önerilmiştir. Ancak Herman'ın oğlu Svante ve ark. (1987) bu yöntemi kemometri (hesaplamalı kimya) alanında yaptıkları çalışmalarla popüler hale getirmiştir. Höskuldsson (1988) ise tüm çalışmaları göz önünde bulundurarak PLSR yönteminin istatistiksel çalışmalarda da kullanılabilceğini göstermiştir.

PLSR yöntemi ile fazla sayıda olan ve aralarında çoklu iç ilişki bulunan açıklayıcı değişkenler, yanıt ve açıklayıcı değişkendeki değişimi büyük ölçüde açıklayan ve aralarında çoklu iç ilişki sorunu olmayan daha az sayıda yeni değişkenlere indirgenir. Bu değişkenler gizli değişken (LV) olarak adlandırılır. Söz konusu indirgeme işlemi için NIPALS algoritması yaygın olarak kullanılır.

PLSR modeli oluşturulmadan önce bileşenler, yükler ve ağırlıklar bulunur.

PLSR'de ağırlık vektörü (4.6) denkleminde  $E_a = X - \sum_{j=1}^a t_j p_j'$  ve buradan da  $E_0 = X$  olmak üzere  $t_a = E_{a-1} w_a = X w_a$  bulunur.  $t_a$  elde edildikten sonra  $t_a = X r_a$  eşitliğinden  $r_a$  hesaplanır. Burada  $a = 1, \dots, h$  ( $h < p$ ) indisi indirgenmiş PLS modelindeki gizli değişken sayısını göstermek üzere

$$T_h = X R_h \quad (4.7)$$

$$P_h = X' T_h (T_h' T_h)^{-1} \quad (4.8)$$

olarak ifade edilir. (4.7) eşitliğinde kullanılan  $R_h$  matrisi ise

$$R_h = W_h (P_h' W_h)^{-1} \quad (4.9)$$

şeklindedir. Burada  $R_h$  ve  $W_h$  ağırlıkları gösterir. PLSR ile elde edilen model

$$\hat{Y}_{PLSR} = T_h (T_h' T_h)^{-1} T_h' Y \quad (4.10)$$

gösterimine sahiptir. (4.10) eşitliğinde  $T_h$  yerine  $XR_h$  ve  $Y$  yerine  $X\hat{\beta}$  yazıldığında

$$\hat{Y}_{PLSR} = XR_h (R_h' X' XR_h)^{-1} R_h' X' X \hat{\beta} \quad (4.11)$$

elde edilir. (4.11) denklemi (2.1) modeline göre düşünüldüğünde PLSR tahmin edicisinin

$$\hat{\beta}_{PLSR} = R_h (R_h' X' XR_h)^{-1} R_h' X' X \hat{\beta} \quad (4.12)$$

ifadesine eşit olduğu görülür. Eşitlik (4.7)'nin (4.8)'de yerine koyulmasıyla elde edilen  $P_h = X'XR_h (R_h' X' XR_h)^{-1}$  ve (4.9) eşitlikleri kullanılarak

$$\begin{aligned} \hat{\beta}_{PLSR} &= R_h \underbrace{(R_h' X' XR_h)^{-1} R_h' X' X}_{P_h'} \hat{\beta} \\ &= R_h P_h' \hat{\beta} \\ &= W_h (P_h' W_h)^{-1} P_h' \hat{\beta} \end{aligned}$$

eşitlik (4.12)'nin daha sade bir biçimine ulaşılır (Phatak ve De Jong, 1997; Yeniay ve Göktaş, 2002).

PLSR'nin beklenen değeri,

$$\begin{aligned} E(\hat{\beta}_{PLSR}) &= E(R_h (R_h' X' XR_h)^{-1} R_h' X' y) \\ &= E(R_h \underbrace{(R_h' X' XR_h)^{-1} R_h' X' X}_{P_h'} \hat{\beta}) \\ &= R_h P_h' \beta \end{aligned}$$

yanlılığı

$$\begin{aligned}
Bias(\hat{\beta}_{PLSR}) &= E(\hat{\beta}_{PLSR}) - \beta \\
&= R_h P_h' \beta - \beta \\
&= (R_h P_h' - I) \beta,
\end{aligned}$$

kovaryans matrisi

$$\begin{aligned}
Cov(\hat{\beta}_{PLSR}) &= Cov(R_h (R_h' X' X R_h)^{-1} R_h' X' Y) \\
&= R_h (R_h' X' X R_h)^{-1} R_h' X' \underbrace{Cov(Y)}_{\sigma^2} X R_h (R_h' X' X R_h)^{-1} R_h' \\
&= \sigma^2 R_h \underbrace{(R_h' X' X R_h)^{-1} R_h' X' X R_h (R_h' X' X R_h)^{-1}}_I R_h' \\
&= \sigma^2 R_h (R_h' X' X R_h)^{-1} R_h'
\end{aligned}$$

ve matris MSE ise

$$\begin{aligned}
MSE(\hat{\beta}_{PLSR}) &= Cov(\hat{\beta}_{PLSR}) + Bias(\hat{\beta}_{PLSR}) Bias(\hat{\beta}_{PLSR})' \\
&= \sigma^2 R_h (R_h' X' X R_h)^{-1} R_h' + (R_h P_h' - I) \beta \beta' (R_h P_h' - I)'
\end{aligned}$$

olarak bulunur.

PLSR yönteminde tahmin yaparken PCR yönteminde olduğu gibi  $X$  matrisinden elde edilen bilgiden faydalanarak, yani  $X$ 'in varyansını maksimum yapacak şekilde seçilen temel bileşenler yerine hem  $X$  hem de  $Y$  matrisindeki bilgileri yansıtan ve daha çok  $Y$ 'deki değişkenlik ile bağlantılı olan gizli değişkenler dikkate alınır. PLSR yönteminde de her bir gizli değişken  $Y$  ve  $X$  arasındaki kovaryansı maksimum yapacak şekilde seçilir.

PLSR yönteminde (4.6) eşitliğindeki  $X$  değişkenine ek olarak  $Y$  değişkeni

$$Y = \hat{U} \hat{Q}' + \hat{F} \quad (4.13)$$

şeklinde ayrıştırılır.  $\hat{U}$ ,  $n \times h$  boyutlu bileşen (skor) matrisi  $\hat{Q}$ ,  $m \times h$  boyutlu yük matrisi ve  $\hat{F}$ ,  $n \times m$  boyutlu artık matrisidir (yani  $h$  tane gizli değişken tarafından  $Y$  matrisinin açıklanamayan kısmını gösterir).  $u$  ile  $t$  arasındaki ilişki aşağıda belirtilen algoritma dikkate alındığında  $\hat{u}_a = b_a t_a$  şeklindedir.

1984 yılında Wold tarafından geliştirilen ve PLSR'nin temel algoritması olan NIPALS algoritması aşağıdaki adımlar ile gösterilir. NIPALS yinelemeli bir algoritma olup algortmada  $a$  ( $a = 1, 2, \dots, h$ ) gizli değişken sayısını göstermektedir. Algoritmanın ilk adımı ölçeklendirilmiş ve merkezleştirilmiş  $X$  ve  $Y$  matrisleri ile başlar (Bulut ve Alın, 2009).

**Adım 1:** Birden fazla yanıt değişken var ise  $Y$  yanıt değişken matrisinin en yüksek varyansa sahip olan sütunu ya da ilk sütunu, yanıt değişken sayısı tek ise doğrudan o değişken sütunu  $u_a$  vektörü olarak alınır.

**Adım 2:**  $X$ 'in  $Y$ 'nin ilgili bileşeni  $u_a$  üzerine regresyonundan  $X$  ve  $u$  arasındaki kovaryansı en büyük yapan  $w$  ağırlık vektörü  $w_a = X'_a u_a / (u'_a u_a)$  ile elde edilir.

**Adım 3:**  $w_a$  vektörü normuna bölünerek boyu 1 olacak şekilde ölçeklendirilir. Yani  $w_a / \|w_a\|$  şeklinde hesaplanır.

**Adım 4:**  $t_a = X_a w_a$  eşitliği ile  $X$ 'in ilgili bileşeni  $t_a$ ,  $w_a$  ağırlık vektörü ile  $X$ 'in doğrusal bir kombinasyonu olacak şekilde hesaplanır.

**Adım 5:**  $t_a$  bileşeninin  $Y$ 'yi modellemedeki katkısını açıklayan  $c_a$  ağırlık vektörü  $c_a = Y'_a t_a / (t'_a t_a)$  ile  $Y$ 'nin  $t_a$  üzerine regresyonundan elde edilir.

**Adım 6:**  $c_a$  ağırlık vektörü normuna bölünerek boyu 1 olacak şekilde ölçeklendirilir. Yani  $c_a / \|c_a\|$  şeklinde hesaplanır.

**Adım 7:**  $Y$  için ilgili bileşen  $u_{a(yeni)}$ ,  $c_a$  ağırlık vektörü ile  $Y$ 'nin doğrusal bir kombinasyonunu oluşturacak şekilde  $Y_a c_a / (c'_a c_a)$  ile hesaplanır.

**Adım 8:** Adım 2'de kullanılan  $u_a$  değeri ile Adım 7'de kullanılan  $u_{a(yeni)}$  değeri arasında bir yakınsama sağlanıp sağlanmadığına bakılır. Söz konusu yakınsama, iki

vektörün farkının normunun  $10^{-6}$  gibi sifira çok yakın bir değer olması ile tespit edilir. Bu yakınsama sağlandığı zaman sonraki adımlara geçilerek algoritma sonlandırılır, aksi takdirde Adım 7'de elde edilen  $u_{a(\text{yeni})}$  değeri Adım 2'de yerine koyularak algoritmaya devam edilir.

**Adım 9:**  $X$ 'in ilgili bileşenin  $t_a$  üzerine regresyonundan, bileşenin açıklayıcı değişken üzerindeki etkisini ifade eden yük vektörü  $p_a$ ,  $X'_a t_a / (t'_a t_a)$  ile elde edilir.

**Adım 10:**  $Y$ 'nin ilgili bileşeni  $u_a$  üzerine regresyonundan, bileşenin yanıt değişken üzerindeki etkisini ifade eden yük vektörü  $q_a$ ,  $Y'_a u_a / (u'_a u_a)$  ile elde edilir.

**Adım 11:** Hem  $X$  hem de  $Y$  için bileşenler ayrı ayrı hesaplandığından bileşenler arasındaki ilişki zayıf olmaktadır. Bu durumu ortadan kaldırmak için her bir bileşen için  $Y$ 'nin ilgili bileşeni  $u_a$ 'nın  $X$ 'in ilgili bileşeni  $t_a$  üzerine regresyonundan elde edilen  $b_a$  katsayısı  $b_a = u'_a t_a / (t'_a t_a)$  ile hesaplanır.

**Adım 12:** Elde edilen bileşenler ve yükler yanıt ve açıklayıcı değişkeni sırasıyla  $X = TP'$  ve  $Y = BTC'$  şeklinde modellemektedir. Bir sonraki bileşeni elde etmek için kullanılacak olan  $X_{a+1}$  ve  $Y_{a+1}$  artık matrisleri  $X_{a+1} \rightarrow X_a - t_a p'_a$  ve  $Y_{a+1} \rightarrow Y_a - b t_a c'_a$  ile hesaplanmaktadır.

Algoritma istenilen gizli değişken sayısı elde edilince sonlandırılır.

## 5. SÜREKLİ REGRESYON

Açıklayıcı değişkenler arasında çoklu iç ilişki olması durumunda RR, PLSR ve PCR yöntemleri EKK yöntemine alternatif olarak önerilen yanlı regresyon yöntemleridir. Sürekli regresyon (CR) ise Stone ve Brooks (1990) tarafından EKK, PLSR ve PCR yöntemlerini birleştiren bir yöntem olarak tasarlanmıştır. CR’de amaç iyi bir öntahmin sağlayacak  $p \times 1$  boyutlu bir  $\beta$  vektörü bulmaktır.  $\beta$  vektörü,  $y$  vektörünün  $Xc$  ( $c_{p \times 1}$ ) açıklayıcı değişkeni üzerinde basit lineer regresyonu ile belirlenir. Burada  $c$  katsayı vektörü EKK, PLSR ve PCR yöntemleri için sırasıyla

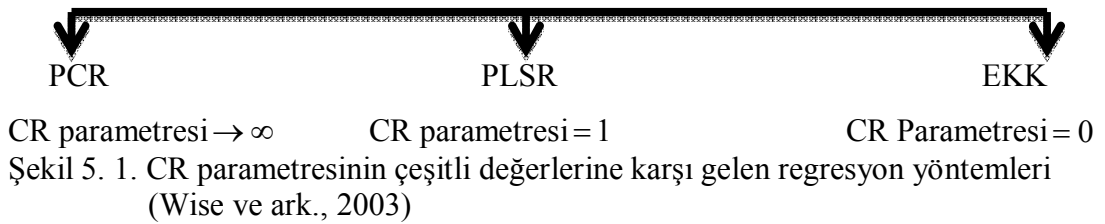
- $y$  ve  $Xc$  arasındaki korelasyonun karesi ( $r^2(y, Xc)$ ) maksimum,
- $\|c\| = 1$  kısıtı altında  $y$  ve  $Xc$  arasındaki kovaryansı maksimum,
- $\|c\| = 1$  kısıtı altında  $Xc$  nin örneklem varyansı (veya  $\|Xc\|^2$ ) maksimum

olacak şekilde seçilir (Björkström ve Sundberg, 1996).

Sürekli regresyon  $\|c\| = 1$  iken  $\gamma \geq 0$  için

$$T(\gamma, c) = (y'Xc)^2 \|Xc\|^{2(\gamma-1)} \propto r^2(y, Xc) \|Xc\|^{2\gamma} \quad (5.1)$$

ifadesini maksimum yapan  $c_\gamma$  ’yı bulmayı amaçlar. Burada CR parametresi sırasıyla  $\gamma = 0$ ,  $\gamma = 1$  ve  $\gamma \rightarrow \infty$  değerlerini aldığıında EKK, PLSR ve PCR tahmin edicilerine karşılık gelir.  $T$  fonksiyonunda tanımlanan  $\gamma$  ’nın özel durumları için EKK, PLSR ve PCR arasındaki ilişki Şekil 5.1’de gösterilmiştir.



Stone ve Brooks (1990) ayrıca  $0 \leq \alpha \leq 1$  sürekli (continuum) aralığında olmak üzere  $\alpha$  ve  $\gamma$  arasında  $\alpha = \gamma/(1+\gamma)$  bağıntısı olduğunu göstererek (5.1) ile verilen ifadeye denk olarak,

$$T = \|y\|^2 r_c^2 S_c^{\alpha/(1-\alpha)} = (c's)^2 (c'Sc)^{\alpha/(1-\alpha)-1} = (c's)^2 (c'Sc)^{\gamma-1} \quad (5.2)$$

eşitliğini önermiştir. Burada  $r_c^2$  ve  $S_c$  sırasıyla  $r_c^2 = \frac{(c's)^2}{\|y\|^2 c'Sc}$  ve  $S_c = c'Sc$  olarak ifade edilir.  $\alpha = 0$  iken EKK,  $\alpha = 1/2$  iken PLSR ve  $\alpha \rightarrow 1$  iken PCR yöntemini ifade eder.

(5.1) veya (5.2)'ye denk olarak Björkström ve Sundberg (1999) ise sürekli regresyonu

$$T_\gamma = R^2(c)V^\gamma(c) \quad (5.3)$$

olarak ifade etmişlerdir. Burada  $R(c)$  korelasyonu,  $V(c)$  ise varyansı göstermektedir.

Genellikle  $y$  yanıt değişkenini tahmin etmek için  $x$  açıklayıcı değişkeni kullanılarak  $\beta'x$  lineer tahmin edicisi oluşturulur. Ancak burada  $\|c\| = 1$  olmak üzere  $x_1, x_2, \dots, x_p$  açıklayıcı değişkenlerinin  $c$  vektörü ile  $c'x = \sum c_j x_j$  lineer kombinasyonu açıklayıcı değişken olarak alınır. Böylece lineer tahmin edici  $\beta'c'x$  şeklinde tanımlanır.

$(\dot{x}(1), \dots, \dot{x}(p), \dot{y})$  verisinden  $(\dot{x}_i, \dot{y}_i)$   $i = 1, \dots, n$  örnekleme seçilmiş olsun ve  $x$  ve  $y$

$$\begin{aligned} y_i &= \dot{y}_i - (\dot{y}_1 + \dots + \dot{y}_n)/n = \dot{y}_i - \bar{\dot{y}} \\ x_i(j) &= \dot{x}_i(j) - \{\dot{x}_1(j) + \dots + \dot{x}_n(j)\}/n \\ x(j) &= \dot{x}(j) - \{\dot{x}_1(j) + \dots + \dot{x}_n(j)\}/n \end{aligned}$$



şeklinde merkezileştirilmiş veriyi gösterebilir.

$t(j) = c'_j \dot{x}$ ,  $j = 1, \dots, g$  verilen açıklayıcı değişkenleri gösterebilir ve ek açıklayıcı değişkenler  $t(g+1) = c'_{g+1} \dot{x}$ ,  $t(g+2) = c'_{g+2} \dot{x}, \dots$  şeklinde oluşturulsun. Verilen açıklayıcı değişkenlere karşılık gelen lineer kombinasyonların hiçbirinin örneklem kovaryansı sıfır değildir. Bu değişkenlerin sonraki açıklayıcı değişkenlerle ilişkisiz olduğu varsayılır. Bu durumda verilen açıklayıcı değişkenlerle ek açıklayıcı değişkenlerin karşılıklı olarak örneklem korelasyonları sıfır olur.  $w$  kabul edilen açıklayıcı değişkenlerin sayısını göstermek üzere EKK yöntemi ile bulunan  $\hat{y}$  yanıt değişkeninin tahmin değeri

$$\begin{pmatrix} t_1(1) & \dots & \dots & t_1(w) & y_1 \\ \vdots & & & \vdots & \vdots \\ \vdots & & & \vdots & \vdots \\ \vdots & & & \vdots & \vdots \\ t_n(1) & \dots & \dots & t_n(w) & y_n \end{pmatrix} = (t(1) \dots t(w) \quad y) = (c'_1 \dot{x} \dots c'_w \dot{x} \quad y)$$

$$\hat{y} = \bar{y} + b_1 t(1) + \dots + b_w t(w) \quad (5.4)$$

olarak gösterilir. Aralarında ilişki olmayan  $t(g+1), \dots, t(w)$   $j = g+1, \dots, w$  açıklayıcı değişkenleri için katsayılar ise

$$\begin{aligned} b_j &= \frac{\sum_{i=1}^n t_i(j) y_i}{\sum_{i=1}^n t_i(j)^2} \\ &= \frac{\sum_{i=1}^n c'_j (x_i - \bar{x}) y_i}{\sum_{i=1}^n [c'_j (x_i - \bar{x})]^2} \end{aligned} \quad (5.5)$$

şeklinde ifade edilir.  $w = g$  olması durumunda modele yeni açıklayıcı değişkenlerin eklenmeyeceği kabul edilir.

$p \times p$  boyutlu  $S = X'X$  matrisi sıfıra eşit olmayan  $m = \min\{n-1, p\}$  özdeğere sahip olsun.  $\|c_{k+1}\| = 1$  ve  $c'_j S c_{k+1} = 0$ ,  $j = 1, \dots, k$  ( $g < k < m-2$ ) kısıtları altında  $T$ 'yi maksimum yapan  $c_{k+1}$  belirlenmek istensin.  $0 < \lambda_1 < \dots < \lambda_m$   $S$ 'nin özdeğerleri ve  $v_1, \dots, v_m$ , ortonormal özvektörler olsun. Stone ve Brooks (1990)  $z_1, \dots, z_m$  skalerler olmak üzere  $c_{k+1} = z_1 v_1 + \dots + z_m v_m$  yazılabileceğini göstermiştir. Dolayısıyla (5.2)'de  $c$  yerine  $c_{k+1}$  yazıldığında

$$T = (c'_{k+1} s)^2 (c'_{k+1} S c_{k+1})^{\gamma-1}$$

olur.  $v'_i S v_i = \lambda_i$  olduğundan ve  $\check{d}_i = s' v_i$  olmak üzere

$$\begin{aligned} c'_{k+1} S c_{k+1} &= (z_1 v_1 + \dots + z_m v_m)' S (z_1 v_1 + \dots + z_m v_m) \\ &= (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2) \end{aligned}$$

$$c'_{k+1} s = z_1 \check{d}_1 + \dots + z_m \check{d}_m$$

tanımlamalarından sonra

$$T = (z_1 \check{d}_1 + \dots + z_m \check{d}_m)^2 (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2)^{\gamma-1} \quad (5.6)$$

bulunur. Ayrıca  $c_{k+1}$  ve  $c'_j S c_{k+1}$  ifadeleri

$$\begin{aligned} \|c_{k+1}\| &= 1 \\ \Rightarrow c'_{k+1} c_{k+1} &= (z_1 v_1 + \dots + z_m v_m)' (z_1 v_1 + \dots + z_m v_m) \\ \Rightarrow z_1^2 \underbrace{v'_1 v_1}_1 + \dots + z_m^2 \underbrace{v'_m v_m}_1 + \sum_i \sum_j z_i z_j \underbrace{v'_i v_j}_0 &= z_1^2 + \dots + z_m^2 = 1 \end{aligned} \quad (5.7)$$

ve  $a_{ij} = \lambda_i c'_j v_i$  olmak üzere ve  $S v_i = \lambda_i v_i$ ,  $i = 1, \dots, m$  özdeğer özvektör ilişkisinden

$$\begin{aligned} c'_j S c_{k+1} &= c'_j S z_1 v_1 + \dots + c'_j S z_m v_m \\ &= a_{1j} z_1 + \dots + a_{mj} z_m = 0 \end{aligned} \quad (5.8)$$

olur.  $\mu_0$  ve  $\mu_j$  lagrange çarpanları olmak üzere, Lagrange çarpanları yöntemi kullanıldığında

$$\begin{aligned} T - \mu_0 (\|c_{k+1}\| - 1) - \mu_j c'_j S c_{k+1} \\ = (z_1 \check{d}_1 + \dots + z_m \check{d}_m)^2 (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2)^{\gamma-1} - \mu_0 (z_1^2 + \dots + z_m^2 - 1) \\ - \mu_j (a_{1j} z_1 + \dots + a_{mj} z_m) \end{aligned} \quad (5.9)$$

yazılabilir. Bu ifadenin  $z_i$ 'ye göre türevi alınsın. Eşitlik (5.6)'nın  $z_i$ 'ye göre türevi alındığında

$$2\check{d}_i (z_1 \check{d}_1 + \dots + z_m \check{d}_m) (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2)^{\gamma-1} + (z_1 \check{d}_1 + \dots + z_m \check{d}_m)^2 (\gamma-1) 2\lambda_i z_i (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2)^{\gamma-2}$$

ve gerekli sadeleştirmeler yapıldığında

$$2 \left[ \underbrace{\check{d}_i (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2)}_{\rho} + \underbrace{(z_1 \check{d}_1 + \dots + z_m \check{d}_m)}_{\tau} (\gamma-1) \lambda_i z_i \right] (z_1 \check{d}_1 + \dots + z_m \check{d}_m) (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2)^{\gamma-2}$$

elde edilir.  $\mu_j (a_{1j} z_1 + \dots + a_{mj} z_m)$  ifadesinin  $z_i$ 'ye göre türevi ise  $\mu_j a_{ij}$ 'dir. Sonuç olarak (5.9)'un  $z_1, \dots, z_m$ 'ye göre türevi

$$\begin{aligned} \left( \frac{\check{d}_i}{\tau} + (\gamma-1) \frac{\lambda_i z_i}{\rho} \right) \rho \tau (z_1 \check{d}_1 + \dots + z_m \check{d}_m) (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2)^{\gamma-2} \\ - \mu_0 z_i - \mu_1 a_{i1} - \dots - \mu_k a_{ik} = 0 \end{aligned} \quad (5.10)$$

olarak bulunur. (5.10)'nun  $z_i$  ile çarpılması ve  $i$  üzerinden toplanması ve (5.7) ve (5.8) eşitliklerinin de dikkate alınması ile  $\mu_0$  bulunur:

$$\sum_{i=1}^m \underbrace{\left( \frac{d_i z_i}{\tau} + (\gamma - 1) \frac{\lambda_i z_i^2}{\rho} \right)}_{\frac{\tau}{\tau + (\gamma - 1) \frac{\rho}{\rho}} = \gamma} \rho \tau (z_1 \check{d}_1 + \dots + z_m \check{d}_m) (\lambda_1 z_1^2 + \dots + \lambda_m z_m^2)^{\gamma - 2} \quad (5.11)$$

$$- \mu_0 \underbrace{\sum_{i=1}^m z_i^2}_1 - \mu_1 \underbrace{\sum_{i=1}^m z_i a_{i1}}_0 - \dots - \mu_k \underbrace{\sum_{i=1}^m z_i a_{ik}}_0 = 0$$

(5.11)'den  $\gamma \tau^2 \rho^{\gamma - 1} = \mu_0$  ve (5.6)'dan  $\gamma T = \mu_0$  bulunur.  $\mu_0$ 'ın (5.10)'da yazılmasıyla

$$\left( \frac{d_i}{\tau} + (\gamma - 1) \frac{\lambda_i z_i}{\rho} \right) T - \gamma T z_i - \mu_1 a_{i1} - \dots - \mu_k a_{ik} = 0 \quad (5.12)$$

elde edilir. (5.8) ve (5.12)'nin birlikte çözülmesiyle çözüme ulaşılır.  $A = (a_{ij})$ ,  $\tilde{D} = \text{diag}(\gamma \rho T + (1 - \gamma) T \lambda_1 \quad \gamma \rho T + (1 - \gamma) T \lambda_2 \quad \dots \quad \gamma \rho T + (1 - \gamma) T \lambda_m)$  matrisleri ve  $z = (z_1 \quad \dots \quad z_m)'$ ,  $d = (\check{d}_1 \quad \dots \quad \check{d}_m)'$ ,  $\varpi = (\rho \mu_1 \quad \rho \mu_2 \quad \dots \quad \rho \mu_m)'$  vektörlerinin tanımlanması ile

$$\tilde{D}z + A\varpi - T \frac{\rho}{\tau} d = \begin{pmatrix} (\gamma \rho T + (1 - \gamma) T \lambda_1) z_1 + a_{11} \rho \mu_1 + a_{12} \rho \mu_2 + \dots + a_{1k} \rho \mu_k - T \frac{\rho}{\tau} \check{d}_1 \\ (\gamma \rho T + (1 - \gamma) T \lambda_2) z_2 + a_{21} \rho \mu_1 + a_{22} \rho \mu_2 + \dots + a_{2k} \rho \mu_k - T \frac{\rho}{\tau} \check{d}_2 \\ \vdots \\ (\gamma \rho T + (1 - \gamma) T \lambda_m) z_m + a_{m1} \rho \mu_1 + a_{m2} \rho \mu_2 + \dots + a_{mk} \rho \mu_k - T \frac{\rho}{\tau} \check{d}_m \end{pmatrix}$$

$$A'z = \begin{pmatrix} a_{11} z_1 + a_{21} z_2 + \dots + a_{m1} z_m \\ a_{12} z_1 + a_{22} z_2 + \dots + a_{m2} z_m \\ \vdots \\ a_{1k} z_1 + a_{2k} z_2 + \dots + a_{mk} z_m \end{pmatrix}$$

eşitlikleri elde edilir. Böylece (5.8) ve (5.12) eşitlikleri

$$\begin{pmatrix} \tilde{D} & A \\ A' & 0 \end{pmatrix} \begin{pmatrix} z \\ \varpi \end{pmatrix} = T \frac{\rho}{\tau} \begin{pmatrix} d \\ 0 \end{pmatrix} \quad (5.13)$$

denklem sistemi ile ifade edilebilir.  $\tilde{D}$  köşegen matris olduğundan  $\begin{pmatrix} \tilde{D} & A \\ A' & 0 \end{pmatrix}$

matrisinin tersi vardır ve (5.13) denklem sisteminin çözümü

$$\begin{aligned} \begin{pmatrix} z \\ \varpi \end{pmatrix} &= T \frac{\rho}{\tau} \begin{pmatrix} \tilde{D} & A \\ A' & 0 \end{pmatrix}^{-1} \begin{pmatrix} d \\ 0 \end{pmatrix} \\ &= T \frac{\rho}{\tau} \begin{pmatrix} \tilde{D}^{-1} - \tilde{D}^{-1} A (A' \tilde{D}^{-1} A)^{-1} A' \tilde{D}^{-1} & \tilde{D}^{-1} A (A' \tilde{D}^{-1} A)^{-1} \\ (A' \tilde{D}^{-1} A)^{-1} A' \tilde{D}^{-1} & - (A' \tilde{D}^{-1} A)^{-1} \end{pmatrix} \begin{pmatrix} d \\ 0 \end{pmatrix} \\ &= T \frac{\rho}{\tau} \begin{pmatrix} (\tilde{D}^{-1} - \tilde{D}^{-1} A (A' \tilde{D}^{-1} A)^{-1} A' \tilde{D}^{-1}) d \\ (A' \tilde{D}^{-1} A)^{-1} A' \tilde{D}^{-1} d \end{pmatrix} \end{aligned}$$

olarak bulunur. Bu durumda  $z_i$   $i = 1, \dots, m$  skalerleri

$M = \tilde{D}^{-1} - \tilde{D}^{-1} A (A' \tilde{D}^{-1} A)^{-1} A' \tilde{D}^{-1}$  olmak üzere  $z = T \frac{\rho}{\tau} M d$  vektörünün elemanları

ile birim normal uzunluklu olacak şekilde  $z = \frac{M d}{\|M d\|}$  ile belirlenir. Dolayısıyla

$z_1, \dots, z_m$  ve buradan da  $c_{k+1}$  belirlenir (Stone ve Brooks, 1990).

CR algoritması ilk kez Wise ve Ricker (1993) tarafından önerilmiştir. Wise ve ark. (2006) PLSR ve PCR yöntemlerinde olduğu gibi CR yöntemi için bir algoritma geliştirmiştir:

**1.adım:** İlk olarak (4.2) eşitliğinde elde edilen tekil değer ayrışımı ele alınır. (4.2) ile verilen  $V$  matrisi PCA'da tanımlanan yük matrisi  $P$ 'ye eşit alınır.

**2.adım:** PCA bileşenleri (skorları)  $T$  ile tekil değer ayrışımında kullanılan model parametreleri  $U$  ve  $D$  arasındaki ilişki  $T = UD$  şeklinde tanımlanır.

**3.adım:** (4.6) denklemi ile gösterilen PCA modeli  $X = TP' + E = UDP' + E$  şeklinde yazılır.

**4.adım:** Tekil değerlerin  $\gamma$  kuvvetini alarak  $D^\gamma$  matrisi oluşturulur. Yeni  $X$  matrisi ise  $X^\gamma$  olarak gösterilir. Buna göre tekil değer ayrışımı  $X^\gamma = UD^\gamma P'$  olarak ifade edilir.

**5.adım:** PLSR'de tanımlanan katsayı vektörü CR yönteminde  $b^{CR}(\gamma) = R_h(R_h'(X^\gamma)'X^\gamma R_h)^{-1}R_h'(X^\gamma)'X^\gamma \hat{\beta}$  olarak tanımlanır (McKay ve ark., 2000) ve  $\hat{\beta}_{scl} = b^{CR}(\gamma)PD_{scl}P'$  şeklinde ölçeklendirilir. Burada  $D_{scl} = D^\gamma ./ D$  olmak üzere  $./$  sembolü tekil değer matrislerinin sıfır olmayan elemanlarının terim terim bölünmesini sağlar.

CR algoritmasında  $\gamma$  kuvveti 0, 1 veya  $\infty$  değerlerini aldığıında model EKK, PLSR veya PCR olur. Adımlarda yakınsama olduğunda iterasyon sonlandırılır.

### 5. 1. Sürekli Regresyon ile En Küçük Kareler Yöntemi Arasındaki İlişki

$p \times p$  boyutlu  $S = X'X$  matrisi sıfıra eşit olmayan  $m = \min\{n-1, p\}$  özdeğerine sahip olduğundan daha önce de bahsedilmişti.  $s$  matrisi ise  $s = X'y$  şeklinde ifade edilsin. Buna göre

$$c_1 = \frac{S^{-s}}{\|S^{-s}\|} = \frac{(X'X)^{-1}X'y}{\|(X'X)^{-1}X'y\|} = \frac{\hat{\beta}}{\|\hat{\beta}\|} \quad (5.12)$$

ifadesi (5.1) eşitliğini maksimum yapar. Bu durumda  $t(1) = c_1'x = \hat{\beta}'x / \|\hat{\beta}\|$  olur ve (5.5) ile verilen katsayılar

$$\begin{aligned}
b_1 &= \frac{\sum_{i=1}^n t_i(1)y_i}{\sum_{i=1}^n (t_i(1))^2} = \frac{\sum_{i=1}^n (\hat{\beta}'x_i/\|\hat{\beta}\|)y_i}{\sum_{i=1}^n (\hat{\beta}'x_i/\|\hat{\beta}\|)^2} = \frac{\frac{1}{\|\hat{\beta}\|} \sum_{i=1}^n \hat{\beta}'x_i y_i}{\frac{1}{\|\hat{\beta}\|^2} \sum_{i=1}^n (\hat{\beta}'x_i)^2} = \|\hat{\beta}\| \frac{\sum_{i=1}^n \hat{\beta}'x_i y_i}{\sum_{i=1}^n (\hat{\beta}'x_i)^2} \\
&= \|\hat{\beta}\| \frac{\hat{\beta}' \sum_{i=1}^n x_i y_i}{(X\hat{\beta})'(X\hat{\beta})} = \|\hat{\beta}\| \frac{(\hat{\beta}')s}{\hat{\beta}' \underbrace{X'X}_S \hat{\beta}} = \|\hat{\beta}\| \frac{(S^-s)'s}{(S^-s)'S(S^-s)} \\
&= \|\hat{\beta}\| \frac{s'S^-s}{s'S^-SS^-s} = \|\hat{\beta}\| \frac{s'S^-s}{s'S^-s} = \|\hat{\beta}\|
\end{aligned} \tag{5.13}$$

olarak bulunur.

(5.4) ve (5.5) denklemleri birlikte ele alındığında

$$\begin{aligned}
\hat{y} &= \bar{y} + b_1 t(1) \\
&= \bar{y} + \|\hat{\beta}\| c_1' x \\
&= \bar{y} + \|\hat{\beta}\| \frac{\hat{\beta}' x}{\|\hat{\beta}\|} \\
&= \bar{y} + \hat{\beta}' x
\end{aligned} \tag{5.14}$$

elde edilir. Bu durumda,  $S$  tekil olmadığında,  $\hat{y} = \beta_0 1 + X\beta + \varepsilon$  modelinde  $\beta$ 'nin EKK tahmin edicisi  $\hat{\beta}$  olur.  $S$ 'nin tekil olup olmadığına bakmaksızın (5.14) denkleminin  $x$  açıklayıcı değişkenleri kullanılmasıyla elde edilen  $\hat{y}$ 'nin EKK tahmini olduğu söylenebilir.

(5.12) ile gösterilen  $c_1$ ,  $y$  ve  $c'x$  arasındaki örneklem korelasyon katsayısının karesi

$$r_c^2 = \frac{\left( \sum_{i=1}^n y_i c_1' x_i \right)^2}{\left( \sum_{i=1}^n y_i^2 \right) \sum_{i=1}^n (c_1' x_i)^2} = \frac{\left( c' \sum_{i=1}^n x_i y_i \right)^2}{\|y\|^2 (c'X)'(c'X)} = \frac{(c'Xy)^2}{\|y\|^2 c' \underbrace{X'X}_S c} = \frac{(c's)^2}{\|y\|^2 c'Sc}$$

maksimum olacak şekilde seçilir (Stone ve Brooks, 1990). Yani

$$c_{EKK} = \max_{c'c=1} \text{corr}^2(y, c'x)$$

olarak ifade edilir (Frank ve Friedman, 1993).

## 5.2. Sürekli Regresyon ile Temel Bileşenler Regresyon Arasındaki İlişki

Stone ve Brooks (1990),  $c_1, c_2, \dots$  vektörlerinin gerdiği uzay  $\langle c \rangle$  olmak üzere,  $\dot{x}_1, \dots, \dot{x}_n$  açıklayıcı değişkenlerinin  $\langle c \rangle$  üzerindeki dik izdüşümlerinin kareleri toplamını

$$S_c = \sum_{i=1}^n (c'x_i)^2 = (Xc)'(Xc) = c' \underbrace{X'X}_S c = c'Sc$$

maksimum yapacak şekilde belirlenir. Burada  $S_c$ 'yi maksimum yapan  $c_1$ ,  $S$  matrisinin en büyük özdeğerine karşılık gelen özvektörü ifade eder.  $c_2$  ise  $S$ 'nin bir sonraki özdeğerine karşılık gelen normalleştirilmiş özvektör olmak üzere  $c_1$ 'de olduğu gibi temel bileşenler analizinde  $t[c]t(1) = c'Sc_1 = 0$  veya  $c'c_1 = 0$  koşulunu sağlayan  $c$  vektörlerinden seçilir.  $c_j, s$ 'nin sayısı  $S_c$  sifıra eşit olacak şekilde  $S$ 'nin rankı özdeğer sayısı olarak ifade edilen  $m = \min\{n-1, p\}$  değerine ulaşıncaya kadar  $c$  vektörleri bulunmaya devam edilir (Wold, 1978).

PCR yönteminin maksimum olmasını sağlayacak açıklayıcı değişken  $k \neq l$  olmak üzere

$$c_{PCR} = \max_{\substack{\{c'V_{c_i=0}\}_1^{n-1} \\ c'c=1}} \text{var}(c'x) \quad (5.15)$$

olarak bulunur (Frank ve Friedman, 1993).



### 5.3. Sürekli Regresyon ile Kısmi En Küçük Kareler Regresyon Arasındaki İlişki

Frank (1987) PLSR yönteminin yanıt değişken ile açıklayıcı değişkenler arasındaki kovaryansı maksimum yapmaya eşdeğer olduğunu belirtmiştir.

$c_1$  değeri,

$$\left(\sum_{i=1}^n y_i c'_1 x_i\right)^2 = \left(c'_1 \sum_{i=1}^n x_i y_i\right)^2 = (c'_1 \underbrace{X'y}_s)^2 = (c'_1 s)^2 \quad (5.16)$$

ile orantılı olarak örnek kovaryansının karesini maksimum yapacak şekilde seçilir ve

$$c_1 = s/\|s\| = X'y/\|X'y\|$$

birinci kanonik kovaryans değişkeni olarak ifade edilir. İkinci kanonik kovaryans değişkeni  $c_2$ ,  $c'_2 S c_1 = 0$  ( $c$ ,  $S$ 'nin  $c_1$ 'e dik olmasıdır) koşulunu sağlayan birim uzunluğa sahip  $c$  için yine (5.16) denklemini maksimum yapacak şekilde seçilir:

$$c_2 \propto s - \left(\frac{s'Ss}{s'S^2s}\right) Ss \quad (5.17)$$

$$\begin{aligned} c'_2 S c_1 &= \left[ s' - \left(\frac{s'Ss}{s'S^2s}\right) s'S \right] S \frac{s}{\|s\|} \\ &= \frac{s'Ss}{\|s\|} - \left(\frac{s'Ss}{s'S^2s}\right) \frac{1}{\|s\|} s'S^2s \\ &= \frac{s'Ss}{\|s\|} - \frac{s'Ss}{\|s\|} \\ &= 0 \end{aligned} \quad (5.18)$$

Bu şekilde  $c_2, c_3, \dots$  kanonik kovaryans değişkenleri elde edilmeye devam edilir.

Ayrıca  $c_1, \dots, c_w$  sıralamasında  $w$  değerinin ne olacağına karar verebilmek için çapraz

geçerlilik kullanılır Stone ve Brooks (1990). Buna göre Wold (1984)  $c_1, \dots, c_w$  değişkenleri ile  $s, Ss, \dots, S^{w-1}s$  ( $1 \leq w \leq m$ ) sıralamasını dikkate alarak

$$\langle c_1, \dots, c_a \rangle = \langle s, Ss, \dots, S^{w-1}s \rangle \quad (5.19)$$

ifadesinin lineer olarak bağımsız olduğunu göstermiş ve (5.19) ile PLSR arasındaki ilişki göz önünde bulundurulduğunda PLSR'nin çeşitli formülasyonları ile karşılaştığını belirtmiştir.

Özet olarak PLSR'nin maksimizasyonu

$$c_{PLSR} = \max_{\substack{\{c'Vc_i=0\}_i^{k-1} \\ c'c=1}} corr^2(y, c'x).var(c'x) \quad (5.20)$$

şeklindedir (Frank ve Friedman, 1993).

#### 5.4. Sürekli Regresyon ile Ridge Regresyon Arasındaki İlişki

Öntahmin performansı dikkate alındığında RR; PLSR ve PCR ile rekabet edebilecek yani iyi sonuç verebilecek bir yöntemdir. RR, ortogonal bileşenlere gerek duymaması ve regresyon vektörüne bir büzülme uygulaması bakımından PLSR ve PCR'den farklıdır, fakat  $k$  ridge sabiti boyunca açık bir sürekliliğe sahip olması nedeni ile birinci faktör PLSR ve PCR ile yakından ilişkilidir. Dolayısıyla Fearn (1983), Stone ve Brooks'un (1990) makalelerinin tartışma kısmında birinci faktör CR (BFCR) ile standart ridge regresyon arasındaki benzerliğe dikkat çekmiş fakat detaylı olarak açıklamamıştır. Daha sonraki çalışmalarda CR ve RR arasındaki ilişki detaylı olarak incelenmiştir.

Sundberg (1993), EKK-PLSR ( $0 < \alpha < \frac{1}{2}$ ) yörüngesi için birinci faktör CR'nin yeniden ölçeklendirilmiş ridge regresyona ( $0 < k < \infty$ ) denk olduğunu göstermiştir. CR ve RR arasındaki ilişkiyi gösteren Sundberg'in (1993) çalışması

pozitif ridge sabiti ile kısıtlıdır ve bu kısıt EKK ile PLSR arasında uzanan yörüngeyi tanımlar (açıklar), PLSR ve PCR arasındaki yörüngeyle ilgilenmez. Buradan hareketle de Jong ve Farebrother (1994) RR için negatif  $k$  değerine izin verilmesi durumunda PLSR-PCR  $(\frac{1}{2} < \alpha < 1)$  yörüngesini RR  $(-\infty < k < -\lambda_{\max})$  ile ilişkilendirmiştir. Böylece CR ile RR arasındaki ilişki CR yöntemi tarafından belirlenen EKK-PLSR-PCR yöntemlerinin tamamına genişletilmiştir. Benzer şekilde Björkström ve Sundberg (1999) de CR ile RR ilişkisini ele almıştır.

Sundberg (1993) ele aldığı aşağıdaki önermeyle CR ve RR arasındaki ilişkiyi ayrıntılı bir şekilde göstermiştir.

**Önerme 5.1 (Sundberg, 1993):** Birinci faktör CR'nin katsayılar vektörü  $b^{CR}(\gamma)$   $(0 \leq \gamma < 1)$  ile standart ridge regresyonun katsayılar vektörü  $\hat{\beta}(k)$   $(k \geq 0)$

$$b^{CR}(\gamma) = \left(1 + \frac{\gamma}{1-\gamma}\right) \hat{\beta}(k) \quad (5.21)$$

bağıntısı ile orantılıdır ve  $\tilde{e}(\gamma) = \frac{b^{CR}(\gamma)' S b^{CR}(\gamma)}{b^{CR}(\gamma)' b^{CR}(\gamma)} = \frac{\hat{\beta}(k)' S \hat{\beta}(k)}{\hat{\beta}(k)' \hat{\beta}(k)}$  ifadesi  $\gamma$  ile birlikte artan  $S$  matrisine ait özdeğerlerin ağırlıklı ortalaması olmak üzere ridge sabiti ve sürekli regresyon parametresi arasında  $k(\gamma) = \tilde{e}(\gamma) \frac{\gamma}{1-\gamma}$  ilişkisi vardır.

**İspat:** (5.2) denkleminin logaritması alınıp  $c'c=1$  kısıtlaması altında Lagrange çarpanı metodu uygulandığında  $2\log(c's) - (1-\gamma)\log(c'Sc) - t(c'c-1)$  elde edilir. Bu ifadenin  $c$ 'ye göre türevi alındığında

$$\frac{s}{c's} - \frac{(1-\gamma)Sc}{c'Sc} - tc = 0 \quad (5.22)$$

ve (5.22) eşitliği soldan  $c'$  ile çarpıldığında,

$$\begin{aligned}\frac{c's}{c's} - (1-\gamma) \frac{c'Sc}{c'Sc} - t \underbrace{c'c}_1 &= 0 \\ \Rightarrow 1 - (1-\gamma) - t &= 0 \\ \Rightarrow 1 - 1 + \gamma - t &= 0 \\ \Rightarrow \gamma &= t\end{aligned}$$

olduğu görülür. RR ve CR arasındaki ilişkiyi gösterebilmek için (5.22) eşitliği  $(c'Sc)$  ile çarpıldığında elde edilen

$$\frac{(c'Sc)s}{c's} - (1-\gamma)Sc - \gamma(c'Sc)c = 0$$

ifadesi  $(1-\gamma)$ 'ya bölüldüğünde

$$\frac{(c'Sc)s}{c's(1-\gamma)} - \frac{(1-\gamma)}{(1-\gamma)}Sc - \frac{\gamma}{1-\gamma}(c'Sc)c = 0.$$

olur. Buradan gerekli düzenlemeler yapıldığında

$$\begin{aligned}\frac{(c'Sc)s}{c's(1-\gamma)} - Sc - \frac{\gamma}{1-\gamma}(c'Sc)c &= 0 \\ \frac{(c'Sc)s}{c's(1-\gamma)} - \left[ S + \underbrace{\frac{\gamma}{1-\gamma}(c'Sc)}_{k = \frac{\gamma}{1-\gamma} \frac{c'Sc}{c'c}} I \right] c &= 0 \\ \frac{(c'Sc)s}{c's(1-\gamma)} - (S + kI)c &= 0\end{aligned}$$

şeklinde yazılır. İşleme devam edildiğinde  $(S + kI)c = \frac{(c'Sc)s}{c's(1-\gamma)}$  eşitliğinden

$$c = \frac{c'Sc}{c's(1-\gamma)} (S + kI)^{-1} s \quad (5.23)$$

elde edilir. Burada  $c's$ ,  $c'Sc$  ve  $(1-\gamma)$  skalerlerini dikkate almadan  $c$  için çözüm yapıldığında

$$c \propto (S+kI)^{-1}s = \hat{\beta}(k) \quad (5.24)$$

elde edilir öyle ki ridge sabiti

$$k = \frac{\gamma}{1-\gamma} \frac{c'Sc}{c'c} \quad (5.25)$$

şeklindedir ve (5.24) dikkate alındığında  $c_i \propto \frac{s_i}{\lambda_i + k}$  veya

$$c_i \propto \frac{\lambda_i^{-1}s_i}{1+k/\lambda_i} \quad (5.26)$$

olarak yazılır. Ayrıca

$$\begin{aligned} \hat{\beta} &= V\hat{\alpha} \\ &= P\Lambda^{-1}P'X'y = P\Lambda^{-1}P's \\ &= P \begin{pmatrix} \lambda_1^{-1} & . & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & \lambda_p^{-1} \end{pmatrix} P' \begin{pmatrix} s_1 \\ . \\ . \\ . \\ s_p \end{pmatrix} = \begin{pmatrix} \lambda_1^{-1}s_1 \\ . \\ . \\ . \\ \lambda_p^{-1}s_p \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ . \\ . \\ . \\ \hat{\beta}_p \end{pmatrix} \end{aligned}$$

kanonik formu kullanıldığında tahmin edici  $\hat{\beta}_i = \lambda_i^{-1}s_i$  olarak yazılır ve (5.26)'daki

ifade  $c_i \propto \frac{\hat{\beta}_i}{1+k/\lambda_i}$  olur.

$$k = \frac{\gamma}{1-\gamma} \frac{c'Sc}{c'c} \text{ ifadesi için kanonik form kullanıldığında}$$

$$c'c = (c_1 \ . \ . \ . \ c_p) \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_p \end{pmatrix} = (c_1^2 + \dots + c_p^2)$$

ve

$$c'\Lambda c = (c_1 \ . \ . \ . \ c_p) \begin{pmatrix} \lambda_1 & . & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & \lambda_p \end{pmatrix} \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_p \end{pmatrix} = \begin{pmatrix} c_1\lambda_1 & . & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & c_p\lambda_p \end{pmatrix} \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_p \end{pmatrix} \\ = (c_1^2\lambda_1 \ . \ . \ . \ c_p^2\lambda_p)$$

eşitliklerinden  $\tilde{e} = \frac{c'Sc}{c'c} = \frac{\sum c_i^2 \lambda_i}{\sum c_i^2}$  yazılır. Dolayısıyla  $k$ ,  $k = \frac{\gamma}{1-\gamma} \tilde{e}$  şeklinde ifade

edilir.  $\gamma$  parametresi sıfıra yakınsadığında (5.1) ile verilen ifade EKK'yı verdiğiinden

$c_i \approx s_i/\lambda_i$  elde edilir, böylece  $\tilde{e} \approx \frac{\sum (s_i/\lambda_i)^2 \lambda_i}{\sum (s_i/\lambda_i)^2}$  olur. Bu durumda (5.23) eşitliği dikkate

alındığında, birinci faktör sürekli regresyon katsayı vektörü  $c_\gamma$  ile RR  $\hat{\beta}(k)$ 'nin orantılı olduğu gösterilmiş olur. Böylece ispat tamamlanmış olur.

$$k = (\gamma/1-\gamma) \frac{c'Sc}{c'c} \text{ olmak üzere sürekli regresyon tahmin edicisi}$$

$$\begin{aligned}
b^{CR}(\gamma) &= (1 + \frac{\gamma}{1-\gamma})(X'X + \gamma/(1-\gamma)(\frac{c'Sc}{c'c}I)^{-1}X'y \\
&= (1 + \frac{\gamma}{1-\gamma})(X'X + \gamma/(1-\gamma)\frac{c'Sc}{c'c}I)^{-1}X'X\hat{\beta} \\
&= (1 + \frac{\gamma}{1-\gamma})[(X'X)^{-1}(X'X + \gamma/(1-\gamma)\frac{c'Sc}{c'c}I)^{-1}]^{-1}\hat{\beta} \\
&= (1 + \frac{\gamma}{1-\gamma})[I + \frac{c'Sc}{c'c}(X'X)^{-1}]^{-1}\hat{\beta} \\
&= (1 + \frac{\gamma}{1-\gamma})Q_{\gamma}\hat{\beta}
\end{aligned}$$

dir. Burada  $Q_{\gamma} = [I + \frac{c'Sc}{c'c}(X'X)^{-1}]^{-1}$  olup yanlılığı

$$\begin{aligned}
Bias(b^{CR}(\gamma)) &= E(b^{CR}(\gamma)) - \beta \\
&= (1 + \frac{\gamma}{1-\gamma})Q_{\gamma}\beta - \beta \\
&= [(1 + \frac{\gamma}{1-\gamma})Q_{\gamma} - I]\beta,
\end{aligned}$$

kovaryans matrisi

$$\begin{aligned}
Cov(b^{CR}(\gamma)) &= Cov[(1 + \frac{\gamma}{1-\gamma})Q_{\gamma}\hat{\beta}] \\
&= (1 + \frac{\gamma}{1-\gamma})^2 \sigma^2 Q_{\gamma}(X'X)^{-1}Q_{\gamma}',
\end{aligned}$$

matris MSE

$$\begin{aligned}
MSE(b^{CR}(\gamma)) &= Cov(b^{CR}(\gamma)) + Bias(b^{CR}(\gamma))Bias(b^{CR}(\gamma))' \\
&= (1 + \frac{\gamma}{1-\gamma})^2 \sigma^2 Q_{\gamma}(X'X)^{-1}Q_{\gamma}' + \left[ (1 + \frac{\gamma}{1-\gamma})Q_{\gamma} - I \right] \beta \beta' \left[ (1 + \frac{\gamma}{1-\gamma})Q_{\gamma} - I \right]'
\end{aligned}$$

ve skaler MSE

$$\begin{aligned}
 sMSE(b^{CR}(\gamma)) &= \text{trace}[MSE(b^{CR}(\gamma))] \\
 &= \text{trace} \left\{ \left(1 + \frac{\gamma}{1-\gamma}\right)^2 \sigma^2 Q_\gamma (X'X)^{-1} Q_\gamma' + \left[ \left(1 + \frac{\gamma}{1-\gamma}\right) Q_\gamma - I \right] \beta \beta' \left[ \left(1 + \frac{\gamma}{1-\gamma}\right) Q_\gamma - I \right]' \right\} \\
 &= \sum_{i=1}^n \left(1 + \frac{\gamma}{1-\gamma}\right) \left(1 + \frac{c'Sc}{c'c} \frac{1}{\lambda_i}\right)^2 \frac{1}{\lambda_i} + \left[ \left(1 + \frac{\gamma}{1-\gamma}\right) Q_\gamma - I \right] \beta \beta' \left[ \left(1 + \frac{\gamma}{1-\gamma}\right) Q_\gamma - I \right]'
 \end{aligned}$$

olarak elde edilir.

Sundberg (1993), CR tahmin edicisini

$$b^{CR}(\gamma) = c \frac{(c's)}{(c'Sc)} \quad (5.27)$$

olarak vermiştir. Fakat (5.26) eşitliği  $c$ 'de ölçek invariant olduğu için  $c = \hat{\beta}(k)$  alınabilir (eşitlik (5.24)). Bu durumda  $c = (S + kI)^{-1} s \Rightarrow s = (S + kI)c$  olduğundan ve  $\frac{c's}{c'Sc} = \frac{c'(S + kI)c}{c'Sc} = \frac{c'Sc}{c'Sc} + k \frac{c'c}{c'Sc} = 1 + \left(\frac{\gamma}{1-\gamma} \frac{c'Sc}{c'c}\right) \frac{c'c}{c'Sc} = 1 + \frac{\gamma}{1-\gamma}$  ifadesi de

dikkate alındığında (5.27) eşitliği  $b^{CR}(\gamma) = \left(1 + \frac{\gamma}{1-\gamma}\right) \hat{\beta}(k)$  olarak yeniden yazılır.

Böylece CR tahmin edici ve RR tahmin edici arasındaki ilişki gösterilmiş olur.

RR tahmin edici ve EKK tahmin edici arasındaki  $\hat{\beta}(k) = (I + kS^{-1})^{-1} \hat{\beta}$  ilişkisi kullanılırsa ve  $k = \frac{\gamma}{1-\gamma} \tilde{e}$  olduğundan  $\hat{\beta}(k) = \left(I + \frac{\gamma}{1-\gamma} \tilde{e}(\gamma) S^{-1}\right)^{-1} \hat{\beta}$  yazılabilir. Dolayısıyla CR ve EKK tahmin edicileri arasında ise

$$\begin{aligned}
 b^{CR}(\gamma) &= \left(1 + \frac{\gamma}{1-\gamma}\right) \left[ I + \frac{\gamma}{1-\gamma} \tilde{e}(\gamma) S^{-1} \right]^{-1} \hat{\beta} \\
 &= [(1-\gamma)I + \gamma \tilde{e}(\gamma) S^{-1}] \hat{\beta} = [I + \gamma(\tilde{e}(\gamma) S^{-1} - I)] \hat{\beta}
 \end{aligned}$$



bağıntısı vardır. Kanonik form kullanıldığında ise  $b_i^{CR}(\gamma) = \frac{\hat{\beta}_i}{1 + \gamma[\tilde{e}(\gamma)/\lambda_i - 1]}$  yazılır.

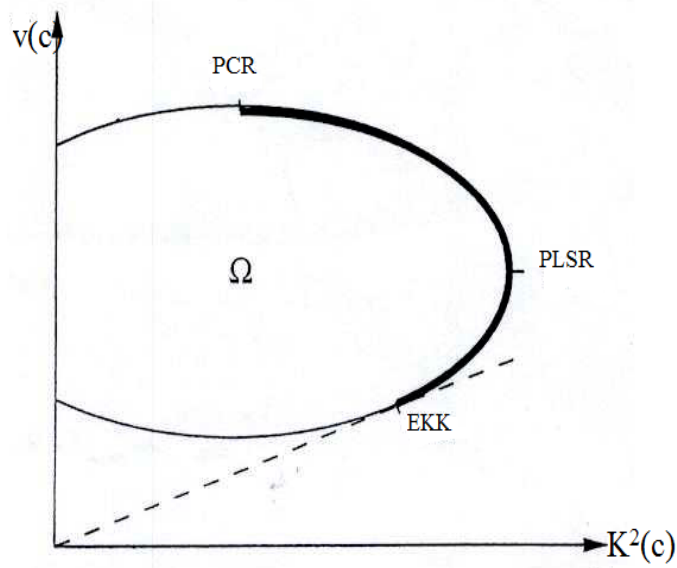
Ridge sabiti  $k$  genellikle pozitif değerler almasına rağmen de Jong ve Farebrother (1994)  $\lambda_{\max}$   $XX'$  matrisinin en büyük özdeğeri olmak üzere, Sundberg'in (1993)  $c_k \propto (XX' + kI)^{-1}X'y$   $\|c_k\|=1$  ile verdiği ilişkinin  $-\infty \leq k \leq -\lambda_{\max}$  aralığında da geçerli olabileceğini göstermiştir. Björkström ve Sundberg (1999), de Jong ve Farebrother'ın (1994) bu genişletmesini dikkate alarak CR ile RR arasındaki ilişkiyi aşağıdaki şekilde açıklamıştır.

Şekil 5.2 ile verilen  $\Omega$  bölgesi  $\|c\|=1$  olduğu için birim küre olarak adlandırılır. Birim küre üzerindeki her bir  $c$  noktası  $(K^2(c), V(c))$  koordinatını, Şekil 5.2 kürenin her yerinde  $V(c) > 0$  olması için  $\lambda_{\min} > 0$  olduğu durumu gösterir.

Şekil 5.2 dikkate alındığında  $k$  sabiti  $[0, \infty]$  ve  $[-\infty, -\lambda_{\max}]$  aralıklarından geçtiği için  $c_k$  birim küre üzerinde ( $\|c_k\|=1$ ) iki sürekli yörünge tanımlar.  $k \rightarrow -\infty$  ve  $k \rightarrow \infty$  limitleri aynı birinci faktör PLSR'ye karşılık geldiğinden, bir ucunda ( $k = 0$  noktasında) EKK olup PLSR aracılığı ile diğer uca yani PCR'ye ( $k \rightarrow -\lambda_{\max}$ ) doğru bir yörünge tanımlanabilir (Şekil 5.2). Aynı zamanda bu yörünge  $\varepsilon > -1$  olmak üzere

$$c_\varepsilon \propto (XX' + \frac{\lambda_{\max}}{\varepsilon} I)^{-1} X'y \quad \|c_\varepsilon\|=1$$

şeklinde de tanımlanabilir. Bu tanımlama ile birlikte  $\varepsilon \rightarrow \infty$  limiti EKK'ya,  $\varepsilon \rightarrow -1$  limiti PCR'ye karşılık gelir. Bu tanımlama PLSR yönteminin  $\infty$ 'dan  $-\infty$ 'a sıçramasını engellemiş olur. Üçüncü bir tanımlama ise iki ayrı yörüngeyi birleştirmek amacıyla  $k' = k/(k + \lambda_{\max})$  parametresi ile yapılır. Bu parametre sürekli regresyon parametresi  $\gamma$ 'nın PLSR, EKK ve PCR için aldığı değerlere karşılık gelmektedir. Yani EKK'de  $k' = 0$ , PLSR'de  $k' = 1$  ve PCR'de  $k' \rightarrow \infty$  değerlerini alır.



Şekil 5. 2. Mümkün olan tüm  $(K^2(c), V(c))$  değerlerinin gösterilmesi (Björkström ve Sundberg, 1999)

De Jong ve Farebrother (1994) CR ve RR yöntemlerini dikkate alarak EKK, PLSR ve PCR yöntemlerinin hangi parametre değerlerine karşılık geldiğini Tablo 5.1 ile özetlemiştir. Tablo 5.1'e göre EKK çözümü  $\gamma = k = 0$  için  $c \propto S^{-1}s$ , PLSR çözümü  $\gamma = 1$  veya  $k \rightarrow \pm\infty$  için  $c \propto s$ , PCR çözümü ise  $\gamma = +\infty$  veya  $k = -\lambda_{\max}$  için burada  $V(c)$  varyansı göstermek üzere  $c \propto V(c)$  olarak tanımlanır.  $(S + kI)$  matrisinin pozitif tanımlı bir matris olması için  $0 \leq \gamma < 1$  olduğunda  $k$  değeri negatif olmalıdır. Buna göre birinci faktör sürekli regresyon (BFCR) tahmin edicisinin standart ridge regresyon tahmin edicisi ile orantılı olduğu söylenebilir. Diğer yandan  $k$  değeri  $(S + kI)$  matrisinin negatif tanımlı bir matris olması için  $\gamma > 1$  olduğunda  $-\lambda_{\max}$ 'dan daha az olmalıdır. Negatif  $k$  değerinin pozitif tanımlı bir matris vermesi için söz konusu matris  $(k^{-1}S + I)$  olarak ifade edilir. Böylece birinci faktör CR tahmin edicisi  $1 < \gamma < +\infty$  aralığı için  $(k^{-1}S + I)^{-1}s$  şeklinde standart olmayan RR tahmin edicisi ile orantılı olur.

Tablo 5.1. EKK, PLSR ve PCR yöntemlerinin sürekli regresyon ve ridge regresyona karşılık gelen parametre değerleri

Yöntem	Kriter	$\alpha$	$\gamma$	$k$	$c$
BF-EKK	Korelasyon	0	0	0	$\infty S^{-1}S$
BF-PLSR	Kovaryans	0.5	1	$\pm \infty$	$\infty S$
BF-PCR	Varyans	1	$\infty$	$-\lambda_{\max}$	$\infty V(c)$

Bu açıklamalardan hareketle, Björkström ve Sundberg (1999) RR ile CR arasındaki ilişkiyi göstermek amacıyla Önerme 5.2'yi ele almıştır.

Önerme 5.2'ye geçmeden önce CR için Stone ve Brooks (1990) tarafından tanımlanan (5.3) eşitliğine göre Björkström ve Sundberg'in (1999)  $\gamma \geq 0$  için  $c_\gamma$  açıklayıcı değişkenini kullanarak yaptığı tanımlama dikkate alınmalıdır. Buna göre ispatı kolaylaştırmak amacıyla korelasyon  $R(c)$  yerine kovaryans  $K^2(c) = (y'Xc)^2$  kullanılacaktır. En büyük kovaryans değeri açıklayıcı değişken olarak alınacaktır. Ayrıca Sundberg (1993), de Jong ve Farebrother (1994) ve Björkström ve Sundberg (1996) çalışmalarına göre burada tanımlanan açıklayıcı değişken ridge açıklayıcı değişkeni ile orantılıdır.

**Önerme 5.2 (Björkström ve Sundberg, 1999):** (5.5) eşitliğine göre tanımlanan  $f(K^2(c), V(c))$  fonksiyonu  $V(c)$  sabit iken  $K^2(c)$ 'de (veya  $R^2(c) \propto K^2(c)/V(c)$ ),  $R^2(c)$  sabit iken  $V(c)$ 'de artan bir fonksiyon olmak üzere

$$c_f = \max_{|c|=1} f(K^2(c), V(c)) \quad (5.28)$$

açıklayıcı değişkeni tanımlanır.  $X'y$  vektörü en büyük özdeğere karşılık gelen özvektörlere dik olmamak üzere  $c_f \propto (X'X + kI)^{-1} X'y$  olur (dik olmama koşulu sağlanmadığı takdirde  $c_f \propto (X'X + kI)^{-1} X'y$  maksimum olmaz). Buna göre  $k \rightarrow 0$  (EKK),  $k \rightarrow \infty$  (PLSR) ve  $k \rightarrow -\lambda_{\max}$  (PCR) limitlerini içeren bir  $k$  değeri vardır.

**İspat:**  $f(K^2(c), V(c))$  fonksiyonunun  $V(c)$  sabit iken  $K^2(c)$ 'de artması için,  $(K^2(c_f), V(c_f))$  noktasının Şekil 5.2'deki  $\Omega$  bölgesine düşmesi gerekir. Bu nedenle " $V(c)$  ve  $|c|^2 = 1$  kısıtları ile  $K^2(c)$ 'nin maksimum yapılması" sayesinde  $c_f$ 'nin mutlaka bir çözümü vardır.

$f(K^2(c), V(c))$  fonksiyonunun  $K^2(c)/V(c)$  sabit iken  $V(c)$ 'de artması için,  $V(c)$ ,  $v_{EKK}$ 'den (EKK açıklayıcı değişkeninin varyansı) küçük  $\lambda_{\max}$ 'dan büyük olamaz.  $V(c) \neq \lambda_{\max}$  ( $V(c) = \lambda_{\max}$  durumu EK A'da gösterilmektedir) ise,  $c = c_f$  noktasındaki Lagrange çarpanları yönteminden  $\mu_1$  ve  $\mu_2$  Lagrange çarpanları olmak üzere

$$K^2(c) - \mu_1(\|c\|^2 - 1) - \mu_2 V(c)$$

elde edilir. Türevin alınıp sıfıra eşitlenmesi ile elde edilen  $2K(c)X'y - 2\mu_1 c - 2\mu_2 X'Xc = 0$  eşitliğinden

$$K(c)X'y = \mu_1 c + \mu_2 X'Xc \quad (5.29)$$

olur.  $K = 0$  durumu  $f(K^2(c), V(c))$  fonksiyonunun minimum değerine karşılık geldiği için dikkate alınmaz ve  $\mu_2 = 0$  durumu birinci faktör PLSR'ye karşılık geldiğinden  $\mu_2 \neq 0$  olduğu durum dikkate alınır. (5.29) eşitliği  $\mu_2$ 'ye bölüldüğünde

$$\frac{K(c)}{\mu_2} X'y = \frac{\mu_1}{\mu_2} c + \frac{\mu_2}{\mu_2} X'Xc$$

bulunur. Sıfır olmayan  $K(c)/\mu_2$  oranı  $\kappa$  değeri ile  $\mu_1/\mu_2$  oranı ise  $k$  değeri olarak gösterilmek üzere

$$(X'X + kI)c = \kappa X'y \quad (5.30)$$

elde edilir.  $(X'X + kI)$  matrisi tekil olmadığından tersi alınarak  $c_f \propto (X'X + kI)^{-1} X'y$  şeklinde yazılır. Böylece RR ile CR arasındaki ilişki gösterilmiş olur.

CR ve RR arasındaki ilişki dikkate alınarak RR'nin maksimizasyonu

$$c_{RR} = \max_{c'=1} r^2(y, c'x) \frac{\text{var}(c'x)}{\text{var}(c'x) + k} \quad (5.31)$$

olur (Frank ve Friedman, 1993).

### 5.5. EKK, PCR, PLSR, RR ve CR Arasındaki İlişki

Çoklu iç ilişki olması durumunda varyans çok büyük değerler alacağından EKK katsayıları tam olarak tanımlanamaz. Böyle bir durumda varyansı küçültmek için büzülme yöntemleri uygulanır. Bu yöntemler gerçek  $\beta$  tahminlerine yakın değerler vermesinin yanısıra yanıt değişken için iyi bir öntahmin verir.

Ridge tahmin edicisi;  $\hat{\beta}(k) = (S + kI)^{-1} s$  şeklindedir. PCR ve PLSR yöntemleri ise değişken seçimi ile açıklayıcı değişken sayısını azaltarak daha az sayıda değişken ile katsayı tahmini yaptığı için RR yöntemi gibi büzülme yöntemleri olarak adlandırılır ve bu yöntemler modelin doğruluk derecesini artırır (Sundberg, 2002).

Özvektörler  $V = \sum_{a=1}^p d_a^2 v_a' v_a$  ve EKK tahmini  $\hat{\beta}$  olmak üzere Frank ve

Friedman'a (1993) göre PCR, PLSR ve RR yöntemlerindeki büzülme yapısı

$$a_*(RR : PCR : PLSR) = \sum_{j=1}^p f_j(RR : PCR : PLSR) \hat{\beta}_j v_j \quad (5.32)$$

şeklinde gösterilir. Ayrıca ortalamann farklı bir gösterimi ile katsayı tahmini

$$\hat{\beta}_j = \text{ort}(y v_j' x) / d_j^2 \quad (5.33)$$

olarak ifade edilir. Burada  $\hat{\beta}_j$   $j$ . özvektör üzerindeki EKK çözümünün izdüşümlerini gösterir.

(5.32) eşitliği RR için

$$f_j(RR) = \frac{d_j^2}{(d_j^2 + k)}, \quad (5.34)$$

PCR için

$$f_j(PCR) = \begin{cases} 1 & d_j^2 \geq d_s^2 \\ 0 & d_j^2 < d_s^2 \end{cases}, \quad (5.35)$$

PLSR için  $h$  gizli değişken sayısı olmak üzere

$$f_{jh}(PLSR) = \sum_{a=1}^h \tilde{\beta}_a d_j^{2a} \quad (5.36)$$

şeklindedir. Burada  $\tilde{\beta}$  herhangi bir vektör ve  $\tilde{w}$  kovaryans vektörü olmak üzere

$\tilde{\beta} = \{\tilde{\beta}_a\}_1^h = \tilde{W}^{-1}\tilde{w}$  'dır. Ayrıca  $w_a = \sum_{j=1}^p \hat{\beta}_j^2 d_j^{2(a+1)}$  ve  $W_{al} = \sum_{j=1}^p \hat{\beta}_j^2 d_j^{2(a+l+1)}$  olarak ifade

edilir. Buna göre PLSR fonksiyonunun gizli değişken sayısına bağlı olduğu söylenebilir.

(5.32) eşitliği RR, PCR ve PLSR için ayrı ayrı yazıldığında RR, PCR ve PLSR yöntemlerindeki büzülme yapılarının  $\{d_j^2\}_1^p$  ve  $\{\hat{\beta}_j^2\}$  değerlerine bağlı olduğu

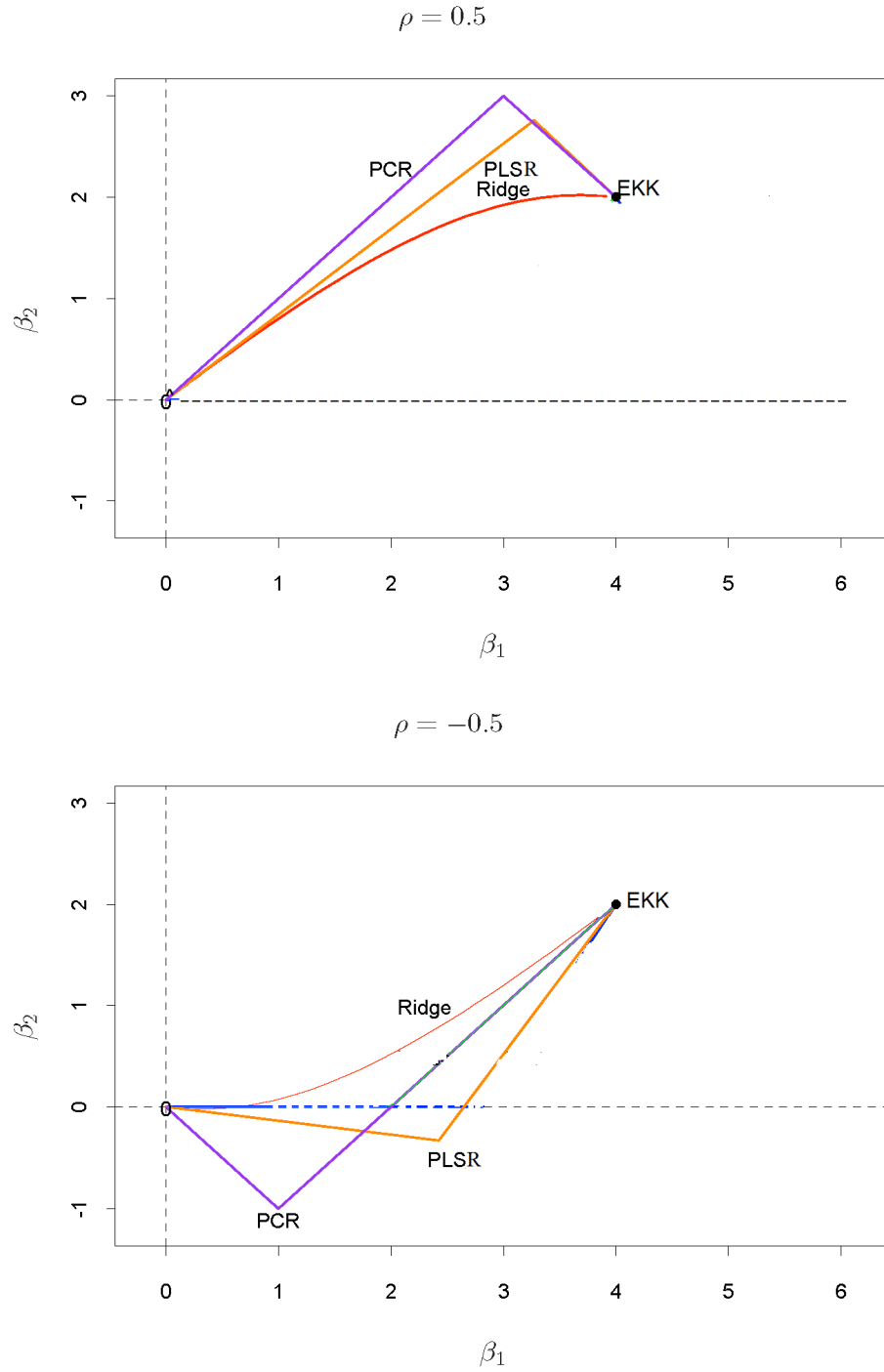
görülmektedir. PLSR;  $\{\hat{\beta}_j\}_1^p$  çözümüne bağlı fakat EKK çözüm uzunluğundan  $|\hat{\beta}|^2$

bağımsızdır. PLSR, RR ve PCR fonksiyonlarında olduğu gibi kolaylıkla tamamlanamamasına rağmen, RR ve PCR fonksiyonları ile kıyaslanabilmesi için

$\{d_j^2\}_1^p$  ve  $\{\hat{\beta}_j\}_1^p$  değerleri kullanılarak tanımlanır.

RR, PCR ve PLSR yöntemleri öntahmin edicinin düşük varyansa sahip olması için bütün çözüm katsayı vektörlerine ( $a_*$ ) ceza uygular. Böylece RR, PLSR ve PCR yöntemlerinin çözüm katsayı vektör uzunlukları EKK çözüm katsayı vektör uzunluğundan daha kısa olur. Cezanın gücü PLSR için gizli değişken sayısı ve PCR için bileşen sayısı arttıkça azalırken, RR için cezanın gücü ridge sabitinin değerinin artmasına bağlı olarak artar.

Hastie ve ark. (2008) PCR, PLSR, RR ve EKK yöntemleri arasındaki ilişkiyi açıklamak amacıyla Şekil 5.3'ü ele almıştır. Burada yanlı regresyon yöntemleri arasındaki ilişki gösterilirken korelasyonu ifade eden  $\rho$  değerlerinin  $\rho = 0.5$  (üstteki şekil) ve  $\rho = -0.5$  (alttaki şekil) ile regresyon katsayılarının  $\beta_1 = 4$  ve  $\beta_2 = 2$  (yani Şekil 5.3'teki EKK noktasının apsis ve ordinat eksenlerindeki değerleri) olduğu varsayılmıştır. Şekil 5.3'de ayarlanmış parametreler ( $\beta_1$  ve  $\beta_2$ ) değiştikçe farklı yöntemlerin katsayı profilleri gösterilmektedir. Ridge katsayıları üzerinde ayarlama yapıldığında RR yöntemi EKK çözümüne doğrudan ulaşırken PCR ve PLSR yöntemlerinde böyle bir durum söz konusu değildir. RR'de herhangi bir katsayıda yapılan ayarlama EKK çözümüne ulaşana kadar katsayıların büzülmesi anlamına gelir. PLSR ve PCR yöntemleri RR'den farklı olarak tanımlanmasına rağmen benzer özellikler gösterir.



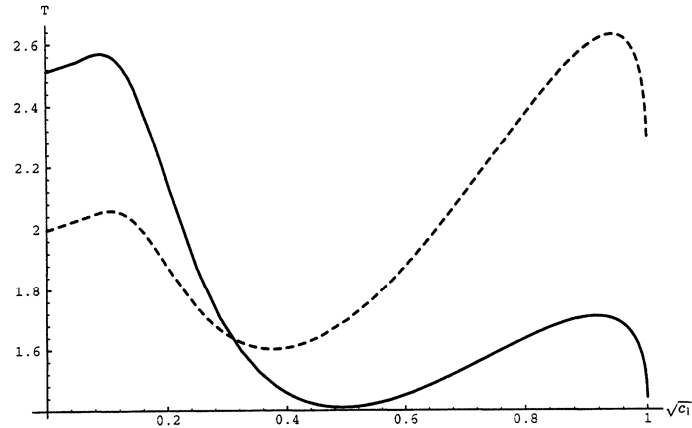
Şekil 5. 3. Gerçek regresyon katsayısı  $\beta = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$  ve aralarındaki korelasyon  $\mp 0.5$  olan iki açıklayıcı değişkenli regresyon için farklı yöntemlerin katsayı profilleri(Hastie ve ark., 2008).



## 6. SÜREKLİ REGRESYONDA SÜREKSİZLİKLER

Björkström ve Sundberg (1996) çalışmaları sırasında CR parametresi  $\gamma$  değıştikçe CR'de süreksizlik meydana gelebileceğini gözlemlemişler ve bir veri kümesinin hangi koşullar altında bu etkiyi yarattığını açıklayabilmek için CR ve RR arasındaki ilişkiyi göz önünde bulundurmuşlardır.

Bölüm 5'te bahsedildiği gibi birinci faktör sürekli regresyon,  $c_\gamma \equiv \max_{|c|=1} \{T(\gamma, c)\}$  fonksiyonunun maksimizasyonunu içermektedir ve  $T$ ,  $c$  ve  $\gamma$ 'nın her ikisinde de sürekli olmasına rağmen  $c_\gamma$ ,  $\gamma$ 'da her zaman sürekli değildir. Björkström ve Sundberg (1996) bu durumu örneklendirmek için  $p=2$  olmak üzere Şekil 6.1'i ele almışlardır.



Şekil 6. 1.  $X'y = (3,1)$  ve  $X'X = \text{diag}(100,0.01)$  olmak üzere  $\gamma = 0.6$  (düz çizgi) ve  $\gamma = 0.7$  (kesikli çizgi) değerleri ile  $c$  değerleri kullanılarak  $T(\gamma, c)$  fonksiyonunun maksimum yapılması (Björkström ve Sundberg, 1996)

Burada  $T(\gamma, c)$ ,  $\gamma = 0.6$  ve  $\gamma = 0.7$  değerleri için  $c$ 'nin bir fonksiyonudur.  $\gamma$  değıştikçe fonksiyona ait yerel maksimum noktaları da değışmektedir. Yani  $T(\gamma, c)$  fonksiyonu  $\gamma = 0.6$  noktasında maksimum iken daha sonra azalmış ve  $\gamma = 0.7$  noktasında maksimum olmuştur ve  $\gamma$ 'nın 0.6 ve 0.7 değerleri arasında bir sıçrama meydana gelmiştir. Buna göre  $T(\gamma, c)$  fonksiyonu  $\gamma$  parametresinin tek bir

değerinde maksimum olmaz. Sonuç olarak  $\gamma$ 'da meydana gelecek hafif bir değişiklik tahmin edicinin ani sıçrama yapmasına neden olur.

Sundberg'in (1993) ele aldığı (5.25) denklemini Björkström ve Sundberg (1996)

$$k = \frac{\gamma}{1-\gamma} \frac{\|Xc_\gamma\|^2}{\|c_\gamma\|^2} \quad (6.1)$$

olarak ifade etmiştir.

$k = k_{CR}(\gamma)$ ,  $\gamma$ 'nın monoton fonksiyonu olmak üzere ridge tahmin edici ile sürekli regresyon tahmin edici arasında  $c_\gamma \propto \hat{\beta}(k)$  ilişkisi vardır. Buna göre tekil değer ayrışımı da kullanılarak

$$XX = VDU'UDV' = VD^2V' = V\Lambda V'$$

$$X'y = (UDV')'y = VDU'y = Vg$$

$$\begin{aligned} c_\gamma &= \hat{\beta}(k) = (XX + kI)^{-1} X'y \\ &= (V\Lambda V' + kI)^{-1} VDU'y \\ &= [V(\Lambda + kI)V']^{-1} VDU'y \\ &= (V')^{-1}(\Lambda + kI)^{-1} V^{-1} VDU'y \\ &= V(\Lambda + kI)^{-1} \underbrace{DU'y}_g \\ &= V(\Lambda + kI)^{-1} g \end{aligned}$$

$$\begin{aligned} c_\gamma' c_\gamma &= [g'(\Lambda + kI)^{-1} V'] [V(\Lambda + kI)^{-1} g] \\ &= \|(\Lambda + kI)^{-1} g\|^2 \end{aligned}$$

$$\begin{aligned}
\|Xc_\gamma\|^2 &= (Xc_\gamma)'(Xc_\gamma) = c_\gamma' X' X c_\gamma = c_\gamma' V \underbrace{D' D}_\Lambda V c_\gamma \\
&= g'(\Lambda + kI)^{-1} \underbrace{V' V}_\Lambda \underbrace{V' V}_\Lambda (\Lambda + kI)^{-1} g \\
&= \left[ g'(\Lambda + kI)^{-1} \Lambda^{1/2} \right] \left[ \Lambda^{1/2} (\Lambda + kI)^{-1} g \right] \\
&= \left\| \Lambda^{1/2} (\Lambda + kI)^{-1} g \right\|^2
\end{aligned}$$

ifadeleri tanımlanır. Bu durumda (6.1) eşitliği,

$$k = \frac{\gamma}{1-\gamma} \frac{\left\| \Lambda^{1/2} (\Lambda + kI)^{-1} g \right\|^2}{\left\| (\Lambda + kI)^{-1} g \right\|^2} \quad (6.2)$$

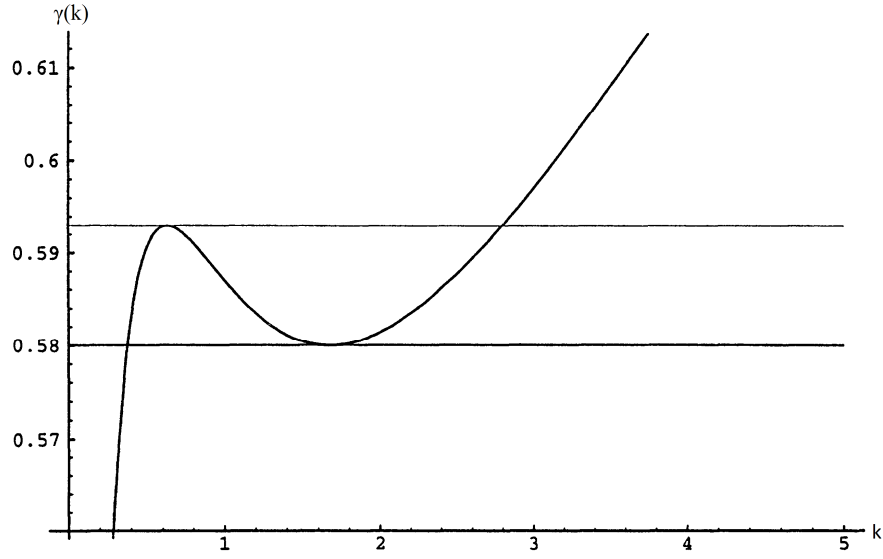
olarak bulunur.  $P(k) = \left\| \Lambda^{1/2} (\Lambda + kI)^{-1} g \right\|^2$  ve  $Q(k) = \left\| (\Lambda + kI)^{-1} g \right\|^2$  polinomları tanımlandığında (6.2) eşitliği

$$k = \frac{\gamma}{1-\gamma} \frac{P(k)}{Q(k)} \quad (6.3)$$

olarak elde edilir. Böylece  $\gamma$  ve  $k$  arasındaki ilişki daha açık bir şekilde görülür. (6.3) eşitliği  $\gamma$  için çözüldüğünde

$$\begin{aligned}
&\Rightarrow kQ(k)(1-\gamma) = \gamma P(k) \\
&\Rightarrow kQ(k) = \gamma[kQ(k) + P(k)] \\
&\Rightarrow \gamma = \gamma(k) = \frac{kQ(k)}{kQ(k) + P(k)}
\end{aligned} \quad (6.4)$$

bulunur. Söz konusu  $\gamma$  değeri için (6.4) eşitliği  $k = k_{CR}(\gamma)$  ifadesini sağlar. Yani  $\gamma(k)$ ,  $k$ 'nın monoton fonksiyonu olduğundan (6.4) eşitliğinin tersi alındığında  $k_{CR}(\gamma)$  fonksiyonu elde edilir. Eşitlik (6.4)'e göre verilen bir  $\gamma$  değeri için  $k$ 'nın birden fazla çözümü vardır. Bu durum Şekil 6.2'de örneklendirilmiştir.



Şekil 6. 2. Ordinatı  $\gamma(k)$  ve apsisi  $k$  olmak üzere (6.2) eşitliğinin gösterimi (Björkström ve Sundberg, 1996).

Şekil 6.2'ye göre  $0.580 \leq \gamma \leq 0.593$  aralığı için  $\gamma = \gamma(k)$  ifadesinin üç farklı çözümü olacağından üç farklı  $k$  değeri vardır. Birden fazla çözümün olması CR açıklayıcı değişkenini belirsiz yapmaz, ancak birim küre üzerindeki ( $\|c\|=1$ )  $T(\gamma, c)$  fonksiyonunun yerel maksimum noktası sonlu sayıdaki adaylar arasında yer alır. Yani Şekil 6.2'de görüldüğü gibi  $k_{CR}(\gamma)$  fonksiyonunda süreksizlik meydana gelir. Diğer taraftan Eşitlik (6.4)'ün her yerde artması  $k_{CR}(\gamma)$  fonksiyonunu sürekli yapar.

Björkström ve Sundberg'in (1996) verdiği Şekil 6.2  $\gamma = 0$  (EKK) ve  $\gamma = 1$  (PLSR) aralığını grafiklendirmiştir. Fakat  $\gamma > 1$  olduğu durumlar için de grafiklendirmenin yapılabileceği ve  $\gamma > 1$  olduğunda  $k$  değerinin  $k \in (-\infty, -\lambda_1)$  aralığında değerler aldığı belirtilmiştir.

Bir veri kümesinde CR tahmin edicisinin sıçramalardan etkilenmemesi için

gerek ve yeter koşul  $\frac{kQ(k)}{kQ(k) + P(k)}$  fonksiyonun monoton veya denk olarak bütün  $k$

değerleri için  $\frac{kQ(k)}{P(k)}$  fonksiyonu artan olacak şekilde  $\lambda_1, \dots, \lambda_p$  özdeğerleri ve

$g_1, \dots, g_p$  katsayılarının var olmasıdır. Bu durumu göstermek için  $\frac{kQ(k)}{P(k)}$  fonksiyonunun türevi

$$\begin{aligned} \frac{\partial(\frac{kQ(k)}{P(k)})}{\partial k} &\Rightarrow \frac{[kQ(k)]'P(k) - [kQ(k)]P'(k)}{P^2(k)} = 0 \\ &\Rightarrow \frac{(Q + kQ')P - kQP'}{P^2} = 0 \\ &\Rightarrow \frac{QP + kPQ' - kQP'}{P^2} = 0 \end{aligned}$$

dikkate alındığında  $\gamma(k)$  fonksiyonunun monoton olması için gerek ve yeter koşul  $QP + kPQ' - kQP' = 0$  eşitliğinin  $\{-\infty, -\lambda_1\} \cup \{0, \infty\}$  ( $\lambda_1 = \lambda_{\max}$  yani  $X'X$  matrisinin en büyük özdeğer) bölgesinde çözümünün olmamasıdır. Ayrıca  $\{-\infty, -\lambda_1\} \cup \{0, \infty\}$  bölgesi  $k$  için  $0 \leq k < \infty$  ve de Jong ve Farebrother'ın (1994) genişlettiği  $-\infty \leq k \leq -\lambda_{\max}$  aralıklarının birleşimini göstermektedir. Bu durum veri kümesinde süreksizlik olup olmayacağı konusunda bir çözüm yolu sunmaktadır.

$QP + kPQ' - kQP' = 0$  eşitliği yeniden düzenlendiğinde:

$$\begin{aligned} P(k) &= \|\Lambda^{1/2}(\Lambda + kI)^{-1} \mathbf{g}\|^2 = \mathbf{g}'(\Lambda + kI)^{-1} \Lambda (\Lambda + kI)^{-1} \mathbf{g} \\ &= (\mathbf{g}_1 \dots \mathbf{g}_p) \begin{pmatrix} (\lambda_1 + k)^{-1} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & (\lambda_p + k)^{-1} \end{pmatrix} \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \lambda_p \end{pmatrix} \begin{pmatrix} (\lambda_1 + k)^{-1} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & (\lambda_p + k)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_p \end{pmatrix} \\ &= (\mathbf{g}_1 \dots \mathbf{g}_p) \begin{pmatrix} \lambda_1(\lambda_1 + k)^{-2} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \lambda_p(\lambda_p + k)^{-2} \end{pmatrix} \begin{pmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_p \end{pmatrix} \\ &= \mathbf{g}_1^2 \lambda_1 (\lambda_1 + k)^{-2} + \dots + \mathbf{g}_p^2 \lambda_p (\lambda_p + k)^{-2} \\ &= \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-2} \end{aligned}$$

$$\begin{aligned}
Q(k) &= \left\| (\Lambda + kI)^{-1} \mathbf{g} \right\|^2 = \mathbf{g}' (\Lambda + kI)^{-1} (\Lambda + kI)^{-1} \mathbf{g} \\
&= (\mathbf{g}_1 \ \dots \ \mathbf{g}_p) \begin{pmatrix} (\lambda_1 + k)^{-1} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & (\lambda_p + k)^{-1} \end{pmatrix} \begin{pmatrix} (\lambda_1 + k)^{-1} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & (\lambda_p + k)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_p \end{pmatrix} \\
&= (\mathbf{g}_1 \ \dots \ \mathbf{g}_p) \begin{pmatrix} (\lambda_1 + k)^{-2} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & (\lambda_p + k)^{-2} \end{pmatrix} \begin{pmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_p \end{pmatrix} \\
&= \mathbf{g}_1^2 (\lambda_1 + k)^{-2} + \dots + \mathbf{g}_p^2 (\lambda_p + k)^{-2} \\
&= \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-2}
\end{aligned}$$

olmak üzere

$$\begin{aligned}
QP + kPQ - kQ^2 &= \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-2} \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-2} \\
&+ k \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-2} \underbrace{\frac{d \left[ \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-2} \right]}{dk}}_{-2 \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-3}} - k \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-2} \underbrace{\frac{d \left[ \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-2} \right]}{dk}}_{-2 \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-3}} \\
&= \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-2} \underbrace{\left[ \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-2} + 2k \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-3} - 2k \lambda_j \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-3} \right]}_{\substack{= \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-2} [1 + 2k(\lambda_j + k)^{-1}] \\ = \sum_{j=1}^p \mathbf{g}_j^2 \lambda_j (\lambda_j + k)^{-3} (\lambda_j + 3k)}} \\
&= \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-2} \sum_{j=1}^p \mathbf{g}_j^2 (\lambda_j + k)^{-3} \underbrace{[\lambda_j (\lambda_j + 3k) - 2k \lambda_j]}_{\lambda_j^2 + k \lambda_j}
\end{aligned}$$

işlemleri yapılarak  $\lambda_j^2 + k \lambda_j = 0$  iken  $k = -\lambda_j$  olarak bulunur. Bu durumda çözüm bölgesi  $\lambda_1 > \lambda_2 > \dots > \lambda_p \geq 0$  sıralamasına göre  $-\lambda_1 < -\lambda_2 < \dots < -\lambda_p \leq 0$  şeklinde tanımlanır.

➤  $k = -\lambda_1$  için

$$\frac{P(k)}{Q(k)} = \frac{\mathbf{g}_1^2 \lambda_1 (\lambda_1 + k)^{-2} + \dots + \mathbf{g}_p^2 \lambda_p (\lambda_p + k)^{-2}}{\mathbf{g}_1^2 (\lambda_1 + k)^{-2} + \dots + \mathbf{g}_p^2 (\lambda_p + k)^{-2}}$$

ifadesinin pay ve paydası  $(\lambda_1 + k)^2 (\lambda_2 + k)^2 \dots (\lambda_p + k)^2$  ile çarpılırsa

$$\frac{P(k)}{Q(k)} = \frac{g_1^2 \lambda_1 (\lambda_2 + k)^2 \dots (\lambda_p + k)^2 + \dots + g_p^2 \lambda_p (\lambda_1 + k)^2 \dots (\lambda_{p-1} + k)^2}{g_1^2 (\lambda_2 + k)^2 \dots (\lambda_p + k)^2 + \dots + g_p^2 (\lambda_1 + k)^2 \dots (\lambda_{p-1} + k)^2}$$

elde edilir.  $k = -\lambda_1$  olarak alındığında

$$\frac{P(k)}{Q(k)} = \frac{g_1^2 \lambda_1 (\lambda_2 - \lambda_1)^2 (\lambda_3 - \lambda_1)^2 \dots (\lambda_p - \lambda_1)^2 + \underbrace{g_2^2 \lambda_2 (\lambda_1 - \lambda_1)^2}_{=0} + \dots + 0}{g_1^2 (\lambda_2 - \lambda_1)^2 (\lambda_3 - \lambda_1)^2 \dots (\lambda_p - \lambda_1)^2 + \underbrace{g_2^2 (\lambda_1 - \lambda_1)^2}_{=0} + \dots + 0} = \lambda_1$$

bulunur.  $\lim_{\gamma \rightarrow \infty} \frac{\gamma}{1-\gamma} = -1$  olduğundan  $\lim_{\gamma \rightarrow \infty} \frac{\gamma}{1-\gamma} \frac{P(k)}{Q(k)} = -\lambda_1$  olur. Dolayısıyla,

$k_{CR}(\gamma \rightarrow \infty) = -\lambda_1$  ile PCR elde edilir.

➤  $k = \mp \infty$  için

$\lim_{k \rightarrow \infty} \frac{P(k)}{Q(k)} = \lim_{k \rightarrow \infty} \frac{P'(k)}{Q'(k)} = 1$  olduğundan  $\lim_{k \rightarrow \infty} \frac{\gamma}{1-\gamma} \frac{P(k)}{Q(k)} = \frac{\gamma}{1-\gamma}$  yazılabilir.

$k = \frac{\gamma}{1-\gamma} \frac{P(k)}{Q(k)}$  ifadesinde  $k$ 'nin sonsuza gidebilmesi için  $1-\gamma$  değeri sıfıra eşit yani

$\gamma = 1$  olması gerekir.  $k_{CR}(\gamma = 1) \rightarrow \pm \infty$  ile PLSR elde edilir.

➤  $k = 0$  için

$k = \frac{\gamma}{1-\gamma} \frac{P(k)}{Q(k)}$  eşitliğinden  $\gamma = 0$  bulunur.  $k_{CR}(\gamma = 0) = 0$  ile EKK elde edilir.





## 7. MODEL SEÇME KRİTERLERİ

Model seçme kriterleri ve çapraz geçerlilik regresyon modelinin tahminleme gücünü göstermektedir. İyi bir modelin kurulabilmesi tamamen doğru değişkenlerin seçimine dayanmaktadır. Böylece model tahmin hatası minimuma inecek ve model istenmeyen değişkenlerden korunmuş olacaktır (Bulut ve Alma, 2011).

*RMSE* model seçme kriteri ayarlama, öntahmin ve çapraz geçerlilik amaçlarına göre Tablo 7.1'deki gibi özetlenmiştir.

Tablo 7. 1. RMSE ve MSE ölçülerinin amaçlarına göre sınıflandırılması (Jorgensen ve Goegebeur, 2007)

Amaç	RMSE	MSE
Ayarlama (Calibration)	RMSEC	MSEC
Ön Tahmin	RMSEP	MSEP
Çapraz Geçerlilik	RMSECV	MSECV

### 7. 1. Hata Kareler Ortalamasının Karekökü

Bir regresyon modeli için

$$RMSE = \sqrt{MS_{Res}(\hat{y})} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (7.1)$$

olarak tanımlanan hata kareler ortalamasının karekökü model belirleme kriteri olarak kullanılır (Naes ve ark., 2002).  $\hat{y}_i$  yanıt değişkenin tahmin değerlerini gösterir.  $n$  ise gözlem sayısıdır. Küçük *RMSE* değerine sahip modeller veriye daha iyi uyum sağlayan modellerdir.

## 7.2. Ayarlama Kümesine Dayalı Geçerlilik

Ayarlama hata kareler ortalamasının karekökü (Root Mean Square Error of calibration)

$$RMSEC = \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2 / (n - a - 1)} \quad (7.2)$$

olarak ifade edilir. Burada  $y_i$  ayarlama kümesinin ölçeklendirilmiş yanıt değişken değerini,  $\hat{y}_i$  yanıt değişkenin tahmin değerini gösterir.  $n$  ayarlama kümesindeki gözlem sayısı,  $a$  ise bileşen sayısını göstermektedir (Naes ve ark., 2002). Açıklayıcı değişken sayısı fazla olduğunda RMSEC değeri büyük değerler alır. RMSEC değeri ayarlama hatası veya ayarlama varyansını gösterdiğinden modelin ayarlama verisine uyumunun ne kadar iyi olduğunu gösterir. Küçük RMSEC değeri modelin iyi olduğunu gösterir.

## 7.3. Öntahmin Testi

Öntahmin hata kareler ortalamasının karekökü (Root Mean Square Error of Prediction)

$$RMSEP = \sqrt{\sum_{i=1}^{n_p} (\hat{y}_i - y_i)^2 / n_p} \quad (7.3)$$

şeklinde ifade edilir (Naes ve ark., 2002). RMSEP değeri öntahmin hatası veya öntahmin varyansını gösterdiğinden öntahmin modelinin niteliğini ifade eder. RMSEP değeri küçük olan öntahmin modeli tercih edilir.  $y_i$  öntahmin veri kümesinin ölçeklendirilmiş değerini,  $\hat{y}_i$  öntahmin veri kümesindeki öntahmin

değerini verir.  $n_p$  öntahmin kümesindeki veri sayısını gösterir. RMSEP değeri regresyon yöntemlerinin öntahmin performansını karşılaştırmak için kullanılır.

#### 7.4. Çapraz Geçerlilik

Çapraz geçerlilik modelleri öngörü yeteneğine göre seçmek için kullanılan bir yöntemdir (Shao, 1993). İlk kez Stone (1974) tarafından önerilmiştir.

Regresyon modelinde yer alacak değişken sayısına karar vermede ve doğrusal regresyon modelinin tahmin geçerliliğini incelemede genellikle çapraz geçerlilik kullanılır (Bulut ve Alma, 2011). Çapraz geçerlilik her defasında bir gözlem dışarıda bırakarak model seçer ve geriye kalan veri noktaları ile model oluşturarak dışarıda kalan gözlemlerde ortalama öntahmin hatasını en aza indirir. (2.2) eşitliğindeki kayıp fonksiyonun kullanılması ile

$$C_{\alpha} = \frac{1}{n} \sum_{i=1}^n L(y_i, \hat{y}_{-i}) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_{-i} - y_i)^2$$

olarak ifade edilir.  $\hat{y}_{-i}$ ,  $i$  – inci gözlem atılarak bulunan  $y$  ’nin tahminini gösterir. Ayrıca burada  $\min C_{\alpha}$  olması şartı aranır.

Grafik çiziminde ise çapraz geçerlilik indeksi

$$I_{\alpha} = 1 - C_{\alpha} / C_{-,g}$$

olarak ifade edilir (Stone ve Brooks, 1990).  $C_{-,g}$  ise çizdirilecek grafiğe göre çeşitli gözlemlerin atılmasıyla hesaplanır. Çapraz geçerlilik indeksi 1 değerini geçemez ve negatif değerler alabilir. Büyük değerler alması ise tahmin yeteneğinin iyi olduğunu gösterir.

Doğrusal regresyon modellerinin yeni gözlemleri öntahmin yeteneğinin bir ölçüsü olarak kullanılan,  $C_{\alpha}$  PCR modelindeki PC’lerin sayısı veya PLS

modelindeki gizli değişkenlerin sayısına göre elde edilen PRESS değeri ile bağlantılıdır.

Bir modelin ön tahmin performansını ölçmek için bir diğer kriter de PRESS istatistiğini minimum yapmaktır ve PRESS istatistiğini hesaplarken birini-dışarıda-bırak (LOOCVE/leave-one-out) çapraz geçerlilik yöntemi kullanılır. Tahminlenen ve gözlemlenen değerler arasındaki farkın kareler toplamı o model için PRESS değerini vermektedir. PRESS kriteri bir model için tahminlenen değerlerin açıklayıcı değişkenin gözlemlenen yanıt değişkeni nasıl tahminlediğini gösteren bir kriterdir. Bu prosedür Allen (1974) tarafından önerilmiş ve Golub, Heath ve Wahba (1979) tarafından geliştirilmiştir. PRESS istatistiğinin küçük değerler alması, küçük değerli toplam hata kareler ortalamasına sahip modeli temsil eder yani regresyon modeli yeni gözlemlerde iyi öntahminler sağlar.

Allen (1971, 1974) tarafından önerilen PRESS istatistiği  $\hat{e}_{(i)} = y_i - \hat{y}_{(i)}$ ,  $i = 1, 2, \dots, n$  ile tanımlanan PRESS rezidüleri bağlıdır. Burada  $\hat{y}_{(i)}$ ,  $i$ . gözlem atılarak diğer  $n - 1$  gözleme uydurulan regresyon modeli ile elde edilen  $y_i$  nin öntahminidir.

$y_i - \hat{y}_i$  EKK rezidü ve  $h_{ii}$ ,  $H = X(X'X)^{-1}X'$  şapka matrisinin  $i$ . köşegen elemanı olmak üzere PRESS rezidü EKK rezidü kullanılarak  $\hat{e}_{(i)} = \frac{y_i - \hat{y}_i}{1 - h_{ii}}$  den de hesaplanabilir.

Böylece PRESS istatistik  $n$  PRESS rezidünün kareleri toplamı

$$PRESS = \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2 = \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{1 - x_i'(X'X)^{-1}x_i} \right)^2$$

olarak bulunur. Burada  $x_i'$ ,  $X$  matrisinin  $i$ . satır vektörüdür. PRESS  $n - 1$  gözlemin her alt kümesini tahmin kümesi olarak kullanır ve diğer gözlemi de ön tahmin kümesini oluşturmak için kullanır. PRESS istatistiği yeni gözlemlerin iyi bir

öntahmini olan modeli belirler. En iyi modelin seçimi için küçük değerli PRESS tercih edilir.

PRESS istatistiği ridge tahmin ediciye

$$PRESS(k) = \sum_{i=1}^n \left[ \frac{\hat{e}_{Ri}(k)}{1 - h_{k,ii}} \right]^2 = \sum_{i=1}^n \left( \frac{y_i - x_i' \hat{\beta}(k)}{1 - x_i'(X'X + kI)^{-1} x_i} \right)^2$$

olarak uyarlanmış olup  $\hat{e}_{Ri}(k)$  belli bir  $k$  değeri için  $i$ . rezidü ve  $h_{k,ii}$   $H_k = X(X'X + kI)^{-1} X'$  nin  $i$ . köşegen elemanıdır. PRESS modelin ön tahmin performansının bir ölçüsü olduğundan  $PRESS(k)$ 'yi minimum yapacak  $k$  değeri seçilir. Fakat  $PRESS(k)$ ,  $k$ 'nin bir konveks fonksiyonu olmadığından  $k$ 'yi seçmek zordur.

PRESS istatistiği kullanılarak çapraz geçerliliğin hata kareler ortalamasının karekökü (Root Mean Square Error of Cross-Validation)  $RMSECV$

$$RMSECV = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_{-i} - y_i)^2} = \sqrt{\frac{PRESS}{n}}$$

olarak hesaplanır. Çapraz geçerlilik gözlemlerin gerçek kitesinden çekilen  $n-1$  gözleme dayalı ayarlama modellerinin ortalama öntahmin hatasının bir tahminini verir (Naes ve ark., 2002). Aynı öntahmin hatasına sahip iki modelden az sayıda bileşen içeren model seçilir. Çünkü az sayıda bileşen içeren model az sayıda parametreye sahip olacağından daha kararlı yapıya sahip olur.



## 8. UYGULAMA

Bu uygulamada daha önce çeşitli yazarlar tarafından kullanılan kimyasal bir veri üzerinde çalışılmıştır. Fearn (1983) ridge regresyon ile ilgili bir çalışmasında öğütülmüş buğdaya (ground wheat) ilişkin protein içeriğinin ölçüm değerleri üzerinde ayarlama yapmak için yakın kızılötesi yansıma aracını (NIR) kullanarak bulunan deney sonuçları ile bir veri kümesi elde etmiştir. Burada altı farklı değışkene ait değerler dalga boyu 1680-2310 arasında değışen buğday örneğindeki NIR radyasyon yansıma ölçüleridir. Bu yansıma ölçüleri,  $R$  yansımayı ifade etmek üzere,  $\log(1/R)$  şeklinde ölçeklendirilerek  $L_1 - L_6$  (log değerlerini gösterir) olarak adlandırılır. Buna göre Fearn'nın (1983) verisinde ayarlama (calibration) deneyi sonuçları kullanılarak 24 birime sahip bir ayarlama elde edilir kümesi ve ayarlamayı test etmek amacıyla 26 birime sahip başka bir öntahmin kümesi bulunur. Yani protein içeriğinin öntahminini yapan log değerlerinin doğrusal bir birleşimini bulmak amacıyla bir ayarlama modeli oluşturulur ve söz konusu ayarlama modeli kullanılarak öntahmin yapılır. Böylece bilinmeyen gözlemlerin protein içeriği tahmin edilebilir. Ayrıca bu tez çalışmasında  $X_a$  ve  $y_a$  gösterimleri ayarlama;  $X_p$  ve  $y_p$  gösterimleri de öntahmin deney sonuçlarını göstermek için kullanılır (Veriler EK B'de verilmiştir).

Bu çalışmada EKK, PCR, PLSR, RR ve CR yöntemlerini inceleyebilmek için "Eigenvector Research" (1997,1998) ve Wise'a (1992,1993,1995) ait komutlar Matlab programında uygulanmıştır. Ayrıca Matlab programı kullanılarak bulunan sonuçlar ile diğer yazarların PLS\_Toolbox kullanarak buldukları sonuçlar karşılaştırıldığında aralarında sayısal farklılık görölse de istatistiksel yorumları benzerlik göstermektedir.

Analizlere başlamadan önce değışkenler standartlaştırılır. Bu nedenle modelde sabit terim yer almaz. Protein (%) ve  $L_1 - L_6$  değerlerinin mevcut olduğu ayarlama kümesi ile çoklu lineer modele ilişkin korelasyon matrisi,

$$X'_a X_a = \begin{bmatrix} 1.0000 & 0.9939 & 0.9958 & 0.9946 & 0.9371 & 0.9887 \\ 0.9939 & 1.0000 & 0.9993 & 0.9801 & 0.9246 & 0.9876 \\ 0.9958 & 0.9993 & 1.0000 & 0.9843 & 0.9337 & 0.9886 \\ 0.9946 & 0.9801 & 0.9843 & 1.0000 & 0.9545 & 0.9890 \\ 0.9371 & 0.9246 & 0.9337 & 0.9545 & 1.0000 & 0.9493 \\ 0.9887 & 0.9876 & 0.9886 & 0.9890 & 0.9493 & 1.0000 \end{bmatrix}$$

olarak bulunur. Korelasyon matrisi incelendiğinde korelasyonların 1'e yakın değerler aldığı görülmektedir. Regresyon katsayılarının tahmini, VIF değerleri ve korelasyon matrisinin özdeğerleri ise Tablo 8.1'de gösterilmiştir. Tablo 8.1'e göre VIF değerleri 10'dan büyük çıkmıştır ve özdeğerler elde edildikten sonra en büyük özdeğerin en küçük özdeğere bölünmesiyle koşul sayısı

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{5.8680}{0.0002} = 29340$$

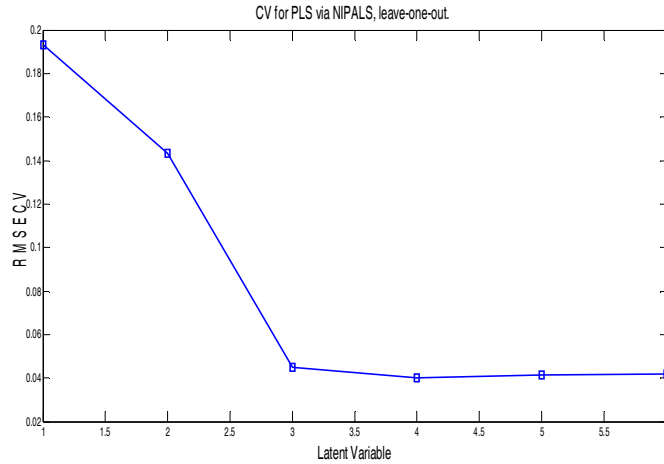
elde edilmiştir. 1000'den büyük olduğu için açıklayıcı değişkenler arasında çok ciddi boyutlarda iç ilişki problemi vardır. Çoklu iç ilişkiyi belirlemede kullanılan tüm kriterler göz önüne alındığında EKK yöntemi ile elde edilen katsayıların güvenilir sonuçlar vermeyeceğine karar verilir. EKK yöntemi ile bulunan modelin veriyi ne derece doğru temsil ettiği konusunda emin olabilmek için bu yönteme alternatif olarak kullanılan PLSR, PCR, RR ve CR yöntemleri Matlab programı yardımıyla veri üzerinde uygulanır. Daha sonra RMSEC, RMSECV ve öntahmin kümesinden elde edilen RMSEP değerleri ile modellerin veriye uyumu ve öntahmin yeteneği karşılaştırılır. Ayrıca uygulamada “birini-dışarıda-bırak” (leave-one-out) çapraz geçerlilik yöntemi kullanılır.



Tablo 8. 1. EKK yöntemi ile bulunan katsayılar, VIF değerleri ve özdeğerler

Değişken	Katsayılar	Standart Hata	t	VIF	Özdeğer
$L_1$	0.6577	1.8663	0.3524	3512.2	5.8680
$L_2$	0.0333	1.6932	0.0197	2890.9	0.1007
$L_3$	5.0248	1.6090	3.1230	2610.3	0.0187
$L_4$	-5.1816	1.3240	-3.9137	1767.5	0.0122
$L_5$	0.3708	0.1865	1.9885	35.1	0.0002
$L_6$	-0.4267	0.5306	-0.8042	283.9	0.0002

PLSR modeli oluşturulmadan önce modelde yer alacak gizli değişkenlerin sayısının belirlenmesi gerekir. Bunun için NIPALS algoritması yardımıyla hesaplanan RMSECV değerlerine karşı gizli değişkenlerin yer aldığı Şekil 8.1 incelendiğinde gizli değişken sayısı dört olduğunda RMSECV en küçük değeri almaktadır. Bu nedenle PLSR modeli oluşturulurken dört gizli değişken dikkate alınır.



Şekil 8. 1. 24 gözlem için PLSR yöntemine ilişkin RMSECV grafiği

PLSR modeli için bulunan RMSECV değerleri ise Tablo 8.2 ile verilmiştir.

Tablo 8. 2. 24 gözlem PLSR modeline ait RMSECV değerleri

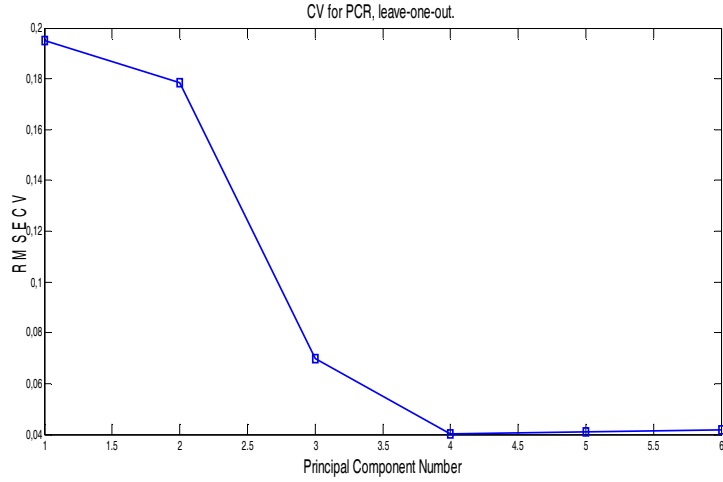
Değişken Sayısı	RMSECV_PLSR
1	0.1934
2	0.1431
3	0.0449
<b>4</b>	<b>0.0399</b>
5	0.0415
6	0.0419

Tablo 8.3, dört gizli değişkene sahip PLSR modelinin açıklayıcı değişkenler ve yanıt değişken için hesaplanan varyans yüzdelerini göstermektedir. Tablo 8.3'e göre dört gizli değişkene sahip PLSR modeli açıklayıcı değişkenlerdeki varyansın %99.99'nu açıklarken yanıt değişkendeki değişimin %97.91'ni açıklamaktadır.

Tablo 8. 3. 24 gözlem için PLSR yöntemi kullanarak oluşturulan regresyon modeli tarafından açıklanan varyans yüzdeleri

LV No	X-Blok		Y-Blok	
	LV	Toplam	LV	Toplam
1	97.77	97.77	22.46	22.46
2	1.55	99.33	40.32	62.78
3	0.46	99.79	35.00	97.77
<b>4</b>	<b>0.20</b>	<b>99.99</b>	<b>0.13</b>	<b>97.91</b>
5	0.00	100.00	0.30	98.20
6	0.00	100.00	0.01	98.21

PCR modelinde kullanılacak temel bileşen sayısı Şekil 8.2 yardımıyla belirlenir. PCR için hesaplanan RMSECV değerlerine karşı temel bileşenlerin yer aldığı Şekil 8.2'de temel bileşen sayısı dört olduğunda RMSECV en küçük değeri almaktadır. Bu nedenle PCR modeli için dört temel bileşen dikkate alınır.



Şekil 8. 2. 24 gözlem için PCR yöntemine ilişkin RMSECV grafiği

PCR modeli için bulunan RMSECV değerleri ise Tablo 8.4 ile gösterilir.

Tablo 8. 4. 24 gözlem için PCR modeline ait RMSECV değerleri

Değişken Sayısı	RMSECV_PCR
1	0.1952
2	0.1786
3	0.0698
<b>4</b>	<b>0.0401</b>
5	0.0411
6	0.0419

Tablo 8.5 dört bileşenli PCR modeli açıklayıcı değişkenler ve yanıt değişken için hesaplanan varyans yüzdelerini gösterir. Tablo 8.5'e göre dört bileşenli PCR modeli açıklayıcı değişkenlerdeki varyansın %99.99'nu açıklarken yanıt değişkenindeki değişimin %97.89'nu açıklamaktadır.

Tablo 8. 5. 24 gözlem için PCR yöntemi kullanılarak oluşturulan regresyon modeli tarafından açıklanan varyans yüzdeleri

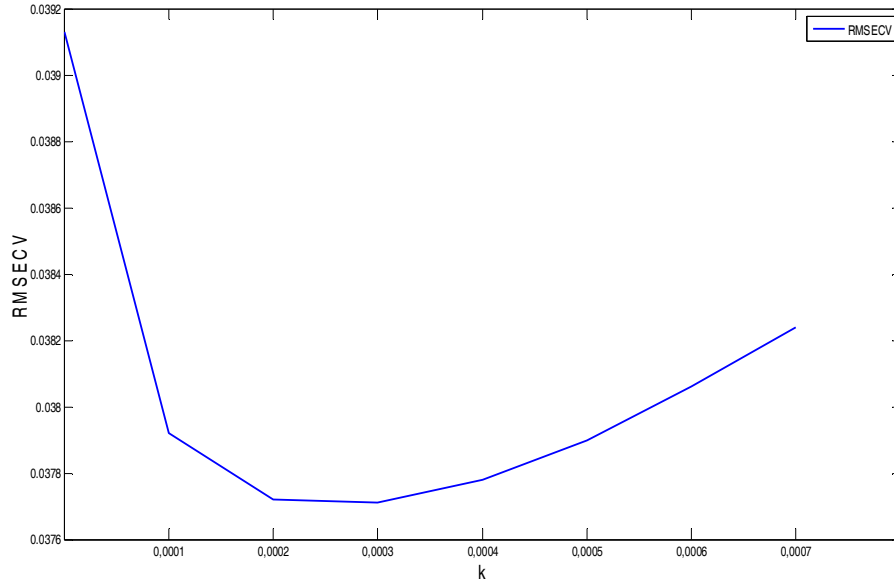
PC No	X-Blok		Y-Blok	
	PC	Toplam	PC	Toplam
1	97.80	97.80	21.50	21.50
2	1.68	99.48	16.71	38.21
3	0.31	99.79	58.33	96.54
<b>4</b>	<b>0.20</b>	<b>99.99</b>	<b>1.35</b>	<b>97.89</b>
5	0.00	100.00	0.12	98.01
6	0.00	100.00	0.21	98.21

Ridge regresyon yönteminde en önemli konu uygun  $k$  değerinin seçimidir. Farklı seçim yöntemleri olmakla birlikte PLSR ve PCR yöntemlerinde modelde kalacak bileşen sayısı RMSECV ile belirlendiğinden  $k$ 'nın değeri de RMSECV kullanılarak belirlenmiştir. Tablo 8.6'da  $k = 0.0003$  olduğu gösterilmektedir.

Tablo 8. 6.  $k$ 'nın çeşitli değerlerine karşı gelen  $PRESS(k)$  ve RMSECV değerleri

<b>k</b>	<b>PRESS_k</b>	<b>RMSECV_Ridge</b>
0	0.03675	0.03913
0.0001	0.03451	0.03792
0.0002	0.03415	0.03772
<b>0.0003</b>	<b>0.03414</b>	<b>0.03771</b>
0.0004	0.03427	0.03778
0.0005	0.03449	0.03790
0.0006	0.03477	0.03806
0.0007	0.03511	0.03824

$RMSECV$  grafiği ise Şekil 8.3'te gösterildiği gibidir. Şekil 8.3'e göre minimum  $RMSECV$  değeri ridge sabiti yaklaşık olarak 0.0003 değerine karşılık gelmektedir.



Şekil 8. 3. Ridge regresyon için  $RMSECV$  grafiği

Bu sonuçlar doğrultusunda Fearn'nın (1993) verisi için

EKK yöntemi uygulandığında

$$\hat{y} = 0.6577L_1 + 0.0333L_2 + 5.0248L_3 - 5.1816L_4 + 0.3708L_5 - 0.4267L_6$$

PCR yöntemi uygulandığında

$$\hat{y} = -0.7552L_1 + 3.0947L_2 + 2.6914L_3 - 3.9080L_4 + 0.4373L_5 - 1.0692L_6$$

PLSR yöntemi uygulandığında

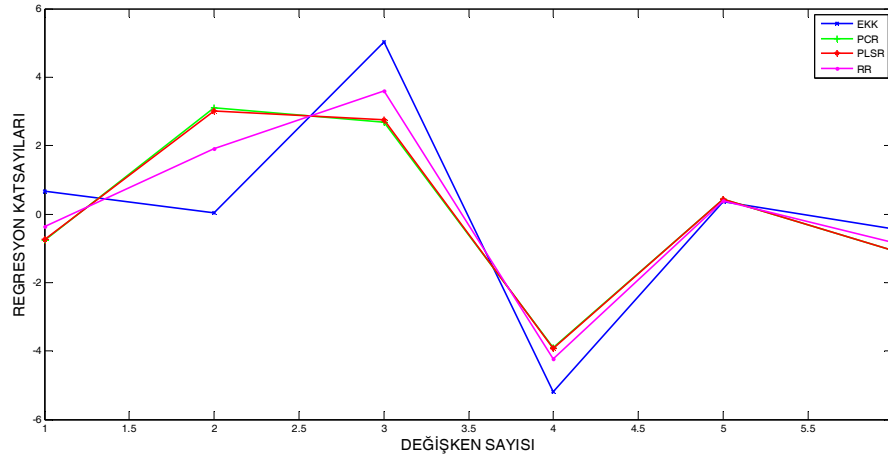
$$\hat{y} = -0.7303L_1 + 3.0148L_2 + 2.7652L_3 - 3.9287L_4 + 0.4354L_5 - 1.0657L_6$$

$k = 0.0003$  için RR yöntemi uygulandığında

$$\hat{y} = -0.3626L_1 + 1.9210L_2 + 3.6021L_3 - 4.2240L_4 + 0.3788L_5 - 0.8301L_6$$

modelleri tahmin edilmiştir.

Şekil 8.4'e göre PCR ve PLSR modellerindeki regresyon katsayıları arasında işaret ve büyüklük bakımından genel olarak çok fazla bir fark olmadığı görülmektedir. RR modelindeki regresyon katsayıları ise işaret bakımından PLSR ve PCR modelleri ile aynı olmasına karşın büyüklük bakımından farklılık göstermektedir. Yanlı regresyon yöntemleri ile elde edilen modellerin regresyon katsayıları EKK modelindeki regresyon katsayılarından farklı çıkmıştır.



Şekil 8. 4. RR, PLSR, PCR ve EKK modellerindeki regresyon katsayıları

Protein verisi için CR modelleri (PLSR, PCR, EKK) Matlab'da  $cr[X_a, y_a, (\gamma)]$  komutu kullanılarak hesaplanmıştır. İlk olarak EKK modelini veren katsayılar elde edilmiştir. CR parametresi sıfır olduğunda birden altıya kadar kaç değişken kullanırsa kullanılsın yine EKK katsayıları elde edilir.

- Sürekli regresyon parametresi 0 iken EKK modeli;

```
>> b = cr[X_a, y_a, 6, (0)]
```

b =

```
0.6577 0.0333 5.0248 -5.1816 0.3708 -0.4267
```

```
0.6577 0.0333 5.0248 -5.1816 0.3708 -0.4267
```

```
0.6577 0.0333 5.0248 -5.1816 0.3708 -0.4267
```

```
0.6577 0.0333 5.0248 -5.1816 0.3708 -0.4267
```

```
0.6577 0.0333 5.0248 -5.1816 0.3708 -0.4267
```

```
0.6577 0.0333 5.0248 -5.1816 0.3708 -0.4267
```

$$\hat{y} = 0.6577L_1 + 0.0333L_2 + 5.0248L_3 - 5.1816L_4 + 0.3708L_5 - 0.4267L_6$$

İkinci olarak RMSECV grafiğine göre dört gizli değişken dikkate alındığı için CR komutunda PLSR katsayıları dört gizli değişken kullanıldığında elde edilir. Ayrıca altı değişkenin tamamı kullanıldığında EKK katsayıları elde edilmiştir.

- Sürekli regresyon parametresi 1 iken PLSR modeli;

```
>> b = cr[Xa, ya, 6, (1)]
```

```
b =
```

```
0.0813  0.0960  0.0936  0.0667  0.0625  0.0786
0.1432  1.4338  1.2056 -1.0916 -1.1900 -0.0728
-0.8653  3.0471  2.6667 -3.9782  0.4031 -0.7836
-0.7303  3.0148  2.7652 -3.9287  0.4354 -1.0657
-0.0014  0.1808  5.2729 -4.7356  0.3213 -0.5601
0.6577  0.0333  5.0248 -5.1816  0.3708 -0.4267
```

$$\hat{y} = -0.7303L_1 + 3.0148L_2 + 2.7652L_3 - 3.9287L_4 + 0.4354L_5 - 1.0657L_6$$

Son olarak RMSECV grafiğine göre dört temel bileşen dikkate alındığı için CR komutunda PCR katsayıları dört değişken kullanıldığında bulunur. PLSR modelinde olduğu gibi altı değişkenin tamamı dikkate alındığında EKK katsayıları elde edilmiştir.

- Sürekli regresyon parametresi  $\infty$  iken PCR modeli;

```
>> b = cr[Xa, ya, 6, (inf)]
```

```
b =
```

```
0.0787  0.0784  0.0786  0.0786  0.0759  0.0786
0.3536  0.5192  0.4212  0.0428 -1.0491  0.1427
-1.1262  3.0114  2.4723 -4.0417  0.3048 -0.1340
-0.7552  3.0947  2.6914 -3.9080  0.4373 -1.0692
-1.7159  2.0353  4.3833 -3.4052  0.2850 -1.0983
0.6577  0.0333  5.0248 -5.1816  0.3708 -0.4267
```

$$\hat{y} = -0.7552L_1 + 3.0947L_2 + 2.6914L_3 - 3.9080L_4 + 0.4373L_5 - 1.0692L_6$$

olarak elde edilir.

Fearn'nın (1983) verisi kullanılarak bulunan modellerin karşılaştırılabilmesi için hata kareler ortalamasının karekökü yani *RMSE* değerlerinin hesaplanması önemlidir. Hesaplanan *RMSE* değerleri (RMSEC, RMSECV, RMSEP) Tablo 8.7'de gösterilmektedir. Ayarlama kümesine uygulanan yöntemlerle elde edilen modellerin öntahmin yeteneği öntahmin kümesi kullanılarak hesaplanan RMSEP değeri ile belirlenir. Buna göre RMSEP değeri en küçük olan model PLSR modelidir. Yani RMSEP değerine göre öntahmin bakımından en iyi model PLSR modelidir. En küçük RMSEC değeri EKK ve en küçük RMSECV değeri RR modellerine aittir. RMSEC değeri bakımından EKK yöntemi PLSR, PCR ve RR yöntemlerinden baskın çıkmıştır. Açıklayıcı değişkenler arasında iç ilişki problemi olduğu için RMSEC, değeri bakımından EKK modeli yerine PLSR, PCR ve RR yöntemlerinin en iyi sonucu vermesi beklenir.

Tablo 8. 7. EKK, PCR, PLSR ve RR modelleri için RMSEC, RMSEP ve RMSECV değerleri

	EKK	PCR (4 PC)	PLSR (4 LV)	RR
RMSEC	0.0273*	0.0297	0.0295	0.0337
RMSEP	0.1437	0.1523	0.1522*	0.1481
RMSECV	0.0419	0.0401	0.0410	0.0377*

Fearn'nın (1983) üzerinde çalıştığı veriler ile ilgili literatürde yapılan çalışmalara göre ikinci veri kümesi (Tablo B.2) ilk veri kümesinden (Tablo B.1) farklıdır. Eğer Tablo B.2'de herhangi bir açıklayıcı değişken için ele alınan herhangi bir gözlem Tablo B.1'de aynı açıklayıcı değişkene ait gözlemlerin aldığı değer aralığının dışında kalan bir değere sahipse söz konusu gözlem Tablo B.1 tarafından kaplanan alanın dışında olmaktadır. Dolayısıyla Tablo B.2'de 26 gözlem yer almaktadır. Bu nedenle RR ve diğer yanlı regresyon yöntemlerinin EKK yöntemine göre öntahmin yeteneği 26 gözlemlili öntahmin kümesinde değerlendirildiğinde beklenmedik sonuçların çıktığı ileri sürülmektedir (Ryan, 1997).

Naes ve ark.'na (1986) göre ayarlama kümesindeki gözlem sayısının ( $n$ ) NIR radyasyon yansıma ölçülerini gösteren değişken sayısına ( $p$ ) oranı



$(n/p = 24/6 = 4)$  çok yüksek bir orandır. NIR verilerinde genellikle deęişken sayısı çok fazla olduęu için 4 gibi yüksek bir  $n/p$  oranı elde edilememektedir.  $n/p$  oranını küçültebilmek amacıyla Naes ve Martens (1985), Stone ve Brooks (1990) ve Björkström ve Sundberg'in (1996, 1999) çalışmaları dikkate alınarak Tablo B.1 ve Tablo B.2'deki ilk 12 gözlem ile analizler yapılacaktır. Böylece  $n/p = 24/12 = 2$  şeklinde daha küçük bir oran elde edilir.

İlk 12 gözlem analizler yapılmadan önce standartlaştırılmıştır. Protein (%) ve  $L_1 - L_6$  deęerlerinin mevcut olduęu ayarlama kümesi ile çoklu lineer modele ilişkin korelasyon matrisi,

$$X_a'X_a = \begin{bmatrix} 1.0000 & 0.9523 & 0.9675 & 0.9790 & 0.8277 & 0.9517 \\ 0.9523 & 1.0000 & 0.9968 & 0.8828 & 0.7895 & 0.9597 \\ 0.9675 & 0.9968 & 1.0000 & 0.9072 & 0.8215 & 0.9633 \\ 0.9790 & 0.8828 & 0.9072 & 1.0000 & 0.8485 & 0.9303 \\ 0.8277 & 0.7895 & 0.8215 & 0.8485 & 1.0000 & 0.8329 \\ 0.9517 & 0.9597 & 0.9633 & 0.9303 & 0.8329 & 1.0000 \end{bmatrix}$$

olarak bulunmuştur. Korelasyon matrisi incelendiğinde korelasyonların 1'e yakın deęerler aldığı görülmektedir. Regresyon katsayılarının tahmini, VIF deęerleri ve korelasyon matrisinin özdeęerleri ise Tablo 8.8'de gösterilmiştir. Tablo 8.8'e göre VIF deęerleri 10'dan büyük deęerler almıştır. Özdeęerler elde edildikten sonra en büyük özdeęerin en küçük özdeęere bölünmesiyle koşul sayısı

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{5.5438}{0.0004} = 1359.5$$

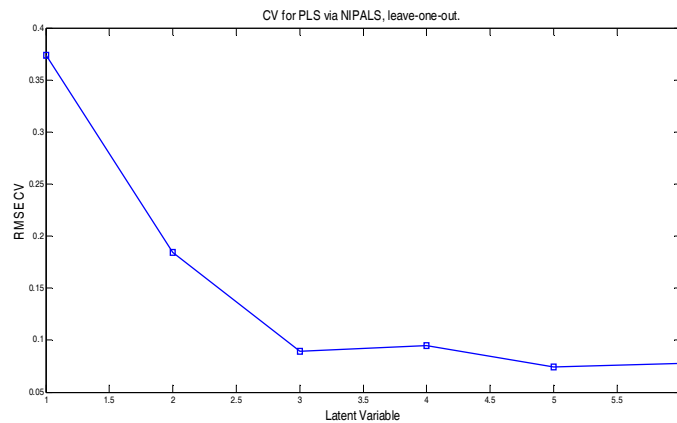
elde edilmiştir. 1000'den büyük olduęu için açıklayıcı deęişkenler arasında iç ilişki problemi vardır.

Tablo 8. 8. EKK yöntemi ile bulunan katsayılar, VIF değerleri ve özdeğerler

Değişken	Katsayılar	Standart Hata	t	VIF	Özdeğer
$L_1$	3.3033	4.5487	2.1329	1155.6	5.5438
$L_2$	0.1275	1.0830	0.1777	565.1	0.2722
$L_3$	0.0961	1.4877	0.0646	1066.3	0.1355
$L_4$	-4.6094	1.0938	-4.2141	576.4	0.0471
$L_5$	0.3572	0.1820	1.9630	16.0	0.0010
$L_6$	0.6560	0.4152	1.5799	83.1	0.0004

İlk olarak PLSR ve PCR yöntemleri için modellerde kalacak bileşen sayısını belirlemek için RMSECV grafikleri ve bu bileşenlere ait varyans yüzdeleri dikkate alınır.

PLSR modeli oluşturulmadan önce modelde yer alacak gizli değişkenlerin sayısının belirlenmesi için NIPALS algoritması yardımıyla hesaplanan RMSECV değerlerine karşı gizli değişkenlerin yer aldığı Şekil 8.5'te gizli değişken sayısı beş olduğunda RMSECV en küçük değeri almaktadır. Bu nedenle PLSR modeli oluşturulurken beş gizli değişken dikkate alınır.



Şekil 8. 5. 12 gözlem için PLSR yöntemine ilişkin RMSECV grafiği

PLSR modeli için bulunan RMSECV değerleri ise Tablo 8.9 ile verilmiştir.

Tablo 8. 9. 12 gözlem için PLSR modeline ait RMSECV değerleri

Değişken Sayısı	RMSECV_PLSR
1	0.3740
2	0.1843
3	0.0896
4	0.0949
<b>5</b>	<b>0.0744</b>
6	0.0780

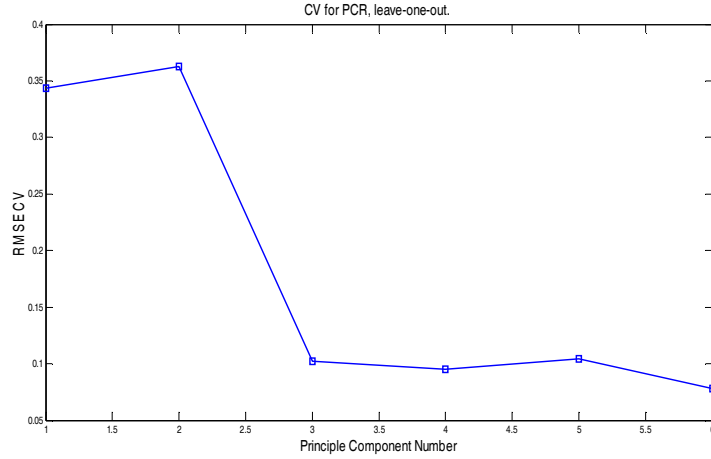
Tablo 8.10 beş gizli değişkene sahip PLSR modelinin açıklayıcı değişkenler ve yanıt değişken için hesaplanan varyans yüzdelerini göstermektedir. Tablo 8.10'a göre beş gizli değişkene sahip PLSR modeli açıklayıcı değişkenlerdeki varyansın %99.99'nu açıklarken yanıt değişkendeki değişimin %98.56'sını açıklamaktadır.

Tablo 8. 10. 12 gözlem için PLSR yöntemi kullanarak oluşturulan regresyon modeli tarafından açıklanan varyans yüzdeleri

LV No	X-Blok		Y-Blok	
	LV	Toplam	LV	Toplam
1	4.68	4.68	85.07	85.07
2	90.09	94.77	2.01	87.08
3	4.40	99.17	9.44	96.52
4	0.81	99.98	0.79	97.31
<b>5</b>	<b>0.01</b>	<b>99.99</b>	<b>1.25</b>	<b>98.56</b>
6	0.01	100.00	0.19	98.75

PCR modelinde kullanılacak temel bileşen sayısı ise Şekil 8.6 yardımıyla belirlenir. PCR için hesaplanan RMSECV değerlerine karşı temel bileşenlerin yer

aldığı Şekil 8.6'da temel bileşen sayısı altı olduğunda RMSECV en küçük değeri almaktadır. Bu nedenle PCR modeli için altı temel bileşen dikkate alınır.



Şekil 8. 6. 12 gözlem için PCR yöntemine ilişkin RMSECV grafiği

PCR modeli için bulunan RMSECV değerleri ise Tablo 8.11 ile verilmiştir.

Tablo 8.11. 12 gözlem için PCR modeline ait RMSECV değerleri

Değişken Sayısı	RMSECV_PCR
1	0.3436
2	0.3624
3	0.1019
4	0.0951
5	0.1044
<b>6</b>	<b>0.0780</b>

Tablo 8.12'den görüldüğü üzere altı bileşenli PCR modeli açıklayıcı değişkenler ve yanıt değişken için hesaplanan varyans yüzdelerini göstermektedir. Tablo 8.12'ye göre altı bileşenli PCR modeli açıklayıcı değişkenlerdeki varyansın %100'nü açıklarken yanıt değişkendeki değişimin % 98.75'ni açıklamaktadır.

Tablo 8. 12. 12 gözlem için PCR yöntemi kullanılarak oluşturulan regresyon modeli tarafından açıklanan varyans yüzdeleri

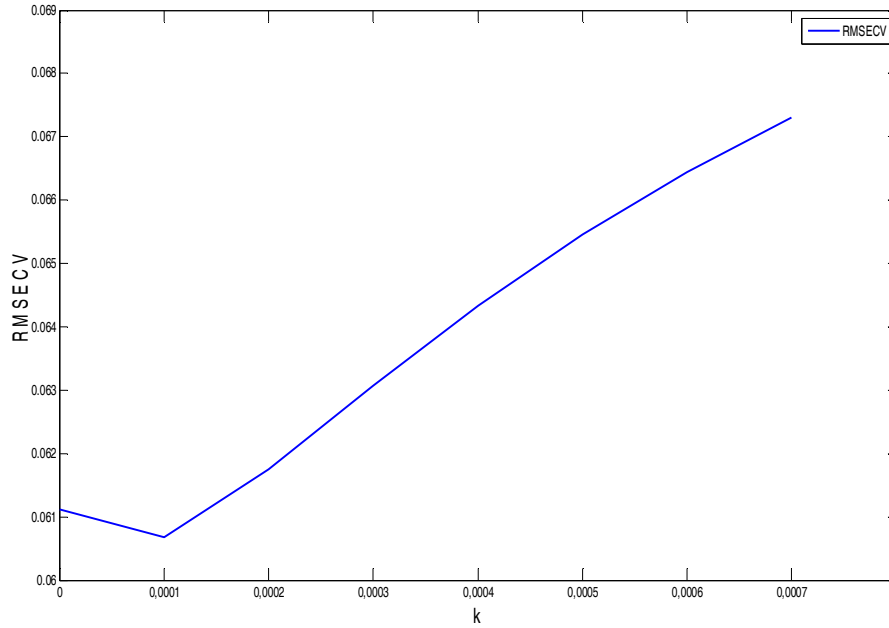
PC No	X-Blok		Y-Blok	
	PC	Toplam	PC	Toplam
1	92.40	92.40	0.00	0.00
2	4.54	96.93	18.74	18.74
3	2.26	99.19	75.85	94.58
4	0.79	99.98	2.66	97.24
5	0.02	99.99	0.76	98.00
<b>6</b>	<b>0.01</b>	<b>100.00</b>	<b>0.75</b>	<b>98.75</b>

PLSR ve PCR yöntemlerinin bileşen sayısı RMSECV ile belirlendiğinden  $k$ 'nın değeri de öntahmin amaçlı bir kriter olan RMSECV kullanılarak belirlenmiştir. Tablo 8.13 bu değerin  $k = 0.0001$  olduğunu göstermektedir.

Tablo 8. 13.  $k$ 'nın çeşitli değerlerine karşı gelen  $PRESS(k)$  ve RMSECV değerleri

<b>k</b>	<b>PRESS_k</b>	<b>RMSECV_Ridge</b>
0	0.04482	0.06111
<b>0.0001</b>	<b>0.04419</b>	<b>0.06068</b>
0.0002	0.04577	0.06175
0.0003	0.04774	0.06307
0.0004	0.04967	0.06433
0.0005	0.05145	0.06546
0.0006	0.05298	0.06644
0.0007	0.05436	0.06731

$RMSECV$  grafiği ise Şekil 8.7'de gösterildiği gibidir. Şekil 8.7'ye göre minimum  $RMSECV$  değeri ridge sabiti olarak yaklaşık 0.0001 değerine karşılık gelmektedir.



Şekil 8. 7. Ridge regresyon için  $RMSECV$  grafiği

EKK ve RR yöntemleri altı değişkeni de kullanırken PLSR ve PCR yöntemleri model için önemli olan bileşenleri seçer. Fakat ilk 12 gözlem ele alındığında PCR için önemli olan bileşenler altı olarak belirlenmiştir. Bu sonuçlar doğrultusunda Fearn'nın (1983) verisi için

EKK yöntemi uygulandığında

$$\hat{y} = 3.3033L_1 + 0.1275L_2 + 0.0961L_3 - 4.6094L_4 + 0.3572L_5 + 0.6560L_6$$

PCR yöntemi uygulandığında

$$\hat{y} = 3.3033L_1 + 0.1275L_2 + 0.0961L_3 - 4.6094L_4 + 0.3572L_5 + 0.6560L_6$$

PLSR yöntemi uygulandığında

$$\hat{y} = 2.2340L_1 - 0.5901L_2 + 1.5199L_3 - 3.9530L_4 + 0.2099L_5 + 0.5029L_6$$

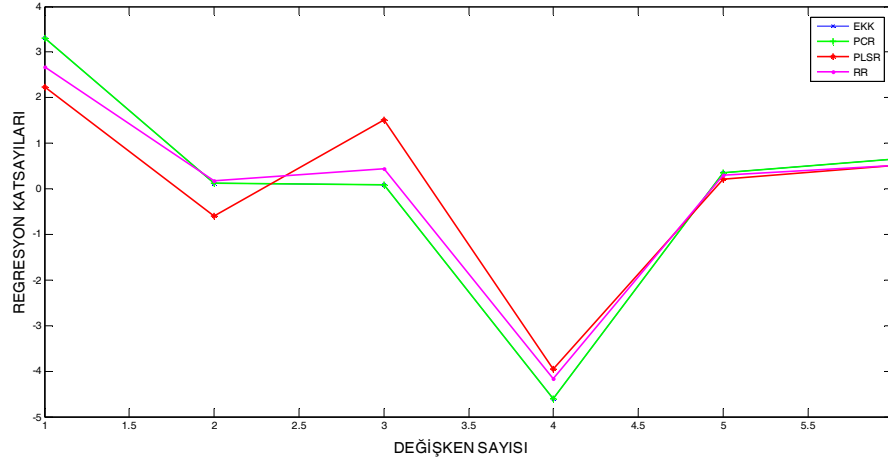
$k = 0.0001$  için RR yöntemi uygulandığında

$$\hat{y} = 2.6746L_1 + 0.1780L_2 + 0.4346L_3 - 4.1623L_4 + 0.3022L_5 + 0.5097L_6$$

modelleri edilir.

Şekil 8.8'e göre EKK ile PCR modelleri aynı çıkmıştır. Çünkü PCR yöntemiyle bulunan  $RMSECV$  grafiğinde en küçük değer altı bileşen modele dahil

edildiğinde bulunmaktadır. PLSR modeli 2. değişkene ait katsayının işareti diğer modellerden farklıdır. Diğer katsayılar ise işaret bakımından aynı ve birbirine yakın değerler almıştır.



Şekil 8.8. RR, PLSR, PCR ve EKK modellerindeki regresyon katsayıları

Protein verisi için CR modelleri (PLSR, PCR, EKK) Matlab'da  $X_a$  ve  $y_a$ , 12 birimlik gözlemleri göstermek üzere  $cr[X_a, y_a, (\gamma)]$  komutu kullanılarak hesaplanmıştır. İlk olarak EKK modelini veren katsayılar elde edilmiştir. CR parametresi sıfır olduğunda birden altıya kadar hangi değişken kullanılırsa kullanılsın yine EKK katsayıları elde edilir.

- Sürekli regresyon parametresi 0 iken EKK modeli;

```
>> b = cr[X_a, y_a, 6, (0)]
```

b =

3.3033	0.1275	0.0961	-4.6094	0.3572	0.6560
3.3033	0.1275	0.0961	-4.6094	0.3572	0.6560
3.3033	0.1275	0.0961	-4.6094	0.3572	0.6560
3.3033	0.1275	0.0961	-4.6094	0.3572	0.6560
3.3033	0.1275	0.0961	-4.6094	0.3572	0.6560
<b>3.3033</b>	<b>0.1275</b>	<b>0.0961</b>	<b>-4.6094</b>	<b>0.3572</b>	<b>0.6560</b>

$$\hat{y} = 3.3033L_1 + 0.1275L_2 + 0.0961L_3 - 4.6094L_4 + 0.3572L_5 + 0.6560L_6$$

İkinci olarak RMSECV grafiğine göre beş gizli değişken dikkate alındığı için CR komutunda PLSR katsayıları beş gizli değişken kullanıldığında elde edilir. Ayrıca altı değişkenin tamamı dikkate alındığında EKK katsayıları elde edilmiştir. Sürekli regresyon parametresi 1 iken PLSR modeli;

$$\gg b = cr[X_a, y_a, 6, (1)]$$

b =

-0.4104	1.1603	0.8939	-1.4395	-0.5130	0.1327
-0.4025	1.2114	0.9410	-1.4602	-0.4726	0.1535
-0.6031	1.2991	1.0428	-1.8387	0.0786	0.0074
-0.4039	1.3337	1.1577	-1.8643	0.0966	-0.3338
<b>2.2340</b>	<b>-0.5901</b>	<b>1.5199</b>	<b>-3.9530</b>	<b>0.2099</b>	<b>0.5029</b>
3.3033	0.1275	0.0961	-4.6094	0.3572	0.6560

$$\hat{y} = 2.2340L_1 - 0.5901L_2 + 1.5199L_3 - 3.9530L_4 + 0.2099L_5 + 0.5029L_6$$

Son olarak RMSECV grafiğine göre altı temel bileşen dikkate alındığı için CR komutunda PCR katsayıları altı değişken kullanıldığında bulunur. Bu nedenle EKK katsayıları ile aynı katsayılar elde edilmiştir.

- Sürekli regresyon parametresi  $\infty$  iken PCR modeli;

$$\gg b = cr[X_a, y_a, 6, (\text{inf})]$$

b =

-0.0006	-0.0006	-0.0006	-0.0006	-0.0006	-0.0006
0.1064	0.3030	0.2167	-0.0991	-0.7173	0.1173
-0.7419	1.2433	0.9338	-1.7952	0.0360	0.3077
-0.4780	1.3831	1.1444	-1.8089	0.0928	-0.3460
0.5354	-0.5618	2.5538	-2.7628	0.0549	0.1178
<b>3.3033</b>	<b>0.1275</b>	<b>0.0961</b>	<b>-4.6094</b>	<b>0.3572</b>	<b>0.6560</b>



$$\hat{y} = 3.3033L_1 + 0.1275L_2 + 0.0961L_3 - 4.6094L_4 + 0.3572L_5 + 0.6560L_6$$

olarak elde edilir.

Tablo 8.14'te RMSEP1 12 gözlemlili öntahmin kümesi ve RMSEP2 ise 26 gözlemlili öntahmin kümesi için hesaplanmıştır. RMSEP1 değeri en küçük olan model PLSR modelidir. RMSEP2 değeri ise RR yönteminde en küçük değeri almıştır. RR ve PLSR modelleri öntahmin bakımından en iyi modellerdir ve hemen hemen aynı performansı göstermektedir. En küçük RMSEC ve RMSECV değerleri ise RR yöntemi ile bulunan modele aittir. Bu durumda Fearn'nın (1983) 12 gözlem içeren verisini en iyi yansıtan modelin RR modeli olduğu söylenebilir. Dolayısıyla en iyi model açıklayıcı değişkenler arasında çoklu iç ilişki olduğunda kullanılan yanlı regresyon yöntemleri kullanılarak bulunan modeller arasından seçilmiştir.

Tablo 8. 14. 12 gözlem için EKK, PCR, PLSR ve RR modelleri için RMSEC, RMSEP ve RMSECV değerleri

	<b>EKK</b>	<b>PCR (6 PC)</b>	<b>PLSR (5 LV)</b>	<b>RR</b>
<b>RMSEC</b>	0.0322	0.0322	0.0346	0.0274*
<b>RMSECV</b>	0.0780	0.0780	0.0744	0.0607*
<b>RMSEP1</b>	0.1749	0.1749	0.1704*	0.1705
<b>RMSEP2</b>	0.1190	0.1190	0.1173	0.1170*



## 9. SONUÇLAR VE ÖNERİLER

Açıklayıcı değişkenler arasında çoklu iç ilişki problemi söz konusu olduğunda EKK tahmin edicisi büyük varyansa sahip olur ve katsayılar gerçek değerlerinden uzaklaşır. Bu durumda RR, PLSR ve PCR yöntemleri EKK'ya alternatif olarak önerilmiştir. RR, rezidü kareler toplamının minimum olabilmesi için tahmin ediciler üzerinde bir ölçek ayarlaması yapar. PLSR yanıt değişkenler ve bu değerlerin uydurulmuş değerleri arasındaki kovaryansı maksimum, PCR açıklayıcı değişkenlerin varyansını maksimum ve EKK yöntemi ise açıklayıcı değişkenler arasındaki korelasyonu maksimum yapmayı amaçlar. Stone ve Brooks (1990) PCR, PLSR ve EKK yöntemlerini kapsayacak şekilde CR yöntemini önermiştir. Yapılan çalışmalarda CR ile RR yöntemi arasında da ilişki olduğu çeşitli yazarlar tarafından gösterilmiştir.

Bu çalışmada CR üzerinde durulmuş ve bu çerçevede yer alan EKK, RR, PLSR ve PCR yöntemleri incelenmiştir. PLSR ve PCR yöntemleri için NIPALS algoritmasına yer verilmiştir. Ayrıca regresyon çözümlemesinde modelin veriye uyumunu ve geçerliliğini test etmek amacıyla RMSE, RMSEC ve RMSECV ölçüleri anlatılmıştır. Literatürde sürekli regresyonla ilgili en sık kullanılan Fearn'nın (1983) verisi ele alınarak teorik incelemeler örneklendirilmiştir. Bu veri üzerinde Matlab programı kullanılarak EKK, RR, PLSR, PCR ve CR yöntemleri ile elde edilen modeller veriye uyum bakımından ve öntahmin bakımından karşılaştırılmıştır. Çalışmanın teorik kısmında PLSR yöntemi için NIPALS algoritması tanımlanmıştır. Buna göre NIPALS algoritmasına ilişkin RMSECV grafiğine yer verilmiştir.

Stone ve Brooks (1990) sürekli regresyonun PCR, PLSR ve EKK yöntemleri ile ilişkili olduğunu önermiştir. Bu önermeyi göstermek amacıyla yapılan analizlerde PLSR, PCR ve EKK yöntemlerine ait komutlar dışında CR komutunda CR parametresi yerine EKK, PLSR ve PCR yöntemlerine karşılık gelen 0, 1 ve  $\infty$  değerleri kullanılarak modeller bulunmuştur.

Fearn'nın (1983) verisi üzerinde başlangıçta çoklu lineer regresyon yöntemi ile korelasyon matrisi, VIF değerleri ve koşul sayısı bulunmuş ve bu değerlere bakarak değişkenler arasında çoklu iç ilişki problemi olduğu belirlenmiştir.

Dolayısıyla EKK yöntemi ile elde edilen katsayılar güvenilir sonuçlar vermeyeceğinden EKK yöntemine alternatif olarak önerilen yöntemlerden PLSR ve PCR yöntemleri için RMSECV grafiği yardımıyla gizli değişken ve temel bileşen sayıları bulunmuştur. RR yöntemi için de PRESS değerleri ve RMSECV değerleri yardımıyla uygun ridge sabiti belirlenmiştir. Bu bilgilerden faydalanarak PCR, PLSR ve RR modelleri oluşturulmuştur. 24 birimlik veri kullanıldığında EKK modeli RMSEC değeri bakımından en iyi sonucu vermiştir. Fakat değişkenler arasında çoklu iç ilişki problemi olduğundan EKK modelinin RMSEC değeri bakımından yanlı regresyon yöntemlerine göre iyi sonuç vermesi beklenen bir durum değildir. Fearn'nın (1983) üzerinde çalıştığı veriler ile ilgili literatürde yapılan çalışmalara göre öntahmin veri kümesi ayarlama veri kümesinden farklıdır. Bu nedenle literatürde Fearn'nın (1983) verisi ile ilgili çalışmalardan esinlenerek ayarlama kümesi için ilk 12 gözlem, öntahmin kümesi için ilk 12 gözlem (RMSEP1) ve 26 gözlem (RMSEP2) ile analiz yapılmıştır. Buna göre RMSECV, RMSEC ve RMSEP2 değerleri bakımından RR modelinin veriyi iyi yansıttığı söylenebilir. Ayrıca RMSEP1 değeri bakımından PLSR modeli iyi çıkmıştır. Öntahmin bakımından PLSR ve RR yöntemleri birbirine yakın performans göstermektedir. Dolayısıyla yanlı regresyon yöntemlerine ait modellerin iyi çıkması ile istenen sonuç elde edilmiştir.

## KAYNAKLAR

- ALLEN, D. M., 1971. Mean Square Error of Prediction as a Criterion for Selection of Variables. *Technometrics*, 13, 469-475.
- ALLEN, D. M., 1974. The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction. *Technometrics*, 16, 125-127.
- BJÖRKSTRÖM, A. and SUNDBERG, R., 1996. Continuum Regression is not Always Continuous. *J. R. Statist. Soc. B*, 58, No:4, pp.703-710.
- BJÖRKSTRÖM, A. and SUNDBERG, R., 1999. A Generalized View on Continuum Regression. *Board of The Foundation of The Skandinavian Journal of Statistics*, Vol: 26, pp. 17-30.
- BULUT, E. and ALIN, A., 2009. Kısmi En Küçük Kareler Regresyon Yöntemi Algoritmalarından Nipals ve PLSR-Kernel Algoritmalarının Karşılaştırılması ve Bir Uygulama. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* Cilt:24, Sayı:2, ss.127-138.
- BULUT, E. ve ALMA, Ö.G., 2011. Kısmi En Küçük Kareler Regresyonu Yardımıyla Optimum Bileşen Sayısını Seçmede Model Seçme Kriterlerinin Performans Karşılaştırması. *Ekonometri ve İstatistik Dergisi*, Sayı: 15, ss. 38-52.
- DE JONG, S. and FAREBROTHER, R. W., 1994. Extending the Relationship Between Ridge Regression and Continuum Regression. *Chemometr. Intell. Lab. Syst.*, 25, pp. 179-181.
- EİGENVECTOR RESEARCH, 1997-1998.
- FARRAR, D. E. and GLAUBER, R. R., 1967. Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics*. 49, 1, pp. 92-107.
- FEARN T., 1983. A Misuse of Ridge Regression in the Calibration of a Near Infrared Reflectance Instrument. *Applied Statistics*, 32, No:1, pp. 73-79.
- FRANK, I. E. and FRIEDMAN, J. H., 1993. A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, Vol: 35, No: 2., pp. 109-148.

- FRANK, I. E., 1987. Intermediate Least Squares Regression Method. *Chemomet. Intell. Lab. Syst.*, 1, 233-242.
- GOLUB, G. H., HEATH, C. G. and WAHBA, W., 1979. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21, 215-223.
- GUJARATI, D. N. (1995). *Basic Econometrics*, 3<sup>rd</sup>Ed., McGraw-Hill, New-York.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J., 2008. *The Elements of Statistical Learning (Data Mining, Inference and Prediction)*. Springer Series in Statistics, Second Edition, Stanford/California.
- HOERL, A. E. and KENNARD, R. W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 1, pp. 55-67.
- HOERL, A. E., 1962. Application of Ridge Analysis to Regression Problems. *Chemical Engineering Progress*, 58, pp. 54-59.
- HÖSKULDSSON, A., 1988. PLSR Regression Methods, *Journal of Chemometrics*, 2, 211-228.
- JORGENSEN, B. and GOEGEBEUR, Y., 2007. *Multivariate Data Analysis and Chemometrics*. Department of Statistics.
- JUDGE, G. G., GRIFFITHS, W. E., HILL, R. C., LÜTKEPOHL, H. and LEE, T. C., 1985. *The Theory and Practice of Econometrics*. John Wiley and Sons, Newyork, 1018p.
- MARQUARDT, D. W., 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation. *Technometrics*, 12, 591-612.
- MASSY, W. F. 1965. Principal Components Regression in Explanatory Statistical Research. *Journal of the American Statistical Association*, 60, pp. 234-246.
- MCKAY, B., WILLIS, M., SEARSON, D. and MONTAGUE, G. (2000). Non-Linear Continuum Regression Using Genetic Programming. *Proceedings of The Genetic and Evolutionary Computation Conference*, 2, Page 1106-1111. Orlando, Florida, USA, Morgan Kaufmann.
- MONTGOMERY, D., PECK, E. A. and VINING, G. G., 2001. *Introduction to Linear Regression Analysis*. John Wiley ans Sons Inc., New York.

- NAES, T. and MARTENS, H., 1985. Comparison Prediction Methods for Multicollinear Data, *Communications in Statistics-Simulation and Computation*, 14(3), 545-576.
- NAES, T. and MARTENS, H., 1989. *Multivariate Calibration*. John Wiley&Sons.
- NAES, T., IRGENS, C. and MARTENS, H., 1986. Comparison of Linear Statistical Methods for Calibration of NIR Enstruments, *Applied Statistics*, 35(2), 195-206.
- NAES, T., ISAKSSON, T., FEARN, T., DAVIES, T., 2002. *A User-Friendly Guide to Multivariate Calibration and Classification*, NIR Publications Chichester, UK.
- NETTER, J., WASSERMAN, W. and KUNTER, M., (1990). *Applied Linear Statistical Models*. 3<sup>rd</sup>Ed., New Jersey.
- ÖZKALE, M. R., 2007. Çoklu İç İlişki İle İlgili Problemler. Çukurova Üniversitesi Fen Bilimleri Enstitüsü. Doktora Tezi.
- PHATAK, A. and DE JONG, S., 1997. The Geometry of Partial Least Squares. *Journal of Chemometrics*, 11, 311-338.
- POLAT, E., 2009. Kısmi En küçük Kareler Regresyonu. Hacettepe Üniversitesi, Yüksek Lisans Tezi.
- RYAN, P. T., 1997. *Modern Regression Methods*. John Wiley&Sons, New York.
- SHAO, J., 1993. Linear Model Selection by Cross-Validation. *Journal of The American Statistical Association*.
- SILVEY, S. D., 1969. Multicollinearity and Imprecise Estimation. *Journal of the Royal Statistical Society, Series B (Methodological)*, 31, 3, pp. 539-552.
- STONE, M. and BROOKS, R. J., 1990. Continuum Regression: Cross-Validated Sequentially Constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares and Principal Components Regression. *J. R. Statist. Soc. B*, 52, pp. 237-269.
- STONE, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of Royal Statistical Society B*. 36:111-147.
- SUNDBERG, R., 1993. Continuum Regression and Ridge Regression. *J. R. Statist. Soc. B*, 55, No:3, pp.653-659.

- SUNDBERG, R., 2002. Shrinkage Regression. John Wiley&Sons. LTD. Chichester.
- WISE, B. M. and RICKER, N. L., 1990. The Effect of Biased Regression on The Identification of Fir and Arx Models. Prepared for Presentation at The 1990 AIChE Annual Meeting. Recent Advances in Process Control.
- WISE, B. M. and RICKER, N. L., 1993. Identification of Finite Impulse Response Models with Continuum Regression. Journal of Chemometrics, Vol.7, 1-14.
- WISE, B. M., 1992-1993-1995.
- WISE, B. M., GALLAGHER, N. B., BRO, R. and SHAVER, J. M., 2003. PLSR\_Toolbox 3.0 for use with Matlab.
- WISE, B. M., GALLAGHER, N. B., BRO, R., SHAVER, J. M., WINDING W. and KOCH R. S., 2006. Chemometrics Tutorial for PLSR\_Toolbox and Solo.
- WOLD, H., 1966. Estimation of Principal Components and Related Models by Iterative Least Squares. New York: Academic Press.
- WOLD, H., 1984. PLS Regression. In Encyclopaedia of Statistical Sciences (eds N. L. Johnson and S. Kotz), vol. 6, pp.581-591. New York:Wiley.
- WOLD, S., 1978. Cross-Validatory Estimation of The Number of Components in Factor and Principal Components Models. Technometrics, 20, 397-405.
- WOLD, S., ESBENSEN, K. and GELADİ, P., 1987. Chemometrics and Intelligent Laboratory Systems. Volume 2, Issues 1-3, Pages 37-52.
- YENİAY, Ö. ve GÖKTAŞ, A., 2002. A Comparison of Partial Least Squares Regression With Other Prediction Methods. Hacettepe Journal of Mathematics and Statistics Volume: 31, pp. 99-111.
- <http://sites.stat.psu.edu/~jiali/course/stat597e/notes2/lreg.pdf>
- <http://www.eigenvector.com>
- <http://www.eigenvector.com/about/BMWcv.html>



## **ÖZGEÇMİŞ**

29.08.1987'de Tarsus'da doğdu. İlk, orta ve lise öğrenimini Tarsus'da tamamladı. 2005 yılında Osmangazi Üniversitesi Fen Edebiyat Fakültesi İstatistik bölümünü kazandı. 2009 yılında bu bölümden mezun oldu. Aynı yıl Çukurova Üniversitesi İstatistik bölümünde yüksek lisansa başladı. 2011 yılından itibaren T.C. Enerji ve Tabii Kaynaklar Bakanlığı'nda çalışıyor.

# **EKLER**



## EK A:

$$V(c) = \lambda_{\max} \text{ Kısıtı}$$

Önerme 5.2’de varyans olarak ifade edilen  $V(c)$  maksimum olduğunda  $V(c) = \lambda_{\max} = \lambda_1$  olur.  $\lambda_1$ ’in çarpan sayısı  $m_1 \geq 2$  olmadığında birim küre üzerinde  $V(c) = \lambda_1$  ifadesini sağlayan sadece  $c = V_1(c)$  noktası vardır. Bu durumda Lagrange çarpanları yöntemi uygulanamaz. Ancak  $g_1 > 0$  için  $V_1(c)$ ,  $k \rightarrow -\lambda_1$  de olduğu gibi açıklayıcı değişken  $c \propto (\Lambda + kI)^{-1} g$ ’nin limitini gösterir.

$\lambda_1$ ’in çarpan sayısı  $m_1 \geq 2$  olduğunda,  $\|c\|=1$  olmak üzere sürekli (continuum)  $c$  noktaları,  $V(c) = \lambda_1$  koşulunu sağlar. Buradan  $K^2(c)$ ,  $j \leq m_1$  için  $c_j \propto g_j$  ve  $j > m_1$  için  $c_j = 0$  değerlerini alarak maksimum yapılır. Ayrıca bu nokta  $j \leq m_1$  için  $g_j > 0$  koşulunu sağlayan’ de olduğu gibi  $c \propto (\Lambda + kI)^{-1} g$ ’e yakınsayan gerçek noktadır. ( $V_1(c), \dots, V_m(c)$  tek olmadığı için bu durum her zaman sağlanır). Böylece  $V(c) = \lambda_1$  olduğunda  $c_j$  açıklayıcı değişkeni önermede bahsedildiği üzere ridge açıklayıcı değişkeninin limiti ile orantılıdır.

**EK B:**

Tablo B.1. 15. Buğdaya ilişkin 24 gözlem için ayarlama deneyi sonuçları

Gözlem	Protein(%)	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$	$L_6$
1	9,23	468	123	246	374	386	-11
2	8,01	458	112	236	368	383	-15
3	10,95	457	118	240	359	353	-16
4	11,67	450	115	236	352	340	-15
5	10,41	464	119	243	366	371	-16
6	9,51	499	147	273	404	433	5
7	8,67	463	119	242	370	377	-12
8	7,75	462	115	238	370	353	-13
9	8,05	488	134	258	393	377	-5
10	11,39	483	141	264	384	398	-2
11	9,95	463	120	243	367	378	-13
12	8,25	456	111	233	365	365	-15
13	10,57	512	161	288	415	443	12
14	10,23	518	167	293	421	450	19
15	11,87	552	197	324	448	467	32
16	8,09	497	146	271	407	451	11
17	12,55	592	229	360	484	524	51
18	8,38	501	150	274	406	407	11
19	9,64	483	137	260	385	374	-3
20	11,35	491	147	269	389	391	1
21	9,70	463	121	242	366	353	-13
22	10,75	507	159	285	410	445	13
23	10,75	474	132	255	376	383	-7
24	11,47	496	152	276	396	404	6

Tablo B.2. 16.Buğdaya ilişkin 26 gözlem için öntahmin kümesi

Gözlem	Protein(%)	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$	$L_6$
1	8,66	486	144	266	393	373	26
2	7,90	485	136	260	393	395	6
3	9,27	482	136	260	388	423	-2
4	11,77	443	112	232	346	355	-18
5	9,70	478	134	257	382	390	-5
6	10,46	449	113	233	351	343	-18
7	10,17	461	121	243	366	378	-14
8	11,10	503	155	280	403	414	6
9	12,03	493	146	271	390	878	-3
10	9,43	368	40	158	275	250	-63
11	8,66	462	114	237	367	331	-19
12	14,44	438	109	229	333	326	-28
13	8,50	478	127	252	384	378	-11
14	10,41	405	73	193	311	305	-44
15	9,72	498	146	273	403	415	0
16	11,69	442	106	226	341	303	-28
17	12,19	457	118	240	354	327	-23
18	11,59	439	103	224	339	325	-29
19	8,76	500	146	272	404	398	5
20	8,60	427	85	207	334	319	-36
21	8,54	479	128	253	384	382	-10
22	9,34	444	102	224	350	333	-27
23	10,09	458	118	239	362	355	-16
24	8,72	518	162	290	426	464	16
25	10,87	465	124	247	369	386	-13
26	10,89	457	120	242	363	411	-15