

**FİRMA PERFORMANS ANALİZİNDE MAKİNE ÖĞRENMESİ:
DÜZENLİLEŞTİRİCİ REGRESYON YÖNTEMLERİ**

Doktora Tezi

Önder DORAK

Eskişehir 2023

**FİRMA PERFORMANS ANALİZİNDE MAKİNE ÖĞRENMESİ:
DÜZENLİLEŞTİRİCİ REGRESYON YÖNTEMLERİ**

Önder DORAK

Doktora Tezi

İşletme Anabilim Dalı

Sayısal Yöntemler Bilim Dalı

Danışman: Prof. Dr. Emel ŞIKLAR

Eskişehir

Anadolu Üniversitesi

Sosyal Bilimler Enstitüsü

2023

JÜRİ VE ENSTİTÜ ONAYI

ÖZET

FİRMA PERFORMANS ANALİZİNDE MAKİNE ÖĞRENMESİ: DÜZENLİLEŞTİRİCİ REGRESYON YÖNTEMLERİ

Önder DORAK

İşletme Anabilim Dalı, Sayısal Yöntemler Bilim Dalı

Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü, 2023

Danışman: Prof. Dr. Emel ŞIKLAR

Bu çalışmada, düzenlileştirici regresyon yöntemlerinden ridge, lasso ve elastic-net yöntemleri ile en küçük kareler tahmincisini kullanan sıradan değişken seçim yöntemlerinin gerçek bir işletme problemindeki değişken seçim performansları ile bu yöntemlerce tahmin edilen doğrusal regresyon modellerinin tahmin başarılarının karşılaştırmalı olarak incelenmesi, yöntemlerin güçlü ve zayıf yönlerinin ortaya konulması amaçlanmıştır. Karşılaştırmalı analiz için finansal oranların firma performansı üzerindeki etkisi incelenmiştir. Böylece likidite, borçluluk ve faaliyet oranlarından hangilerinin firma performansını etkiledikleri ve bu finansal oranlar ile firma performansı arasındaki ilişkinin yapısının belirlenmesi amaçlanmıştır. Araştırmaya 2012-2019 yılları arasında Borsa İstanbul'da işlem gören imalat firmaları dahil edilmiş ve firma performansının ölçütü olarak kârlılık oranlarından aktif getiri oranı (ROA) ve öz kaynak getiri oranı (ROE) belirlenmiştir. Analizlerden elde edilen bulgulara göre lasso ve elastic-net yönteminin değişken seçimi konusunda daha kararlı davrandıkları ve en iyi tahmin başarısına sahip modelin lasso regresyon yönteminin belirlediği değişken kümesinin en küçük kareler yöntemi ile tekrar uyumlandırılması ile elde edilen model olduğu sonucuna ulaşılmıştır. Ayrıca bulgulara göre aktif kârlılığı etkileyen en önemli oranlar aktif devir hızı ve kaldıraç oranı iken; öz kaynak kârlılığı üzerinde en çok etkisi olan oranlar duran varlıklar/öz kaynaklar ile banka kredileri/toplam varlıklar oranlarıdır. Sonuçlara göre imalat firmaları performanslarını artırmak için borca olan bağımlılıklarını azaltmalı, stok miktarları ile duran varlıklarını minimum seviyede tutmalıdırlar.

Anahtar Sözcükler: Düzenlileştirici yöntemler, Lasso, Ridge, Firma performansı, Finansal oranlar

ABSTRACT

MACHINE LEARNING IN FIRM PERFORMANCE ANALYSIS: REGULARIZATION METHODS

Önder DORAK

Department of Business Administration, Quantitative Methods Program

Anadolu University, Graduate School of Social Sciences, March 2023

Supervisor: Prof. Dr. Emel ŞIKLAR

In this study, it is aimed to compare linear regression models estimated by regularization methods (ridge, lasso and elastic-net) and ordinary variable selection methods that use ordinary least squares estimator in terms of the variable selection and prediction performances on a real business problem. For the comparative analysis the effect of financial ratios on firm performance is analyzed. Thus, this thesis also seeks to determine which liquidity, debt and efficiency ratios affect the firm performance and to present the mathematical structure of this relationship. Manufacturing firms listed and traded in Borsa İstanbul during the 2012-2019 period are sampled for the analysis and among profitability ratios return on assets (ROA) and return on equity (ROE) are chosen as proxies of firm performance. According to the findings, lasso and elastic-net methods have more stable behavior on variable selection. The model with the highest estimation success is the model whose variables are selected by lasso and then the coefficients are estimated by ordinary least squares. Also, the asset turnover ratio and leverage ratios affect the return on asset the most. On the other hand, total assets/equity and bank credits/total assets ratios are the most influential financial ratios over return on equity. In conclusion, manufacturing firms should decrease their dependency on their debt and keep their stock and long-term assets at minimum in order to enhance their firm performance.

Keywords: Regularization methods, Lasso, Ridge, Firm performance, Financial ratios

TEŐEKKÖR

Öncelikle tezimin hayata geemesinde görüşleri ve tecrübesiyle değerli katkıları sunan ve yol gösteren tez danışmanım Prof. Dr. Emel ŐIKLAR'a teşekkürü bir bor bilirim.

Tezimin birçok aşamasında bilgisi ve uzmanlığıyla yardımlarını esirgemeyen değerli hocam Prof. Dr. Fikret ER'e en derin Őükranlarımı sunarım.

Her zaman yanımda olan ve tez yazım sürecinde istemeden ihmal ettiğim sevgili eşim Ece Nur DORAK'a sonsuz teşekkür ederim.

05/05/2023

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Anadolu Üniversitesi tarafından kullanılan “bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Önder DORAK

İÇİNDEKİLER

	<u>Sayfa</u>
BAŞLIK SAYFASI.....	i
JÜRİ VE ENSTİTÜ ONAYI	ii
ÖZET	iii
ABSTRACT	iv
TEŞEKKÜR.....	v
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	vi
İÇİNDEKİLER.....	vii
ŞEKİLLER DİZİNİ	x
TABLolar DİZİNİ	xiii
SİMGELER VE KISALTMALAR DİZİNİ.....	xv
GİRİŞ.....	1
1. İSTATİSTİKSEL ÖĞRENME	7
1.1. İstatistiksel Model	13
1.2. Yanlılık – Varyans İlişkisi	16
1.3. Yeniden Örnekleme ve Hiperparametre Optimizasyonu.....	21
1.3.1. Geçerlilik kümesi yaklaşımı.....	22
1.3.2. k-katlı çapraz geçerlilik	25
1.4. Optimizasyon Yöntemleri.....	29
1.4.1. Konveks optimizasyon.....	30
1.4.2. Koordinat iniş algoritması.....	32
2. DOĞRUSAL REGRESYON ANALİZİ	34
2.1. Doğrusal Regresyon Yönteminin Varsayımları	37
2.2. En Küçük Kareler Yöntemi	39
2.2.1. En küçük kareler tahmincilerinin özellikleri.....	43
2.2.2. Çoklu bağıntı.....	46

2.2.3.	Belirlilik katsayısı	50
2.2.4.	Doğrusal regresyon analizinde hipotez testleri.....	51
2.3.	Çok Değişkenli Regresyon Modellerinde Değişken Seçimi	54
2.3.1.	Tüm olası alt kümeler yöntemi.....	55
2.3.2.	İleriye doğru seçim yöntemi	57
2.3.3.	Geriye doğru seçim yöntemi	57
3.	DÜZENLİLEŞTİRİCİ YÖNTEMLER	59
3.1.	Bridge Regresyon	60
3.1.1.	Bridge tahmincisinin özellikleri	62
3.1.2.	Bridge tahmincisinin hesaplanması	63
3.1.3.	Bridge tahmincisinin varyansı.....	64
3.1.4.	Ceza ve ayar parametrelerinin seçimi	65
3.2.	Ridge Regresyon.....	66
3.2.1.	Ridge tahmincisi ve özellikleri.....	66
3.2.2.	Ridge tahmincisinin hata kareleri ortalaması	69
3.2.3.	Ridge tahmincisinin hesaplanması.....	71
3.2.4.	Ridge izi ve ceza parametresinin seçimi	73
3.3.	Lasso Regresyon	74
3.3.1.	Lasso tahmincisi ve özellikleri.....	75
3.3.2.	Lasso çözümü	76
3.3.3.	Lasso çözümünün hesaplanması	80
3.3.4.	Serbestlik derecesi	82
3.4.	Elastic Net	82
3.4.1.	Naive elastic net	84
3.4.2.	Elastic net tahmincisi	86
4.	FİRMA PERFORMANS ANALİZİ	89
4.1.	Araştırmanın Evreni ve Örneklemi.....	103

4.2.	Veri Ön Analizi.....	107
4.3.	Aktif Kârlılık Analizi	112
4.3.1.	En küçük kareler yöntemi	112
4.3.2.	Düzenleştirici yöntemler analizi.....	116
4.3.3.	Modellerin karşılaştırılması	125
4.4.	Öz Kaynak Kârlılık Analizi.....	134
4.4.1.	En küçük kareler yöntemi	134
4.4.2.	Düzenleştirici yöntemler analizi.....	138
4.4.3.	Modellerin karşılaştırılması	146
	SONUÇ	157
	KAYNAKÇA	164

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 1.1. Denetimli öğrenme süreci	10
Şekil 1.2. Denetimli öğrenme problemi	11
Şekil 1.3. Öğrenme probleminin bileşenleri	12
Şekil 1.4. Yanlılık-varyans ilişkisi	20
Şekil 1.5. Model karmaşıklığının eğitim ve test hatasına etkisi	21
Şekil 1.6. İkili geçerlilik kümesi yaklaşımı	23
Şekil 1.7. Üçlü geçerlilik kümesi yaklaşımı	24
Şekil 1.8. 5-katlı çapraz geçerlilik yöntemi ile model değerlendirme	26
Şekil 1.9. k-katlı çapraz geçerlilik ile model seçimi	28
Şekil 3.1. $t=1$ durumunda farklı q değerleri için kısıt bölgeleri	61
Şekil 3.2. Newton-Raphson yönteminin bridge regresyona uygulanması	64
Şekil 3.3. Ridge regresyon ve yanlılık-varyans ilişkisi	70
Şekil 3.4. Ridge ve lasso tahmincilerinin karşılaştırılması	78
Şekil 3.5. Lasso ve ridge tahminlerinin gösterimi	79
Şekil 3.6. Elastic Net tahmininin gösterimi	84
Şekil 3.7. Ortogonal durumda naive elastic net	86
Şekil 4.1. Firma performansının alanları	93
Şekil 4.2. İmalat firmalarının saçılım grafiği	108
Şekil 4.3. Net satışlar ile toplam varlıkların yıllara ve sektörlere göre karşılaştırılması	108
Şekil 4.4. Korelasyon grafiği	109
Şekil 4.5. Aktif kârlılık ve öz kaynak kârlılığı yoğunluk grafikleri	111
Şekil 4.6. EKK aktif kârlılık modelinin tanı grafikleri	114

Şekil 4.7. Lasso ve ridge regresyon yöntemlerinin ÇG ile optimum ceza parametre değerleri	117
Şekil 4.8. Çapraz geçerlilik yöntemi ile elastic-net hiperparametrelerin değişim grafiği	117
Şekil 4.9. Düzenleştirici yöntemlerin katsayı tahminlerinin çubuk grafikleri ...	120
Şekil 4.10. Lasso katsayıların değişimi – Lasso açıklanan değişim	121
Şekil 4.11. Lasso regresyon bağımsız değişkenlerin önem sıralaması	121
Şekil 4.12. Ridge katsayıların değişimi– Ridge açıklanan değişim	122
Şekil 4.13. Ridge regresyon değişkenlerin önem sıralaması	122
Şekil 4.14. Elastic-net katsayıların değişimi– Elastic-net açıklanan değişim	123
Şekil 4.15. Elastic-net regresyon değişkenlerin önem sıralaması	123
Şekil 4.16. Geriye doğru seçim yöntemi ÇG HKO sonuçlarının düzenleştirici yöntemler ÇG HKO sonuçları ile karşılaştırılması	127
Şekil 4.17. Lasso-EKK ÇG HKO sonuçlarının değişken seçim yöntemleri ile karşılaştırılması	128
Şekil 4.18. Belirlilik katsayılarının kutu grafikleri	129
Şekil 4.19. EKK öz kaynak kârlılığı modelinin tanı grafikleri	136
Şekil 4.20. Lasso ve ridge regresyon yöntemlerinin ÇG ile optimum ceza parametre değerleri	138
Şekil 4.21. Çapraz geçerlilik yöntemi ile elastic-net hiperparametrelerin değişim grafiği	139
Şekil 4.22. Düzenleştirici yöntemlerin katsayı tahminlerinin çubuk grafikleri ...	141
Şekil 4.23. Lasso katsayıların değişimi – Lasso açıklanan değişim	142
Şekil 4.24. Lasso regresyon bağımsız değişkenlerinin önem sıralaması	142
Şekil 4.25. Ridge regresyon katsayıların değişimi – Ridge açıklanan değişim	143
Şekil 4.26. Ridge regresyon bağımsız değişkenlerinin önem sıralaması	144
Şekil 4.27. Elastic-net katsayıların değişimi – Elastic-net açıklanan değişim	144

Şekil 4.28. Elastic-net regresyon bağımsız değişkenlerinin önem sıralaması	145
Şekil 4.29. Tüm olası alt kümeler ÇG HKO sonuçlarının düzenleştirici yöntemler ÇG HKO sonuçları ile karşılaştırılması	148
Şekil 4.30. Lasso-EKK ÇG HKO sonuçlarının değişken seçim yöntemleri ile karşılaştırılması	149
Şekil 4.31. Belirlilik katsayılarının kutu grafikleri	149

TABLULAR DİZİNİ

	<u>Sayfa</u>
Tablo 4.1. Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti	98
Tablo 4.2. (Devam) Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti	99
Tablo 4.3. (Devam) Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti	100
Tablo 4.4. (Devam) Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti	101
Tablo 4.5. (Devam) Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti	102
Tablo 4.6. Likidite oranları	103
Tablo 4.7. Finansal yapı oranları	104
Tablo 4.8. Faaliyet oranları	105
Tablo 4.9. Kârlılık oranları	105
Tablo 4.10. İmalat firmalarının alt sektörlerle ve yıllara göre dağılımı	106
Tablo 4.11. Finansal oranların tanımlayıcı istatistikleri	110
Tablo 4.12. En küçük kareler aktif kârlılık doğrusal regresyon modeli	113
Tablo 4.13. Tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim doğrusal regresyon modellerinin katsayı tahminleri	115
Tablo 4.14. Aktif kârlılık EKK modellerinin değerlendirilmesi	116
Tablo 4.15. Lasso, elastic-net ve ridge regresyon modellerinin katsayı tahminleri	119
Tablo 4.16. Lasso-EKK ve ElasticNet-EKK katsayı tahminleri	124
Tablo 4.17. Aktif kârlılık düzenleştirici modellerin değerlendirilmesi	125
Tablo 4.18. Doğrusal regresyon modellerin eğitim ve test hataları	126

Tablo 4.19. 80/20 eğitim/test ayrımı sonrası aktif kârlılık modellerinin katsayı tahminleri	130
Tablo 4.20. 60/40 eğitim/test ayrımı sonrası aktif kârlılık modellerinin katsayı tahminleri	131
Tablo 4.21. 80/20, 70/30 ve 60/40 eğitim/test ayrımlarına ilişkin değerlendirme sonuçları	132
Tablo 4.22. En küçük kareler öz kaynak kârlılığı doğrusal regresyon modeli	135
Tablo 4.23. Tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim doğrusal regresyon modellerinin katsayı tahminleri	137
Tablo 4.24. Öz kaynak kârlılığı EKK modellerinin değerlendirilmesi	138
Tablo 4.25. Lasso, elastic-net ve ridge regresyon modellerinin katsayı tahminleri	140
Tablo 4.26. Lasso-EKK modelinin katsayı tahminleri	145
Tablo 4.27. Düzenleştirici yöntemlerin değerlendirilmesi	146
Tablo 4.28. Doğrusal regresyon modellerin eğitim ve test hataları	147
Tablo 4.29. 80/20 eğitim/test ayrımı sonrası öz kaynak kârlılığı modellerinin katsayı tahminleri	151
Tablo 4.30. 60/40 eğitim/test ayrımı sonrası öz kaynak kârlılığı modellerinin katsayı tahminleri	153
Tablo 4.31. 80/20, 70/30 ve 60/40 eğitim/test ayrımlarına ilişkin değerlendirme sonuçları	154

SİMGELER VE KISALTMALAR DİZİNİ

ÇG	: Çapraz Geçerlilik
EKK	: En Küçük Kareler
Enet	: Elastic-Net
HKO	: Hata Kareleri Ortalaması
HKT	: Hata Kareleri Toplamı
GKO	: Genel Kareler Ortalaması
GKT	: Genel Kareler Toplamı
KKT	: Karush-Kuhn-Tucker
OMH	: Ortalama Mutlak Hata
RKO	: Regresyon Kareleri Ortalaması
RKT	: Regresyon Kareleri Toplamı
TCMB	: Türkiye Cumhuriyet Merkez Bankası
TOAK	: Tüm Olası Alt Kümeler

GİRİŞ

Fen ve beşerî alanlarda incelemeye tabi fenomenlerin özü itibariyle sahip oldukları belirsizliğin açıklanması ve anlamlandırılmasında istatistik disiplinine başvurulması artık bilimsel bir normdur. Çeşitli yaklaşımlar ve uygulamalar istatistiğin tanımına ilişkin farklı manifestolar sunsalar da istatistik alanı belirsizlikle başa çıkmak, veriyi anlamlandırmak, karar vermeye yardımcı olmak, geleceği tahmin etmek, bilinmeyen hakkında çıkarımda bulunmak gibi birçok anlama içkindir. İlk yıllarda ağırlıklı olarak fen ve sanayi alanındaki deneylerde başvuru istatistik alanının günümüzde her anımızda durmaksızın üretilen veri ile olan bağlantısı çok daha görünür hale gelmiştir. Günümüzde istatistik bilim dalı verideki trendleri, örüntüleri, ilişkileri ortaya çıkarmak ve veriyi anlamak ile ilişkilendirilen bir disiplindir.

Gelişen bilgi ve veri saklama teknolojileri ile birlikte her alanda benzeri görülmemiş miktarda veri üretilmeye başlanmıştır. İşletme alanı da bundan muaf değildir. Perakende işletmelerinde gerçekleşen her bir satışta, dünyanın her yerinde her saniye gerçekleştirilen finansal işlemlerde, bir alışveriş sitesinde gezinen müşterilerin sayfadaki öğelere her tıklayışında veri ortaya çıkmakta ve kaydedilmektedir. Hızına yetişilemeyen bu veri artışı istatistik problemlerinin yapısını ve istatistiğe olan yaklaşımı çarpıcı biçimde dönüştürmüştür.

Ortaya çıkan büyük ve karmaşık verinin anlaşılmasına yardımcı olabilecek yöntemlere duyulan ihtiyaç istatistik biliminde bir devrime yol açmış ve istatistik problemlerindeki hesaplama zorluğunun üstesinden gelmek amacıyla istatistik ve bilgisayar bilimlerini bir araya getirmiştir. İstatistik esas olarak veriden hangi çıkarımların elde edileceği ile ilgilenirken; çok boyutlu ve karmaşık veri analizi ile birlikte bir veri kümesinden maksimum enformasyonu daha kolay, doğru ve etkin bir biçimde elde edebilecek algoritmaların neler olduğu sorusu ortaya atılmıştır. İşte istatistik ve bilgisayar biliminin kesişiminde yer alan makine öğrenmesi, veri analizinin gereksinimlerini giderebilecek araçlar kümesini sunan bir disiplini ifade etmektedir. Makine öğrenmesi olarak literatürde yer edinen bu istatistiksel hesaplama sistemleri, belirli görevlerdeki performanslarını deneyimlerle geliştirir ve böylece veriden öğrenirler.

Aralarında bir takım kavramsal farklılıklar bulunan yapay zeka, derin öğrenme, örüntü tanıma, veri madenciliği gibi terimler de makine öğrenmesi alanında yaygın olarak kullanılmaktadır. Görece yeni bir kavram olan istatistiksel öğrenme ise Hastie vd. (2009)

tarafından ortaya atılmıştır. Hastie vd. (2009)'ye göre istatistiksel öğrenme, karmaşık veri kümelerini anlamak ve modellemek amacıyla geliştirilmiş birtakım araçlar kümesini ifade etmektedir.

İstatistiksel öğrenme epeyce yeni bir kavram olsa da bu alanda yer alan kavramların birçoğu modern istatistiğin başlangıcına dayandırılmaktadır. Doğrusal regresyon, doğrusal diskriminant analizi, lojistik regresyon yöntemlerini kapsayan istatistiksel öğrenme yöntemlerini 1970'lerde Nelder ve Wedderburn genelleştirilmiş doğrusal modeller olarak tanımlamıştır. Bilgisayarların hesaplama kapasitelerinin artmasıyla birlikte doğrusal olmayan yöntemler, sınıflandırma, regresyon ağaçları, genelleştirilmiş toplamsal modellerin tamamı bu sınıfa dahil edilmiş ve nihai olarak denetimli ve denetimsiz modellemeye ve tahmine odaklı istatistiksel öğrenme alanı istatistiğin bir alt alanı olarak kabul edilmiştir (James, Witten, Hastie ve Tibshirani, 2013, s. 6)

İstatistiksel modeller, belirsizliğe sahip karmaşık olayların veriden yola çıkarak basitleştirilerek açıklanması ve yorumlanmasını sağlarlar. İstatistiksel modeller arasında en eski ve en yaygın kullanılanlardan birisi de doğrusal regresyon modelidir. Her ne kadar regresyon analizinin temelleri bilgisayar öncesi dönemlere uzanıyor olsa da günümüzde hala doğrusal regresyon modellerinin kullanılması için birçok neden vardır. Doğrusal modeller her şeyden önce girdilerin çıktılar üzerindeki etkilerini karar vericiler için yeterli ve yorumlanabilir biçimde açıklama kabiliyetine sahiplerdir. Ayrıca küçük hacimli örneklemelerin kullanıldığı araştırma problemlerinde doğrusal olmayan ve/veya parametrik olmayan yöntemlerden daha başarılı tahminler gerçekleştirebilirler.

Bir doğrusal regresyon modeli, bir bağımlı (açıklanan) değişken ile bağımsız (açıklayıcı) değişkenler arasındaki ilişkiyi matematiksel olarak modellemeyi amaçlar. Böylece bağımsız değişkenler ile bağımlı değişken arasında bir ilişkinin var olup olmadığı belirlenir; eğer anlamlı bir ilişki varsa ilişkinin matematiksel yapısı ortaya konur. Böylece doğrusal regresyon modeli değişkenler arasındaki ilişkinin yorumlanmasını ve verilen bir veri kümesinden yararlanarak bağımlı değişken değerini tahmin etmeyi sağlar. Modelin yorumlanması ve bağımlı değişkenin tahmin edilmesi modeldeki parametrelerin tahmin edilmesi sayesinde gerçekleştirilir. Katsayıların tahmin edilmesi için en eski yöntem en küçük kareler yöntemidir. En küçük kareler tahmincisi seçilmesi mümkün tüm yansız tahminciler arasından minimum varyanslı tahmincidir. Bu yüzden en iyi yansız tahminci olarak tanımlanır. Basitliği ve elde edilen modellerin kolay yorumlanabilmesi sebebiyle sıklıkla tercih edilse de ileride açıklanacağı üzere birçok

varsayıma sahip olduğundan dolayı değişen varyanslılık ve çoklu bağıntı gibi sorunlar mevcut olduğunda yöntemin varsayımları ihlal edilir ve elde edilen sonuçlar geçerli kabul edilmez. Bununla birlikte örneklem hacmi bağımsız değişken sayısından fazla büyük değilse değişkenlik artar ve model başarısız tahminlerde bulunur.

İstatistiksel öğrenme alanında karşılaştığımız önemli problemlerden bir tanesi de çok boyutlu bir istatistik problemi için hangi boyutların (değişkenlerin) önemli olduğuna karar vermektir. Bu problem literatürde değişken seçimi problemi olarak adlandırılır. Hangi girdi değişkenlerinin gözlemlerin sınıflandırılmasında, kümelenmesinde veya bir çıktı değişkeni üzerinde etkili olduğuna karar verilmesinde geleneksel olarak araştırmacının uzmanlık bilgisine veya literatürdeki çalışmalardan yararlanılmaktadır. Bu yaklaşım çok boyutlu bir istatistik probleminin çözümüne kaçınılmaz olarak öznel bir boyut kazandırmaktadır. Ne var ki sınıflandırma ve tahmin modellerinde yer alacak bağımsız değişkenlerin belirlenmesi istatistik alanında yeni araştırılan bir problem değildir.

En küçük kareler yöntemi doğrusal regresyon analizinde sıklıkla kullanılsa da değişken seçimini gerçekleştiremez. Modellerde yer alacak değişkenlerin belirlenmesinde sıklıkla adimsal regresyon ve tüm olası alt kümeler yöntemlerine istatistik alanında uzun yıllardır başvurulmaktadır. Geleneksel olarak adlandırabileceğimiz bu yöntemler çeşitli ölçütlere dayalı olarak yinelemeli bir biçimde değişken seçimini gerçekleştirir ve elde ettikleri nihai bağımsız değişken kümesini en küçük kareler yöntemi ile uyumlandırarak nihai doğrusal regresyon modelini tahmin eder. Bu ölçütlerden sıklıkla kullanılanlar arasında düzeltilmiş R^2 , Akaike bilgi kriteri, Bayes bilgi kriteri ve Mallow'un C_p kriteri sayılabilir (James ve diğerleri, 2013, s. 212). Bununla birlikte çapraz geçerlilik ve bootstrap gibi yeniden örnekleme yöntemleri de bu yöntemler tarafından kullanılmaktadır (Raschka, 2018) (Genç, 2020).

Geleneksel seçim yöntemlerine ve en küçük kareler regresyonuna alternatif olarak katsayıları bir ceza parametresi ile kısıtlayan ve düzenleyen, literatürde cezalı veya daraltıcı olarak da adlandırılan düzenleştirici yöntemler de yaygın kullanıma sahiptir. Önde gelen düzenleştirici yöntemler ridge, lasso ve elastic-net yöntemleridir.

Hoerl ve Kennard (1970a) çalışmalarında çoklu bağıntı sorununun üstesinden geldiğini öne sürdükleri ridge regresyon yöntemini önermiştir. Bir ceza parametresi yardımıyla katsayı tahminlerini sıfıra yaklaştırarak yanlı ancak daha düşük varyanslı

tahminler üreten ridge regresyon en küçük kareler yöntemine göre daha başarılı tahminler elde etmektedir. Ancak ridge regresyon kullandığı ℓ_2 norm ceza parametresi uyarınca katsayı değerlerini sıfıra yaklaştırsa da hiçbir zaman sıfıra eşitleyemediği için eş zamanlı olarak değişken seçimini gerçekleştirememektedir.

Tibshirani (1996) ise çalışmasında ℓ_1 norm ceza parametresinin kullanıldığı lasso regresyon yöntemini önermiştir. Ridge regresyona benzer bir biçimde katsayı tahminlerini bir ceza parametresine bağlı olarak sıfıra yaklaştıran lasso regresyon yanı sıra ancak daha düşük varyanslı katsayı tahminleri gerçekleştirerek modelin tahmin başarısını artırmayı hedefler. Ridge regresyondan farklı olarak ℓ_1 norm ceza parametresini kullanan lasso regresyonda katsayıların tahmin değerleri sıfıra eşitlenebilmektedir. Böylece lasso regresyon eş zamanlı olarak modelin katsayılarını tahmin etmekte ve değişken seçimini gerçekleştirmektedir. Son derece popüler bir yöntem haline gelen ve çeşitli uzantıları da yıllar içinde önerilen lasso regresyon gözlem sayısından daha fazla değişkeni seçememekte ve çoklu bağıntı durumunda ridge regresyondan daha başarısız tahminler vermektedir.

Zou ve Hastie (2003) ise çalışmalarında ℓ_1 ve ℓ_2 uzaklıklarına ödünleşimli bir biçimde yer verdikleri bir ceza parametresi kullanmışlar ve böylece ridge ile lasso regresyonun güçlü yönlerini bir araya getirdikleri elastic-net yöntemini önermişlerdir. Bu sayede hem değişken seçimini hem de çoklu bağıntı durumunda başarılı tahminler elde etmeyi amaçlamışlardır.

Bu çalışmada tüm olası alt kümeler, ileriye doğru seçim, geriye doğru seçim, ridge regresyon, lasso regresyon ve elastic-net yöntemleri ile tahmin edilen doğrusal regresyon modellerinin gerçek bir veri kümesinden yararlandığı bir işletme problemindeki performanslarının, güçlü ve zayıf yönlerinin karşılaştırılması olarak analiz edilmesi amaçlanmıştır. Bu amaçla işletme ve finans literatüründe yaygın olarak araştırılan finansal oranların firma performansı üzerindeki etkisinin incelenmesi problemi seçilmiştir. Böylece firma performansını etkileyen finansal oranların belirlenmesi ve belirlenen finansal oran ile firma performansı arasında mevcut olan ilişkinin matematiksel ilişkinin yapısının ortaya konulması amaçlanmıştır.

İşletmenin amacının kârını maksimize etmektir düşüncesinden hareketle firma performansının bir ölçütü olarak sıklıkla kârlılık oranlarından yararlanılmaktadır. Bu çalışmada da firma performansının ölçütü olarak birer kârlılık oranı olan aktif getiri oranı

(ROA) ve öz kaynak getiri oranı (ROE) seçilmiştir. Aktif getiri oranı firmanın varlıklarını ne ölçüde kâra dönüştürdüğünü gösterirken; öz kaynak getiri oranı ise firmanın öz kaynaklarını ne ölçüde kâra dönüştürdüğünü gösterir. Bu iki finansal oran araştırmada bağımlı değişkenler olarak kabul edilmiştir. 4. bölümde yukarıda kısaca değinilen doğrusal regresyon yöntemleri ile incelenecektir. Araştırmada bağımsız değişkenler olarak 33 adet finansal oran belirlenmiştir. Söz konusu tüm oranların hesaplanmasında işletmelerin halka açık finansal raporlarında beyan edilen finansal kalemler kullanılmış ve TCMB'nin önerdiği finansal formüllerden yararlanılmıştır. Araştırmaya Borsa İstanbul'da 2012-2019 yılları arasında imalat sektöründe faaliyet gösteren işletmeler dahil edilmiştir.

Yapılan birçok ampirik çalışmada düzenlileştirici yöntemler çeşitli simülasyonlar ya da yazılımlarda yer alan açık veri kümeleri üzerinde karşılaştırılmalı olarak veya ayrı ayrı gerçek işletme problemlerinde kullanılsa da bu yöntemlerin gerçek işletme problemlerinde karşılaştırmalı analizleri literatürde fazla yer almamaktadır. Bununla birlikte firma performansının belirleyicileri olan finansal oranlar genellikle önceki çalışmalara ve uzmanların öznel yaklaşımlarına bağlı olarak belirlenmektedir. Dolayısıyla firma performansını etkileyen finansal oranların eş zamanlı, objektif olarak seçilmesi ve modelde yer alan finansal oranların katsayılarının belirlenmesi bildiğimiz kadarıyla Türkiye'deki alanyazında ilk kez gerçekleştirilmektedir. Tüm bu bulguların Türkiye'deki finans alanına da katkı sağlayacağı düşünülmektedir. Bu nedenlerden dolayı bu çalışma alandaki mevcut araştırma eksikliğinin giderilmesine yardımcı olacak ve gelecekteki araştırmacılar ile karar vericilere yardımcı olacaktır.

Çalışmanın 1. bölümünde istatistiksel öğrenme kavramı ve veriden öğrenme kavramına ilişkin temel paradigmlar tanıtılacaktır. Bununla birlikte çalışmanın ilerleyen kısımlarında sıklıkla başvurulan bir takım temel istatistik ve istatistiksel öğrenme kavramları açıklanmıştır.

2. bölümde doğrusal regresyon analizinin varsayımları, en küçük kareler yönteminin özellikleri ve çok değişkenli regresyon modellerinde geleneksel olarak değişken seçimi amacıyla kullanılan tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemleri kısaca açıklanmıştır.

3. Bölümde düzenlileştirici yöntemlerin temel prensipleri ve bu yöntemlerden bridge, ridge, lasso ve elastic net yöntemlerinin özellikleri izah edilmiştir.

4. Bölümde öncelikle firma performansının tanımı yapılacak ve finansal oranların kullanım alanları tanıtılmıştır. Sonrasında finansal oranlar yardımıyla firma performansının tahminine ilişkin geçmişte yapılan çalışmaların bir özeti sunulmuştur. Bunu takiben araştırmada kullanılan finansal oranlar formülleri ile birlikte sunulmuş ve verinin ön analizi yapılmıştır. Gerçek bir veriden yararlanılarak finansal oranların firma performansı üzerindeki etkilerinin belirlenmesi problemi özelinde, ele alınan regresyon yöntemlerinin karşılaştırmalı analizleri, bu yöntemlerin tahmin performansları, güçlü ve zayıf yönleri değerlendirilmiş ve tartışılmıştır.

1. İSTATİSTİKSEL ÖĞRENME

Geçtiğimiz yüzyılda öncelikle bilgisayarların ortaya çıkışı ve zamanla bilgisayarların veri depolama ve hesap yapma kapasitelerinin artmasıyla birlikte istatistik problemlerinin yapısı da değişime uğradı. Göreceli olarak küçük örneklemelere ilişkin problemlerin çözülmesi ile ilgilenen istatistik alanı, artan veri miktarı ile birlikte büyük hacimli ve çok yüksek boyutlu problemlerin çözülmesi sorunu ile karşı karşıya geldi. İşletme, iktisat, biyoloji, tıp ve daha birçok beşerî ve fen alanlarında çok büyük miktarda üretilen veriyi işleyerek, verideki önemli örüntüleri, trendleri, ilişkileri ortaya çıkarmak için çeşitli algoritmalar geliştirilmiş ve bu süreç veriden öğrenme olarak adlandırılmıştır. Böylece literatüre veri madenciliği, makine öğrenmesi ve derin öğrenme gibi kavramlar dahil olmuştur.

Söz konusu öğrenme kavramı şöyle açıklanabilir: “Bir bilgisayar programı belirli bir T görevinin P performansını E deneyiminden yararlanarak geliştiriyorsa bu program öğreniyor denir.” (Mitchell, 1997, s. 2). Bahsi geçen T görevi sınıflandırma, kümeleme, regresyon, anomali saptama gibi çok çeşitli görev türlerinden birisi olabilir. P performansı ise genellikle yapılan doğru tahminlerin yanlış tahminlere oranı olarak hesaplanan ancak ölçümü problemin ve görevin yapısına bağlı olarak değişen bir başarı ölçütüdür. E deneyimi ise bir veri kümesi üzerinde bir algoritma yardımıyla yinelemeli olarak gerçekleştirilen işlemler bütünü olarak tanımlanabilir (Goodfellow, Bengio ve Courville, 2016, s. 99).

Rosenblatt’ın (1958) önerdiği perceptron, öğrenme süreçlerinin matematiksel analizinin başlangıcı ve ilk öğrenme makinesi olarak kabul edilir. Perceptronların yaygın olarak tanınması sonucunda gerçek hayat problemlerinin çözülmesi amacıyla yapay sinir ağları, karar ağaçları gibi çok sayıda öğrenme algoritması ve uzman sistemlerin geliştirilmesi ile istatistiksel öğrenme teorisinin temelleri atılmıştır (Vapnik, 2000, s. 7).

Öğrenme teorisinin yapısı uygulamalı problemlerin çözümü için geliştirilmiş birçok teoriden ayrılmaktadır. Bir uygulamalı problemin çözümünde şu üç temel adım izlenir:

1. Problemin matematiksel olarak ifade edilmesi
2. Problemin çözümünü verecek genel bir ilkenin formülleştirilmesi
3. Bu genel ilkeye dayanarak bir algoritmanın geliştirilmesi

Bu süreçte en fazla uğraş üçüncü adımda verilir. Ancak öğrenme süreçlerinin uygulandığı problemlerde bir hayli sayıda algoritmanın halihazırda geliştirilmiş olması ancak buna

karşın öğrenme sürecini genel çerçevesiyle açıklayacak bir matematiksel ifadenin ve genel bir ilkenin yokluğu araştırmacıları öğrenme teorisini geliştirmeye itmiştir (V Vapnik, 1991). Buna göre, geliştirilen tüm bu makinelerin ortak özelliklerini ve hepsinin uyguladığı genel bir tümevarımsal çıkarım ilkesinin var olup olmadığını inceleyen istatistiksel öğrenme teorisinin temel amacı istatistiksel çıkarım problemi için genel bir çerçeve sunmaktır (Bousquet, Boucheron ve Lugosi, 2004, s. 169).

Bilgisayarların hesaplama kapasitelerinin artması ve veri tabanı teknolojilerinin gelişmesi ile birlikte veriden yararlı bir model öğrenmek için çok sayıda yöntem geliştirilmiştir. Bu yöntemler arasında çok değişkenli regresyon, sınıflandırma gibi geleneksel istatistiksel yöntemler; mühendislikte kullanılan örüntü tanıma, sinyal işleme; yapay zekâ, sinir ağları, bulanık mantık ve genetik algoritmalar gibi bilgisayar bilimi alanında kullanılan yöntemler de yer almaktadır. Tüm bu öğrenme yaklaşımları üç temel paradigmaya dayanmaktadır (Cherkassky ve Mulier, 2007, s. 15):

1. Parametrik Tahmin
2. Model Tanımlama/Fonksiyon Yakınsama
3. Risk Minimizasyonu

Parametrik tahmin yaklaşımında bilinmeyen bağımlılığın şeklinin bilindiği varsayılır ve yaklaşımın amacı mevcut veriden yararlanarak bu varsayılan şekli tanımlayan modelin parametrelerini tahmin etmektir. 1970'lerden itibaren gelişen teknoloji sayesinde veri güdümlü esnek modellerin geliştirilmesi ile birlikte fonksiyon yakınsama yaklaşımı geliştirilmiştir. Aslında fonksiyon yakınsama yaklaşımı birinci yöntemdeki parametrik form ile ilgili bazı katı varsayımların esnetilerek parametrik tahmin yaklaşımının genişletilmiş bir halidir. Bilinmeyen model esnek parametrize etme kuralları ile belirlenir. Amaç birinci yöntemde olduğu gibi gerçek modelin doğru bir tahminini gerçekleştirmektir. Üçüncü yöntem olan risk minimizasyonunun teorik çerçevesi ise Vapnik ve Chervonenkis (1998; 2000) tarafından geliştirilmiştir ve literatürde VC teorisi ya da istatistiksel öğrenme teorisi olarak da adlandırılmaktadır. Bu çerçevede amaç tahmin riskini en aza indiren ve böylece başka örneklerde tahmin başarıları (genelleme kabiliyeti) en yüksek olan modelleri belirlemektir (Cherkassky ve Ma, 2009, s. 959).

Bahsi geçen bu öğrenme paradigmaları kendi kavramlarına ve matematiksel teorilere sahip olsalar da öğrenme algoritmaları uygulanırken bu yaklaşımlar arasındaki

farklılıklar silikleşmektedir. Bunun ilk nedeni birçok öğrenme algoritması farklı paradigmlar altında açıklanabilmektedir. Örneğin en küçük kareler yaklaşımı parametrik tahmin yaklaşımında yer almakla birlikte, risk minimizasyonu yaklaşımı ile de ifade edilebilir. İkinci neden ise mühendislik, biyoloji gibi alanlardaki çeşitli problemlerin çözümü için önce öğrenme algoritmalarının geliştirildikten sonra teorik yapısının oluşturulmasıdır. Bu nedenle belirli öğrenme algoritmaları için farklı paradigmlar altında teorik çerçeveler geliştirilmiştir (Cherkassky ve Ma, 2009, s. 959) .

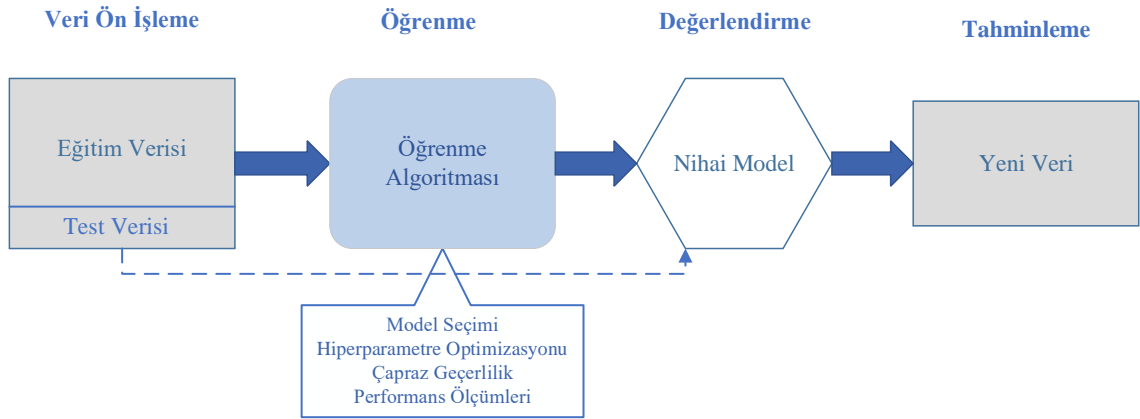
Hangi yaklaşımdan yararlanırsa yararlınsın veri güdümlü tüm öğrenme yöntemlerinin amacı veriden yola çıkarak faydalı modeller elde etmektir. Bir modelin faydası üç kritere bağlı olarak ölçülür. Bunlardan ilki tahmin doğruluğudur. Mevcut veriye dayalı olarak oluşturulan modelin başka veri kümelerinde de yüksek doğruluğa sahip tahminler gerçekleştirmesi istenir. İkinci kriter ise gerçek modelin doğru bir tahminini elde etmektir. Gerçeğe ne kadar yakın bir model tahmin edilirse, elde edilen modellerle başka veri kümelerinde de o kadar başarılı tahminler gerçekleştirileceği için ilk iki kriter birbirleriyle ilişkilidir. Üçüncü fayda kriteri ise modelin açıklayıcı olmasıdır. Veri kümesindeki ilişkilerin anlaşılır ve kolayca yorumlanabilir modeller olması istenir. Üç kriterin tamamını aynı anda sağlamak her zaman mümkün olmayabilir. İlk iki fayda kriterini esas alan yöntemler tahminleyici öğrenme yöntemleri, üçüncü kriteri esas alan yöntemler ise tanımlayıcı öğrenme yöntemleri olarak adlandırılır. Öğrenme literatüründe tahminleyici yöntemler aynı zamanda denetimli öğrenme yöntemleri, tanımlayıcı yöntemler ise denetimsiz öğrenme yöntemleri olarak da adlandırılırlar (Cherkassky ve Mulier, 2007, s. 16).

Bu çalışmanın konusu olan regresyon analizi bir denetimli öğrenme yöntemi olduğu için izleyen kesimde yalnızca fonksiyon yakınsama paradigmasına dayalı olarak elde edilen denetimli/tahminleyici öğrenme yaklaşımı ele alınacaktır. Denetimli öğrenmede bir x girdi vektörü ile y yanıt vektörü arasındaki ilişkinin matematiksel yapısı sonlu sayıdaki bir veri kümesinden yararlanılarak bir fonksiyon yardımıyla açıklanmaya çalışılır. Sınıflandırma ve regresyon yöntemleri bu grupta yer almaktadır. Eğitim kümesinde yanıt değerleri bilindiği için bu öğrenme türü denetimli olarak adlandırılır. Denetimsiz öğrenmede ise girdi değerleri öğrenme sistemine sunulur ancak öğrenme sürecinde çıktı değerleri bilinmez. Denetimsiz öğrenmenin amacı x ve y gibi iki vektör arasındaki ilişkinin matematiksel yapısını açıklamaktan ziyade, eğitim kümesindeki

girdilerin olasılık dağılımını yaklaşık olarak belirlemek ya da veride yer alan bilinmeyen yapıları ortaya çıkarmaktır (Cherkassky ve Mulier, 2007, s. 3).

Denetimli öğrenme süreci Şekil 1.1'de gösterildiği üzere birbirini izleyen dört adımdan oluşur:

1. Veri ön işleme
2. Öğrenme
3. Değerlendirme
4. Tahminleme

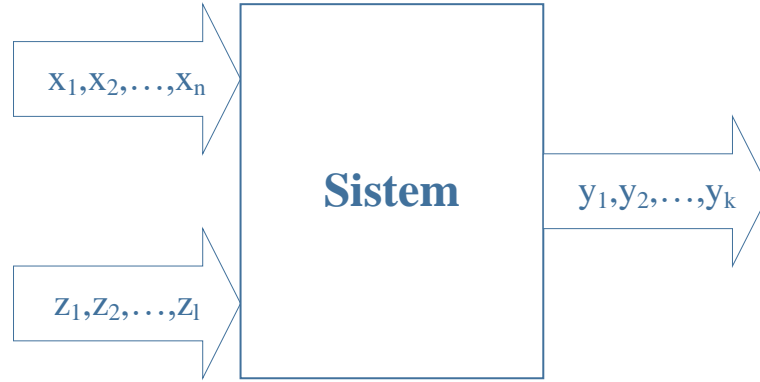


Şekil 1.1. Denetimli öğrenme süreci (Raschka, 2015)

Veri ön işleme adımında ham veri öğrenme algoritmasının kullanımına sunulmak üzere hazır hale getirilir. Veri ön işleme süreci entegrasyon, temizleme, dönüştürme ve indirgeme adımlarından oluşur (Akpınar, 2014, s. 88). Bu adımda toplanan gözlemler sırasıyla öğrenme ve değerlendirme adımlarında kullanılmak üzere eğitim kümesi ve test kümesi olarak iki kısma ayrılır. Öğrenme aşamasında eğitim kümesinden yararlanılarak model seçimi ve hiperparametre optimizasyonu gerçekleştirilerek nihai modele ulaşılar. Daha sonra değerlendirme aşamasında nihai model test verisinden yararlanılarak modelin başarısı değerlendirilir. Son aşama olan tahminleme aşamasında ise nihai model yeni tahminlerin gerçekleştirilmesinde kullanılır (Raschka, 2015, s. 12).

Denetimli öğrenmede araştırılan problem değişken olarak adlandırılan çok sayıdaki ölçülebilen nicelikler ile nitelendirilir. Değişkenler bağımsız değişkenler ve bağımlı değişkenler olmak üzere ikiye ayrılırlar. Bağımsız değişken aynı zamanda istatistikte açıklayıcı değişken ve makine öğrenmesinde de girdi değişken olarak da adlandırılır. Bağımlı değişken ise istatistik alanında açıklanan değişken veya yanıt değişkeni, makine öğrenmesinde ise çıktı değişkeni biçiminde isimlendirilir. Tahminleyici öğrenme

yaklaşımında amaç girdi ve çıktı değişkenleri arasında var olduğu düşünülen matematiksel ilişkiyi veriden yola çıkarak belirlemek ve daha sonrasında belirlenen bu ilişkiden yararlanarak yalnızca girdi değişkenlerinin değerleri verildiğinde bunlara karşılık gelen çıktı değerlerini tahmin etmektir (Cherkassky, Friedman ve Wechsler, 2012, s. 1). Denetimli öğrenme probleminin şeması Şekil 1.2.'de gösterilmiştir.

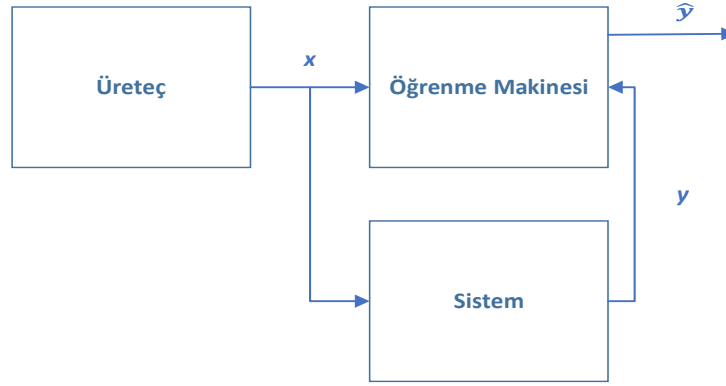


Şekil 1.2. Denetimli öğrenme problemi (Cherkassky ve diğerleri, 2012)

Şekil 1.2.'de görülebileceği üzere girdi ve çıktı değişkenleri arasındaki bu bilinmeyen yapı sistem olarak adlandırılır. Sistemin sol tarafında girdi değişkenleri bulunur. $X = \{x_1, x_2, \dots, x_n\}$ girdileri ölçülebilen ve gözlemlenebilen girdi değişkenlerini ifade ederken $Z = \{z_1, z_2, \dots, z_l\}$ girdileri ise $Y = \{y_1, y_2, \dots, y_k\}$ çıktıları etkileyen ancak gözlemlenemeyen ya da ölçülemeyen değişkenlerdir (Cherkassky, Friedman ve Wechsler, 2012, s. 2). Araştırılan problemde değişkenler arasındaki bağımlılığın matematiksel yapısı ve şekli bilinmemektedir. Bazı girdi değişkenler hakkındaki bilgi eksikliği çıktı değişkenlerinde belirsizliğe yol açacağından dolayı gözlemlenen X girdi değişkenleri ile Y çıktı değişkenleri arasında bir istatistiksel bağımlılık mevcuttur. Girdi değerlerinin bilinmeyen bir $P(x)$ olasılık dağılımından rassal olarak seçildiği varsayılır. Gözlemlenen girdilerin $\{x_1, x_2, \dots, x_n\}$ etkisi koşullu olasılık dağılımı $P(y|\mathbf{x})$ ile tanımlanır. Bir x girdi değeri verildiğinde y 'nin değeri $P(y|\mathbf{x})$ olasılığı ile gerçekleşir (Cherkassky ve Mulier, 2007, s. 8).

Şekil 1.2.'den yola çıkarak girdi ve çıktı değerleri arasında nedensel bir ilişkinin var olduğu düşünülebilir. Ancak istatistiksel bağımlılığın olması değişkenler arasında mutlaka nedensel bir ilişkinin var olduğu anlamına gelmez. Diğer bir ifadeyle girdi-çıkıtı değişkenleri arasındaki bağımlılığın doğru bir biçimde tahmin edilmesi için nedenselliğin var olması şart değildir (Cherkassky, Friedman ve Wechsler, 2012, s. 3).

Vapnik (1991) öğrenme sürecini üç bileşen ile açıklamıştır. Bunlardan ilki $x \in X$ rassal vektör üretici G 'dir. G üretici (*generator*), sabit ama bilinmeyen bir $P(x)$ olasılık dağılım fonksiyonundan çekilen birbirinden bağımsız ve aynı dağılıma sahip x rassal vektörlerini verir. İkinci bileşen her bir x girdi vektörü için bilinmeyen bir $P(y|x)$ koşullu dağılım fonksiyonuna göre y yanıt vektörünü üreten S operatörüdür. S (*system*) operatörü matematiksel eşitliği ve şekli bilinmeyen ancak sabit ve mevcut olduğu bilinen gerçek hedef fonksiyonudur: $f : X \rightarrow Y$. Sonucu bileşen ise önceden a priori olarak belirlenen bir fonksiyonlar kümesini veri kümesine uygulayabilen öğrenme makinesi LM 'dir (*learning machine*) (Vapnik, 1991, s. 832). Burada çözümü istenen problem, bir fonksiyon kümesi içinden S operatörünün yanıt vektörüne en yakın değerleri veren fonksiyonun belirlenmesidir. Üreteç, öğrenme makinesi ve sistem arasındaki ilişki Şekil 1.3.'de gösterilmiştir:



Şekil 1.3. Öğrenme probleminin bileşenleri (Cherkassky ve Mulier, 2007)

Öğrenme probleminin ilk bileşeni olan üreteç X girdi kümesinden $P(x)$ olasılığı ile rassal $x \in \mathfrak{R}^d$ vektörünü verir. Sistem ise verilen $P(x)$ olasılığı ile rassal olarak seçilen her bir x vektörü için $P(y|\mathbf{x})$ koşullu olasılığına göre y değerlerini verir. Elde edilen bu ikili gözlemler kümesi öğrenme makinesinin eğitilmesinde kullanılacağı için eğitim kümesi olarak adlandırılır ve $D = (x_1, y_1), \dots, (x_n, y_n)$ biçiminde gösterilir. Buna göre eğitim kümesindeki her bir gözlemin ortak bir dağılım fonksiyonundan $P(x, y) = P(x)P(y|\mathbf{x})$ rassal ve birbirinden bağımsız şekilde çekildikleri varsayılır. Öğrenme makinesi ise önceden belirlenen bir fonksiyonlar kümesini (hipotez uzayı H) eğitim verisine uygular (Cherkassky ve Mulier, 2007, s. 18).

Öğrenme sürecinin sonucunda elde edilen ve girdi ve çıktı vektörleri arasındaki esas bağımlılığı yaklaşık olarak tahmin eden bu fonksiyona yaklaşık fonksiyon $f(x)$ ya da istatistiksel öğrenme literatüründe *hipotez* adı verilir (Kecman, 2001, s. 125). $f(x)$ fonksiyonunu tahmin etmek için öncelikle incelenecek fonksiyonlar kümesinin belirlenmesi gerekir. Gerçek f fonksiyonuna yaklaşık bir tahmini gerçekleştirmek için yararlanılabilecek olası tüm fonksiyonlar (hipotezler) kümesine *hipotez uzayı* denir ve H ile gösterilir. Örneğin, bir doğrusal regresyon modeli için parametrelerince doğrusal ve toplamsal olan tüm fonksiyonlar kümesi regresyon modelinin hipotez uzayını oluşturur. Basit doğrusal regresyon modelini ele alırsak modelin parametreleri reel sayılar kümesinde tanımlı olduğu için H sonsuz sayıda model içerir (Jones, Desmarais ve Zorn, 2017, s. 9).

Bir öğrenme probleminde hedef fonksiyonu ve eğitim kümesi araştırmacının kontrolünde olmayıp problem tarafından verilirler. Araştırmacı problemin ve eğitim kümesindeki değişkenlerin türüne en uygun öğrenme algoritmasını ve hipotez uzayını seçer. Seçilen öğrenme algoritması ve hipotez uzayı birlikte öğrenme modelini oluştururlar (Abu-Mostafa, Magdon-Ismail ve Lin, 2012, s. 5).

1.1. İstatistiksel Model

Şekil 1.1.'de gösterilen tahminleyici öğrenme problemine ilişkin matematiksel model

$$y_k = g_k(x_1, \dots, x_n, z_1, \dots, z_l), \quad k = 1, \dots, K \quad (1.1)$$

biçiminde gösterilir. y_k k . çıktı değerini ve g_k ise y_k 'nin değişen değerlerini belirleyen tüm olası girdilerin deterministik bir fonksiyonudur. Ancak (z_1, z_2, \dots, z_l) değişkenleri gözlemlenemediği ya da ölçülemediği için (1.1)'deki matematiksel model istatistiksel model ile değiştirilir:

$$y_k = f_k(x_1, \dots, x_n) + \varepsilon_k, \quad k = 1, \dots, K \quad (1.2)$$

f_k gözlemlenen değişkenlerin tek değerli deterministik bir fonksiyonu, ε_k ise stokastik bileşendir (Cherkassky ve diğerleri, 2012, s. 4). (1.2)'deki eşitlikte her bir k . çıktı için ilişkiler ayrı olarak temsil edilmektedir. Bu eşitlikten k indisini kaldırır ve girdi kümesini $X = \{x_1, x_2, \dots, x_n\}$ biçiminde gösterirsek istatistiksel modelin genel temsili şöyle ifade edilir:

$$Y = f(X) + \varepsilon \quad (1.3)$$

İstatistiksel model yapısı gereği bir belirsizliğe sahiptir. Eşitlik (1.3) incelendiğinde sabit bir sayıyı hata terimi ε 'ye ekleyip aynı sabiti $f(X)$ 'ten çıkararak çıktı değişkeni Y nin değerlerini sabit tutabildiğimiz görülmektedir. Bu tanımsal muğlaklığı çözmek için hata terimlerinin beklenen değerinin sıfıra eşit olduğu ($E(\varepsilon) = 0$) ve X değerlerinden bağımsız olduğu varsayılır. Böylece $f(X)$ verilen herhangi bir $X=x$ değeri için Y nin ortalama değerine eşit olur (Cherkassky ve diğerleri, 2012, s. 5):

$$f(X) = E[Y|X] \quad (1.4)$$

Denetimli öğrenme yaklaşımlarının tamamının amacı gerçek hedef fonksiyonu $f(X)$ 'e en yakın yaklaşık tahmin fonksiyonu $\hat{f}(X)$ 'i elde etmektir. Denetimli öğrenme yaklaşımında girdi-çıkı ilişkisi örnekler yardımıyla öğrenilmeye çalışılır. Bu süreçte araştırılan sistemin girdi değerleri ve karşılık gelen çıktı değerleri tekrarlı biçimde incelenir. Sistemin girdileri aynı zamanda öğrenme makinesinin de girdileridir. Öğrenme algoritması her bir x_i değerine karşılık $\hat{f}(x_i)$ tahminlerini verir. Daha sonra tekrarlı olarak y_i gözlemleri ile $\hat{f}(x_i)$ tahminleri arasındaki farkı inceleyerek hedef fonksiyonuna en yakın yaklaşık tahmin fonksiyonunu belirlemeye çalışır (Cherkassky ve diğerleri, 2012, s. 5).

$\hat{f}(X)$ yaklaşık fonksiyonunun $f(X)$ 'e ne kadar yaklaşık olduğunu belirlemek için bir ölçüt gereklidir. Bu ölçüt literatürde hata fonksiyonu, kayıp fonksiyonu, maliyet fonksiyonu ya da risk olarak adlandırılır. Kayıp fonksiyonu hipotez uzayındaki her bir fonksiyonun hedef fonksiyona ne kadar yakın olduğunu ölçer. y_i gözlemleri ile $\hat{f}(x_i)$ tahminleri arasındaki fark ya da kayıp, yaklaşık tahmin fonksiyonunun seçilmesinde genel bir kriter olarak belirlenir. Kayıp fonksiyonu öğrenme makinesi tarafından sağlanan yaklaşık fonksiyonun niteliğini gösteren bir ölçüdür. Literatürde çok sayıda kayıp fonksiyonu $L(Y, f(X))$ mevcut olmakla birlikte bu tezin konusu olan regresyon analizinde en sık olarak kareli hata (ℓ_2 norm) $(y_i - \hat{f}(x))^2$ ve mutlak hata (ℓ_1 norm) $|y_i - \hat{f}(x)|$ kayıp fonksiyonları kullanılır (Abu-Mostafa ve diğerleri, 2012, s. 28).

Kareli hata kayıp fonksiyonu olarak seçildiğinde tahminlerin beklenen hatası hata kareleri ortalaması (HKO) ile ifade edilir:

$$Hata = \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{f}(x_i)) \quad (1.5)$$

$$HKO = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}(x_i))^2 \quad (1.6)$$

Eşitlik (1.6)'da verilen HKO yalnızca eğitim kümesinden yararlanılarak hesaplandığı için bu hataya eğitim hatası adı verilir.

Eğitim kümesinden yola çıkılarak eğitilen bir modelin yeni bir veri kümesi için ne kadar başarılı tahminler yaptığının belirlenmesi için eğitim hatası bir kriter olarak kullanılmaz. Regresyon problemi için kayıp fonksiyonunu kareli hata olarak ele alırsak eğitim kümesindeki tüm gözlemlerin doğru tahmin edilmesi durumunda eğitim kaybı sifıra eşit olur. Asimptotik sürekli azalan bir fonksiyon olan eğitim hatasının minimizasyonu modelin genelleştirme kabiliyetinin ölçülmesinde bu yüzden bir kriter olamaz (Kroese ve diğerleri, 2019). Bir fonksiyonun başarılı tahminler gerçekleştirip gerçekleştirmediği, fonksiyonun eğitim kümesinden farklı bir veri kümesindeki performansına bakılarak ölçülür. Eğitim veri kümesinden farklı, birbirinden bağımsız ve aynı dağılıma sahip gözlemlerden oluşan bu kümeye test kümesi denir ve eğitim kümesinden elde edilen modelin performansının ölçülmesi amacıyla kullanılır. Sabit bir eğitim kümesi ve F hipotez uzayından öğrenme yoluyla elde edilen tahmin modeli $\hat{f}(X)$ olmak üzere test hatası ya da genelleştirme hatası

$$Hata_{Test} = E \left[L(Y, \hat{f}(X)) | D \right] \quad (1.7)$$

biçiminde gösterilir. X ve Y eğitim kümesinden bağımsız ancak aynı ortak dağılımdan rassal olarak seçilmiş test kümesi gözlemleridir. D ise sabit eğitim kümesidir. Dolayısıyla test hatası belirli bir eğitim kümesi için geçerli olan tahmin hatasını belirtir (Hastie, Tibshirani ve Friedman, 2008, s. 220).

Test hatasını hesaplarırken kayıp fonksiyonu olarak (1.6)'da olduğu gibi kareli hata kullanılırsa test kümesi gözlemleri (x_0, y_0) için hata kareleri ortalaması

$$HKO_{Test} = E \left(y_0 - \hat{f}(x_0) \right)^2 \quad (1.8)$$

olur. Görüleceği üzere test hatası, eğitim kümesine bağlı olduğu için yalnızca eğitim kümesindeki bilgiye dayalı olarak tahminlerin hatasını belirlemek etkili bir yöntem

olmadığından dolayı çapraz geçerlilik, bootstrapping vb. yöntemlerden yararlanılarak test hatasının beklenen değeri hesaplanır:

$$Hata = E[Hata_{Test}] \quad (1.9)$$

Beklenen hata çok sayıda eğitim kümesinden yararlanarak $\hat{f}(X)$ 'i tahmin ettiğimizde elde edeceğimiz farklı test hatalarının ortalamasını ifade eder. Bu sayede $\hat{f}(X)$ 'i hesaplayan eğitim kümesindeki rassallığın ortalaması alınmış olunur (Hastie ve diğerleri, 2008, s. 257).

Bir öğrenme yönteminin genelleme performansı eğitim kümesinden farklı bir veri kümesinde sahip olduğu tahmin başarısının bir ölçüsüdür. Bir öğrenme sürecinde birbirini takip eden iki temel amaç mevcuttur. Bunlardan ilki hipotez uzayından en iyi modeli seçmek için farklı modellerin performanslarının tahmin edilmesidir. Öğrenme sürecinin sonucunda hipotez uzayındaki farklı modellerin tahmin başarılarının kıyaslanarak en iyi modelin belirlenmesi adımı model seçimi olarak adlandırılır. İkinci amaç ise seçilen nihai modelin yeni bir veri kümesindeki genelleme hatasını ölçmektir. Belirlenen modelin farklı bir veri kümesindeki tahmin başarısının belirlenmesi sürecine ise model değerlendirilmesi denir. Modelin uyumlandırılması için eğitim kümesinden yararlanılır. Seçilmesi mümkün modeller arasından seçim yapmak için çapraz geçerlilik yönteminden ve son olarak nihai modelin genelleme hatasını değerlendirmek için test kümesinden yararlanılır (Hastie ve diğerleri, 2008, s. 222).

Model seçimi ve model değerlendirmesi süreçlerini açıklayabilmek adına izleyen kısımda model karmaşıklığı, yanlılık-varyans ilişkisi, aşırı uyum ve eksik uyum kavramlar açıklanmıştır.

1.2. Yanlılık – Varyans İlişkisi

Bir $\hat{\theta}$ tahmincisinin yanlılığı, θ parametresinin gerçek değeri ile tahmincinin beklenen değeri $E[\hat{\theta}]$ arasındaki farka eşittir:

$$Yanlılık(\hat{\theta}) = E[\hat{\theta}] - \theta \quad (1.10)$$

Eğer yanlılık sıfıra eşit ise $\hat{\theta}$ tahmincisi θ parametresinin yansız bir tahmincisidir. Varyans ise tahmincinin veri kümesinin bir fonksiyonu olarak gerçekleştirdiği tahminlerin ne kadar değişkenlik gösterdiğinin bir ölçüsüdür (Raschka, 2018, s. 6):

$$Var(\hat{\theta}) = E\left[\left(\hat{\theta} - E[\hat{\theta}]\right)^2\right] \quad (1.11)$$

Sıfır ortalama ($E\{\varepsilon\} = 0$) ve σ^2 varyanslı ε hata terimlerine sahip bir $Y = f(X) + \varepsilon$ hedef fonksiyonunu $D = (x_1, y_1), \dots, (x_n, y_n)$ eğitim kümesinden yararlanarak $\hat{f}(X)$ fonksiyonu ile yaklaşık olarak tahmin ettiğimizi varsayalım. Buna göre tahminlerin hatası $(Y - f(X))^2$ ile ölçülür ve kayıp fonksiyonunun enküçüklenmesi ile en iyi $\hat{f}(X)$ tahmin modeli elde edilir. y_i gözlem değerleri her zaman gürültüye (ε) sahip olacağı için uyumlandıracağımız tüm modeller kaçınılmaz olarak bir hataya sahip olacaklardır.

Bir $\hat{f}(x_i)$ tahmininin doğruluğu iki farklı hataya bağlı olarak ölçülür: indirgenebilir hata ve indirgenemeyen hata. İndirgenebilir hata, $\hat{f}(X)$ tahmin fonksiyonunun hedef fonksiyonu $f(X)$ 'e eşit olmamasından dolayı ortaya çıkan hatadır. $\hat{f}(X)$ tahmin fonksiyonunun doğruluğu, probleme en uygun istatistiksel öğrenme yönteminin uygulanması sayesinde artırılabilir olduğundan dolayı bu hataya indirgenebilir hata adı verilmiştir. En uygun öğrenme yönteminin kullanılması ile hedef fonksiyonunun tam doğru bir tahminini elde etsek bile yine de tahminlerimiz bir miktar hata içerecektir. Bu hata, Y 'nin aynı zamanda hata terimi ε 'nin bir fonksiyonu olmasından kaynaklanır ve indirgenemeyen hata olarak adlandırılır. Bu yüzden hata teriminin değişkenliği tahminlerin doğruluğunu etkileyecektir (James ve diğerleri, 2013, s. 18).

Veri kümesinin $E(\varepsilon) = 0$ ve $Var(\varepsilon) = \sigma^2$ varsayımı ile $Y = f(X) + \varepsilon$ modelinden elde edildiğini varsayalım. x_i değerlerinin önceden bilindiği (rassal olmadığı) varsayımı altında x_0 için beklenen hata, diğer bir ifade ile test (genelleştirme hatası) indirgenebilir ve indirgenemeyen hatanın toplamsal bir fonksiyonu olarak ifade edilebilir (James ve diğerleri, 2013, s. 19):

$$\begin{aligned} E\left(Y - \hat{f}(X)\right)^2 &= E\left[f(X) + \varepsilon - \hat{f}(X)\right]^2 \\ &= E\left[f(X) - \hat{f}(X)\right]^2 + E\left[\varepsilon^2\right] \\ &= \underbrace{\left[f(X) - \hat{f}(X)\right]^2}_{\text{İndirgenebilir}} + \underbrace{Var(\varepsilon)}_{\text{İndirgenemez}} \end{aligned} \quad (1.12)$$

İndirgenemeyen hata Y 'nin tahminlerinin doğruluğunun üst sınırını oluşturur. $f(X)$ fonksiyonunun tahmini eğitim kümesinden yararlanılarak indirgenebilir hatanın enküçüklenmesi ile elde edilir:

$$\hat{f}(X) = \arg \min_{g(x)} \sum_{i=1}^N [y_i - g(x_i)]^2 \quad (1.13)$$

Problemin çözümünde karşılaşılan temel zorluk eğitim kümesinin sonlu sayıda gözlemden oluşmasıdır. Eğer eğitim kümesi sonsuz sayıda gözlemden oluşsaydı $\varepsilon = 0$ olacağından dolayı verilen herhangi bir $X=x$ değeri için hedef fonksiyonu doğrudan hesaplanabilirdi. Ancak problem sınırlı sayıdaki gözlemlere dayanarak çözüldüğü için sonsuz sayıda çözümü vardır ve (1.13) çözümü tek çözüm olmadığı için denetimli öğrenme problemi kötü konumlanmış (*ill-posed*) bir problemdir (Cherkassky ve diğerleri, 2012, s. 9).

Eşitlik (1.12)'i çoğaltma kuralı yardımıyla

$$\begin{aligned} E\left[\left(Y - \hat{f}(x_0)\right)^2\right] &= E\left[\left(f(x_0) + \varepsilon - \hat{f}(x_0) + E[\hat{f}(x_0)] - E[\hat{f}(x_0)]\right)^2\right] \\ &= \left(f(x_0) - E[\hat{f}(x_0)]\right)^2 + E\left[\left(E[\hat{f}(x_0)] - \hat{f}(x_0)\right)^2 + E[\varepsilon^2]\right] \quad (1.14) \\ &= \text{Yanlılık}^2(\hat{f}(x_0)) + \text{Varyans}(\hat{f}(x_0)) + \sigma_\varepsilon^2 \end{aligned}$$

biçiminde bileşenlerine ayırırsak beklenen hatanın; yanlılığın karesi, varyans ve hata teriminin varyansının toplamsal bir fonksiyonu olduğu görülür. Kareli yanlılık ve tahminlerin varyansı birlikte (1.12)'de gösterilen indirgenebilir hatayı (HKO) oluştururlar (Genç, 2020, s. 11). σ_ε^2 indirgenemeyen hatadır ve gerçek hedef fonksiyonunu bilsek bile her zaman var olacaktır. Kareli yanlılık $f(x_0)$ 'ın ortalaması ile $\hat{f}(x_0)$ tahmininin beklenen değeri arasındaki farktır. Varyans ise modelin tahminlerinin değişkenliğinin ölçüsüdür (Hastie ve diğerleri, 2008, s. 38).

Varyans ve yanlılık arasındaki ilişki model karmaşıklığı ile ilgilidir. Model karmaşıklığı (esnekliği), öğrenme yönteminin sonlu bir veri kümesindeki keyfi bağımlılıkları tahmin etme kapasitesidir. Bu sebepten dolayı istatistiksel öğrenme literatüründe kapasite olarak da anılır. Doğrusal regresyon yönteminde model karmaşıklığı etkili parametre sayısı ile diğer bir ifade ile serbestlik derecesi ile ölçülür. Parametrik yöntemlerde uyulması gereken son derece sıkı varsayımlar olduğundan dolayı

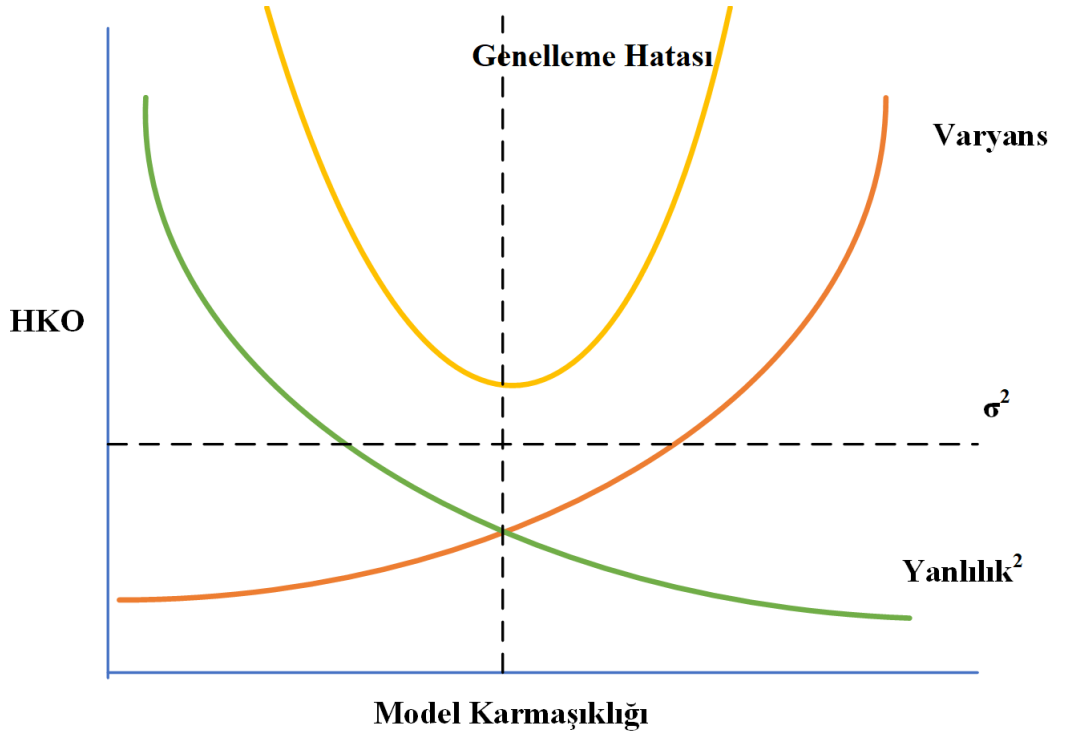
bağımlılığın gerçek parametrik şekli bilinmediğinde doğru tahmin modelinin geliştirilmesinde başarısız olmaları son derece muhtemeldir. Öte yandan parametrik olmayan yöntemler ise sonlu sayıdaki bir örneklemin kullanıldığı çok boyutlu problemlerin çözümünde başarısız olmaktadır. Sonlu bir veri kümesinden yararlanarak yararlı çözümler elde edebilmek için esnek (uyarlanabilir) yöntemler geliştirilmiştir. Bu yöntemler herhangi sürekli bir fonksiyona ilişkin yaklaşık bir çözüme ulaşmak için geniş ve esnek bir yaklaşık fonksiyonlar kümesinden yararlanırlar. Ancak mevcut verinin sonlu olması sebebiyle tek bir çözüm elde edebilmek için bu geniş fonksiyonlar kümesi bir şekilde kısıtlandırılmalıdır. Sonlu bir veriden yola çıkarak geniş bir fonksiyonlar kümesinden tek bir çözüme ulaşabilmek için bir çerçeve sunan çok sayıda yöntem geliştirilmiştir. Bunlar arasında düzenlileştirici (ceza) yöntemler ve yapısal risk minimizasyonu, Bayeşçi çıkarım en yaygın olanlardır (Cherkassky ve Mulier, 2007, s. 38). Çalışmanın 3. Bölümünde düzenlileştirici yöntemler ayrıntılı olarak ele alınacaktır.

Sonlu örneklemden yola çıkarak elde edilen ve en iyi tahmin performansına sahip model optimum karmaşıklığa sahip olan modeldir. Bu modeli seçmek amacıyla test hatasını en aza indirecek optimum model karmaşıklığını belirlemek için yanlılık-varyans ilişkisinden yararlanılır. Genelleştirme hatasını (test HKO) tahmin etmek için açıkça görüldüğü üzere eğitim hatasından (eğitim HKO) yararlanılır. Ancak eğitim hatası test hatasının iyi bir tahmini olmadığı için optimum model karmaşıklığının açıklanmasında kullanılamaz (Hastie ve diğerleri, 2008, s. 38).

Kareli yanlılığın ve varyansın toplamsal bir fonksiyonu olan genelleştirme hatası Şekil 1.4.'te görüleceği üzere konveks bir fonksiyondur. Tahmin edilen modelin yüksek bir tahmin performansına sahip olması isteniyorsa genelleştirme hatası en aza indirilmelidir. Genelleştirme hatasını en aza indirebilmek için varyans ve yanlılık eş zamanlı olarak azaltılmalıdır. Varyans ve yanlılık sıfırdan büyük değerler oldukları için genelleştirme hatası da her zaman sıfırdan büyük olur. Şekil 1.4.'te görüleceği üzere test HKO değeri her zaman indirgenemeyen hatadan (σ_ε^2) büyüktür.

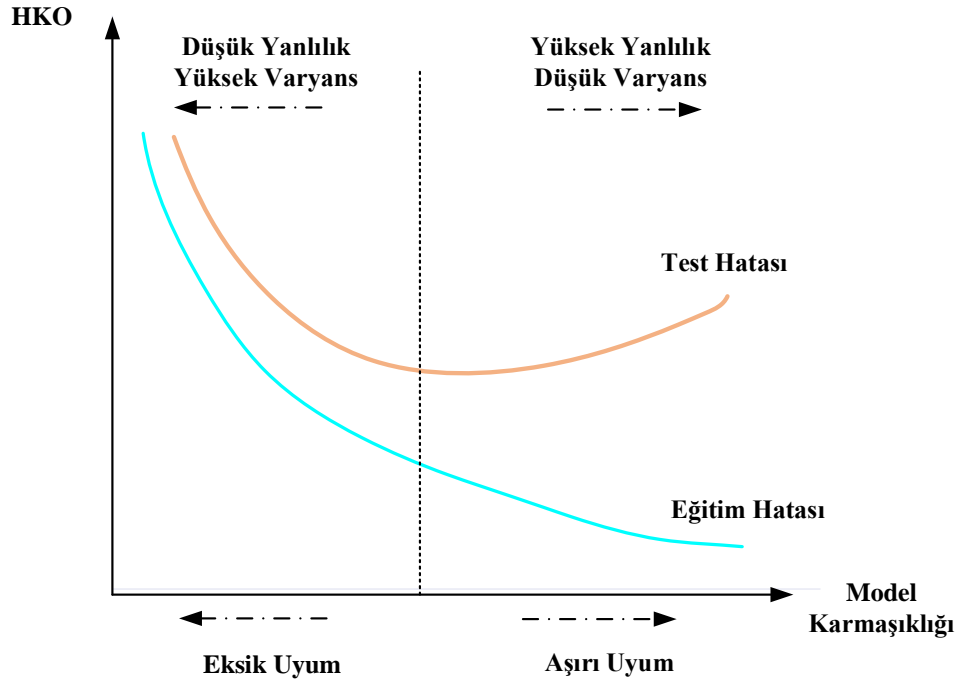
Öğrenme sürecinde kullanılan eğitim kümesi değiştikçe tahmin edilen model de değişiklik gösterir. Eğitim kümesi değiştikçe $f(X)$ 'in tahminin çok fazla değişkenlik göstermesi istenmez. Ancak modelin kapasitesi arttıkça tahmin fonksiyonun uyum kabiliyeti arttığından dolayı veri kümesindeki gözlemlere daha fazla uyumlanır. Bu sebepten dolayı eğitim kümesindeki tek bir gözlem değiştirildiğinde dahi $f(X)$ 'in

tahmininde çok büyük deęişiklikler gözlemlenebilir. Bu sebepten dolayı kapasitesi yüksek istatistiksel yöntemlerin varyansı daha yüksek, kapasitesi düşük yöntemlerin de varyansı daha düşüktür. Öte yandan yanlılık ise hedef fonksiyonun şekline ve kullanılan istatistiksel yöntemin şekline baęlı olarak deęişen bir ölçüttür. Gerçekte doğrusal olmayan bir fonksiyon doğrusal bir yöntemle tahmin edilmek istendiğinde elde edilen $\hat{f}(X)$ bir miktar yanlılığa sahip olacaktır (James ve dięerleri, 2013, s. 35).



Şekil 1.4. Yanlılık-varyans ilişkisi (Genç, 2020)

Genel olarak yüksek kapasiteye sahip yöntemler daha az yanlı modeller verirler. Bu iki ölçütün niceliğinde gerçekleşen deęişimlerin oranı test hatasının deęerini belirler. Modelin karmaşıklığını artırdığımızda başlangıçta yanlılıktaki azalma varyanstaki artışa göre çok daha hızlıdır. Böylece test HKO bir minimum deęere doğru azalır. Ancak bir noktadan sonra kapasitenin artırılmasının yanlılıktaki azalma oranı varyanstaki artış oranına göre daha düşük olduğundan test hatası artmaya başlar. Dolayısıyla en düşük test hatasına sahip modeli belirleyebilmek için optimum kapasiteyi belirlememiz gerekir (James ve dięerleri, 2013, s. 35).



Şekil 1.5. Model karmaşıklığının eğitim ve test hatasına etkisi (Hastie ve diğerleri, 2008)

Model karmaşıklığı arttıkça eğitim ve test hatasının değişimi Şekil 1.5.'te gösterilmiştir. Eğitim hatasının genel olarak test hatasından daha düşük olması beklenir. Bundan dolayı genelleştirme performansının belirlenmesinde eğitim hatasının kullanılması iyimser bir yaklaşım olarak kabul edilir. Modelin karmaşıklığı arttıkça gözlemlere daha fazla uyum sağlaması aşırı uyum olarak adlandırılır. Sonuç olarak aşırı uyum modelin eğitim kümesine iyi uyumlanması ancak test verisine uyum becerisinin düşük olmasıdır. Aşırı uyum Şekil 1.5.'teki grafiğin sağ tarafında görülebilir. Test ve eğitim hatası arasındaki fark model karmaşıklığı arttıkça artmaktadır. Hem eğitim hatası hem de test hatasının yüksek olmasına ise eksik uyum denir ve düşük kapasiteye sahip modellerde gözlemlenir. Eksik uyum, modelin kabiliyetinin düşük olmasından kaynaklı olarak veri kümesindeki gözlemlere uyumlanma ve bağımlılıkları ortaya çıkarma becerisinin az olması anlamına gelir. Dolayısıyla hem test hatası hem eğitim hatası yüksek olur. Eksik uyum ise Şekil 1.5.'te verilen grafiğin sol tarafında gözlemlenebilir. En iyi uyum optimum model karmaşıklığına bağlıdır (Goodfellow ve diğerleri, 2016, s. 113).

1.3. Yeniden Örnekleme ve Hiperparametre Optimizasyonu

Öğrenme algoritmalarının birçoğu öğrenme sürecinde genelleştirme performansını iyileştirmek için algoritmanın davranışlarını kontrol etmeye yarayan ayar parametrelerine sahiptirler. Bu ayar parametreleri aynı zamanda hiperparametre olarak da adlandırılırlar.

Hiperparametreler seçilen bir öğrenme algoritması için hipotez uzayını kısıtlayarak modelin kapasitesini düzenler ve böylece yanlılık-varyans arasındaki doğru denge noktasının belirlenmesine yardımcı olur. Düzenleştirici regresyon yöntemleri özelinde konuyu ele alacak olursak belirlenen hiperparametreler modeldeki değişken seçimini gerçekleştirir. Bu sebepten öğrenme sürecinde hiperparametre optimizasyonunun gerçekleştirildiği adım model seçimi olarak da adlandırılır.

Önceki kısımda modelin uyumlandırılması sürecinde eğitim kümesinin ve tahminlerin başarısının ölçülmesinde de test kümesinin kullanıldığından bahsedilmişti. Ayrıca yanlılık ve varyans arasındaki ilişki de öğrenme algoritmalarının sabit hiperparametrelere sahip olduğu varsayımı altında incelenmiştir. Buna ek olarak hiperparametrelerin de bir optimizasyon sürecinde belirlenmesi gerekir. Öğrenme algoritmasının bulacağı nihai modelin parametrelerinden farklı olarak hiperparametreler öğrenme süreci başlamadan önce a priori olarak belirlenir. Ancak sonsuz farklı seçenek arasından hiperparametrelerin araştırmacı tarafından belirlenmesi son derece güç olacağından hiperparametre optimizasyonu adı verilen bir süreç öğrenme sürecine dahil edilerek iç içe bir öğrenme süreci tasarlanır (Goodfellow ve diğerleri, 2016, s. 119).

Öğrenme algoritması eğitim kümesi üzerinde çalışırken hiperparametre değerlerini değiştirmek farklı modellerin elde edilmesiyle sonuçlanır. Farklı hiperparametre değerleri tarafından belirlenen bir fonksiyonlar kümesinden en başarılı tahminler üreten modelin seçilmesi sürecine model seçimi adı verilir. Dolayısıyla hiperparametrenin belirlenmesi de model seçimi sürecinin bir parçasıdır (Raschka, 2018, s. 22).

Sonlu sayıda bir veriden yararlanarak model seçimi ve genelleme hatasının tahmin edilebilmesi için yeniden örnekleme yöntemlerinden yararlanır. Yeniden örnekleme yöntemleri bir veri kümesinden tekrarlı biçimde örneklemelerin seçilmesi ve bir modelin her bir örnekleme uyumlandırılarak modelden daha fazla bilginin elde edilmesini sağlarlar. Yeniden örnekleme yöntemleri sayesinde optimum hiperparametre değerleri belirlenir, aşırı uyum sorunundan kaçınılır ve modelin bağımsız bir veri kümesindeki performansı değerlendirilerek genelleştirme özelliği değerlendirilir (Genç, 2020, s. 16)

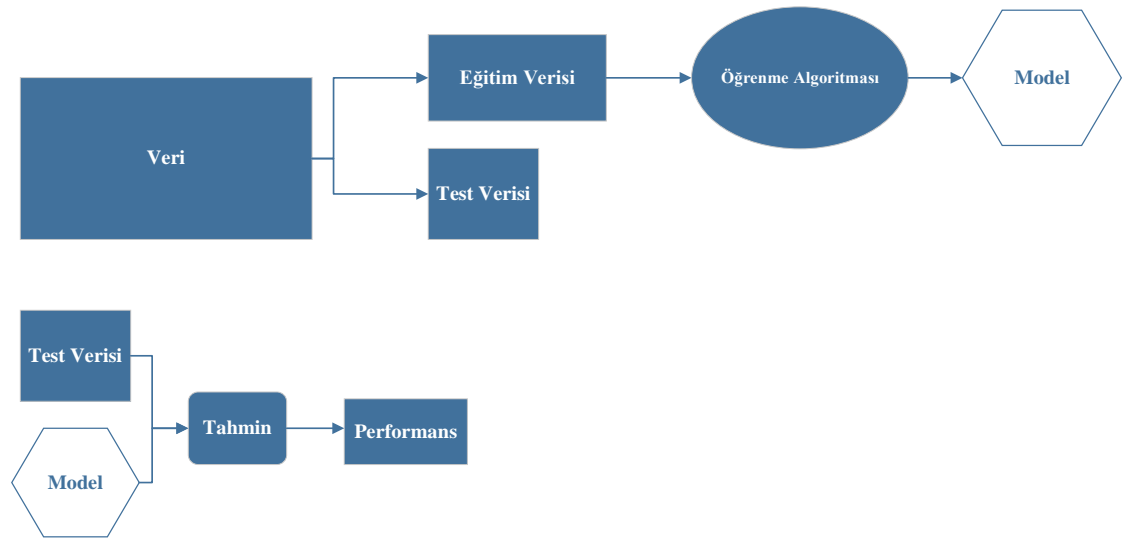
1.3.1. Geçerlilik kümesi yaklaşımı

Veri kümesinin eğitim ve test kümesi olarak ikiye ayrılarak modelin seçilmesi amacıyla eğitim kümesinin ve elde edilen modellerin performanslarının değerlendirilmesi için test kümesinin kullanılması geçerlilik kümesi yaklaşımı olarak adlandırılır. Geçerlilik

kümesi yaklaşımını açıklayabilmek için öncelikle hiperparametrelerin sabit olduğu ya da hiperparametre tahminin gerekmediği bir öğrenme problemini ele alalım. Bu durumda veri rassal olarak eğitim kümesi ve test kümesi biçiminde ikiye ayrılır. Önceki kısımlarda ele alınan konular bu varsayım altında açıklanmıştır.

Verinin eğitim ve test olarak ikiye bölünmesi araştırmacıyı verinin ne kadarının eğitim ve ne kadarının test için ayrılması gerektiği sorusuyla baş başa bırakır. Bununla ilgili literatürde net bir uzlaşma olmamakla birlikte 60/40, 80/20 ve hatta veri hacmi çok büyük ise 90/10 gibi ayrımlara gidilebilmektedir. Ancak genel olarak bütün yeniden örnekleme yöntemlerinde ikiye ayrılan verinin büyük kısmının eğitim ve küçük kısmının test amacıyla ayrılması konusunda literatürde genel bir uzlaşma mevcuttur (Raschka, 2018, s. 8).

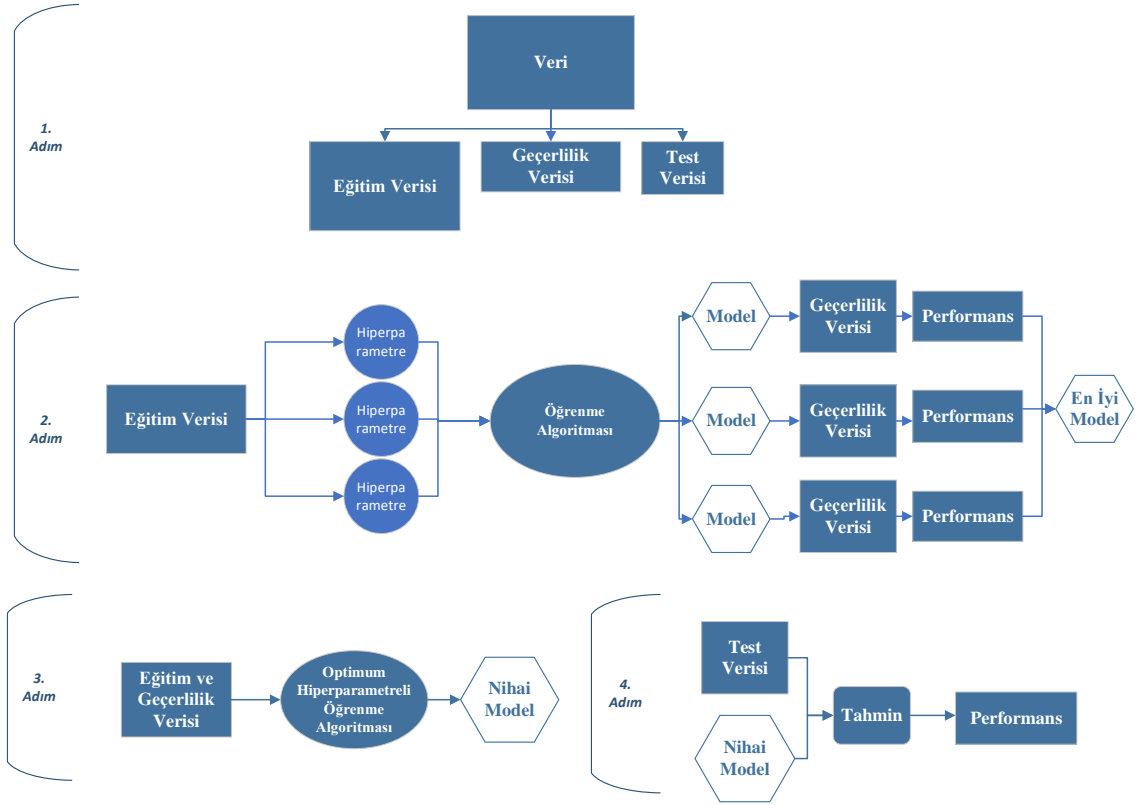
Veri iki kısma ayrıldıktan sonra probleme uygun bir öğrenme algoritması seçilerek önceden belirlenen hiperparametreler ile birlikte eğitim kümesinde öğrenme algoritması çalıştırılır ve model elde edilir. Daha sonra model test verisinde sınanarak modelin test hatası yardımıyla genelleme performansı belirlenir. Test hatasının yansız olarak belirlenebilmesi için öğrenme adımında test verisinin kullanılmaması son derece önemlidir.



Şekil 1.6. İkili geçerlilik kümesi yaklaşımı (Raschka, 2018)

Yukarıda açıklanan geçerlilik kümesi yaklaşımını sabit hiperparametrelerin yer aldığı bir öğrenme algoritması yardımıyla modelin parametrelerinin tahmini ve elde edilen modelin değerlendirilmesi amacıyla kullanılmaktadır. Ancak önceden belirtildiği gibi ilerleyen kesimlerde açıklanacak olan ridge regresyon, lasso regresyon gibi

düzenleştirici yöntemlerin bazı öğrenme algoritmalarının hiperparametreleri sabit değildir ve bu ayar parametrelerinin (λ_1, λ_2) optimum değerleri de veriden öğrenme yoluyla belirlenmesi gerekir. Dolayısıyla hiperparametre optimizasyonu için de bir veri kümesine ihtiyaç duyulur. Elimizde çok sayıda gözlem mevcut ise veri kümesi eğitim, geçerlilik ve test olarak üç gruba ayrılır. Eğitim verisi modelin uyumlandırılması, doğrulama verisi model seçimi ve test verisi de seçilen modelin genelleştirme hatasının değerlendirilmesi için kullanılır. Bu yönteme üçlü geçerlilik kümesi yöntemi adı verilir ve bu yöntemin adımları Şekil 1.7.'de gösterilmiştir (Raschka, 2018, s. 22).



Şekil 1.7. Üçlü geçerlilik kümesi yaklaşımı (Raschka, 2018)

Üçlü geçerlilik kümesi yönteminde ilk olarak farklı hiperparametre değerlerine sahip öğrenme algoritmaları kullanılarak eğitim verisindeki gözlemlere modeller uyumlandırılır. Daha sonra bu modellerin performansları geçerlilik kümesi yardımıyla değerlendirilir. Bu aşamanın sonunda en iyi performansa sahip modelin hiperparametre değerleri optimum değerler olarak belirlenir ve böylece model seçimi gerçekleştirilir. Üçüncü adımda eğitim ve geçerlilik kümeleri birleştirilir ve belirlenen optimum hiperparametre değerlerini kullanarak öğrenme algoritması çalıştırılır ve nihai model tahmin edilir. Elde edilen nihai modelin genelleştirme performansı test verisi yardımıyla değerlendirilerek öğrenme süreci sonlandırılır (Raschka, 2018, s. 23).

Bu yöntemde ikili geçerlilik kümesi yaklaşımının aksine geçerlilik kümesi yardımıyla hiperparametreler belirlenir. Eğer hem optimum hiperparametrelerin hem de en iyi modelin belirlenmesinde yalnızca eğitim kümesinden yararlanılırsa hiperparametreler mümkün en yüksek kapasiteyi seçeceği için aşırı uyum sorunu ile karşılaşılır (Goodfellow ve diğerleri, 2016, s. 119). Öte yandan test verisi de modelin seçilmesi sürecinde asla kullanılmaz. Test verisi öğrenme sürecine dahil edilirse modelin genelleştirme hatası gerçekte olduğundan daha düşük olarak tahmin edilir ve böylece iyimser bir biçimde gerçekte olduğundan daha başarılı bir model tahminine ulaşıldığı sonucuna varılır (Hastie ve diğerleri, 2008, s. 222).

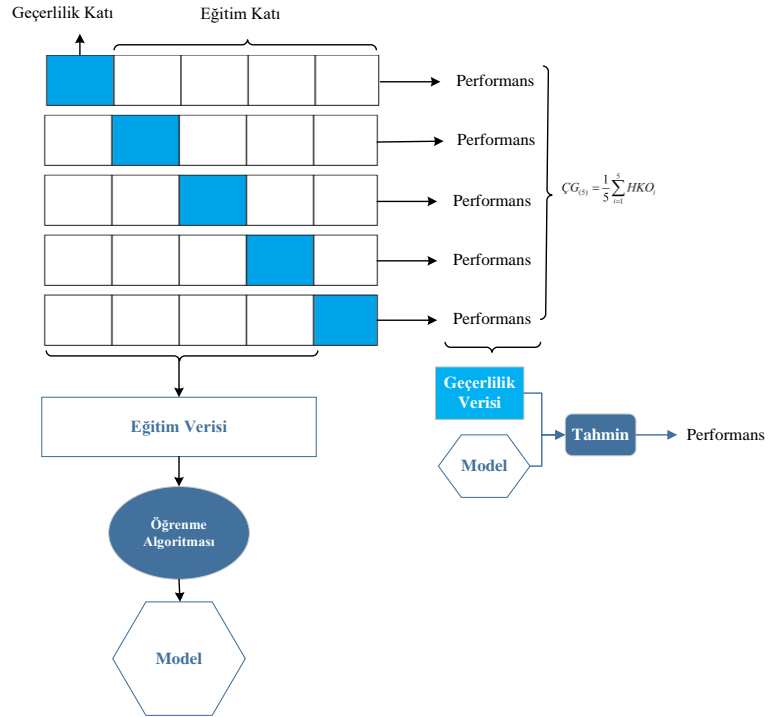
Geçerlilik kümesi yöntemi her ne kadar pratik bir yöntem olsa da bazı dezavantajları mevcuttur. Bu tür yöntemlerde her ne kadar gözlemler rassal olarak eğitim, test ve geçerlilik kümelerine atansalar da test hatası gözlemlerin yer aldığı kümeye bağlı olarak değişkenlik gösterir. Özellikle veri hacmi küçük olduğunda daha yüksek bir değişkenliğin ortaya çıkması beklenir. İkinci dezavantaj ise veri kümesinin üç parçaya ayrılırken eğitim kümesinde yeterince gözlemin ayrılmaması durumunda ortaya çıkar. Eğer öğrenme algoritmasına gereğinden daha az eğitim verisi ayrılırsa bu durumda belirlenen modelin kapasitesine ulaşmadığı sonucuna varılır. Her bir yeni gözlemin öğrenme sürecine katkı sağlayacağını göz önüne alırsak veri kümesindeki bazı gözlemlerin eğitim kümesine dahil edilmemesi durumunda modelin sahip olacağı genelleştirme performansı verinin tamamının kullanıldığında elde edilecek genelleştirme performansından her zaman daha düşük olacaktır. Dolayısıyla bu yöntemde elde edilen genelleştirme performansı her zaman kötümser bir yanlılığa sahiptir. Diğer bir ifade ile test hatası olduğundan fazla tahmin edilir (Raschka, 2018, s. 10).

1.3.2. k-katlı çapraz geçerlilik

k-katlı çapraz geçerlilik yöntemini açıklarken öncelikle hiperparametrelerin sabit olduğunu varsayalım. k-katlı çapraz geçerlilik yönteminde veri kümesi eşit hacimli ve birbirinden farklı k adet kümeye ayrılır. Daha sonra bir küme test için ve k-1 adet küme ise bir bütün olarak eğitim için belirlenir. Her bir adımda ayrı tutulan kümede HKO değeri hesaplanır ve bu sayede modelin performansı tahmin edilir. Bu işlem k defa tekrarlanır ve böylece her adımda farklı bir grup test kümesi görevi görür. İşlem sonucunda k adet model ve k adet performans tahmini (HKO) elde edilir. k-katlı çapraz geçerlilik tahmini

k adet HKO değerinin ortalamasına eşittir ve çapraz geçerlilik hatası olarak adlandırılır (James ve diğerleri, 2013, s. 183):

$$CG_{(5)} = \frac{1}{5} \sum_{i=1}^5 HKO_i \quad (1.15)$$



Şekil 1.8. 5-katlı çapraz geçerlilik yöntemi ile model değerlendirme (Raschka, 2018)

k-katlı çapraz geçerlilik yönteminin uygulanabilmesi için öncelikle k grup sayısının belirlenmesi gerekir. k-katlı çapraz geçerlilik yönteminin özel bir durumu $k = n$ durumudur. Bu durumda her bir tekrarda yalnızca tek bir gözlem değeri test değeri olarak belirlenirken $n - 1$ adet gözlem değeri eğitim kümesi olarak belirlenir. Bu durumda çapraz geçerlilik süreci n adet tekrardan oluşur. Bu çapraz geçerlilik yöntemine birini hariç tutma (leave-one-out) adı verilir. Birini hariç tutma yönteminin çapraz geçerlilik hatası (1.15) 'ten yola çıkarak şöyle elde edilir:

$$CG_{(n)} = \frac{1}{n} \sum_{i=1}^n HKO_i \quad (1.16)$$

Açıktır görüleceği üzere birini hariç tutma yönteminin hesaplama maliyeti özellikle n çok büyük ise oldukça yüksektir.

k-katlı çapraz geçerlilik yönteminin bir diğer örneği ise önceki kısımda ele alınan ikili geçerlilik kümesi yöntemidir. Her ne kadar ikili geçerlilik yönteminde eğitim kümesi genel olarak test kümesinden daha büyük olarak belirlense de (genelde 2/3 eğitim ve 1/3

test kümesi olacak biçimde) $k = 2$ katlı çapraz geçerlilik yöntemi teoride eğitim ve test kümesinin eşit hacimde olduğu ikili geçerlilik yöntemine karşılık gelmektedir. İdeal k değerinin kaç olması gerektiği bu iki uç örneği ele alarak değerlendirilebilir. Bunun için yine yanlılık-varyans ilişkisinden yararlanmamız gerekmektedir (James ve diğerleri, 2013, s. 185).

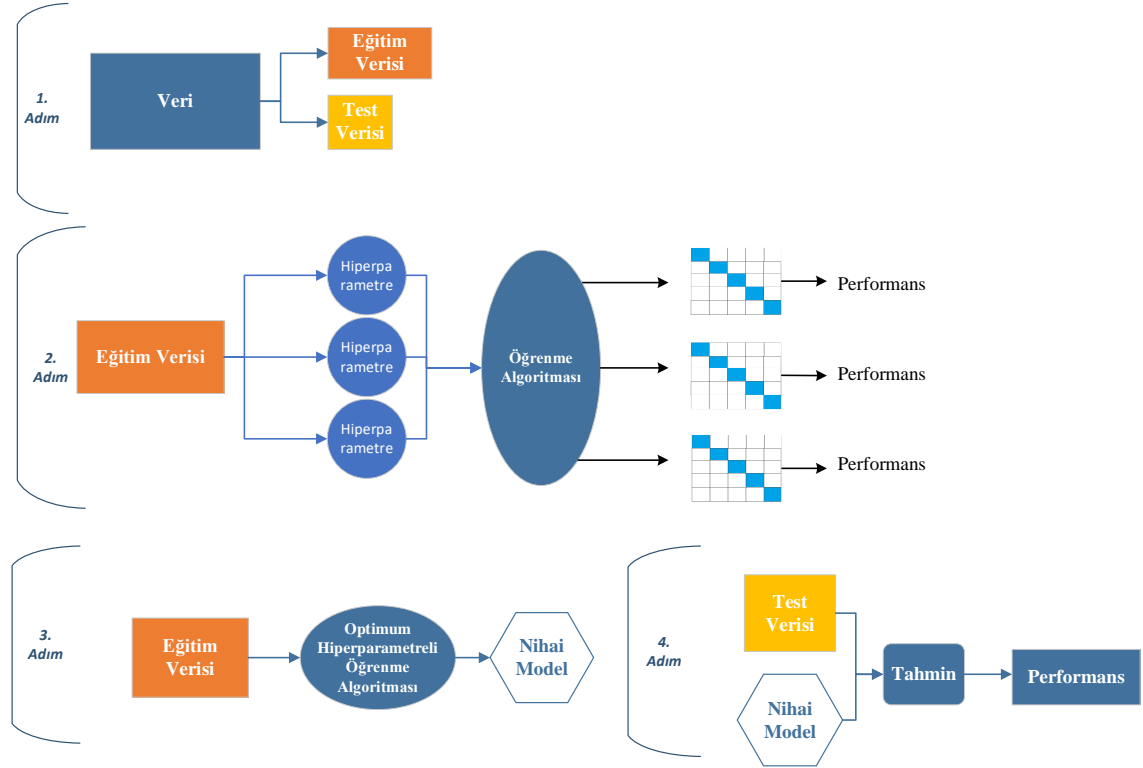
Daha önce belirtildiği üzere geçerlilik kümesi yöntemi mevcut gözlemlerin verinin tamamını kullanmak yerine yalnızca bir kısmını eğitim verisi olarak kullandığı için test hatasının değerini olduğundan daha fazla tahmin etme eğilimindedir. Buradan yola çıkarak birini hariç tutma yönteminde $n-1$ adet gözlem eğitim kümesinde yer aldığı için test hatasının neredeyse yansız bir tahminini sunmaktadır. Yanlılığın azaltılması söz konusu olduğunda birini hariç tutma yönteminin $k < n$ için k -fold çapraz geçerliliğe üstünlük sağladığı söylenebilir (James ve diğerleri, 2013, s. 185).

Öte yandan birini hariç tutma yöntemi k -katlı çapraz geçerlilik yöntemine göre daha fazla varyansa sahiptir. Birini hariç tutma yöntemi uygulandığında her biri neredeyse birebir aynı eğitim kümelerinde uyumlandırılmış modellerin çıktılarının ortalaması hesaplanır. Eğitim kümeleri neredeyse birebir aynı oldukları için bu n adet çıktı kendi aralarında pozitif yönlü ilişkilidirler. Öte yandan $k < n$ için k -katlı çapraz geçerlilik uygulandığında k adet daha az korelasyona sahip modelin çıktılarının ortalaması hesaplanır. Yüksek korelasyona sahip niceliklerin ortalaması daha yüksek varyansa sahip olacağından birini hariç tutma yönteminin varyansının daha yüksek olması beklenir (James ve diğerleri, 2013, s. 186). Sonuç olarak k arttıkça çapraz geçerlilik tahmininin yanlılığı azalırken varyansı ve hesaplama maliyeti artar. Ancak veri kümesi küçük ise k değerinin çok küçük ($k = 2$ ya da $k = 3$) olması durumunda da örneklemeden kaynaklı olarak yüksek varyans ortaya çıkabilir (Raschka, 2018, s. 30). Tüm bunları göz önüne aldığımızda k için tek bir doğru değer olmamakla birlikte genelde $k = 5$ ya da $k = 10$ olarak belirlenmektedir (James ve diğerleri, 2013, s. 186).

k -katlı çapraz geçerlilik yöntemi belirlenen bir modelin performansını değerlendirmenin yanında model seçimi olarak adlandırılan değişken seçimini gerçekleştirmek amacıyla da kullanılmaktadır. Model seçimi k -katlı çapraz geçerlilik yönteminin bir parçası olarak gerçekleştirildiğinde test kümesi ayrı tutulduğu için genelleştirme performansının tahmininin yanlılığı azalır. Ancak eğitim kümesinin hacmi azalacağı için test hatasının daha yüksek olması bekleneceğinden genelleştirme

performansının kötümser bir biçimde (gerçekte olduğundan daha düşük) tahmin edilmesine yol açar (Refaeilzadeh, Tang ve Liu, 2007) (Raschka, 2018, s. 30).

k-katlı çapraz geçerlilik yöntemi içerisinde model seçimi Şekil 1.9.'da gösterilmiştir:



Şekil 1.9. k-katlı çapraz geçerlilik ile model seçimi (Raschka, 2018)

k-katlı çapraz geçerlilik yöntemi ile model seçimi aşağıdaki adımların izlenmesi ile yapılır:

1. Veri kümesi rassal olarak eğitim ve test kümesi olarak ikiye ayrılır.
2. Her bir k. adımda bir grup geçerlilik kümesi ve geri kalan k-1 grup eğitim kümesi olarak belirlenerek k-katlı çapraz geçerlilik uygulanır. Eğitim kümesinde farklı hiperparametre değerlerine karşılık gelen k adet model ve performans tahmini elde edilir.
3. İkinci adımda belirlenen en iyi performans (en düşük HKO'ya) sahip hiperparametre değerleri optimum olarak kabul edilir ve birinci adımda belirlenen tüm eğitim kümesi kullanılarak öğrenme algoritması yardımıyla nihai model belirlenir.

4. Birinci adımda belirlenen nihai modelin genelleştirme performansı test kümesi yardımıyla elde edilir (Raschka, 2018, s. 31).

1.4. Optimizasyon Yöntemleri

Bu başlıkta konveks problemlerin çözümü için temel optimizasyon kavramları ve algoritmalar açıklanacaktır. Ele alınan optimizasyon yöntemleri en küçük kareler, ridge, lasso ve elastic-net regresyonlarının çözüm aşamasında kullanılan yöntemlerdir.

Bir optimizasyon problemi

$$\min_{x \in \mathbb{R}^n} f(x) \quad (1.17)$$

biçiminde ifade ediliyorsa buna kısıtsız optimizasyon problemi denir. Amaç fonksiyonu $f: \mathbb{R}^n \rightarrow \mathbb{R}$ biçiminde tanımlanır, $x = (x_1, x_2, \dots, x_n)^T$ vektörü ise problemin optimizasyon değişkenidir. Bu tür problemlerin çözümü bazı durumlarda tam olarak bulunabilse de birçok durumda algoritmalar yardımıyla sonsuz adımda yaklaşık olarak çözülür.

İlgili optimizasyon değişkeni belirli kısıtlamalara tabi ise bu optimizasyon problemine kısıtlı optimizasyon denir:

$$\begin{aligned} & \min f(x) \\ & \text{kısıtlar;} \\ & g_i(x) \leq 0, \quad i \in I \\ & h_i(x) = 0, \quad i \in Q \end{aligned} \quad (1.18)$$

Kısıtlı optimizasyon problemlerinde değişkenlerin çözümü eşitlik ve/veya eşitsizlik kısıtlarınca tanımlanan bir uygun çözüm bölgesinde yer alır. Eşitlik (1.18)'de yer alan f amaç fonksiyonunu, g_i eşitsizlik kısıtlarını ve h_i ise eşitlik kısıtlarını temsil eder. Eğer amaç fonksiyonu ve tüm kısıt fonksiyonları doğrusal ise bu kısıtlı optimizasyon problemine doğrusal programlama problemi denir. Amaç fonksiyonu ikinci dereceden bir fonksiyon ve tüm kısıtlar doğrusal ise optimizasyon problemine kuadratik programlama adı verilir. Amaç fonksiyonu ile tüm eşitsizlik kısıtları konveks ve eşitlik kısıtları doğrusal ise bu probleme konveks programlama denir. Problemin uygun çözüm bölgesi

$$S = \{x \mid g_i(x) \leq 0, \quad i \in I; \quad h_i(x) = 0, \quad i \in Q\} \quad (1.19)$$

biçiminde tanımlanır. f amaç fonksiyonunu minimum yapan x^* noktası $x \in S$ uygun noktaları arasında aranır. Bu durumda (1.18)'deki amaç fonksiyonu

$$\min_{x \in S} f(x) \quad (1.20)$$

biçiminde ifade edilir (Tezel Özturan, 2019, s. 241). Basitlik sağlaması açısından bundan sonraki kısımda tüm kısıtların eşitsizlik biçiminde olduğu varsayılacaktır ($h_i = \emptyset$).

1.4.1. Konveks optimizasyon

Bir $f : \mathbb{R}^n \rightarrow \mathbb{R}$ fonksiyonunun konveks bir fonksiyon olabilmesi için gerek ve yeter koşul $S \subseteq \mathbb{R}^n$ konveks kümesindeki tüm $0 \leq \alpha \leq 1$ ve $\forall x, y \in S$ için

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \quad (1.21)$$

olmasıdır. Geometrik olarak bu eşitsizlik $(x, f(x))$ ve $(y, f(y))$ noktalarını birleştiren doğru parçasının f fonksiyonunun grafiğinin üstünde yer aldığı anlamına gelir. Buna göre bir konveks fonksiyon, global minimum olmayan yerel bir minimum noktaya sahip olamaz (Tezel Özturan, 2019, s. 45).

S konveks bir küme ve f fonksiyonu S üzerinde konveks bir fonksiyon olduğunda (1.18) problemdeki g_i fonksiyonları konveks ise bu problem konveks bir optimizasyon problemidir:

$$\begin{aligned} & \min f(x) \\ & \text{kısıtlar;} \\ & g_i(x) \leq 0, i \in I \end{aligned} \quad (1.22)$$

Amaç fonksiyonu f türevlenebilir bir fonksiyon ise $x^* \in S$ 'nin global optimum olması için gerek ve yeter koşul her $x \in S$ için

$$\nabla f(x^*)(x - x^*) \geq 0 \quad (1.23)$$

olur. Bu koşulun yeterliliği her $x \in S$ için

$$f(x) \geq f(x^*) + \nabla f(x^*)(x - x^*) \geq f(x^*) \quad (1.24)$$

biçiminde gösterilir.

Konveks problemle ilişkili önemli bir fonksiyon Lagrange fonksiyonudur:

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i \quad (1.25)$$

Eşitlikteki $\lambda \geq 0$ ağırlıkları Lagrange çarpanları olarak adlandırılırlar ve $g_i(x) \leq 0$ kısıtı ihlal edildiğinde x 'e bir ceza uygularlar. Lagrange dualite teorisine göre amaç

fonksiyonunun minimumunu $f^* = \min_{x \in S} L(x; \lambda^*)$ veren bir $\lambda^* \geq 0$ değeri mevcuttur. Sonuç olarak (1.22) problemindeki $g_i(x^*) \leq 0$ kısıtlarını sağlayan x^* optimum çözümü aynı zamanda

$$0 = \nabla_x L(x^*; \lambda^*) = \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* g_i(x^*) \quad (1.26)$$

eşitliğini de sağlar (Hastie, Tibshirani ve Wainwright, 2015, s. 96).

Eşitlik (1.22)'de verilen optimizasyon problemi güçlü dualite koşulunu sağlıyorsa x vektörünün bu problemin optimum çözümü olabilmesi için gerek ve yeter koşul Karush-Kuhn-Tucker (KKT) koşullarını sağlamasıdır. KKT koşulları optimum Lagrange çarpanları vektörünü ($\lambda^* \geq 0$) optimum asal vektör $x^* \in \mathbb{R}^s$ ile ilişkilendirir. KKT koşulları aşağıdaki gibidir:

$$\begin{aligned} g_j(x^*) &\leq 0, \quad \forall_j \in I \\ \lambda_j^* g_j(x^*) &= 0, \quad \forall_j \in I \\ \lambda_j^* &\geq 0, \quad \forall_j \in I \\ \nabla_x L(x^*; \lambda^*) &= 0 \end{aligned} \quad (1.27)$$

Birçok optimizasyon probleminde konveks ancak türevlenemeyen fonksiyon yer alır. Bu durumda amaç ve kısıt fonksiyonlarında gradyanlar yer aldıkları için (1.23) ve (1.26)'da verilen optimallik koşulları doğrudan uygulanamazlar. Bu tür konveks problemlerin çözümünde subgradyan kullanılır. $x \in S$ ve $y \in \mathbb{R}^n$ vektörleri olmak üzere $S \subseteq \mathbb{R}^n$ konveks kümesinde tanımlı bir $f : S \rightarrow \mathbb{R}$ konveks fonksiyonu için

$$f(y) \geq f(x) + (y - x)^T z \quad (1.28)$$

ise $z \in \mathbb{R}^n$ vektörü f fonksiyonunun x noktasındaki subgradyanıdır. f fonksiyonunun x noktasındaki subgradyanlar kümesine f fonksiyonunun x noktasındaki subdiferansiyeli denir ve $\hat{\partial}f(x)$ ile gösterilir. Konveks ancak türevlenemeyen problemler için (1.26) koşulunda yer alan gradyanlar subgradyanlar ile değiştirilip

$$0 \in \hat{\partial}f(x^*) + \sum_{i=1}^m \lambda_i^* \partial g_i(x^*) \quad (1.29)$$

biçiminde düzenlenerek KKT koşulları uygulanabilir (Hastie ve diğerleri, 2015, s. 99).

1.4.2. Koordinat iniş algoritması

En küçük kareler ve ridge regresyon problemlerinin aksine ileriki bölümlerde gösterileceği üzere ortogonal tasarım ve tek değişkenli durumlar haricinde ℓ_1 normuna sahip lasso ve elastic-net problemleri kapalı form çözümüne sahip değildirler. Bu tür problemlerin çözümü için çeşitli algoritmalarından yararlanılmaktadır.

Koordinat iniş algoritması, bir fonksiyonun minimumunu bulmak amacıyla koordinat yönleri boyunca minimizasyon işlemi gerçekleştiren bir optimizasyon algoritmasıdır. Her adımda bir koordinat seçim kuralına göre bir koordinat ya da koordinat bloğu belirlenir. Daha sonra diğer tüm koordinatlar ya da koordinat blokları sabit tutularak karşılık gelen koordinat hiperdüzleminde tam ya da hatalı bir biçimde minimizasyon gerçekleştirilir. Lasso ve lasso'nun türleri ayrılabilir olma özelliği sebebiyle bu problemlerin çözümü koordinat minimizasyon algoritmasının uygulanmasına uygundur (Hastie ve diğerleri, 2015, s. 109).

Koordinat iniş algoritmasının konveks bir optimizasyon probleminin global minimumunu yaklaşık olarak bulabilmesi için yeter koşul f fonksiyonunun sürekli türevlenebilir ve kesin konveks olmasıdır. Öte yandan f amaç fonksiyonu

$$f(x_1, \dots, x_p) = g(x_1, \dots, x_p) + \sum_{j=1}^p h_j(x_j) \quad (1.30)$$

biçiminde ayrılabilirlik özelliğine sahip olsun. Burada $g: \mathbb{R}^p \rightarrow \mathbb{R}$ konveks ve türevlenebilir bir fonksiyon, $h: \mathbb{R} \rightarrow \mathbb{R}$ ise konveks ancak türevlenebilir olması zorunlu olmayan fonksiyondur. Bu tür konveks ancak türevlenemeyen kısıtlara sahip konveks problemler ayrılabilirlik özelliğine sahiplerse koordinat iniş algoritması ile global minimum noktasına yaklaşık bir çözüm elde edilir. Burada önemli olan özellik türevlenemeyen $\sum_{j=1}^p h_j(x_j)$ bileşeninin ayrılabilir olmasıdır.

Algoritma her adımda belirli bir koordinat seçim kuralına göre tek bir koordinat seçerek tek bir koordinat üzerinde tek değişkenli minimizasyon gerçekleştirir. Algoritma bir k iterasyonunda belirlenen bir koordinat için çözüm değerlerini aşağıdaki gibi günceller:

$$\begin{aligned}
x_1^{(k+1)} &= \arg \min_{x_1} f(x_1, x_2^{(k)}, x_3^{(k)}, \dots, x_p^{(k)}) \\
x_2^{(k+1)} &= \arg \min_{x_2} f(x_1^{(k+1)}, x_2, x_3^{(k)}, \dots, x_p^{(k)}) \\
&\vdots \\
x_p^{(k+1)} &= \arg \min_{x_p} f(x_1^{(k+1)}, x_2^{(k+1)}, x_3^{(k+1)}, \dots, x_p)
\end{aligned}
\tag{1.31}$$

2. DOĞRUSAL REGRESYON ANALİZİ

Regresyon ifadesi istatistik biliminde ilk kez Francis Galton (Galton, 1886) tarafından kullanılmıştır. Galton, kalıtım üzerine yaptığı çalışmada (1877) bezelyeleri incelemiş ve hacmi büyük olan bezelyelerden elde edilen yeni bezelyelerin ebeveynlerinden daha küçük hacimli, küçük olan bezelyelerden elde edilen yeni bezelyelerin ise ebeveynlerinden daha büyük hacimli olduğunu tespit etmiştir. 1886 yılındaki çalışmasında uzun ebeveynlerin kendilerinden daha kısa ve kısa ebeveynlerin kendilerinden daha uzun çocukları olduğu sonucuna varan Galton, bu durumu sıradanlığa (ortalamaya) doğru gerileme/yakınsama (*regression towards mediocrity*) olarak adlandırmıştır.

Günümüzde regresyon ifadesi bağıntı, bağlanım anlamında kullanılmaktadır. Regresyon analizi ise bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki nedensel ilişkinin varlığını matematiksel bir denklem yardımıyla modellemeyi amaçlar. Modeldeki bağımlı değişken (Y), bağımsız değişkenlerin (X_j) yardımıyla açıklanmaya çalışılan değişkendir. Öte yandan bağımsız değişken ise bağımlı değişkeni etkileyen ve onu açıklamakta kullanılan değişkendir.

Regresyon analizinin başlıca kullanım amaçları şunlardır:

- i. Bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasında var olduğu düşünülen ilişkiyi bir denklem yardımıyla ifade etmek
- ii. Bağımsız değişkendeki bir birimlik değişimin bağımlı değişkende nasıl bir değişime yol açtığını belirlemek
- iii. Bağımsız değişkenlerin değerlerinden yararlanarak bağımlı değişkenin değerini tahmin etmek
- iv. Hangi bağımsız değişkelerin bağımlı değişken üzerinde etkisi olduğunu saptamak (Şıklar, 2000, s. 4), (Alpar, 2013, s. 417).

Elimizde n adet X_i, Y_i ikili gözlem değerleri $i = (1, \dots, n)$ olsun. Eğer ki elde edeceğimiz model Y ile X arasındaki ilişkiyi doğru açıklayabiliyorsa bağımsız değişkenin X_i değerinden yararlanarak bu değere karşı gelen bağımlı değişken Y_i 'nin değeri tahmin edilebilir (Pardoe, 2012, s. 36).

Bir regresyon modelinin genel yapısı (2.1)'de gösterilmiştir:

$$\text{Bağımlı değişken} = \text{Model fonksiyonu} + \text{Hata terimi} \quad (2.1)$$

Regresyon, parametrik bir yöntem olduğu için “model fonksiyonu”nun yapısı ve şekli önceden bilinmektedir. Doğrusal regresyon analizinde bu fonksiyon parametreleri açısından doğrusaldır. Model fonksiyonu, bağımsız değişkenleri (X_k) ve tahminlenmesi gereken parametreleri içerir ve X_i 'nin değerlerinden yararlanarak Y_i 'nin beklenen değerini matematiksel bir eşitlikle ifade eder (Draper ve Smith, 1998, s. 17).

X 'e bağlı olarak Y 'nin değerleri tahmin edilmek istendiğinde, model fonksiyonu bağımlı değişkenin değerini tam olarak veremez. Bunun ilk sebebi matematiksel fonksiyonda Y 'deki değişimin tamamını açıklayan tüm bağımsız değişkenlerin modelde yer almasının imkânsız olmasıdır. Ayrıca X_i değerlerini belirleyen ölçü aletlerinden ve insanlardan kaynaklanan hatalardan dolayı Y_i 'nin gerçek değeri model tarafından belirlenememektedir. Bu sebeple regresyon denkleminde “hata terimi” adı verilen bir bileşen yer almaktadır (Birkes ve Dodge, 1993, s. 10).

(2.1)'de gösterilen “model fonksiyonu” Y 'nin koşullu beklenen değerini veren bir eşitlikle açıklanır. Bu eşitlik ortalama fonksiyonu olarak da adlandırılır:

$$E(Y | X = x) = x\text{'in değerlerine bağlı bir fonksiyon} \quad (2.2)$$

(2.2)'de belirtilen ortalama fonksiyonunun sol tarafında yer alan $E(Y | X = x)$ ifadesi “ $X = x$ sabit değerine bağlı olarak Y 'nin beklenen değeri” biçiminde okunur. Eşitliğin sağ tarafındaki fonksiyon ise problemin ve fonksiyonun yapısına bağlı olarak farklılık gösterir. Doğrusal regresyon analizinde ortalama fonksiyonu (2.3)'de gösterildiği gibidir:

$$E(Y | X = x) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_j X_{ij} \quad (i=1, \dots, n) \quad (j=1, \dots, k) \quad (2.3)$$

Eşitlik (2.3)'te verilen β_k değerleri regresyon katsayıları olarak da adlandırılan fonksiyonun parametrelerini oluşturur ve önceden bilinemediği için tahmin edilmeleri gerekir. Bir doğrusal regresyon modelinde k adet bağımsız değişken için ve bir de sabit terim için olmak üzere toplamda $k + 1$ adet regresyon katsayısı yer alır.

(2.1)'de belirtilen “hata terimi” varyans fonksiyonu ile ifade edilir. Doğrusal regresyon modelinde her bir x değeri için varyansın sabit olduğu varsayılır ve şöyle ifade edilir:

$$\text{Var}(Y | X = x) = \sigma^2 \quad (2.4)$$

Buradan Y'nin beklenen değeri (ortalaması) X'in bir fonksiyonu olmasına karşın Y'nin varyansı X'in değerlerine bağlı olmaz. Hata terimleri birbirleriyle ilişkisiz olduklarından dolayı, bağımlı değişkenin değerleri de birbirleriyle ilişkisizdir.

Ortalama fonksiyonu ve varyans fonksiyonundan yararlanarak çoklu doğrusal regresyon modeli aşağıdaki gibi elde edilir:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (i=1, \dots, n) \quad (2.5)$$

Eşitlik (2.5)'te yer alan β_0 değeri $x = 0$ olduğunda Y_i 'nin alacağı değeri gösterir. $\beta_0, \beta_1, \dots, \beta_k$ regresyon katsayıları ise ait oldukları X_{ik} bağımsız değişkenindeki bir birimlik değişimin Y'de nasıl bir değişime yol açacağını belirtir. Bir doğrusal regresyon modelinde bir tane bağımsız değişken varsa bu model basit doğrusal regresyon modeli, birden fazla bağımsız değişken varsa çoklu doğrusal regresyon modeli olarak adlandırılır.

Çoklu doğrusal regresyon modellerinin matris formunda gösterimi ise şöyledir:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.6)$$

Bu denklemde yer alan matrisler şunlardır:

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}_{n \times p}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{p \times 1}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} \quad (2.7)$$

Söz konusu modelde örneklem hacmi n ve bağımsız değişken sayısı k ile gösterilmiştir. Doğrusal modelde β_0 sabit katsayısı da yer aldığından k adet bağımsız değişken ile birlikte \mathbf{X} matrisinin sütun sayısı $k+1=p$ adettir. Buna göre \mathbf{y} matrisi bağımlı değişkenin gözlemlerine ilişkin $n \times 1$ boyutlu sütun vektörü; \mathbf{X} bağımsız değişkenlere ilişkin n adet gözlem ile k adet bağımsız değişken ve sabit regresyon katsayısı için birler sütununu içeren $n \times p$ boyutlu matris; $\boldsymbol{\beta}$ vektörü $p \times 1$ boyutlu tahmin edilen regresyon katsayıları vektörü ve $\boldsymbol{\varepsilon}$ ise $n \times 1$ boyutlu rassal hatalar vektörüdür (Aydın, 2014, s. 77). $n \times p$ boyutlu \mathbf{X} matrisi regresyon matrisi olarak da adlandırılır ve sütunları doğrusal bağımsızdır (Seber ve Lee, 2003, s. 36).

Bağımlı değişken ve hata terimi rassal değişkenler oldukları için \mathbf{y} ve $\boldsymbol{\varepsilon}$ vektörleri rassal vektörlerdir. Bağımsız değişken değerleri bilinen sabit değerler olduğundan \mathbf{X} matrisi bilinen sabitler matrisi; $\boldsymbol{\beta}$ vektörü ise tahmin edilen bilinmeyen sabitler vektörüdür. Bu yüzden, $\mathbf{X}\boldsymbol{\beta}$ terimi $\boldsymbol{\varepsilon}$ rassal hatalar vektörüne eklenen bir sabitler dizisi veya kümesidir. Buna göre \mathbf{y} 'nin beklenen değeri

$$E(\mathbf{y}) = E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = E(\mathbf{X}\boldsymbol{\beta}) + E(\boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta} \quad (2.8)$$

varyansı ise

$$Var(\mathbf{y}) = Var(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = Var(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n \quad (2.9)$$

eşittir.

2.1. Doğrusal Regresyon Yönteminin Varsayımları

Doğrusal regresyon modeli, değişkenler arasındaki doğrusal ilişkiyi gerçeğe yakın bir biçimde modelleyebilmek için birtakım varsayımlara dayanır. Doğrusal regresyonun varsayımları üç başlık altında toplanabilir:

- Hata terimi ile ilgili varsayımlar
- Bağımsız değişkenler ile ilgili varsayımlar
- Modelin kendisi ile ilgili varsayımlar

Regresyon analizinde hataların olasılık dağılımı her bir X_i değeri için sıfır ortalamaya sahiptir. Diğer bir ifade ile hata terimlerinin beklenen değeri sıfırdır ($E(\varepsilon_i) = 0$). Bu varsayım hata terimlerinin rassal bir değişken olması ile ilişkilidir. Hata terimlerinin sıfır ortalamaya sahip olması modelde sistematik bir hatanın olmadığını ifade eder. Hata terimleri rassal olmadığında ve sıfır ortalamaya sahip olmadığında modelde sistematik hata olduğu sonucuna varılır ve regresyon katsayıları yanlış tahmin edilir (Chatterjee ve Simonoff, 2013, s. 8).

Hata terimleri sıfır ortalamalı ve σ^2 varyanslı normal dağılıma sahiptir:

$$\varepsilon_i \sim N(0, \sigma^2) \quad (2.10)$$

Hata terimlerinin normal dağılıma sahip olması güven aralıklarının hesaplanması ve hipotez testlerinin gerçekleştirilmesi için önemlidir.

Hata teriminin olasılık dağılımı her bir X değeri için sabit varyanslıdır ($V(\varepsilon_i) = \sigma^2$). Buna göre değişkenlik anakütlenin bazı örneklemelerinde daha az veya daha çok olamaz. Seçilmesi mümkün tüm örneklemelerden elde edilen gözlemlerden elde edilen varyans değerleri her örneklem ve anakütle için sabittir. Sabit varyansın olmaması değişen varyans olarak adlandırılır. Değişen varyansın mevcudiyeti durumunda standart hatalar olması gerekenden daha yüksek değer alır ve böylece güven aralıkları olması gerekenden daha geniş olacağından regresyon katsayılarının hipotez testi sınavında yanlış sonuçlara ulaşılır (Pardoe, 2012, s. 60).

Hata terimleri farklı gözlem değerleri için birbirinden bağımsızdır. Buna göre hata terimleri arasında otokorelasyon yoktur. Bir hata teriminin değeri hakkında bir bilgiye sahip olmak, bir başka hata teriminin değerinin belirlenmesi hakkında bize bir bilgi vermez. Bu sebeple regresyon analizinde her bir gözlem diğer bir gözlemden bağımsız olarak elde edilmelidir. Otokorelasyon durumu genellikle zaman serilerinde görülür (Pardoe, 2012, s. 60):

$$Kov(\varepsilon_j, \varepsilon_h) = 0 \quad (2.11)$$

Her X_i değeri için elde edilen hata terimi ilgili bağımsız değişkenden bağımsızdır. Buna göre X_{ik} bağımsız değişkeni ile hata terimi ε_i 'nin kovaryansı sıfıra eşittir:

$$Kov(X_{ik}, \varepsilon_i) = 0 \quad (2.12)$$

Bağımsız değişken ile hata terimi arasında bir ilişki olmaması varsayımı ($Kov(X_{ij}, \varepsilon_j) = 0$) bağımsız değişken X 'in stokastik bir değişken olmamasını gerektirir. İstatistiksel olarak anakütleden çekilebilecek tüm örneklemeler için aynıdır ve sabittir. Ancak hata terimi ve bağımlı değişkenin değerleri rassal nedenlerden dolayı örneklemden örnekleme değişir (Şıklar, 2000, s. 11).

Bağımsız değişken ile bağımlı değişken arasında doğrusal bir ilişki vardır. Apsise göre gözlemlerin değeri arttıkça ordinata göre de artmaktadır. Gözlem noktaları, bir eğrinin değil bir doğrunun etrafında kümelenirler (Marill, 2004). Bir regresyon modelinin doğrusal olması ile ifade edilmek istenen şey, regresyon modelinin katsayıları itibarıyla doğrusal olmasıdır. Buna göre $Y_i = \beta_0 + \beta_1 X + \beta_2 X^2 + e$ modeli her ne kadar doğrusal değil gibi görünse de katsayıları itibarıyla doğrusal bir modeldir. Söz konusu bu modelde

yer alan $X = X_1$ ve $X^2 = X_2$ biçiminde ifade edilirse $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$ biçiminde doğrusal bir model olarak elde edilir (Birkes ve Dodge, 1993, s. 4)

Çoklu doğrusal regresyon yönteminde bağımsız değişkenler birbirinden bağımsızdır. Basit doğrusal regresyon yönteminde tek bir bağımsız değişken olacağından bu varsayım basit doğrusal regresyon yönteminde aranmaz. Eğer bir örneklemden elde edilen iki veya daha fazla bağımsız değişkene ait gözlemler arasında tam çoklu bağıntı varsa, en küçük kareler tahmincisi tek eşsiz bir çözüm üretmez. Böylece regresyon katsayılarının tahmini imkânsız hale gelir. Bağımsız değişkenler arasında tam çoklu bağıntı olmasa bile tahmin edilen regresyon katsayılarının varyansı artar. Bu sebepten dolayı standart hata da büyük olacağından katsayıların güven aralıkları olması gerekenden daha geniş olur ve böylece katsayılar istatistiksel olarak anlamsız olmaya eğilim gösterirler (Aydın, 2014, s. 376).

Bağımsız değişkenlerin değerleri hatasız olarak ölçülmüştür. Bu değerlerin hatalı olarak ölçülmesi bağımsız değişken ile hata terimi arasında bir ilişki olmasına ve (2.12) 'de gösterilen kovaryansın sıfırdan farklı olmasına yol açar (Şıklar, 2000, s. 12).

Doğrusal regresyon analizi ile ilgili bir diğer varsayım da modelin belirlenme hatasına sahip olmamasıdır. Bu varsayım regresyon fonksiyonun yanlış seçildiğinde veya modele yanlış bağımsız değişkenler eklendiğinde ihlal edilmiş olur. Örneğin, gerçekte değişkenler arasında doğrusal olmayan bir ilişki mevcutken doğrusal regresyon analizinin uygulanması ve doğrusal regresyon modelinin uyumlandırılması modelin yanlış belirlenmesine yol açar (Berry, 1993, s. 30).

2.2. En Küçük Kareler Yöntemi

Regresyon analizinde regresyon katsayılarının tahmin edilmesinde ve böylece regresyon modelinin hesaplanmasında en sık olarak kullanılan yöntem Carl Friedrich Gauss tarafından 1795 yılında geliştirilen en küçük kareler yöntemidir.

Eğer μ ortalamalı ve σ^2 varyanslı normal dağılıma sahip bir anakütleden n birimlik bir örneklem seçilirse, anakütle ortalaması μ hakkında örneklem verisinden yararlanarak iki farklı yolla çıkarımda bulunulabilir. Bu yöntemlerin ilki, μ 'nün tahmininin ($\hat{\mu}$) örneklem dağılımından yararlanmak, ikincisi ise doğrusal modelden faydalanmaktır.

Anakütle ortalamasının en iyi tahmincisi örneklem ortalamasıdır. n hacimli bir örnekte yer alan Y_i ($i = 1, \dots, n$) gözlemlerinin örneklem ortalaması \bar{Y} şöyle hesaplanır:

$$\bar{Y} = \frac{\sum Y_i}{n} \quad (2.13)$$

Örneklem ortalaması \bar{Y} 'nin örnekleme dağılımının ortalaması anakütle ortalaması μ 'ye ve standart sapması da standart hataya $\frac{\sigma}{\sqrt{n}}$ 'e eşittir. Bu sebepten dolayı örnekleme dağılımından faydalanılarak bilinmeyen parametre μ hakkında çıkarımda bulunulur.

Rassal bir Y değişkeninin davranışını ve diğer değişkenlerle olan ilişkisini açıklayabilmek için ise değişkenin dağılımının parametrelerini (μ) içeren cebirsel bir denklem yardımıyla bir model kurulur. Eğer bu model istatistiksel bir model ise modelde gözlem değerlerinin Y üzerindeki değişkenliğini temsil eden bir bileşen (σ^2) de yer alır. Doğrusal regresyon analizinde, parametrelerin doğrusal ve toplamsal bir fonksiyonu olarak kurulan bu modele doğrusal model adı verilir:

$$Y_i = \mu + \varepsilon_i \quad (2.14)$$

Y_i : bağımlı değişkenin i . gözlem değeri

μ : bağımlı değişkenin anakütle ortalaması

ε_i : i . gözleme ait hata terimi

Eşitliğin sağ tarafında yer alan μ , modelin deterministik kısmını oluşturur. Eğer modelde hiçbir değişkenlik yoksa ($\sigma^2 = 0$), tüm değerler μ 'ye eşit olur. Bu yüzden ε_i 'nin beklenen değerinin (ortalamasının) sıfıra eşit olması, Y 'nin beklenen değerinin modelin deterministik kısmına eşit olmasını gerektirir.

ε_i ise modelin stokastik ya da rassal bileşenidir. Bu değerler gözlem değerlerinin ortalamadan sapmalarına eşittir ($Y_i - \mu$). Diğer bir ifade ile gözlem değerlerinin değişkenliğini temsil ederler. Hata terimi olarak adlandırılan bu bileşen sayesinde modelin deterministik kısmının anakütleyi ne kadar iyi temsil ettiği belirlenir. Anakütle parametresi σ^2 de hata terimi ε_i 'nin varyansıdır ve hata terimlerinin yayılımının bir ölçüsüdür. Varyansın küçük olması hata terimlerinin sıfıra yakın olduğu ve gözlemlenen

değerlerin anakütle ortalaması etrafında toplandığı anlamına gelir. Bu yüzden hata terimi model uyumunun ölçülmesinde kullanılır ve varyansın küçük olması modelin iyi uyumlu olduğunu gösterir.

Denklem (2.14) sayesinde örneklem verisinden yararlanarak istatistiksel çıkarımda bulunulabilir. Regresyon analizindeki ilk amaç modelin deterministik kısmının belirlenmesidir. Bu amaçla gözlem değerlerine en iyi biçimde uyan modeli belirleyecek en iyi $\hat{\mu}$ tahmincisi belirlenir. Model uyumunun en iyi ölçüsü ise gözlemlenen değerler (Y_i) ile tahmin edilen ortalama ($\hat{\mu}$) arasındaki farkların (gözlemlerin ortalamadan sapmaların) karelerinin toplamıdır. Bu ölçüye hata kareleri toplamı adı verilir ve i . gözleme ait bir hata tahmini (ε_i) şöyle hesaplanır:

$$\varepsilon_i = Y_i - \mu \quad (2.15)$$

En küçük kareler yöntemi ile amaçlanan hata kareler toplamının karesini en küçükleyen $\hat{\mu}$ tahmincisinin belirlenmesidir (Freund, Wilson ve Sa, 2006, s. 11):

$$\sum e_i^2 = \sum (Y_i - \mu)^2 \quad (2.16)$$

Denklem (2.16)'da ifade edilen $\hat{\mu}$ tahmincisi X_i gözlemleri için model tarafından belirlenen bağımlı değişkenin teorik değeri Y ile ifade edilir. e_i ise hata teriminin tahmincisidir. Ortalama fonksiyonu olarak adlandırılan ve (2.3)'te belirtilen bağımlı değişkenin beklenen değeri göz önüne alındığında $\hat{\mu}$ tahmincisinin yerine,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_j X_{ij} \quad (2.17)$$

denklemini yazıldığında (2.16) denklemini

$$\sum_{i=1}^n e_i^2 = \sum (Y_i - Y_i) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \dots - \beta_j X_{ij})^2 \quad (2.18)$$

biçiminde ifade edilir. Tahmin edilmek istenen regresyon katsayıları (2.18) denkleminde $\hat{\beta}_j$ 'lere göre kısmi türevleri alınıp sıfıra eşitlenmesiyle hesaplanır (Freund ve diğerleri, 2006, s. 11).

$\hat{\beta}_0$ ve $\hat{\beta}_1$ gibi iki bilinmeyenli iki denklem sisteminden yararlanarak Y 'nin X 'e göre basit doğrusal regresyon denklemi şöyle ifade edilir:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 \quad (2.19)$$

(2.19)'daki gibi bir basit doğrusal regresyon modelini ele alırsak regresyon katsayıları (2.18)'de verilen denklem (2.20)'deki şekli alır. (2.20) denkleminin $\hat{\beta}_0$ ve $\hat{\beta}_1$ 'e göre kısmi türevleri alınıp sıfıra eşitlenerek (2.19)'daki basit doğrusal regresyon denkleminin regresyon katsayıları aşağıdaki gibi tahmin edilir:

$$\sum_{i=1}^n (Y_i - \hat{Y}_i) = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i})^2 \quad (2.20)$$

$$\frac{\partial e}{\partial \hat{\beta}_0} = -2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i}) = 0 \quad (2.21)$$

$$\frac{\partial e}{\partial \hat{\beta}_1} = -2 \sum X_{1i} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i}) = 0$$

(2.21)'deki denklemden normal denklemler elde edilir:

$$\begin{aligned} \sum Y_i &= \hat{\beta}_0 n + \hat{\beta}_1 \sum X_i \\ \sum X_i Y_i &= \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 \end{aligned} \quad (2.22)$$

Normal denklemlerin çözümü ile en küçük kareler tahmincileri elde edilir:

$$\begin{aligned} \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \end{aligned} \quad (2.23)$$

(2.23)'deki x_i ve y_i sırasıyla X ve Y değişkenlerinin aritmetik ortalamadan sapmalarıdır. Böylece β_0 ve β_1 katsayılarının tahminleri $\hat{\beta}_0$ ve $\hat{\beta}_1$ olarak elde edilir (Aydın, 2014, s. 16)(Şıklar, 2000, s. 14).

(2.6)'da belirtilen çoklu doğrusal regresyon modelinin matris gösteriminin en küçük kareler yöntemi ile çözümü için tahmin edilmesi gereken fonksiyon şöyledir:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (2.24)$$

$\boldsymbol{\beta}$: $p \times 1$ boyutlu katsayıların tahmin vektörü

\mathbf{e} : hata terimlerinin tahmin vektörü

(2.18) denklemi matris formunda ifade edilirse,

$$\begin{aligned}
\sum e_i^2 &= \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{y})'(\mathbf{y} - \mathbf{y}) \\
&= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}
\end{aligned} \tag{2.25}$$

eşitliği elde edilir (Şıklar, 2000, s. 65). (2.25) eşitliğinin $\boldsymbol{\beta}$ 'ya göre kısmi türevi alınırsa normal denklemlerin matris gösterimi

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} \tag{2.26}$$

biçiminde elde edilir. \mathbf{X} matrisinin sütunlarının doğrusal bağımsız olduğu varsayımı altında (2.26)'nın tek bir çözümünü sağlayan $\boldsymbol{\beta}$ vektörü mevcuttur. Eşitlik (2.26)'nın çözülmesiyle $\boldsymbol{\beta}$ yansız tahminci vektörü (2.27)'ye eşit olur:

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.27}$$

$\hat{\mathbf{y}}$ vektörü ise

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.28}$$

denkleme eşittir. Böylece hata terimi $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$

$$\begin{aligned}
\mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y} \\
\mathbf{I}: &\text{Birim matris} \\
\mathbf{H}: &\text{Şapka matrisi}
\end{aligned} \tag{2.29}$$

olarak ifade edilir (Chatterjee ve Simonoff, 2013, s. 8).

2.2.1. En küçük kareler tahmincilerinin özellikleri

Kısım 2.1'de bahsedilen varsayımlar sağlandığında Gauss-Markov teoremine göre regresyon modelinin katsayılarının en küçük kareler tahmincileri iki temel özelliğe sahip olurlar: Yansızlık ve etkinlik. Bir anakütle parametresinin θ tahmincisinin sonsuz sayıda tekrarlı rassal örneklemelerden (örnekleme dağılımından) elde edilen ortalaması, tahminlenmek istenen parametreye eşit ise ($E(\theta) = \theta$) bu tahminciye yansız tahminci adı verilir. Doğrusal modelin doğru olduğu varsayımı altında $\hat{\boldsymbol{\beta}}$ tahmininin beklenen değeri

$$E(\hat{\beta}) = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \quad (2.30)$$

olur. $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ olduğundan (2.30) eşitliği

$$\begin{aligned} E(\hat{\beta}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon)] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon] \\ &= \beta \end{aligned} \quad (2.31)$$

olur. Hataların beklenen değeri sıfır ($E(\varepsilon) = 0$) ve $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$ olduğu için $\hat{\beta}$ tahmini β 'nin doğrusal yansız tahmincisidir (Montgomery, Peck ve Vining, 2013, s. 79).

$$E(\beta | \mathbf{X}) = \beta \quad (2.32)$$

β 'nin varyans özelliği, kovaryans matrisi ile ifade edilir.

$$Kov(\beta) = E\left\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\right\} \quad (2.33)$$

$p \times p$ boyutlu simetrik bir matris olan kovaryans matrisinin j . köşegen elemanı $\hat{\beta}_j$ 'nin varyansını, ij . elemanı ise $\hat{\beta}_i$ ve $\hat{\beta}_j$ arasındaki kovaryansını gösterir. (2.33)'ün çözümü

$$\begin{aligned} Var(\beta) &= Var[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Var(\mathbf{y})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (2.34)$$

biçiminde elde edilir. $E(\mathbf{e} | \mathbf{X}) = 0$, $Var(\mathbf{e} | \mathbf{X}) = \sigma^2\mathbf{I}_n$ ve hata terimleri birbirinden bağımsız olduğunda, en küçük kareler tahmincisi (β), regresyon katsayıları vektörü (β)'nin en iyi doğrusal yansız tahmincisidir ve varyansı

$$Var(\beta | \mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \quad (2.35)$$

'e eşittir (Montgomery ve diğerleri, 2013, s. 80).

σ^2 çoğu kez bilinmediğinden aralık tahminleri ve hipotez testlerinin yapılabilmesi için $\hat{\sigma}^2$ tahmincisi kullanılır. Sabit varyanslılık varsayımına göre her bir Y_i gözleminin varyansı σ^2 'ye eşittir ve tüm hata terimleri için aynıdır ($V(\varepsilon_i) = \sigma^2$). Regresyon

analizinde Y_i değerleri, X_i değerlerine bağlı olarak farklı ortalamaya sahip farklı olasılık dağılımlarından gelirler. Bu yüzden regresyon analizinde varyansın tahmini her Y_i gözlem değerinin kendi değerinin ortalama tahmin değerinden (teorik değerinden) \hat{Y}_i çıkarılması ve karelerinin toplanmasıyla hesaplanır ve hata kareleri toplamı olarak adlandırılır (Kutner, Nachtsheim, Neter ve Li, 2005, s. 25).

$$\text{Hata Kareleri Toplamı} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} \quad (2.36)$$

$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ olduğuna göre hata kareleri toplamı aşağıdaki gibi olur:

$$\begin{aligned} HKT &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned} \quad (2.37)$$

$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ olduğuna göre

$$HKT = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \quad (2.38)$$

biçiminde elde edilir. Regresyon modelinde k adet regresyon katsayısı elde edildiğine göre hata kareleri toplamı $n - k$ serbestlik derecesine sahiptir. Hata kareleri toplamını serbestlik derecesine böldüğümüzde hata kareleri ortalaması elde edilir:

$$\hat{\sigma}^2 = HKO = \frac{HKT}{n - k} \quad (2.39)$$

Görüldüğü üzere hata karelerinin ortalamasının beklenen değeri σ^2 'ye eşittir. Ancak bu özellikler, tek bir örneklemden elde edilecek ve en küçük kareler yöntemine göre hesaplanacak herhangi bir tahminin anakütle parametresine tam olarak eşit olacağı anlamına gelmez. Anakütleden elde edilen tekrarlı örneklemlerin dağılımının –örnekleme dağılımı ortalaması anakütle parametresine cebirsel olarak eşittir. Buna göre, bir regresyon parametresinin en küçük kareler tahmincisi, doğrusal ve yansız tahminciler arasından en az varyanslı (etkin) dağılıma sahiptir. Bu özelliklerden yola çıkarak en küçük kareler tahmincisi en iyi (etkin) doğrusal yansız tahminci olarak adlandırılırlar (Berry, 1993, s. 18). Ancak buradan her yansız tahmincinin tüm yanlı tahmincilerden daha iyi (etkin) olduğu sonucuna ulaşılamaz. En küçük kareler yönteminden farklı bir yöntemle yanlı ancak daha etkin (daha düşük varyanslı) bir tahminci elde edilebilir. Bu yüzden bir

tahmincinin ne kadar başarılı olduğunu ve de sonuç olarak bir modelin eldeki veri kümesine ne kadar uyumlu olduğunu incelerken, söz konusu tahmincinin hem yanlılık hem de varyans açısından değerlendirilmesi gerekir (Berry, 1993, s. 21). 3. bölümde düzenleyici yöntemler konusunda bu konuya daha ayrıntılı değinilecektir.

2.2.2. Çoklu bağıntı

İki veya daha fazla bağımsız değişken birbirine doğrusal bağımlı ise çoklu bağıntı söz konusudur. \mathbf{x} matrisinin tüm sütunlarının birbirinden bağımsız olması ve aralarında ilişki olmaması, \mathbf{x} 'in sütunlarının birbirine dik (ortogonal) olması anlamına gelir. Ancak diklik çoğu zaman rastlanılan bir durum değildir ve tasarlanmış deneylerde görülür. Bu sebeple sütunları dik olmayan bir \mathbf{x} matrisinin sütunları arasında az veya çok bir bağımlılık olacağından dolayı çoklu bağıntı sorunu bir derece sorundur (Aydın, 2014, s. 365).

Gunst ve Webster (2007, s. 279) çalışmalarında çoklu bağıntının üç temel kaynağının olduğunu belirtmişlerdir. Bunlar;

- i. Aşırı tanımlanmış model
- ii. Örnekleme yöntemleri
- iii. Modelde ya da anakütledeki fiziksel kısıtlardır.

Aşırı tanımlı ya da geniş tanımlı model, modelde gözlem sayısından daha fazla bağımsız değişkenin olmasını ifade eder. Böyle bir durumda model, bazı değişkenler modelden çıkarılarak yeniden tanımlanır. Bunun için bağımsız değişken kümesinin ideal bir altkümesinin seçilmesi için ön analizler yapılmalıdır. Problemin ikinci olası kaynağı ise araştırmacının bilerek veya bilmeyerek bağımsız değişkenler uzayının yalnızca bir alt uzayından örneklem seçmesidir. Çoklu bağıntının üçüncü olası kaynağı ise model ya da anakütledeki fiziksel kısıtlardır. Böyle bir durumda örnekleme yönteminden bağımsız biçimde araştırma için önemli olan değişkenler arasındaki ilişkiden kaynaklı olarak çoklu bağıntı ortaya çıkabilir (Gunst ve Webster, 2007, s. 278).

En küçük kareler tahmincisi (2.27)'de verildiği üzere $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ biçimindedir. Eşitlikten görülebileceği üzere $\hat{\beta}$ tahmin vektörünün elde edilebilmesi için $n \times p$ boyutlu \mathbf{x} matrisinin rankı ile $\mathbf{x}'\mathbf{x}$ ilişki matrisinin rankının p 'den (değişken sayısından) küçük olmaması gerekir. İlişki matrisinin rankı p 'den küçük olduğunda $\mathbf{x}'\mathbf{x}$ matrisinin

determinantı sıfıra eşit olacak $|\mathbf{X}'\mathbf{X}|=0$ ve matrisin tersi alınamayacaktır. Tam çoklu bağıntı olarak adlandırılan bu durumda EKK tahmincisi $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ denkleminde yararlanılamaz (İmir, 1986, s. 35).

\mathbf{x} matrisinin j . sütunu \mathbf{X}_j olsun. Eğer tamamı sıfır olmayan t_1, t_2, \dots, t_p sabitleri varsa ve

$$\sum_{j=1}^p t_j \mathbf{X}_j = \mathbf{0} \quad (2.40)$$

koşulu sağlanıyorsa $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p$ vektörleri doğrusal olarak bağımlıdır. (2.40) denklemi \mathbf{x} 'in sütunlarının herhangi bir altkümesi için mevcutsa o halde $\mathbf{x}'\mathbf{x}$ matrisinin rankı p 'den küçüktür ve tersi alınamaz (Montgomery ve diğerleri, 2013, s. 286).

Gauss-Markov kuramına göre bilinmeyen $\boldsymbol{\beta}$ 'nin EKK tahmincileri \mathbf{y} 'nin doğrusal fonksiyonlarıdır ve $\mathbf{x}'\mathbf{x}$ ilişki matrisi birim matrise yakın olduğunda doğru sonuçlar üretir. Ancak $\mathbf{x}'\mathbf{x}$ birim matrise yakın olmadığında EKK tahmincileri hatalara karşı duyarlı olacağından elde edilen tahminlerin hatası da büyüyecektir. Çünkü açıklayıcı değişkenler arasındaki ilişki arttıkça ilişki matrisi $\mathbf{x}'\mathbf{x}$ tekil matrise doğru yaklaşır ve böylece $\mathbf{x}'\mathbf{x}$ matrisinin elemanlarının değeri artar. Buna göre $\hat{\boldsymbol{\beta}}$ 'nin varyans-kovaryans matrisi $(Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ 'nin değeri de artar ve böylece regresyon katsayılarının standart hataları da gerçekte olması gerekenden daha büyük olur.

Çoklu bağıntının EKK tahmincileri üzerindeki etkilerini açıklayabilmek için iki değişkenli bir doğrusal regresyon modelini ele alalım. Değişkenleri ölçek birimlerinden artırmak için birim uzunluk ölçeğine göre standartlaştırılmış olsun.

$$x_{ij} = \frac{X_{ij} - \bar{X}_j}{\sqrt{s_{jj}}} \quad i=1,2,\dots,n \quad j=1,2,\dots,k \quad (2.41)$$

(1.41)'de verilen s_{jj} düzeltilmiş kareler toplamıdır ve $s_{jj} = \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2$ 'ye eşittir.

Bağımlı değişken ise

$$y_i = \frac{Y_i - \bar{Y}}{\sqrt{KT}} \quad (2.42)$$

formülüyle standartlaştırılır ve kareler toplamı (KT) $\sum_{i=1}^n (Y_i^2 - n\bar{Y}) = \sum_{i=1}^n (Y_i - \bar{Y})^2$,ye eşittir. Buna göre elde edilen doğrusal model şöyledir:

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (2.43)$$

Bu standartlaştırma sonunda bağımlı ve bağımsız değişkenler sıfır ortalama ve birim uzunluğa sahip olur. Bu değişkenler yardımıyla oluşturulan regresyon modelinin EKK normal denklemleri aşağıdaki gibidir:

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \quad (2.44)$$

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix} \quad (2.45)$$

(1.45)'teki $\mathbf{X}'\mathbf{X}$ matrisinin tersi

$$\mathbf{C} = \mathbf{X}'\mathbf{X}^{-1} = \begin{bmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{bmatrix} \quad (2.46)$$

olur (Montgomery ve diğerleri, 2013, s. 288). Eşitlik (2.46)'daki denklemden görüleceği üzere $|r_{12}| \rightarrow 1$ oldukça tüm matris elemanlarının değerleri sonsuza yaklaşır. İki değişken arasında tam bağıntı olmaması durumunda bile ters matrisin öğelerinin çok büyük çıkmasına yol açar. $|r_{12}| \rightarrow 1$ oldukça $Var(\hat{\beta}_j) = C_{jj}\sigma^2 \rightarrow \infty$ 'a yaklaşacağından $\boldsymbol{\beta}$ parametre tahminlerinin standart hataları da buna göre artar. Bunun sonucunda t değerleri olduğundan daha küçük çıkar ve 2. tip hataya yol açabilir. Ayrıca $|r_{12}| \rightarrow 1$ oldukça iki değişkenin kovaryansı da $Kov(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\sigma^2 \rightarrow \pm\infty$ 'a yaklaşır (Montgomery ve diğerleri, 2013, s. 289).

Çoklu bağıntının $\boldsymbol{\beta}$ 'nın EKK tahmin değerlerini de artırmaktadır. $\boldsymbol{\beta}$ 'nın üzerindeki etkisini gösterebilmek adına $\hat{\boldsymbol{\beta}}$ vektörünün varyans-kovaryans matrisini ve beklenen değerden uzaklığını inceleyelim.

$$\begin{aligned} (i) \quad Var(\hat{\boldsymbol{\beta}}) &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \\ (ii) \quad L_1 &= \hat{\boldsymbol{\beta}} \text{ ile } \boldsymbol{\beta} \text{ arasındaki uzaklık} \end{aligned} \quad (2.47)$$

(2.47)'ye göre kareli uzaklık (2.48)'e eşit olur.

$$L_1^2 = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \quad (2.48)$$

Buna göre kareli uzaklığın beklenen değeri $E(L_1^2)$ aşağıdaki gibi elde edilir:

$$E(L_1^2) = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sum_{j=1}^p E(\hat{\beta}_j - \beta_j)^2 = \sigma^2 \text{Tr}(\mathbf{X}'\mathbf{X})^{-1} \quad (2.49)$$

(2.49)'daki eşitlikte yer alan Tr matrisin izidir ve $\mathbf{X}'\mathbf{X}$ ilişki matrisinin köşegen elemanlarının toplamına eşittir. Çoklu bağıntı var durumunda \mathbf{X} ortogonallikten uzaklaşacağı için en az bir j . özdeğeri küçülecek ve $\hat{\boldsymbol{\beta}}$ tahmini ile gerçek parametre $\boldsymbol{\beta}$ arasındaki uzaklık artacaktır. $\mathbf{X}'\mathbf{X}$ matrisinin bazı özdeğerleri küçük olur. Buna göre (standartlaştırılmış ölçekte) tahmin edilen katsayı vektörünün karesinin büyüklüğü $\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}$ 'nin beklenen uzunluğu, asıl katsayı vektörünün karesi olan $\boldsymbol{\beta}'\boldsymbol{\beta}$ 'dan uzundur. Bu özellikler $\mathbf{X}'\mathbf{X}$ birim matristen kötü koşullandırılmış (*ill-conditioned*) bir matrise doğru yaklaştıkça $\hat{\boldsymbol{\beta}}$ 'nin belirsizliğinin artacağını gösterir. Bir matrisin köşegen elemanlarının toplamı (izi) özdeğerlerin toplamına eşittir. $\mathbf{X}'\mathbf{X}$ matrisinin özdeğerlerini

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k = \lambda_{\min} > 0 \quad (2.50)$$

biçiminde ifade edersek (2.49) eşitliğini özdeğerler yardımıyla aşağıdaki gibi ifade edebiliriz (A. E. Hoerl ve Kennard, 1970a, s. 56):

$$E(L_1^2) = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j} \quad (2.51)$$

Eşitlik (2.51)'e göre eşitlik (2.49)

$$E(\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}'\boldsymbol{\beta} + \sigma^2 \text{Tr}(\mathbf{X}'\mathbf{X})^{-1} = \boldsymbol{\beta}'\boldsymbol{\beta} + \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} \quad (2.52)$$

biçiminde de gösterilebilir. Sonuç olarak en küçük kareler yöntemi mutlak değerce gerçekte olandan daha büyük katsayı tahminleri vermektedir. Denklem (2.52)'ye göre $\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}$ 'nin beklenen değeri her zaman yüksek olacaktır. Ancak tekil bir matrise yakın olan $\mathbf{X}'\mathbf{X}$ matrisinin özdeğerleri sıfıra yaklaşacağından dolayı katsayı vektörünün tahmini yukarı yönlü aşırı bir yanlılığa sahip olur. Ayrıca katsayıların işaretleri de değişebilir (R. W. Hoerl, 2020, s. 423).

Sonuç olarak çoklu bağıntı, regresyon katsayı tahminlerinin değerce ve işaretçe gerçektekinden oldukça farklı olmasına neden olmaktadır. Ayrıca katsayı tahminleri örneklemeden örnekleme farklılıklar gösterebilir. Buna ek olarak gerçekte olduğundan çok daha farklı regresyon katsayı tahminlerinin yer aldığı bir modelden elde edilen \hat{Y} tahminlerinin de tutarlılığı ve doğruluğu azalacaktır (Şıklar, 2000, s. 81).

Bir regresyon modelinde çoklu bağıntının giderilmesi için aşağıdaki yöntemlerden yararlanır:

- Daha fazla bilgi toplama
- Değişken seçimi
- Bağımsız değişkenlerin kümeleştirilmesi
- Yanlı tahmin yöntemleri

Veri kümesindeki gözlem sayısının yetersiz olması çoklu bağıntıya yol açabileceğinden ilk olarak gözlem sayısı artırılarak parametre tahminlerinin varyanslarının küçülmesi sağlanabilir. İkinci olarak aralarında bağıntı olduğu gözlemlenen değişkenler arasından bağımlı değişken üzerinde daha az açıklayıcılığa sahip değişken(ler)in modelden çıkarılması yoluyla tahminlerin varyansı azaltılabilir. Buna ek olarak aynı şeyi açıkladığı düşünülen ve birbirleriyle bağlantılı değişkenlerden yalnızca biri seçilebilir. Bunun sonucunda serbestlik derecesi $(n - k)$ artacağından tahminlerin standart hataları azalacak ve regresyon katsayılarının etkinliği artacaktır. Son olarak kuramsal açıdan modeldeki değişkenler değiştirilmek istenmiyor veya çok boyutlu bir veri kümesindeki değişkenlerden bazılarının modelden çıkarılması yüksek hesaplama maliyetleri içeriyorsa yanlı tahmin yöntemlerine başvurulabilir. Bu yöntemlerden en sık başvuru alanlar temel bileşenler regresyonu, ridge regresyon ve özdeğerler regresyonudur (Şıklar, 2000, s. 84). Hoerl ve Kennard 'ın (1970a, 1970b) geliştirdiği ridge regresyon yöntemi Bölüm 3.2'de incelenecektir.

2.2.3. Belirlilik katsayısı

Belirlilik katsayısı, bağımlı değişkendeki değişimin ne kadarının bağımsız değişkenler tarafından açıklandığının bir ölçüsüdür ve R^2 sembolü ile gösterilir.

$$R^2 = \frac{RKT}{GKT} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.53)$$

(2.53)'den görüleceği üzere belirlilik katsayısı, model tarafından açıklanan değişimin toplam değişime oranıdır. Çoklu doğrusal regresyon analizi için çoklu belirlilik katsayısı ise matris formunda şöyle hesaplanır:

$$\mathbf{R}^2 = \frac{\mathbf{RKT}}{\mathbf{GKT}} = \frac{\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{n}\bar{y}^2}{\mathbf{y}'\mathbf{y} - \mathbf{n}\bar{y}^2} \quad (2.54)$$

Belirlilik katsayısı, sıfır ile bir arasında değerler alır. R^2 'nin sıfıra yakın olması, bağımlı değişkendeki değişimin çok az bir kısmının bağımsız değişkenler tarafından açıklandığını; bire yakın olması ise bağımlı değişkendeki değişimin büyük bir kısmının bağımsız değişkenler tarafından açıklanabildiğini gösterir (James ve diğerleri, 2013, s. 70).

Çoklu doğrusal regresyon analizinde belirlilik katsayısının değeri, modele her yeni bağımsız değişken eklendiğinde artar çünkü belirlilik katsayısının payının değeri artarken payı aynı kalmaktadır. Bu sakıncayı gidermek için ise düzeltilmiş belirlilik katsayısından yararlanır:

$$\bar{R}^2 = 1 - \left[(1 - R^2) \frac{n-1}{n-k} \right] \quad (2.55)$$

Belirlilik katsayısının karekökü ($\sqrt{R^2} = R$) korelasyon katsayısı olarak adlandırılır. Çoklu doğrusal regresyon analizinde hesaplanan korelasyon katsayısına çoklu korelasyon katsayısı adı verilir. Çoklu korelasyon katsayısı, bağımlı değişken ile bağımsız değişkenler arasındaki doğrusal ilişkinin derecesini gösterir. Ancak uygulamada daha çok belirlilik katsayısından yararlanılmaktadır (Şıklar, 2000, s. 55)

2.2.4. Doğrusal regresyon analizinde hipotez testleri

2.2.4.1. Regresyon modelinin anlamlılık testi

Regresyon modeli, bağımlı değişken ile bağımsız değişkenler arasında anlamlı bir ilişki olup olmadığının belirlenmesi için anlamlılık testine tabi tutulur. Bu hipotez testindeki hipotezler şunlardır:

$$\begin{aligned}
H_0 : \beta_0 = \beta_1 = \dots = \beta_k \\
H_1 : \beta_j, \text{ en az bir } j \text{ için}
\end{aligned}
\tag{2.56}$$

Sıfır hipotezi, regresyon katsayılarının tamamının sıfıra eşit olduğunu (tüm katsayıların anlamsız olduğunu) gösterirken; karşıt hipotez en az bir regresyon katsayısının sıfırdan farklı bir değere sahip olduğunu gösterir. Sıfır hipotezinin reddedilmesi, bağımsız değişkenlerden en az birisinin modele anlamlı bir katkısı olduğu anlamına gelir. Diğer bir ifade ile bağımlı ve bağımsız değişkenler arasında $1-\alpha$ güvenilirlik düzeyinde doğrusal bir bağıntı kurulabilir. Bununla birlikte H_0 'ın kabul edilmesi tüm β_j 'lerin sıfırdan farklı olduğunu değil, en az birinin sıfırdan farklı olduğunu gösterir. Bu hipotez testini gerçekleştirmek için varyans analizinden yararlanılır (Montgomery ve diğerleri, 2013, s. 84).

Bir bağımlı değişkendeki toplam değişim (2.54)'deki fonksiyonla ifade edilir:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 - \sum_{i=1}^n (Y_i - \bar{Y})^2
\tag{2.57}$$

Bu fonksiyondaki bileşenler şu şekilde de ifade edilmektedir:

$$\text{Genel Kareler Toplamı} = \text{Regresyon Kareleri Toplamı} - \text{Hata Kareleri Toplamı} \tag{2.58}$$

Genel kareler toplamı (GKT), bağımlı değişkendeki toplam değişimi; regresyon kareleri toplamı (RKT) model tarafından açıklanan değişimi; hata kareleri toplamı (HKT) ise model tarafından açıklanamayan değişimi ifade eder. (2.57)'deki denklem matris formunda şöyle gösterilir:

$$\mathbf{y}'\mathbf{y} - \mathbf{n}\bar{y}^2 = (\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{n}\bar{y}^2) - (\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})
\tag{2.59}$$

Hipotez testinin gerçekleştirilmesi varyans analizi ile mümkündür. Bunun için F istatistiği hesaplanmalıdır (Aydın, 2014, s. 94)

$$F = \frac{RKO}{HKO} = \frac{RKT / k}{HKT / (n - k - 1)} = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{n}\bar{y}^2) / k}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}) / (n - k - 1)}
\tag{2.60}$$

(2.60)'da gösterildiği üzere tümel F testinde regresyon kareleri ortalaması (RKO), hata kareleri ortalamasına (HKO) oranlanır. Anakütle varyansının beklenen değeri HKO'ya eşit olduğuna göre doğrusal regresyon analizinin varsayımları sağlandığında (2.60)'da gösterilen F testinin paydası σ^2 'ye eşit olur. Eğer H_0 hipotezi doğru ise RKO'nun

beklenen değeri de σ^2 'ye eşit olur. ($E(RKT / k) = E(RKO) = \sigma^2$) Buna göre eğer bağımlı değişken ile bağımsız değişkenler arasında anlamlı bir doğrusal ilişki tespit edilemezse F testi (gerekli serbestlik derecesi ve anlamlılık düzeyinde) 1 veya 1'e yakın bir değer alır. Ancak F istatistik değeri, F tablosundan elde edilecek ilgili serbestlik derecesine ve anlamlılık düzeyine ait F kritik değerden daha büyük ($F > F_{(\alpha, k, n-k-1)}$) ise RKO'nun beklenen değeri σ^2 'den anlamlı biçimde büyük olacaktır ($E(RKT / k) = E(RKO) > \sigma^2$). Bu durumda sıfır hipotezi reddedilir ve bağımlı değişken ile bağımsız değişkenler arasında anlamlı bir doğrusal ilişkinin mevcut olduğu sonucuna varılır (James ve diğerleri, 2013, s. 76).

2.2.4.2. Regresyon katsayılarının anlamlılık testi

Doğrusal regresyon analizinde F -testi sonucunda modelde en az bir bağımsız değişkenin bağımlı değişken ile anlamlı bir doğrusal ilişkiye sahip olduğu sonucuna varıldıktan sonra hangi değişkenlerin anlamlı bir ilişkiye sahip olduğu sorusu ortaya çıkar. Bu amaçla, bir regresyon katsayısının anlamlılığını sınamak için t-testinden (kısmi F testi) yararlanır. Bir β_j katsayısının anlamlılığını sınamak için gerekli hipotezler şöyle kurulur:

$$\begin{aligned} H_0 : \beta_j &= 0 \\ H_1 : \beta_j &\neq 0 \end{aligned} \quad (2.61)$$

Kısmi F testi, denklemdeki her bir değişkenin, diğer değişkenlerin etkileri sabit tutulduğunda, bağımlı değişkeni açıklamadaki önemini belirlemek ve değişkenlerin bağımlı değişkeni açıklamadaki önemlilik derecelerini saptamak amacıyla gerçekleştirilir. Sıfır hipotezinin kabul edilmesi, j . bağımsız değişkenin modele anlamlı bir katkısı olmadığını ifade eder ve değişken modelden çıkarılabilir. Bu hipotez için test istatistiği

$$t_h = \frac{\beta_j}{s(\beta_j)} \quad (2.62)$$

biçiminde hesaplanır. Burada $s(\beta_j)$, β_j 'nin standart hatasıdır ve varyans-kovaryans matrisinde $(\sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ yer alan β_j 'nin varyansının karekökünün alınmasıyla hesaplanır. (2.62)'de hesaplanan t_h değeri, istenen anlamlılık düzeyinde ve gerekli

serbestlik derecesine göre t tablosundan elde edilen kritik t_k değerinden büyükse sıfır hipotezi reddedilir ve katsayının anlamlı olduğu sonucuna varılır. Bu sonuç, diğer bağımsız değişkenler modelde yer aldığında j . bağımsız değişkenin modele anlamlı bir katkı sağladığı biçiminde yorumlanır (Montgomery ve diğerleri, 2013, s. 88).

2.3. Çok Değişkenli Regresyon Modellerinde Değişken Seçimi

Çoklu doğrusal regresyon çözümlemesinde veriyi en iyi tanımlayacak ve bağımlı değişkeni en iyi açıklayacak değişkenlerin belirlenmesi değişken seçimi ya da en iyi alt küme seçimi olarak adlandırılır. Bir regresyon modeline, bağımlı değişken ile ilişkili olduğu düşünülen tüm değişkenlerin dahil edilmesi fikri her ne kadar cazip gelse de böyle bir tercihin bazı sakıncaları olacaktır. X açıklayıcı değişkenler ile Y bağımlı değişkeni arasındaki ilişkiyi gösteren gerçek matematik fonksiyonu bilinmediği müddetçe elde edilen veri yalnızca değişkenler arasındaki ilişkiyi temsil edecek bir modeli elde etmek için değil, aynı zamanda mevcut çok sayıda değişkenler arasından en iyi temsile sahip olanların belirlenmesi amacıyla da kullanılır.

Bir regresyon modelinin belirlenmesinde birbirleriyle çatışan iki amaç bir arada yer almaktadır: (i) Her bir bağımsız değişken ek bilgi içerdiğinden dolayı Y 'nin tahmin değerlerinde bilgi kaybı olmamasını istediğimiz için olabildiğince fazla sayıda bağımsız değişkenin modelde yer alması istenir. (ii) Bağımsız değişken sayısı arttıkça Y 'nin tahmin değerlerinin varyansı artacağından dolayı modelde olabildiğince az sayıda bağımsız değişken olması arzu edilir (Montgomery ve diğerleri, 2013, s. 328).

Cox ve Snell (1974), çalışmalarında tahmin ve tahmin modelleri için değişken seçimi konusunda farklı zorluklar olduğunu vurgulamışlardır. Verilen yeni bir X için Y 'nin tahmininin yapıldığı analizlerde hangi bağımsız değişkenlerin daha yüksek etkiye sahip olduğundan ziyade tahminin doğruluğu önemlidir. Birbirinden tamamen farklı ancak eşit tahmin doğruluğuna sahip iki model arasında kalındığında, basitlik amacıyla daha az değişkene sahip model ya da ölçülmesi daha kolay olan değişkenlerin yer aldığı model tercih edilir. Ancak $f(X)$ regresyon fonksiyonunun bir tahmini olan $\hat{f}(X)$ fonksiyonunun bulunmak istendiği açıklayıcı modellerde ise bağımlı değişken üzerinde önemli etkilere sahip olan bağımsız değişkenlerin belirlenmesi önemlidir.

Her iki durumda da ölçülmüş tüm değişkenlerin modele eklenmesi hem basitlik açısından hem de modele her bir yeni değişken eklendiğinde hata kareleri ortalaması

yükseleceğinden dolayı doğru bir seçim olmayacaktır. Öte yandan gözleme dayalı ve değişkenler arasındaki ilişkiyi matematiksel bir fonksiyon ile açıklamayı hedefleyen çalışmaların analizinde karşılaşılan zorluklar iki nedenden kaynaklanmaktadır: Ölçülmüş ve ilgili değişkenlerden bazılarının analizden çıkarılması ve açıklayıcı değişkenler arasındaki doğrusal bağıntı. Özellikle çok sayıda açıklayıcı değişken modele alınacaksa açıklayıcı değişkenler arasında bağıntı olması kaçınılmazdır. Bu durumda katsayıların yorumu zorlaşır ve birbirinden farklı modeller veriye benzer derecede iyi uyumlanabilirler. Ayrıca, değişkenlerin seçiminde değişkenlerin sağladığı ek bilgi göz önüne alınır. Hangi değişkenler birbirini ikame edebilir ya da hangi değişkenlerin modelde yer alması zorunludur gibi soruların cevaplanması gerekir (Cox ve Snell, 1974, s. 53).

Tümel F testinin sonucunda sıfır hipotezi reddedildiğinde kısmi F testi aracılığıyla hangi β_j 'lerin anlamsız olduğu belirlenerek modele katkısı olmayan değişkenlerin belirlenebileceğinden kısım 2.2.4'te bahsedilmişti. Ancak kısmi F testlerinin incelenerek değişkenlerin seçilmesi tüm bağımsız değişkenlerin birbirine dik olması durumunda mümkündür. Örneğin k adet değişkenin yer aldığı bir modelde $\beta_1 = 0$ ve $\beta_2 = 0$ hipotezleri kısmi F testinde kabul edilmiş olsun. Daha sonra modelden X_1 çıkarıldığında ve $k - 1$ değişkenle model tekrar kurulduğunda $\beta_2 = 0$ hipotezi reddedilebilir. Bağımsız değişkenlerin arasında makul ölçüde de olsa bir ilişkinin olması ve k adet değişkenin yer aldığı bir değişken kümesinden 2^k adet farklı model kurulabileceğinden dolayı kısmi F testlerine dayanarak değişken seçiminin işlem maliyeti oldukça yüksektir.

Yukarıda açıklanan sebeplerden dolayı en küçük kareler regresyon analizinde değişkenlerin seçilmesi için yinelemeli yöntemler geliştirilmiştir. Bu yöntemler izleyen kısımda açıklanacaktır.

2.3.1. Tüm olası alt kümeler yöntemi

Tüm olası alt kümeler yönteminde, doğrusal regresyon modellerinde değişken seçimi ve parametre tahmini için kullanılan en eski yöntemlerden biridir. Hocking ve Leslie (1967, s. 531), en iyi değişken alt küme seçim probleminin iki yönlü olduğunu belirtmişlerdir. Problemin ilk ve en önemli yönü birbirine oldukça benzer iki altküme arasından hangisinin seçileceğinin belirleyecek bir seçim ölçütünün belirlenmesidir.

Ancak deęişken sayısı çok fazla olduęunda bazı seçim ölçütlerinin uygulanması hesaplama maliyeti açısından uygun olmayabilir. Bu durum da bizi problemin ikinci yönüne götürmektedir. Bu da iki regresyon modelini karşılaştırmada kullanılacak ölçütün düşük hesaplama maliyetine sahip olmasıdır.

En iyi alt küme seçimi yönteminin adımları aşağıdaki gibidir:

1. M_0 bağımsız deęişkenin olmadığı yalnızca sabit katsayının yer aldığı ve her bir gözlem için örneklem ortalamasını tahmin eden sıfır modeli olsun.

2. Her $p = 1, \dots, k$ için;

a) k adet deęişken için mümkün tüm $\binom{k}{p}$ modelleri uyumlandırılır.

b) $\binom{k}{p}$ modelleri arasından en düşük hata kareleri toplamına sahip model

seçilir ve M_p olarak isimlendirilir.

3. M_0, \dots, M_k modelleri arasından en iyi seçim ölçütü deęerine sahip model en iyi alt küme modeli olarak seçilir (James ve dięerleri, 2013, s. 205).

Algoritmanın ikinci adımında 2^k olası regresyon model sayısı, hiçbir bağımsız deęişkenin olmadığı sıfır modeli ile birlikte, $k + 1$ adete indirilmiş olur. Hata kareleri toplamı fonksiyonu monoton azalan bir fonksiyon olduęundan dolayı, hata kareleri toplamlarına bakarak $k + 1$ model arasından en iyi tek bir modelin seçilmesi mümkün deęildir. Modeldeki deęişken sayısı arttıkça hata kareleri toplamı aratacağından dolayı en düşük HKT'ye sahip model tüm deęişkenlerin yer aldığı M_k modeli olacaktır. Benzer şekilde monoton artan bir fonksiyon olan R^2 belirlilik katsayısına göre en iyi model M_k modeli olacaktır.

Bu nedenle $k + 1$ model arasından en iyi modelin belirlenmesi için yaygın olarak literatürde Mallow'un C_p istatistięi, düzeltilmiş R^2 , Akaike bilgi kriteri, Bayes bilgi kriteri ve çapraz geçerlilik yöntemlerinden biri kullanılmaktadır (James ve dięerleri, 2013, s. 208). Bu çalışmada çapraz geçerlilik yönteminden yararlanılacaktır.

2.3.2. İleriye doğru seçim yöntemi

İleriye doğru seçim yönteminde hiçbir değişkenin yer almadığı sıfır modelinden başlanarak nihai modele ulaşana kadar her adımda bir değişkeni modele ekleyerek en iyi değişken alt kümesi belirlenir. Tüm olası alt kümelerin aksine daha düşük bir hesaplama maliyetine sahiptir. Ancak modele ilk eklenen değişkenin hangi değişken alt kümesinin seçileceğinde önemli ölçüde bir etkisi vardır. Dolayısıyla ilk başta seçilen değişkenin yanlış belirlenmesi seçilen değişken alt kümesinin temsil gücünü azaltır.

İleri doğru seçim yöntemi iç içe geçmiş bir model dizisini kullanarak en iyi modele doğru ilerleyen bir algoritmaya sahiptir ($\emptyset \dots \subset M_{p-1} \subset M_p \subset M_{p+1}$). İlk olarak sıfır modeli ile başlar ve her seferinde bir değişken ekleyerek ilerler. Özellikle p büyük olduğunda tüm olası regresyonlar yöntemine göre daha düşük bir hesaplama maliyetine sahiptir.

İlk olarak her bir X_j değişkeni için basit doğrusal regresyon modelleri hesaplanır. Bu modeller arasından istatistiksel olarak anlamlı en yüksek F istatistiğine sahip değişken modele ilk olarak seçilir. İlk değişken belirlendikten sonra en yüksek kısmi F değerine sahip olan bir diğer değişken modele eklenir.

1. M_0 bağımsız değişkenin olmadığı yalnızca sabit katsayının yer aldığı ve her bir gözlem için örneklem ortalamasını tahmin eden $y = \beta_0$ sıfır modeli olsun.
2. Her $p = 1, \dots, k$ için;
 - a) M_p modelindeki değişken sayısını bir artıran tüm $(k - p)$ adet model göz önüne alınır.
 - b) $(k - p)$ model arasından en düşük HKT'ye veya en büyük belirlilik katsayısına sahip) en iyi model belirlenir ve M_{p+1} olarak adlandırılır.
3. M_0, \dots, M_k modelleri arasından en iyi seçim ölçütü değerine sahip model en iyi model olarak seçilir (James ve diğerleri, 2013, s. 210).

2.3.3. Geriye doğru seçim yöntemi

Geriye doğru seçim yönteminde, ileriye doğru seçim yöntemindeki sürecin tam tersi uygulanır. İlk olarak tüm değişkenler modele eklenir. Daha sonra her adımda en yüksek

p değerine sahip (istatistiksel olarak en az anlamlı) olan değişkenler her bir adımda modelden çıkarılır (Montgomery ve diğerleri, 2013, s. 345).

1. M_p tüm p bağımsız değişkenlerin yer aldığı tam model olsun.
2. Her $p = k, k-1, \dots, 1$ için;
 - a) Toplamda $p-1$ bağımsız değişken için M_p modelindeki bağımsız değişkenlerden biri hariç tümünü içeren tüm p modelleri göz önüne alınır.
 - b) p model arasından (en düşük HKT'ye veya en büyük belirlilik katsayısına sahip) en iyi model belirlenir ve M_{p-1} olarak adlandırılır.
3. M_0, \dots, M_k modelleri arasından en iyi seçim ölçütü değerine sahip model en iyi model olarak seçilir (James ve diğerleri, 2013, s. 211).

3. DÜZENLİLEŞTİRİCİ YÖNTEMLER

EKK tahminlerinin hata karelerinin toplamının en küçüklenmesi ile elde edildiği 2. Bölüm’de açıklanmıştır. EKK yöntemi regresyon problemini konveks bir en küçükleme problemine dönüştürdüğü için sağladığı hesaplama kolaylığından dolayı literatürde yaygın olarak kullanılmasına karşın iki önemli dezavantajı vardır. Birincisi, EKK tahminleri en iyi doğrusal yansız tahminler olsa da modeldeki bağımsız değişken sayısı arttıkça yüksek varyansa sahip olurlar. Tahminlerin doğruluğunu artırmak için regresyon katsayı tahminlerinin doğruluğu $\hat{\beta}_j$ ’lerin düzenleyici yöntemler ile daraltılarak ya da bazı katsayıların tahminleri doğrudan sıfıra eşitlenerek artırılabilir. Böylece bir miktar yanlılık kabul edilerek, tahminlerin varyansı azaltılır ve toplam tahmin doğruluğu artırılır. İkinci önemli dezavantajı ise modeldeki değişken sayısı arttıkça EKK yöntemi ile en iyi değişken kümesini belirlemek zorlaşmaktadır (Robert Tibshirani, 1996, s. 267). EKK yönteminde bir değişkenin bağımlı değişken ile zayıf ilişkisi olsa dahi katsayısı hiçbir zaman sıfıra eşit olamayacağından dolayı değişken seçimi için geliştirilen bazı algoritmalar mevcuttur. Bunlardan en yaygın olarak kullanılanları ve bu yöntemlerin dezavantajları Bölüm 2.3.’te açıklanmıştır.

EKK yönteminde karşılaşılan bir diğer sorun da çoklu bağıntıdır. $n \times k$ boyutlu \mathbf{x} değişken matrisinin rankı değişken sayısına eşittir ($rank(\mathbf{X}) = k < n$). \mathbf{x} matrisinin rankı ile $\mathbf{x}'\mathbf{x}$ k ’den küçük olduğunda çoklu bağıntı sorunu ortaya çıkacaktır. Öte yandan \mathbf{x} matrisi doğrusal bağıntıya yaklaştıkça $Var(\hat{\beta}_{EKK}) = (\mathbf{X}'\mathbf{X})^{-1} \sigma^2$ varyans tahmini de artar. Doğrusal bağıntı sebebiyle HKO artar ve $\hat{\beta}_{EKK}$ tahminlerinin doğruluğu azalır.

Yukarıda sıralanan sebeplerden dolayı geliştirilen literatürde daraltıcı regresyon yöntemleri veya ceza yöntemleri olarak da adlandırılan düzenleyici yöntemler geliştirilmiştir. Çoklu bağıntı durumunda daha iyi tahminler elde etmek için Hoerl ve Kennard (1970b, 1970a) ridge regresyon yöntemini, Tibshirani (1996) regresyon katsayılarını daraltırken aynı anda değişken seçimini gerçekleştiren lasso regresyonu, Breiman (1995) en iyi değişken kümesinin belirlenmesi için nonnegative garrotte yöntemini geliştirmişlerdir. Frank ve Friedman (1993) ise çalışmalarında optimum bir ceza parametresinin belirlenmesinin regresyon tahminlerinin doğruluğunu artıracaklarını belirtmişler ve daha sonra literatürde bridge regresyon olarak adlandırılan genel düzenleyici regresyon modelini önermişlerdir. İzleyen kesimde düzenleyici

yöntemler arasında bridge regresyon, ridge regresyon, lasso regresyon ve elastic-net yöntemleri açıklanmıştır.

3.1. Bridge Regresyon

En küçük kareler, tüm olası alt kümeler ve ridge regresyon yöntemlerini kısmi en küçük kareler ve asal bileşenler regresyon yöntemleriyle karşılaştırmalı olarak inceleyen Frank ve Friedman (1993), hata kareleri toplamını en küçükleyen fonksiyon bir ceza fonksiyonu ile kısıtlandırıldığında modelin tahmin doğruluğunun arttığını genel bir çerçeveye oturarak açıklamışlardır.

Bridge regresyonun çözümünde hata kareleri toplamı bir ceza fonksiyonu ile kısıtlandırılır. $q > 0$ ve $t \geq 0$ için bridge regresyon problemi denklem (3.1)'deki gibi ifade edilir.

$$\min_{\beta} \left\{ \sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^k \beta_j X_{ij} \right)^2 \right\} \quad (3.1)$$

$$\sum_{j=1}^k |\beta_j|^q \leq t$$

$\sum_{j=1}^k |\beta_j| \leq t$ kısıtı ℓ_q norm kısıtı olarak da yazılabilir: $\|\beta\|_q^q \leq t$. (3.1) denklemi matris-vektör notasyonu ile de gösterilebilir. $\mathbf{y} = (y_1, \dots, y_N)$ $N \times 1$ boyutlu bağımlı değişken vektörü, \mathbf{X} $N \times k$ boyutlu bağımsız değişken matrisi olsun. Bu durumda denklem (3.1)'deki optimizasyon problemi şöyle ifade edilir:

$$\min_{\beta} \left\{ \|\mathbf{y} - \beta_0 \mathbf{1} - \mathbf{X}\beta\|_2^2 \right\} \quad (3.2)$$

$$\|\beta\|_q^q \leq t$$

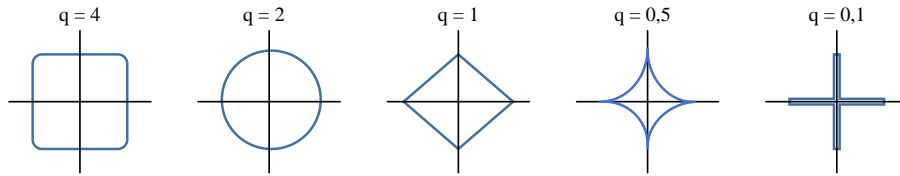
$\mathbf{1}$ N adet birden oluşan vektörü, $\|\cdot\|_2$ ise Öklid uzaklığını temsil etmektedir. t kısıtı regresyon katsayılarının tahminlerini kısıtlayan parametredir. t , sifıra yaklaştıkça ($t \rightarrow 0$) regresyon parametre tahminleri de sifıra doğru yaklaşacaktır. t , $+\infty$ 'a yaklaştıkça ($t \rightarrow +\infty$) optimum çözüm alanı genişler ve parametre tahminleri büyür. t , yeterince büyük olduğunda (3.2) probleminin çözümünde elde edilen $\hat{\beta}$ tahminleri EKK tahminlerine eşit olur (Fu, 1998, s. 398).

(3.1)'de verilen denklemin dual formu $q > 0$ ve $t \geq 0$ için denklem (3.3)'te verilmiştir.

$$\beta_{bridge} = \min_{\beta_0, \beta_j} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^k |\beta_j|^q \right\} \quad (3.3)$$

(3.1) ve (3.3)'de verilen denklemler arasında bir fark olmadığından aynı çözüme sahiptirler. (3.1)'de verilen problem kısıtlı regresyon olarak adlandırılırken, (3.3)'deki problem cezalı regresyon olarak adlandırılır. Her bir $0 \leq \lambda \leq +\infty$ değeri için her zaman bir $t \geq 0$ sabiti mevcuttur (Fu, 1998, s. 398). λ ayar parametresi $+\infty$ 'a yaklaştıkça ($\lambda \rightarrow +\infty$) ceza katsayısı büyür ve katsayı tahminleri sıfıra yaklaşır. Ayar parametresinin değeri sıfıra yaklaştıkça ($\lambda \rightarrow 0$) katsayı tahmin değerleri büyür. $\lambda = 0$ olduğunda problem EKK problemine dönüşür ve çözümü EKK tahminlerine eşit olur. Ceza parametresi çapraz geçerlilik vb. bir hiperparametre optimizasyonu yöntemiyle belirlenmektedir (Fu, 1998, s. 398).

Şekil 3.1.'de $t = 1$ durumunda farklı q değerleri için kısıt bölgeleri gösterilmiştir. q arttıkça ceza parametresinin kısıt alanının genişlediği görülebilir. q sıfıra eşit olduğunda kısıt alanı ortadan kalkar (3.3) eşitliği en küçük kareler problemine dönüşür. $0 < q \leq 1$ durumunda değişken seçimi gerçekleşir. $q > 1$ olduğunda değişken katsayıları daraltılır ancak regresyon katsayıları hiçbir zaman sıfıra eşit olamayacağından değişken seçimi gerçekleşmez (Park ve Yoon, 2011, s. 3507).



Şekil 3.1 $t = 1$ durumunda farklı q değerleri için kısıt bölgeleri (Hastie ve diğerleri, 2015)

(3.3) denklemi $q = 1$ için lasso regresyonunu $q = 2$ için ridge regresyonunu ve $q = 0$ için tüm olası alt kümeler regresyonunu verir. Bu sebepten dolayı bridge regresyon daraltıcı (düzenleştirici) yöntemlerin genel bir formu olarak görülebilir. Dikkat edilirse $q = 1$ durumunda elde edilen lasso regresyon konveks bir kısıt alanı sunarken, değişken seçiminin de eş zamanlı olarak gerçekleştirilmesini sağlamaktadır. Şekil 3.1'den görülebileceği üzere (3.3) denkleminin ridge ve lasso versiyonlarında oluşturulan kısıt

alanlarında problem konveks bir problem olarak korunduğundan dolayı çok sayıda değişkene sahip regresyon problemleri için uygun yöntemlerdir. Öte yandan tüm olası alt kümeler yöntemi ise problemi, konveks olmayan ve kombinasyonlar içeren bir optimizasyon problemine dönüştürür. (Hastie ve diğerleri, 2015, s. 22).

3.1.1. Bridge tahmincisinin özellikleri

Fu (1998), çalışmasında verilen herhangi bir $q \geq 1$ değeri için bridge regresyonun çözüm yöntemlerini incelemiştir. Denklem (3.3)'te verilen minimizasyon problemini göz önüne alırsak bu eşitliği bir G fonksiyonu ile ifade edebiliriz: $G(\beta, X, y, \lambda, q) = HKT + \lambda \sum |\beta|^q$. G fonksiyonu β 'ya göre konvekstir ve $\|\beta\|$ Öklid uzaklığı sonsuza yaklaştıkça ($\|\beta\| \rightarrow +\infty$) G fonksiyonu da sonsuza yaklaşır ($G \rightarrow +\infty$). Buna göre $\hat{\beta} = \arg \min_{\beta} G(\beta, X, y, \lambda, q)$ eşitliğini sağlayan bir $\hat{\beta}$ tahmini mevcuttur ve böylece G fonksiyonu enküçüklenebilir. G 'nin β_j 'ye göre ($\beta_j \neq 0, j=1, \dots, p$) kısmi türevleri $S_j = (\beta, X, y) = \partial HKT / \partial \beta_j$ ve $d(\beta_j, \lambda, q) = \lambda q |\beta_j|^{q-1} \text{sgn}(\beta_j)$ sıfıra eşitlenirlerse ($\partial G / \partial \beta_j = 0$) (3.3)'te verilen denklem (3.4)'teki lineer denklem sistemi yardımıyla çözülebilir:

$$\left\{ \begin{array}{l} S_1(\beta, X, y) + d(\beta_1, \lambda, q) = 0 \\ \vdots \\ S_p(\beta, X, y) + d(\beta_p, \lambda, q) = 0 \end{array} \right\} \quad (3.4)$$

β , p boyutlu B uzayında bir vektör; X , $n \times p$ boyutlu bir matris ve y , n boyutlu bir R^n örneklem uzayında yer alan bir vektör olsun. Sabit $X, y, \lambda \geq 0$ ve $q \geq 1$ için $S_j(\beta, X, y): B \rightarrow R$, $\beta \rightarrow S_j(\beta, X, y), j=1, \dots, p$ gerçel fonksiyonları ile $d(\beta_j, \lambda, q) = \lambda q |\beta_j|^{q-1} \text{sgn}(\beta_j)$ fonksiyonları tanımlansın. $S = (S_1, \dots, S_p)^T$ olduğunda (3.4) denklemi için bir çözüm elde edilir (Fu, 1998, s. 399).

$q > 1$ ve $\lambda > 0$ değerleri için S , β 'ya göre sürekli ve türevlenebilir bir fonksiyon ve Jacobiyen matrisi ($\partial S / \partial \beta$) yarı pozitif tanımlı matris olmak üzere;

- 1) (λ, q) 'da sürekli olduğunda (3.4) denklemi $\hat{\beta}(\lambda, q)$ tek çözümüne sahiptir

2) $\hat{\beta}(\lambda, q)$ çözümünün limiti $q \rightarrow 1+$ 'da mevcuttur. Limit çözümünü $\hat{\beta}(\lambda, +1)$ ile gösterirsek $\lim_{q \rightarrow 1+} \hat{\beta}(\lambda, q) = \hat{\beta}(\lambda, +1)$ olur (Fu, 1998, s. 400).

3.1.2. Bridge tahmincisinin hesaplanması

Herhangi bir $q > 1$ ve $\lambda > 0$ için bridge regresyon problemini çözmek için yinelemeli yeniden ağırlıklandırılmış en küçük kareler (iteratively reweighted least-squares - IRLS) yönteminden yararlanır.

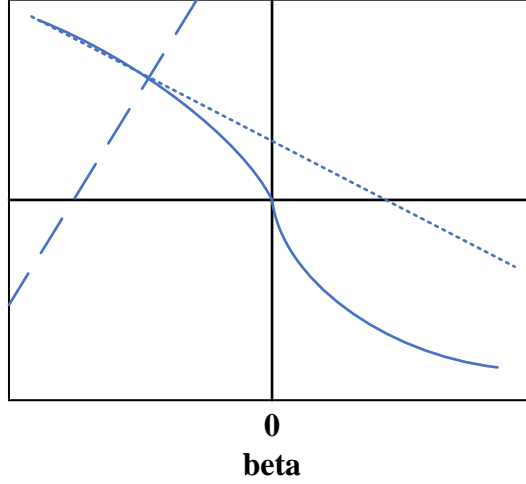
β^{-j} , β_j hariç tüm β 'ları içeren $p-1$ boyutlu bir vektör olsun. Bu durumda (3.4) 'teki j . denklem aşağıdaki gibi ifade edilir:

$$S_j(\beta_j, \beta^{-j}, X, y) = -d(\beta_j, \lambda, q) \quad (3.5)$$

(3.5)'teki denklemin sol tarafı sabit bir β^{-j} için $2x_j^T x_j \beta_j + \sum_{i \neq j} 2x_j^T x_i \beta_i - 2x_j^T y$ 'ye eşittir ve β_j 'nin $2x_j^T x_j$ pozitif eğimine sahip doğrusal bir fonksiyonudur. Eşitliğin sağ tarafı ise β_j 'ye göre doğrusal değildir ve $-\lambda q |\beta_j|^{q-1} \text{sgn}(\beta_j)$ 'ye eşittir. $-\lambda q |\beta_j|^{q-1} \text{sgn}(\beta_j)$ fonksiyonu $q > 1$ için sürekli, türevlenebilir ve monoton azalandır. Ancak $1 < q < 2$ aralığında $\beta_j = 0$ noktasında türevlenemez. (3.5) denklemini $q > 1$ için tek bir çözüme sahiptir ancak $q = 1$ durumunda ya tek çözüme sahiptir ya da çözümü yoktur.

Newton-Raphson yöntemi ile $q > 1$ için bridge tahmincisi hesaplanabilir. Ancak $q < 2$ için $\beta_j = 0$ noktasında d fonksiyonu türevlenebilir olmadığı için yaklaşık bir çözüme ulaşmak adına Fu (1998, s. 400) tarafından Düzenlenmiş Newton-Raphson (Modified Newton-Raphson M-N-R) algoritması önerilmiştir. Buna göre $q < 2$ durumunda çözüm ile orijin arasındaki bir noktada $-d$ fonksiyonunun bir kısmı Şekil 3.2'de görüldüğü gibi tanjant eğrisi ile değiştirilir. Daha sonra Newton-Raphson yöntemi uygulanarak çözüm elde edilir. Şekilde $-d$ kalın çizgi ile gösterilmiştir. Kesikli çizgi ise S_j fonksiyonu ve noktalı çizgi düzeltilmiş $-d$ fonksiyonudur.

M-N-R



Şekil 3.2. Newton-Raphson yönteminin bridge regresyona uygulanması (Fu, 1998)

Algoritmanın $q > 1$ için geçerli adımları aşağıdaki gibidir:

- i. $\hat{\beta}_0 = \hat{\beta}_{EKK} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$ başlangıç çözümü olsun.
- ii. m. adımda her bir $j = 1, \dots, p$ için $S_0 = S_j(0, \hat{\beta}^{-j}, X, y)$ olsun. $S_0 = 0$ ise $\hat{\beta}_j = 0$ 'dır. $q \geq 2$ durumunda (3.5) denklem sisteminden $\hat{\beta}_j$ çözümünü elde etmek için Newton-Raphson yöntemini uygula. $q < 2$ ise $-d$ fonksiyonunu tanjantına göre düzenle ve sonra Newton-Raphson yöntemini $\hat{\beta}_j$ tek çözümünü elde etmek için düzenlenmiş $-d$ fonksiyonunun yer aldığı (3.5) denkleme uygula. Tüm $\hat{\beta}_j$ 'leri güncelledikten sonra yeni bir tahminci oluştur: $\hat{\beta}_m = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$
- iii. ii. adımı $\hat{\beta}_m$ global minimuma ε uzaklığında yakınsayana kadar tekrar et (Fu, 1998, s. 400).

3.1.3. Bridge tahmincisinin varyansı

Bridge tahmincisi, (3.4) problemi için tek çözüm olduğundan ve hemen hemen her zaman sıfırdan farklı olduğu için varyansı, delta yöntemini kullanarak (3.6)'dan elde edilebilir.

$$\text{var}(\hat{\beta}) = (X^T X + D(\hat{\beta})|_{y_0})^{-1} X^T \text{var}(y) X (X^T X + D(\hat{\beta})|_{y_0})^{-1} \quad (3.6)$$

(3.6)'da verilen $D(\hat{\beta}) = \text{diag}(\lambda q(q-1) |\hat{\beta}_j|^{q-2} / 2)$ 'dir ve y_0 ise örneklem uzayında keyfi bir sabit noktadır. Varyansın tahmini eşitlikte $\hat{\beta}|_{y_0}$ yerine $\hat{\beta}$ 'nin koyulması ve $\text{var}(y)$ 'yi kendi tahmini ile değiştirilmesiyle elde edilebilir.

$F_j = S_j(\hat{\beta}, X, y) + d(\hat{\beta}_j, \lambda, q)$ olmak üzere $\mathbf{F} = (F_1, \dots, F_p)^T$ olsun. Gauss dağılımı için kısmi türevler $\partial F / \partial y = -2X^T$ ve $\partial F / \partial \hat{\beta} = 2X^T X + 2D(\hat{\beta})$ olsun. Kapalı fonksiyon teoreminin $\mathbf{F} = 0$ 'a uygulanmasıyla

$$\frac{\partial \hat{\beta}}{\partial y} = - \left(\frac{\partial F}{\partial \hat{\beta}} \right)^{-1} \frac{\partial F}{\partial y} \quad (3.7)$$

elde edilir. Delta yönteminin y 'nin bir fonksiyon olarak $\hat{\beta}$ 'ya uygulanmasıyla $\hat{\beta}$ 'nin varyansı (3.6)'daki gibi elde edilir.

Varyansın (3.6)'daki formülünü EKK için incelersek, EKK regresyonunda $\lambda = 0$ olacağından $D(\hat{\beta})$ sıfır matrisi olur. Böylece $\hat{\beta}$ 'nin varyansı $\text{var}(\hat{\beta}_{EKK}) = (X^T X)^{-1} X^T \text{var}(y) X (X^T X)^{-1}$ olur. Bu sonuç EKK tahmincisinin varyansına $\text{var}(\hat{\beta}_{EKK})$ 'ne eşittir.

Ridge regresyon için $q = 2$ olacağından $D(\hat{\beta}) = \lambda I$ olur. Bu durumda eşitlik (3.6) $\text{var}(\hat{\beta}_{ridge}) = (X^T X + \lambda I)^{-1} X^T \text{var}(y) X (X^T X + \lambda I)^{-1}$ eşitliğine dönüşür. $\hat{\beta}$ 'nin varyansı da ridge tahmincisinin varyansına $\text{var}(\hat{\beta}_{ridge})$ eşittir.

Lasso bazı regresyon katsayılarını sıfıra eşitlediği için ($\hat{\beta}_j = 0$) delta yöntemi lasso regresyona uygulanamaz. Ancak bootstrap ya da jackknife yöntemleri ile lasso tahmincisinin varyansı hesaplanabilir (Fu, 1998, s. 403). Lasso ve ridge yöntemlerinin özellikleri ile birlikte varyans hesaplamaları izleyen kısımlarda açıklanacaktır.

3.1.4. Ceza ve ayar parametrelerinin seçimi

Ceza parametresi q ve ayar parametresi λ 'nın seçilmesi için çapraz geçerlilik yönteminden yararlanır. Ridge regresyon ($q = 2$) ve lasso regresyon ($q = 1$) yöntemleri için çapraz-geçerlilik yöntemi ile λ ayar parametresinin hesaplanması, ilgili regresyon yöntemleri için sırasıyla Bölüm 3.2. ve Bölüm 3.3'te açıklanacaktır.

3.2. Ridge Regresyon

1970 yılındaki çalışmalarında A. E. Hoerl ve Kennard'ın önerdikleri ridge regresyon yöntemi, çoklu bağıntı sorunu nedeniyle ortogonal olmayan problemlerin sıklıkla gözlemlendiği fizik, kimya ve mühendislik gibi alanlardan elde edilen verilere en küçük kareler yönteminin uygulanmasının anlamsız olması ve güvenilir sonuçlar vermemesi üzerine bir çözüm olarak önerilmiştir. Devam eden yıllarda önemli tartışmalara yol açan yöntem, daha sonra düzenleştirici yöntemler başlığı altında sınıflandırılacak lasso, elastic-net gibi diğer yöntemlerin önünü açmış, çok boyutlu problemlerin çözümü konusunda istatistiksel analizlerde birçok önemli perspektifler kazandırmışlardır. İstatistik ve makine öğrenmesi literatüründe önemli bir mihenk taşı olarak anılan ridge regresyon yöntemi ayrıca araştırmacıları varyans-yanlılık ödünleşmesi konusunda da araştırmalar yapma yönünde motive etmiştir.

Gauss-Markov teoremine göre bilinmeyen β vektörünün tahmini yansız tahminciler arasından en küçük varyansa sahiptir. Gauss-Markov tahmin süreci olarak da adlandırılan EKK yöntemi, $\mathbf{x}'\mathbf{x}$ matrisi birim matrise (ortogonale) yakın olduğunda başarılı tahminler üretmektedir. Ancak EKK tahminleri hatalara duyarlı olduğundan dolayı $\mathbf{x}'\mathbf{x}$ ilişki matrisi neredeyse hiçbir zaman ortogonal değildir. 1. bölümde olumsuz etkileri açıklanan çoklu bağıntı sorunu nedeniyle uygulamada ortogonal olmayan problemlerle sıklıkla karşılaşılan fizik, kimya ve mühendislik gibi alanlardan elde edilen verilerde EKK'nin kullanılmasının anlamsız ya da mümkün olmaması sebebiyle Hoerl ve Kennard (1970a) (1970b) çalışmalarında ridge regresyon yöntemini önermişlerdir. Öncesinde çoklu bağıntı durumunda yaygın olarak bazı değişkenler modelden çıkarılırken, ridge regresyonun başlangıçtaki bağımsız değişkenler kümesini modelde tutarak ürettiği çözüm literatürde geniş kabul görmüştür (McDonald, 2009, s. 93).

3.2.1. Ridge tahmincisi ve özellikleri

Hoerl ve Kennard'ın (1970a)'ın modele bir miktar yanlılık ekleyerek EKK tahminlerinin genel tutarsızlığını ve tahmin katsayılarının değerindeki mutlakça artışları kontrol etmeyi hedeflemişlerdir. Ridge regresyon yönteminde β vektörünün tahmininde kullanılan hata kareleri toplamına bir ceza parametresi eklenerek en küçük kareler problemi kısıtlı optimizasyon problemine dönüştürülür. Ridge regresyon, bridge

regresyonun tahmincisi olan (3.1) denkleminin ℓ_2 normuna sahiptir ve aşağıdaki gibi gösterilir:

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\}, \quad (3.8)$$

$$\sum_{j=1}^p \beta_j^2 \leq s$$

(3.8)'deki denklemin dual formu ise şöyledir:

$$\beta_{bridge} = \min_{\beta_0, \beta_j} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (3.9)$$

EKK yönteminde olduğu gibi ridge regresyon da veriye en iyi uyan modeli hata kareleri toplamını enküçükleyerek tahmin etmeye çalışır. Ancak buna ek olarak $\lambda \sum_j \beta_j^2$ ceza parametresine sahiptir. Ayar parametresi λ 'nın değeri bridge regresyonda olduğu gibi sıfıra eşit ya da sıfırdan büyüktür ($\lambda \geq 0$). $\lambda = 0$ olduğunda (3.9)'da verilen ridge regresyon problemi EKK problemine dönüşür ve elde edilen $\hat{\beta}_j$ tahminleri en küçük kareler regresyon katsayı tahminlerine eşit olur. $\lambda \rightarrow \infty$ durumunda ise cezanın etkisi artar ve sonsuza yaklaştıkça katsayı tahminlerinin değeri sıfıra yaklaşır. EKK yönteminde tek bir regresyon katsayı matrisi ($\hat{\beta}$) elde edilirken, ridge regresyonda λ değerine bağlı olarak çok sayıda regresyon katsayı matrisi elde edilir. Bu nedenle doğru λ değerinin belirlenmesi çok önemlidir. λ değerinin belirlenmesinde çapraz geçerlilik yönteminden yararlanır (Hastie ve diğerleri, 2015, s. 63).

Hataların beklenen değerinin sıfır olduğu ($E(\varepsilon) = 0$) ve varyansın sabit olduğu ($E(\varepsilon\varepsilon') = \sigma^2 \mathbf{I}_n$) standart çoklu doğrusal regresyon modelini ($\mathbf{y} = \mathbf{X}\beta + \varepsilon$) ele alalım. Burada \mathbf{x} ($N \times p$) boyutlu tam ranka sahip matristir. 2. Bölümde tartışılan standart EKK katsayı tahminleri ölçek değişmezliğine sahipken ridge regresyon tahminleri ölçek değişmezliğine sahip değildir. EKK yönteminde X_j değişkeninin ölçeğinden bağımsız olarak $X_j \hat{\beta}_j$ değişmez. Ancak ridge regresyon yönteminde tahmin katsayıları $X_j \hat{\beta}_j^{ridge}$ yalnızca λ parametresine bağlı olarak değil, j . değişkenin ölçeğine bağlı olarak da değişecektir. Bu yüzden ridge regresyon yöntemi, değişkenler standartlaştırıldıktan sonra

uygulanır. Böylece $\frac{1}{N} \sum_{i=1} y_i = 0$, $\frac{1}{N} \sum_{i=1} X_{ij} = 0$ ve $\frac{1}{N} \sum_{i=1} X_{ij}^2 = 1$ elde edilir ve bütün değişkenler aynı ölçeğe sahip olurlar. Bu standartlaştırma işleminin sonunda sabit katsayı β_0 sifira eşit olacaktır. Değişkenlerin standartlaştırılmasından sonra (3.9) denklemi bir $\lambda \geq 0$ değeri için Lagrange formunda

$$\min_{\beta} \left\{ \frac{1}{2N} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_2^2 \right\} \quad (3.10)$$

biçiminde gösterilebilir. Lagrange dualitesine göre (3.10) ile (3.8) arasında birebir benzerlik vardır. (3.8)'deki her bir s parametresine karşılık gelen bir λ değeri (3.10) denklemi için mevcuttur. (3.10) eşitliğinin β 'ya göre kısmi türevi alındığında ridge tahmincisi elde edilir:

$$\begin{aligned} \hat{\beta}^{\text{ridge}} &= (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{y} \\ &= \mathbf{W}\mathbf{X}'\mathbf{y} \end{aligned} \quad (3.11)$$

Burada \mathbf{I} $p \times p$ boyutlu birim matris, λ ise çapraz geçerlilik yöntemi ile veri kümesinden belirlenen sifira eşit veya büyük ($\lambda \geq 0$) ayar parametresidir. EKK yönteminde \mathbf{X} matrisin rankı sütun sayısından küçük olduğunda ($\text{rank}(\mathbf{X}) < p$) $\mathbf{X}'\mathbf{X}$ matrisinin tersi alınamazken, ridge regresyon modelindeki λ sifirdan büyük olduğunda (3.11)'deki $\mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$ matrisinin tersi mevcuttur (Emmert-Streib ve Dehmer, 2019, s. 363).

(3.11) 'de verilen ridge tahmincisi EKK tahmincisinin doğrusal bir dönüşümüdür. EKK tahmincisi ile olan ilişkisi:

$$\begin{aligned} \hat{\beta}^{\text{ridge}}(\lambda) &= (\mathbf{I}_p + \lambda(\mathbf{X}'\mathbf{X})^{-1})^{-1} \hat{\beta}^{\text{EKK}} \\ &= \mathbf{Z} \hat{\beta}^{\text{EKK}} \end{aligned} \quad (3.12)$$

biçiminde tanımlanabilir. $\lambda = 0$ olduğunda $\mathbf{Z} = \mathbf{I}$ olacağından ridge tahmincisi EKK tahmincisine eşit olur ($\hat{\beta}^{\text{ridge}}(0) = \hat{\beta}^{\text{EKK}}$). Böylece $E(\hat{\beta}^{\text{ridge}}) = E(\mathbf{Z}_\lambda \hat{\beta}) = \mathbf{Z}_\lambda \beta \neq \beta$ olduğu için ridge tahmincisi $\hat{\beta}^{\text{ridge}}(\lambda)$, β 'nın yanlı bir tahmincisidir (Montgomery ve diğerleri, 2013, s. 306).

3.2.2. Ridge tahmincisinin hata kareleri ortalaması

$\hat{\beta}^{RR}$ 'yi incelemek için ridge tahmincisi $\hat{\beta}^{RR}$ ile gerçek katsayı vektörü β arasındaki uzaklığın karesini $E[L_1^2(\lambda)]$ ele alalım. Hata kareleri ortalaması varyans ve yanlılığın karesinin toplamsal bir fonksiyonu olduğuna göre;

$$\begin{aligned}
 E[L_1^2(\lambda)] &= E[(\hat{\beta}^{RR} - \beta)'(\hat{\beta}^{RR} - \beta)] \\
 &= \sigma^2 \text{Tr}[(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}] + \lambda^2 \beta'(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-2} \beta \\
 &= \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + \lambda)^2} + \lambda^2 \beta'(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-2} \beta \\
 &= \gamma_1(\lambda) + \gamma_2(\lambda)
 \end{aligned} \tag{3.13}$$

elde edilir. $\lambda_1, \lambda_2, \dots, \lambda_p$ $\mathbf{X}'\mathbf{X}$ matrisinin özdeğerleridir. Ayırıştırmanın sonucunda elde edilen $\gamma_1(\lambda)$ ve $\gamma_2(\lambda)$ sırasıyla ridge tahmincisinin varyansı ve yanlılığının karesidir. (3.13)'ten görülebileceği üzere ceza parametresi λ arttıkça yanlılık $\gamma_2(\lambda)$ artarken; varyansı $\gamma_1(\lambda)$ azalır. $\gamma_2(\lambda)$, $\mathbf{Z}\beta$ ile β arasındaki kareli uzaklık olduğundan dolayı $\lambda = 0$ için $\mathbf{Z} = \mathbf{I}$ olur. Yanlılık-varyans ödünleşmesi olarak adlandırılan bu durumda varyanstaki azalışın yanlılığın karesindeki artıştan daha az olduğu ridge regresyon modelini veren bir λ değeri belirlenmelidir. Bu ödünleşmeye dayanarak belirlenen optimum λ parametresi ile elde edilen ridge regresyon tahmincisinin hata kareleri ortalaması EKK tahmincisinin varyansından daha az olacaktır (Montgomery ve diğerleri, 2013, s. 306). Hoerl ve Kennard (1970a) çalışmalarında EKK tahmincisi $\hat{\beta}$ 'nın varyansından daha düşük hata kareleri ortalamasına sahip bir $\hat{\beta}^{ridge}$ farklı tahmincisini veren, sıfırdan farklı bir λ değerinin her zaman mevcut olduğunu göstermişlerdir.

$\gamma_1(\lambda)$ ise parametre tahminlerinin toplam varyansıdır ve ridge tahmincisi ile EKK tahmincisi arasındaki ilişkiyi gösteren

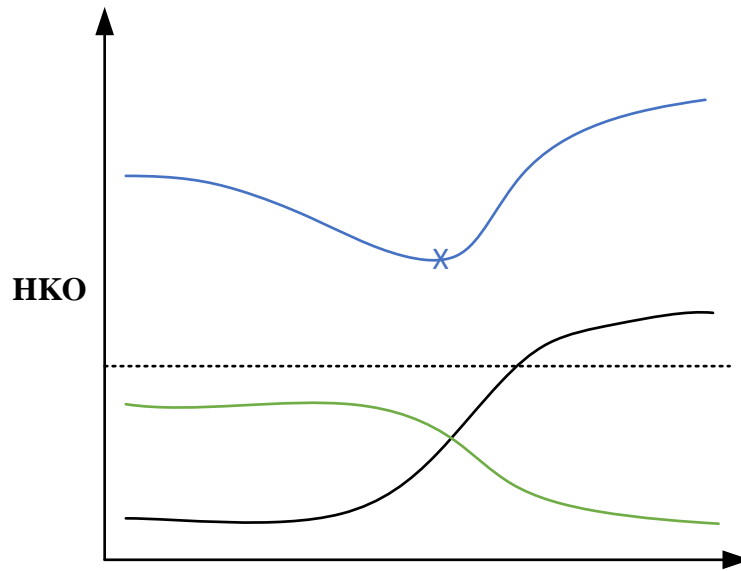
$$\hat{\beta}^{ridge}(\lambda) = \mathbf{Z}\hat{\beta}^{EKK} = \mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \tag{3.14}$$

denkleminde yola çıkılarak $\hat{\beta}^{ridge}$ 'in kovaryans matrisi

$$\begin{aligned}
 \text{Var}(\hat{\beta}^{ridge}) &= \sigma^2 (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \\
 &= \mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \text{Var}(\mathbf{y}) \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{Z}' \\
 &= \sigma^2 \mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{Z}'
 \end{aligned} \tag{3.15}$$

olur. Tüm $\hat{\beta}^{ridge}$ tahmincilerinin varyanslarının toplamı (3.15)'teki köşegen elemanların toplamına eşittir (Hoerl ve Kennard, 1970a, s. 60).

Şekil 3.3'te 45 gözlem ve 50 bağımsız değişkenden oluşan bir veri kümesine ilişkin varyans-yanlılık ilişkisi gösterilmiştir. Kareli yanlılık siyah, varyans yeşil ve ridge regresyon test hata kareleri ortalaması mavi ile gösterilmiştir. Kesikli yatay çizgi mümkün en küçük HKO değeridir. Grafiğin sol tarafındaki değerler $\lambda = 0$ durumunu yani EKK tahminine ait varyans, yanlılık ve test verisine ilişkin hata kareleri ortalaması değerlerini gösterir. $\lambda = 0$ için yanlılık sifıra eşit iken varyans yüksektir. λ arttıkça ridge tahmincisinin katsayı değerleri daralmakta, bunun sonucunda yanlılık artarken varyans azalmaktadır. λ , 0-10 arasında iken yanlılıktaki bir miktar artışa karşın, varyansta büyük bir azalış gerçekleştiğinden dolayı hata kareleri ortalaması da giderek azalmaktadır. Ancak bir noktadan sonra ($\lambda > 10$) varyanstaki azalış hızı, yanlılığın artış hızından daha az olduğu için test verisinin HKO değeri artmaktadır. HKO eğrisi üzerinde çarpı ile gösterilen nokta test verisinin HKO değerinin en az olduğu noktadır (James ve diğerleri, 2013, s. 220).



Şekil 3.3. Ridge regresyon ve yanlılık-varyans ilişkisi (James ve diğerleri, 2013)

Toplam varyans $\gamma_1(\lambda)$, λ 'nın düzgün azalan bir fonksiyonudur. λ 'ya göre birinci dereceden türevi $\gamma_1'(\lambda)$, $\lambda \rightarrow 0^+$ ve $\lambda_p \rightarrow 0$ iken eksi sonsuza yaklaşır: $\gamma_1'(\lambda) \rightarrow -\infty$. Bu durum Şekil 3.3'te gözlemlenmektedir. Kareli yanlılık $\gamma_2(\lambda)$ ise (3.13)'e göre λ 'nın

düzgün artan bir fonksiyonudur. Λ , $\mathbf{X}'\mathbf{X}$ 'nin özdeğerler matrisi ve \mathbf{P} dikey dönüşümü ise $\mathbf{X}'\mathbf{X} = \mathbf{P}'\Lambda\mathbf{P}$ olur. O halde $\gamma_2(\lambda) = \lambda^2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-2}\boldsymbol{\beta}$ denklemi

$$\gamma_2(\lambda) = \sum_1^p \alpha_i^2 / (\lambda_i + \lambda)^2, \quad \boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta} \quad (3.16)$$

biçiminde gösterilebilir. Her i için $\lambda_i > 0$ ve $\lambda \geq 0$ olduğuna göre (3.16)'daki her bir $(\lambda_i + \lambda)$ elemanı pozitifdir ve toplamda tekillik yoktur. Bu durumda $\lambda \geq 0$ için $\gamma_2(\lambda)$ sürekli bir fonksiyon olur. Tüm $\lambda > 0$ için (3.16) denklemi

$$\gamma_2(\lambda) = \sum_1^p \alpha_i^2 / [1 + (\lambda_i/\lambda)]^2 \quad (3.17)$$

biçiminde ifade edilebilir. λ arttıkça (λ_i/λ) fonksiyonu düzgün azalır ve $\gamma_2(\lambda)$ 'ın her bir terimi düzgün artan bir fonksiyondur (Hoerl ve Kennard, 1970a, s. 62).

Hoerl ve Kennard (1970a, s. 62) çalışmalarında EKK tahmincisinin HKO'sundan daha düşük HKO'ya sahip bir ridge tahmincisi veren bir $\lambda > 0$ değerinin her zaman olduğunu göstermişlerdir:

$$E[L_1^2(\lambda)] < E[L_1^2(0)] = \sigma^2 \sum_1^p (1/\lambda_i) \quad (3.18)$$

3.2.3. Ridge tahmincisinin hesaplanması

Denklem (3.11)'de verilen $\hat{\boldsymbol{\beta}}_\lambda^{\text{ridge}} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$ ridge tahmincisinin hesaplanması için $\mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$ matrisinin tersinin alınması gerekir. Ancak bu işlem zor ve hesaplama maliyeti yüksek olduğu için tekil değer ayrıştırmasından (Singular Value Decomposition-SDV) yararlanılarak $\hat{\boldsymbol{\beta}}^{\text{ridge}}$ hesaplanabilir ve ridge tahmincisine ilişkin ek bilgiler de elde edilebilir (Aydın, 2014, s. 471).

$n \times p$ boyutlu \mathbf{X} matrisinin tekil değer ayrıştırması şöyledir:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}' \quad (3.19)$$

Burada \mathbf{U} $n \times n$ boyutlu sol tekil değerleri içeren (\mathbf{u}_i); \mathbf{V} ise $p \times p$ boyutlu sağ tekil değerleri (\mathbf{v}_i) içeren dik matrislerdir. Buna göre $\mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}$ ve $\mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}$ eşitlikleri sağlanır. \mathbf{U} 'nun sütunları \mathbf{X} 'in sütun uzayını ve \mathbf{V} 'nin sütunları da \mathbf{X} 'in satır uzayını veren matristir. \mathbf{D} ise $n \times p$ boyutlu blok köşegen matristir. \mathbf{D} matrisinin sol üst ($\text{rank}(\mathbf{X}) \times \text{rank}(\mathbf{X})$) boyutlu köşegen matrisinin elemanları \mathbf{X} matrisinin tekil

değerleridir ($d_{11} \geq d_{22} \geq \dots \geq d_{pp} \geq 0$) ve matrisin diğer elemanları sıfıra eşittir. Eğer bir veya birden fazla $d_j = 0$ ise \mathbf{X} tekil matris olur.

Tekil değer ayrıştırması kullanılarak EKK tahmincisi

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{\text{EKK}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= (\mathbf{V}\mathbf{D}'\mathbf{D}\mathbf{V}')^{-1} \mathbf{V}\mathbf{D}'\mathbf{U}'\mathbf{y} \\ &= \mathbf{V}(\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}'\mathbf{U}'\mathbf{y}\end{aligned}\quad (3.20)$$

biçiminde hesaplanabilir. $(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}$ $p \times n$ boyutlu matristir. Bu matrisin $p \times p$ boyutlu sol üst köşesinde sıfırdan farklı tekil değerlerin tersleri yer alan köşegen matris vardır. Ridge tahmincisi ise tekil değer ayrıştırması yöntemiyle

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{\text{Ridge}} &= (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{y} \\ &= \mathbf{V}(\mathbf{D}'\mathbf{D} + \lambda\mathbf{I})^{-1} \mathbf{D}'\mathbf{U}'\mathbf{y}\end{aligned}\quad (3.21)$$

biçiminde elde edilebilir. $\mathbf{D}_{jj} = d_{jj}$, $j=1, \dots, p$ \mathbf{D} matrisinin sol üst köşesinde p adet tekil değer olsun. EKK ve ridge sonuçlarını karşılaştırsak $\lambda > 0$ için EKK'nin tekil değerleri d_{jj}^{-1} , ridge regresyonun tekil değerlerinden $d_{jj} / (d_{jj}^2 + \lambda)$ büyüktür: $1/d_{jj} \geq d_{jj} / (d_{jj}^2 + \lambda)$. Sonuç olarak ridge regresyon \mathbf{X} matrisinin tekil değerlerini küçültmektedir (van Wieringen, 2015, s. 9).

EKK uyum vektörü tekil değer ayrıştırması yardımıyla

$$\begin{aligned}\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{EKK}} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= \mathbf{U}\mathbf{U}'\mathbf{y}\end{aligned}\quad (3.22)$$

biçiminde elde edilir. \mathbf{U} 'nun ortonormal tabanına göre \mathbf{y} 'nin koordinatları $\mathbf{U}'\mathbf{y}$ 'dir. Ridge regresyonun uyum vektörü (3.23)'ye eşittir:

$$\begin{aligned}\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{Ridge}} &= \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}'\mathbf{y} \\ &= \mathbf{U}\mathbf{D}(\mathbf{D}^2 + \lambda\mathbf{I})^{-1} \mathbf{D}\mathbf{U}'\mathbf{y} \\ &= \sum_{j=1}^p \mathbf{u}_j \frac{d_j^2}{d_j^2 + \lambda} \mathbf{u}_j' \mathbf{y}\end{aligned}\quad (3.23)$$

\mathbf{u}_j \mathbf{U} matrisinin sütunlarıdır. $\lambda \geq 0$ için $d_j^2 / (d_j^2 + \lambda) \leq 1$ olur. Ridge regresyon da EKK yönteminde olduğu gibi \mathbf{y} 'nin koordinatlarını \mathbf{U} 'nun ortonormal tabanına göre hesaplar ancak daha sonra bu koordinatları $d_j^2 / (d_j^2 + \lambda)$ çarpanlarıyla küçültür. Dolayısıyla daha

küçük d_j^2 değerlerine sahip baz vektörlerinin koordinatları daha fazla daraltılır. Küçük d_j tekil değerler, \mathbf{X} 'in sütun uzayındaki düşük varyansa sahip asal bileşen doğrularına karşılık gelirler ve ridge regresyon en çok bu bileşenleri daraltır. Sonuç olarak ridge regresyon bağımlı değişken üzerinde açıklayıcılığı az olan bağımsız değişkenlerin katsayılarını, açıklayıcılığı yüksek olan değişkenlere kıyasla daha fazla küçültür (Hastie ve diğerleri, 2015, s. 66).

Tekil değer ayrıştırması yöntemiyle ridge regresyonun serbestlik derecesini hesaplayabiliriz. EKK regresyonunda uyum vektörü $\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y}$ denklemi yardımıyla tahmin edilmektedir. EKK regresyonunda kullanılan serbestlik derecesi şapka matrisinin izine eşittir: $tr(\mathbf{H})$. \mathbf{X} matrisi tam ranka sahip olduğunda ($rank(\mathbf{X}) = p$) şapka matrisinin izi p 'ye eşit olur: $tr(\mathbf{H}) = p$. Dolayısıyla EKK regresyonunun serbestlik derecesi tahmin edilmek istenen parametre sayısı olan p 'ye eşittir

Ridge regresyonun şapka matrisi ise $\mathbf{H}(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'$ denklemine eşittir. $\mathbf{H}(\lambda)$ şapka matrisinin izi hesaplandığında ridge regresyonun serbestlik derecesi elde edilir (van Wieringen, 2015, s. 20):

$$\begin{aligned} df(\lambda) &= tr[\mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'] \\ &= tr(\mathbf{H}(\lambda)) \\ &= \sum_{j=1}^p \frac{d_j^2}{d_j^2 + \lambda} \end{aligned} \quad (3.24)$$

λ 'nın düzgün azalan fonksiyonu olan $df(\lambda)$ ridge regresyon uyumunun etkili serbestlik derecesidir. $\lambda = 0$ olduğunda $df(\lambda) = p$ 'dir. $\lambda \rightarrow \infty$ oldukça $df(\lambda) \rightarrow 0$ olur. Elde edilen etkili serbestlik derecesine standartlaştırılmış \mathbf{X} matrisinde yer almayan sabit katsayı için bir serbestlik derecesi daha eklenir (Hastie ve diğerleri, 2015, s. 68).

3.2.4. Ridge izi ve ceza parametresinin seçimi

Ridge izi Hoerl ve Kennard'ın (1970a, s. 58) önerdiği belirli bir aralıktaki λ değerlerine karşılık gelen doğrusal ridge regresyon modelinin tüm $\hat{\boldsymbol{\beta}}_j^{ridge}$ değerlerini eş zamanlı olarak gösteren bir grafikdir. Böylece λ değerindeki değişimin katsayı tahminleri üzerindeki etkileri gözlemlenebilir. Ayrıca ridge regresyon açıklayıcılığı düşük olan

bağımsız değişkenlerin katsayı tahmin değerlerini sıfıra doğru daha hızlı küçülttüğü için grafikteki regresyon katsayıları incelenerek belirli bir λ değeri ve bu λ değerine karşılık gelen model öznel olarak araştırmacı olarak belirlenebileceğinden, ridge izi değişken seçimi probleminin çözümünde de kullanılabilir (Gruber, 1998, s. 133).

Ridge izi grafiğinde yatay ekseninde ceza parametresi λ ve dikey ekseninde $\hat{\beta}_j^{ridge}$ katsayıları yer alır. $\lambda \rightarrow \infty$ 'a yaklaştıkça ($\lambda \rightarrow \infty$) $\hat{\beta}_j^{ridge}$ katsayıların değeri sıfıra doğru yaklaşmaya başlar. Belirli bir λ değerine ulaşıldığında katsayı değerindeki değişim yavaşlar ve neredeyse değişim durma noktasına gelir. Çoklu bağıntı mevcutsa regresyon katsayıları λ 'nın küçük değerleri için çok hızlı değişim gösterirler. Bu sebepten dolayı ridge izi yardımıyla sonuçların duyarlılığı incelenebilir. Ancak $\hat{\beta}_j^{ridge}$ katsayıları hiçbir zaman sıfıra eşit olmayacağı için ridge regresyon yöntemi değişken seçimini gerçekleştiremez. Öte yandan katsayıların değerindeki değişim yavaşladığı nokta araştırmacı tarafından belirlenerek ridge tahmincisinin hesaplanmasında kullanılan λ parametresi belirlenebilir.

Hoerl ve Kennard (1970a, s. 65) ridge izini kullanarak λ parametresinin belirlenmesinde aşağıdaki hususlardan yararlanılabileceğini belirtmişlerdir:

- λ 'nın belirli bir değerinde katsayılar durağanlaşır ve model ortogonal sistemin genel özelliklerine sahip olmalıdır.
- Katsayılar mutlaka ekonomik yorumu olmayan değerlere sahip olmamalıdır.
- Çoklu bağıntı sebebiyle EKK tahmincisinde ($\lambda = 0$) elde edilen yanlış katsayı işaretleri λ değeri arttıkça uygun işarete sahip olmalıdır.
- Hata kareleri toplamı makul bir değerde tutulmalı ve minimum hata kareleri toplamından çok büyük olmamalıdır.

3.3. Lasso Regresyon

Lasso (Least Absolute Shrinkage and Selector Operator – En Küçük Mutlak Daralma ve Seçim Operatörü) regresyon, Robert Tibshirani (1996) tarafından EKK yöntemine alternatif olarak geliştirilmiştir. Tahminlerin doğruluğunu artırırken aynı zamanda değişken seçimini eş zamanlı olarak gerçekleştirmesi ve sahip olduğu hesaplama kolaylığından dolayı çok boyutlu regresyon problemlerinin çözümünde yaygın olarak kullanılmaktadır. Tibshirani çalışmasında lasso regresyonun iki sebepten dolayı

EKK yöntemine alternatif olarak tercih edilebileceğini belirtmiştir: Tahmin doğruluğu ve yorumlanabilirlik.

EKK tahminleri genelde düşük yanlılığa ancak yüksek varyansa sahiptir. Ancak regresyon katsayılarının daraltılması ve bazılarının sifıra eşitlenmesi ile tahminlerin doğruluğu artırılabilir. Böylece yanlılık bir miktar artmasına karşın varyans büyük bir miktarda azaltılır.

Lasso regresyon çok sayıdaki bağımsız değişken kümesinden, bağımlı değişken üzerinde en yüksek etkiye sahip olan daha az değişkene sahip bir alt kümeyi seçmektedir. Böylece 2. bölümde tanıtilan değişken seçme yöntemlerinden herhangi birini kullanmaya gerek olmadan daha kolay yorumlanabilir bir regresyon modeli elde edilir.

3.3.1. Lasso tahmincisi ve özellikleri

Lasso regresyon yöntemi Kısım 3.1.'de tanıtilan bridge regresyonun tahmincisi olan (3.1) denkleminin ℓ_1 normuna sahiptir ve aşağıdaki gibi gösterilir:

$$\hat{\beta}^{lasso} = \arg \min_{\beta_0, \beta_j} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\}, \quad (3.25)$$

$$\sum_{j=1}^p |\beta_j| \leq s$$

Denklem (3.8) ile denklem (3.25) karşılaştırıldığında ridge regresyondaki β_j^2 teriminin yerini lasso regresyonda $|\beta_j|$ teriminin aldığı görülür. Lasso regresyondaki bu ifade ℓ_1 norm olarak adlandırılır. (3.25) eşitliğinin dual formu ise şöyledir:

$$\beta_{lasso} = \min_{\beta_0, \beta_j} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (3.26)$$

Ayrıca (3.25)'te verilen formülasyon matris-vektör notasyonu ile de ifade edilir. N satırdan oluşan yanıt vektörü $\mathbf{y} = (y_1, \dots, y_N)$ ile ve $N \times k$ boyutlu bağımsız değişken matrisi \mathbf{X} ile gösterildiğinde optimizasyon problemi

$$\min_{\beta_0, \beta_j} \left\{ \frac{1}{2N} \|\mathbf{y} - \beta_0 \mathbf{1} - \mathbf{X}\beta\|_2^2 \right\}, \quad (3.27)$$

$$\|\beta\|_1 \leq s$$

biçiminde ifade edilebilir (Hastie ve diğerleri, 2015, s. 8).

Ridge regresyon yönteminde olduğu gibi lasso regresyon yönteminde de \mathbf{X} matrisinin sütunları ($\frac{1}{N} \sum_{i=1}^N x_{ij} = 0$ ve $\frac{1}{N} \sum_{i=1}^N x_{ij}^2 = 0$ olacak biçimde) standartlaştırılır. Aksi takdirde lasso yönteminin çözümleri bağımsız değişkenlerin ölçü birimlerine bağlı olacaktır. Çünkü EKK yönteminin aksine lasso yöntemi ölçek değişmezliğine sahip değildir. Ayrıca yanıt vektörünün değerleri de sıfır ortalamaya sahip olacak şekilde standartlaştırılır: $\frac{1}{N} \sum_{i=1}^N y_i = 0$. Buna göre β_0 sabit terimi (3.27)'deki optimizasyondan çıkarılabilir. Standartlaştırılmış veriden elde edilen lasso çözümlerinden yola çıkılarak sabit katsayının tahmini orijinal veri kümesi için

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^p \bar{x}_j \hat{\beta}_j \quad (3.28)$$

biçiminde hesaplanır. Böylece (3.27)'de verilen lasso problemi verilen bir $\lambda \geq 0$ için Lagrange formunda (3.29)'da gösterildiği biçimde elde edilir:

$$\min_{\beta} \left\{ \frac{1}{2N} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \quad (3.29)$$

Eşitlik (3.27)'deki s kısıt parametresinin (3.29)'daki karşılığı λ ayar parametresidir (Hastie ve diğerleri, 2015, s. 8).

Lasso probleminin özellikleri şunlardır (R. J. Tibshirani, 2013, s. 3) :

- i. Konveks bir problem olduğu ve herhangi bir artış yönüne sahip olmadığı için lasso probleminin en az bir çözümü vardır.
- ii. Her $\hat{\beta}$ lasso çözümü aynı $X\hat{\beta}$ uyum değerini verir.
- iii. Eğer $\lambda > 0$ ise her $\hat{\beta}$ lasso çözümü aynı ℓ_1 normuna $\|\hat{\beta}\|_1$ sahiptir.

3.3.2. Lasso çözümü

Lasso probleminin çözümü için gerek ve yeter koşullar Karush-Kuhn-Tucker koşullarıdır (Hastie ve diğerleri, 2015, s. 9):

$$X^T (y - X\hat{\beta}) = \lambda \gamma \quad (3.30)$$

$$\gamma_j \in \begin{cases} \{sign(\hat{\beta}_j)\}, & \hat{\beta}_j \neq 0 \\ [-1,1], & \hat{\beta}_j = 0 \end{cases} \quad j=1,\dots,p \quad (3.31)$$

Yukarıdaki eşitlikte verilen $\gamma \in \mathbb{R}^p$, $f(x) = \|x\|_1$ fonksiyonunun $x = \hat{\beta}$ noktasındaki subgradyantıdır. γ , $\hat{\beta}_j$ sıfırdan farklı ise $\hat{\beta}_j$ 'nin işaretini alan bilinmeyen bir nicelik; $\hat{\beta}_j$ sıfıra eşit ise $[-1,1]$ kapalı aralığında bir değerdir (Hastie ve diğerleri, 2015, s. 9). Buna göre $\hat{\beta}$ yalnızca ve yalnızca belirli bir γ için (3.30) ve (3.31) koşullarını sağladığında (3.29)'un bir çözümü olabilir. Söz konusu gerek ve yeter koşullar

$$-\frac{1}{N} \mathbf{X}^T (\mathbf{y} - \mathbf{X}\beta) + \lambda sign(\beta) = 0 \quad (3.32)$$

biçiminde de ifade edilebilir (Hastie ve diğerleri, 2015, s. 9).

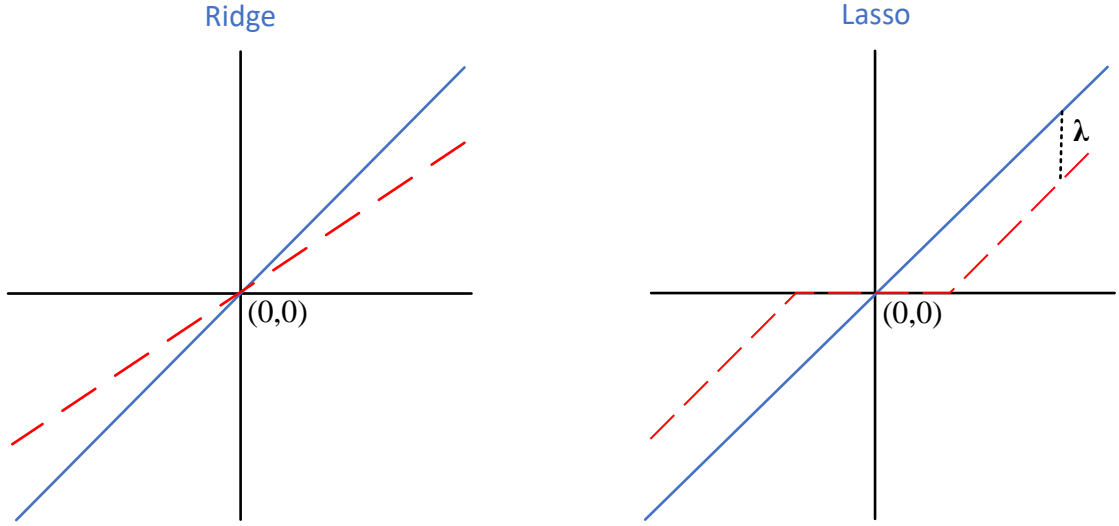
ℓ_1 normunun yapısı itibari ile lasso optimizasyon problemi (3.29) türevlenemeyen bir fonksiyon olduğu için ridge regresyondaki gibi kapalı formda bir çözümü yoktur. Ancak \mathbf{X} matrisi ortonormal olduğunda lasso minimizasyon probleminin kapalı formda çözümü hesaplanabilir. İzleyen kesimde sırasıyla basit doğrusal regresyon ($p=1$) ve ortonormal tasarım için lasso çözümü incelenecektir. Tek bir bağımsız değişkenin olduğu durumda lasso çözümü EKK tahmininin yumuşak eşikli (soft-thresholded) halidir:

$$\hat{\beta}^{lasso}(\lambda) = sign(\hat{\beta}^{EKK}) \left(\left| \hat{\beta}^{EKK} \right| - \lambda \right)_+ \quad (3.33)$$

$$\hat{\beta}^{lasso}(\lambda) = \begin{cases} \hat{\beta}^{EKK} - \lambda, & \hat{\beta}^{EKK} > 0 \text{ ve } \lambda < \left| \hat{\beta}^{EKK} \right| \text{ ise} \\ \hat{\beta}^{EKK} + \lambda, & \hat{\beta}^{EKK} < 0 \text{ ve } \lambda < \left| \hat{\beta}^{EKK} \right| \text{ ise} \\ 0, & \lambda \geq \left| \hat{\beta}^{EKK} \right| \text{ ise} \end{cases} \quad (3.34)$$

Sonuç olarak $\lambda \geq \left| \hat{\beta}^{EKK} \right|$ olduğunda $\hat{\beta}_j$ katsayısının tahmini sıfıra eşit olur. Böylece lasso regresyon değişken seçimini de gerçekleştirir (Friedman, Hastie, Höfling ve Tibshirani, 2007, s. 303).

Ortonormal bir \mathbf{X} matrisi için ridge tahmincisi ($\hat{\beta}_j^{EKK} / (1 + \lambda)$) ve lasso tahmincisinin karşılaştırılması Şekil 3.4'te verilmiştir. 45°'lik açı ile verilen gri doğru kısıtsız tahminciye gösterirken, kesikli kırmızı ile verilenler kısıtlı ridge ve lasso tahminlerini göstermektedir.



Şekil 3.4. Ridge ve lasso tahmincilerinin karşılaştırılması (Hastie ve diğerleri, 2008)

Şekilden görülebileceği üzere ridge regresyon, EKK tahminin önceden belirlenen bir λ değerine bağlı olarak orantılı olarak düzenlenleştirirken; lasso regresyon yumuşak eşikleme (soft-thresholding) adı verilen bir süreci izleyerek her bir katsayıyı sabit λ faktörü kadar azaltır ve sıfıra yaklaştırır (Hastie ve diğerleri, 2015, s. 69)(J. Friedman ve diğerleri, 2007, s. 303).

\mathbf{X} , $p = n$ durumunda $n^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{I}_{p \times p}$ eşitliğini sağlayan tasarım matrisi olsun . Ortonormal durumda bağımsız değişkenlerin aralarında bir ilişkinin olmadığından dolayı $i \neq j$ için $\mathbf{X}_i^T\mathbf{X}_j = 0$ olur. Bu durumda lasso tahmincisi yine EKK tahmincisinin yumuşak eşikli versiyonudur (Bühlmann ve van de Geer, 2011, s. 11):

$$\hat{\beta}_j(\lambda) = \text{sign}(\hat{\beta}_j^{EKK}) \left(|\hat{\beta}_j^{EKK}| - \lambda \right)^+ \quad (3.35)$$

Çok değişkenli ve ortonormal olmayan genel durum için lasso tahmincisinin kapalı çözümü yoktur. \mathbf{X} tam ranka sahip olduğunda $\mathbf{X}^T\mathbf{X}$ matrisinin tersi alınabilir. \mathbf{X}_{-j} , j . sütun hariç tüm sütunları içeren \mathbf{X} matrisi ve β_{-j} ise β_j haricindeki parametre vektörüdür. KKT koşulunu uygularsak

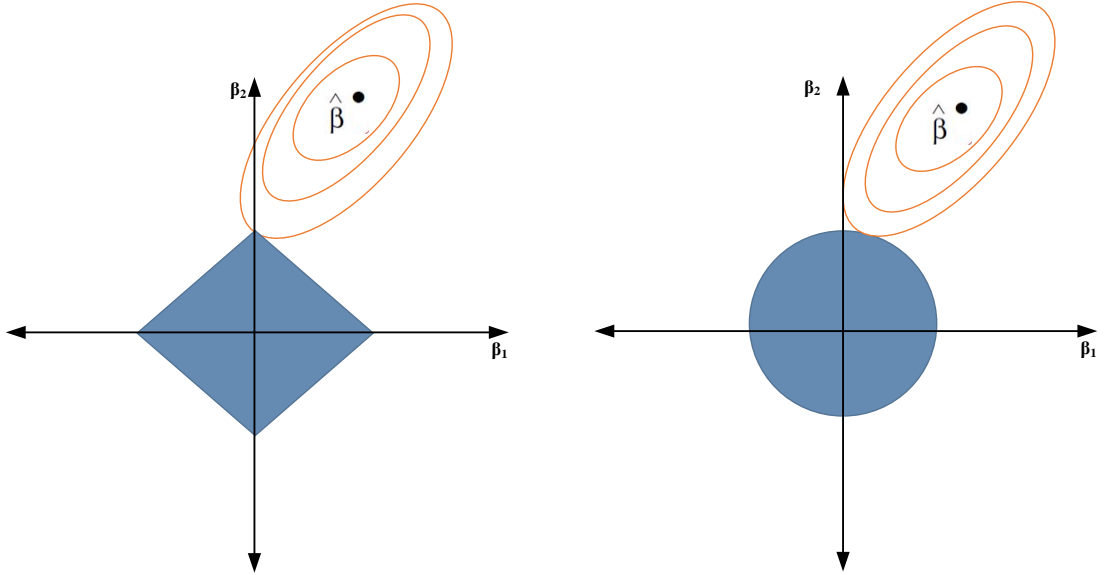
$$-\frac{1}{N}\mathbf{X}_j^T(\mathbf{y} - \mathbf{X}_{-j}\beta_{-j} - \hat{\beta}_j\mathbf{X}_j) + \lambda \text{sign}(\hat{\beta}_j) = 0 \quad (3.36)$$

$$-\frac{1}{N}\mathbf{X}_j^T\mathbf{y} + \frac{1}{N}\mathbf{X}_j^T\mathbf{X}_{-j}\beta_{-j} + \frac{1}{N}\hat{\beta}_j\mathbf{X}_j^T\mathbf{X}_j + \lambda \text{sign}(\hat{\beta}_j) = 0 \quad (3.37)$$

$$\hat{\beta}_j(\lambda) = \frac{1}{N}\mathbf{X}_j^T(\mathbf{y} - \mathbf{X}_{-j}\beta_{-j}) - \lambda \text{sign}(\hat{\beta}_j) \quad (3.38)$$

lasso çözümünün kapalı bir çözüme sahip olmadığı görülür. β_j parametresinin çözümü diğer $\beta_{i \neq j}$ katsayılarına bağlıdır. Ortonormal durumda $\mathbf{X}_j^T \mathbf{X}_{-j}$ ifadesi ortogonallik sebebiyle ortadan kalktığı için lasso kapalı çözüme sahiptir (Gauraha, 2018, s. 455).

Ortonormal olmayan durumu $p = 2$ durumu için tekrar ele alırsak lasso regresyon ve ridge regresyon yöntemlerini karşılaştırarak lasso regresyonun neden değişken seçme özelliğine sahip olduğunu gösterebiliriz. Şekil 3.5'in sol tarafında lasso regresyon ve sağ tarafında ridge regresyon tahminleri gösterilmiştir. EKK tahmini $\hat{\beta}$ ile gösterilmiş ve hata kareleri toplamının eliptik konturları turuncu ile gösterilmiştir. Soldaki grafikte yer alan kare biçimindeki alan lasso regresyonun $|\beta_1| + |\beta_2| \leq s$ kısıtlayıcı çözüm alanını, sağdaki grafikteki daire ise ridge regresyonun $\beta_1^2 + \beta_2^2 \leq s$ kısıtlayıcı çözüm alanını gösterir. Her iki yöntem de hata kareleri toplamı eliptik konturunun kısıtlayıcı alana ilk değdiği noktayı çözüm olarak kabul etmektedir. Dolayısıyla s yeterince büyük ise lasso ve ridge tahminleri EKK tahminlerine eşit olacaktır.



Şekil 3.5. Lasso ve ridge tahminlerinin gösterimi (Hastie ve diğerleri, 2008)

Şekilden görüleceği üzere lasso regresyonun kısıtlayıcı alanı köşelere sahip olduğu için HKT konturu kısıt alanının köşesine değdiğinde bir parametrenin değeri (bu örnekte β_1 katsayısı) sıfıra eşit olur ve böylece başlangıçtaki bağımsız değişken kümesinden daha az sayıda değişkene sahip bir model çözüm olarak elde edilir. Ancak ridge kısıtlayıcı

alanının köşeleri olmadığı için parametrelerin değeri sıfıra çok yakın bir değer olsa da sıfıra eşit olmaz (Hastie ve diğerleri, 2015, s. 72) (Robert Tibshirani, 1996, s. 271).

3.3.3. Lasso çözümünün hesaplanması

Lasso problemi (3.29) konveks kısıta sahip bir kuadratik programlama problemidir ve problemin çözümü için birçok kuadratik programlama yöntemi mevcuttur. En yaygın kullanılan yöntemler arasında LARS (Least Angle Regression) (Efron, Hastie, Johnstone ve Tibshirani, 2004), ADMM (Alternating Direction Method of Multipliers) (Boyd, Parikh, Chu, Peleato ve Eckstein, 2011), Yakınsal Gradyant İniş (Proximal Gradient Descent) (Chen, Lin, Kim, Carbonell ve Xing, 2012) yer almaktadır. Bu çalışmada (Friedman, Hastie ve Tibshirani, 2010) tarafından önerilen Döngüsel Koordinat İniş (Cyclical Coordinate Descent) algoritması kullanılacağından dolayı yalnızca bu yöntem açıklanacaktır.

Koordinat iniş, her iterasyonda tek bir koordinat seçen ve daha sonra bu koordinat üzerinde tek değişkenli minimizasyon gerçekleştirerek β^t 'den β^{t+1} 'e ilerleyen bir algoritmadır (Hastie ve diğerleri, 2015, s. 110). t . adımda seçilen bir k koordinatı için bilinmeyen vektörü

$$\begin{aligned}
 \beta_1^{t+1} &= \arg \min_{\beta_1} f(\beta_1, \beta_2^t, \beta_3^t, \dots, \beta_p^t) \\
 \beta_2^{t+1} &= \arg \min_{\beta_2} f(\beta_1^{t+1}, \beta_2, \beta_3^t, \dots, \beta_p^t) \\
 &\vdots \\
 \beta_p^{t+1} &= \arg \min_{\beta_p} f(\beta_1^{t+1}, \beta_2^{t+1}, \beta_3^{t+1}, \dots, \beta_p)
 \end{aligned} \tag{3.39}$$

biçiminde güncellenerek optimum sonuca doğru ilerlenir ($t = 0, 1, \dots$) (Genç, 2020, s. 23).

Konveks bir fonksiyonun global optimum noktasının koordinat iniş yöntemi ile yaklaşık olarak hesaplanabilmesi için f fonksiyonunun her bir koordinatta sürekli türevlenebilir ve kesin konveks olması yeterlidir. Dolayısıyla türevsiz optimizasyon problemlerinde koordinat minimizasyonu yöntemi optimum olmayan sonuçlar verecektir. Ancak Tseng'in (2001) çalışmasında gösterdiği üzere, konveks ancak türevsiz bir f maliyet fonksiyonu, türevlenebilir ve türevsiz olmak üzere bileşenlere ayrılabilirliğiyle koordinat iniş yöntemi global optimuma yakınsak çözümler üretmektedir. Bu özelliğe ayrılabilirlik koşulu adı verilir.

f toplamsal ayrıştırılabilir bir maliyet fonksiyonu olsun:

$$f(\beta_1, \dots, \beta_p) = g(\beta_1, \dots, \beta_p) + \sum_{j=1}^p h_j(\beta_j) \quad (3.40)$$

Burada verilen $g: \mathbb{R}^p \rightarrow \mathbb{R}$ türevlenebilir ve konveks bir fonksiyondur. Tek değişkenli $h_j: \mathbb{R}^p \rightarrow \mathbb{R}$ fonksiyonları ise konvekstir ancak türevlenebilir olmaları şart değildir (Hastie ve diğerleri, 2015, s. 110). Lasso programının bileşenleri, (3.40) eşitliğinde $g(\beta) = \frac{1}{2N} \|\mathbf{y} - \mathbf{X}\beta\|_2^2$ ve $h_j(\beta_j) = \lambda |\beta_j|$ biçiminde yerine yazılırsa lasso programının da ayrılabilirlik koşulunu sağlayan bir maliyet fonksiyonuna sahip olduğu görülür. Türevsiz $\sum_{j=1}^p h_j(\beta_j) = \lambda \sum_{j=1}^p |\beta_j|$ bileşeni (3.40)'de gösterilen ayrılabilirlik özelliğini sağladığı için koordinat iniş algoritması ile lasso probleminin global optimum çözümüne yakınsak bir sonuç elde edilir (Hastie ve diğerleri, 2015, s. 110).

Lasso probleminin çözümünde bağımsız değişkenler sabit fakat keyfi bir sıra ile tekrarlı olarak güncellenirler. j. adımda tüm diğer katsayıların mevcut değerleri sabitlenmek suretiyle, belirlenen koordinatta amaç fonksiyonu en küçüklenerek β_j 'nin değeri güncellenir. Kısmi artıkları ($r_i^{(j)} = y_i - \sum_{k \neq j} x_{ik} \hat{\beta}_k$) bağımlı değişken olarak kabul edersek enküçüklenecek (3.29) amaç fonksiyonu aşağıdaki biçimde gösterilir:

$$\frac{1}{2N} \sum_{i=1}^N (y_i - \sum_{k \neq j} x_{ik} \hat{\beta}_k)^2 + \lambda \sum_{k \neq j} |\beta_k| + \lambda |\beta_j| \quad (3.41)$$

Denklem (3.41)'yi enküçüklediğimizde lasso regresyon için j. katsayı tahmini

$$\hat{\beta}_j = \frac{S_\lambda \left(\frac{1}{N} \sum_{i=1}^N r_i^{(j)} x_{ij} \right)}{\frac{1}{N} \sum_{i=1}^N x_{ij}^2} \quad (3.42)$$

olur. Denklemdeki $S_\lambda(x) = \text{sign}(x)(|x| - \lambda)_+$ yumuşak eşikleme operatörüdür (Hastie ve diğerleri, 2015, s. 112). Daha önce açıklandığı üzere (3.29) denklemde hem x_i hem de y_i vektörleri sıfır ortalamaya sahip olacak biçimde standartlaştırılmıştır. Buna ek olarak değişkenler bir de birim varyansa sahip olacak biçimde standartlaştırılırsa, bu durumda (3.42) denklemi

$$\hat{\beta}_j = S_\lambda(\tilde{\beta}_j) \quad (3.43)$$

formunda ifade edilir. $\tilde{\beta}_j$, j . deęişkenin kısmi artığının basit doğrusal regresyon katsayısıdır. Buna göre (3.41)'de verilen güncelleme $j = 1, 2, \dots, p, 1, 2, \dots$ yakınsak bir çözüme ulaşana kadar devam ettirilir (Hastie ve dięerleri, 2015, s. 112).

3.3.4. Serbestlik derecesi

Stein'in yansız risk tahmini teorisine göre verilen herhangi bir uyum yönteminin serbestlik derecesi

$$df(\hat{\mu}) = \frac{1}{\sigma^2} \sum_{i=1}^n Cov(y_i, \hat{\mu}_i) \quad (3.44)$$

biçiminde ifade edilir (Ryan Tibshirani ve Wasserman, 2015, s. 3). Verilen bir \mathbf{x} sabit vektörü yardımıyla μ ortalama ve σ^2 varyanslı normal dağılıma sahip \mathbf{y} yanıt vektörü ($y \sim (\mu, \sigma^2 I)$) $\hat{\mu}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ yönteminden yararlanarak tahmin edilsin. Eđer $\hat{\mu}$ doğrusal bir tahminci ($\hat{\mu} = \mathbf{S}\mathbf{y}$) ve \mathbf{S} matrisi \mathbf{y} 'den bağımsız $cov(\hat{\mu}, \mathbf{y}) = \sigma^2 \mathbf{S}$ olacağından serbestlik derecesi \mathbf{S} matrisinin izine eşit olur: $df(\hat{\mu}) = tr(\mathbf{S})$ (Zou, Hastie ve Tibshirani, 2007, s. 2).

Lasso regresyonun ayar parametresi λ 'nın $[+\infty, 0]$ aralığındaki deęerine baęlı olarak aktif küme büyüklüęü deęiştikçe, lasso regresyonun serbestlik derecesi \mathbf{X} matrisinin tam ranka sahip olması ve modele giren bir bağımsız deęişkenin modeli bir daha terk etmedięi varsayımı altında aktif küme büyüklüęüne eşittir (Genç, 2020, s. 45) (Efron ve dięerleri, 2004).

$$df(\lambda) = E|A| \quad (3.45)$$

3.4. Elastic Net

Elastic Net yöntemi Zou ve Hastie (2003) tarafından geliştirilmiş bir düzenleştirci regresyon yöntemidir. Ridge regresyon ve lasso regresyonun ceza parametrelerini tek bir fonksiyonda birleştiren elastic-net bir α deęeriyle ağırlıklandırılan ℓ_1 ve ℓ_2 normlarını aynı fonksiyonda bir araya getiren hibrit bir yöntemdir. Bağımsız deęişkenler arasında yüksek korelasyon mevcut olduğunda lasso regresyon yönteminin ayar parametresi olan λ , $[0, +\infty]$ aralığında deęiştikçe lasso yolu boyunca modele giren ve çıkan deęişkenler kararsız bir yapı gösterirler. Elastic-Net tahmincisi çoklu baęıntı durumunda daha kararlı davranan, daha düşük tahmin hatasına sahip olan ve lasso regresyon gibi deęişken

seçimini de eş zamanlı olarak gerçekleştiren sürekli bir tahmincidir (Hastie ve diğerleri, 2015, s. 57).

Zou ve Hastie (2003) çalışmalarında lasso regresyonun başarısız olduğu üç farklı senaryodan söz etmişlerdir:

- Lasso regresyon $p > n$ olduğunda doğrusal bir regresyon modeli tahmin etmesine karşın modelde en fazla n adet değişkene yer vermektedir.
- Bir grup bağımsız değişken arasında yüksek korelasyon mevcut ise lasso regresyon bu değişkenlerden genelde birini modele alırken diğerlerini model dışında bırakmakta ve katsayılarını aynı oranda düzenlememektedir.
- $n > p$ durumunda bağımsız değişkenler arasında yüksek korelasyon mevcut ise ridge regresyon lasso regresyona göre daha başarılı tahminde bulunmaktadır.

Elastic-Net yönteminde de \mathbf{X} matrisinin sütunları $\left(\frac{1}{N} \sum_{i=1}^N x_{ij} = 0\right)$ ve $\frac{1}{N} \sum_{i=1}^N x_{ij}^2 = 0$ olacak biçimde) ve yanıt vektörü elemanları sıfır ortalamaya $\left(\frac{1}{N} \sum_{i=1}^N y_i = 0\right)$ sahip olacak şekilde standartlaştırılır. Zou ve Hastie'nin (2003) naive elastic net olarak adlandırdıkları fonksiyon $\lambda_1, \lambda_2 \geq 0$ için aşağıdaki gibidir:

$$\hat{\beta}_{Naive}(\lambda_1, \lambda_2) = \arg \min_{\beta} L(\lambda_1, \lambda_2, \beta) = \arg \min_{\beta} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \right\} \quad (3.46)$$

$\lambda_1 = 0$ olduğunda Eşitlik (3.46)'deki problem ridge regresyon, $\lambda_2 = 0$ olduğunda lasso regresyon problemine dönüşür. $\lambda_2 > 0$ olduğunda $\hat{\beta}_{EN} = \hat{\beta}_{EN}(\lambda_1, \lambda_2)$ tek bir çözüme sahiptir ve kesin konvektir. $\alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2}$ olarak ifade edilirse bu durumda denklem (3.46)

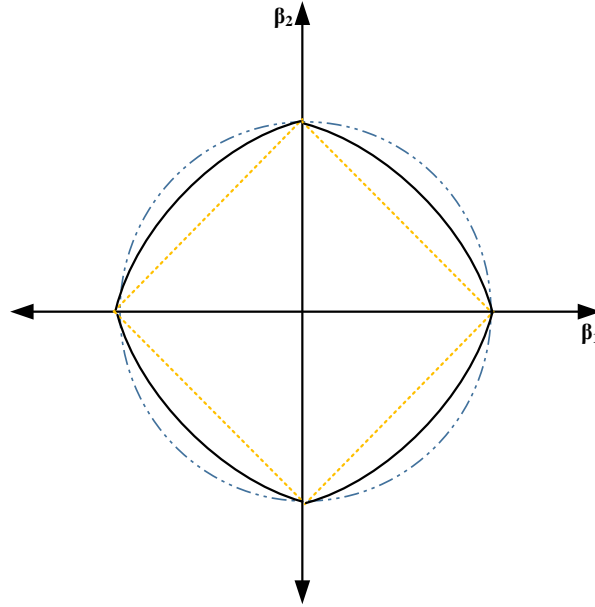
aşağıdaki gibi gösterilebilir:

$$\hat{\beta}_{Naive} = \arg \min_{\beta} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + (1-\alpha) \|\beta\|_1 + \alpha \|\beta\|_2^2 \right\} \quad (3.47)$$

Görüldüğü üzere elastic netin ceza parametresi sırasıyla α ve $(1-\alpha)$ ile ağırlıklandırılmış ℓ_1 ve ℓ_2 normlarının toplamsal bir fonksiyonudur. α parametresi $[0,1]$ aralığında değerler alır. $\alpha = 1$ olduğunda elastic net problemi ridge regresyona,

$\alpha = 0$ olduğunda lasso regresyona dönüşür. $\alpha \in [0,1]$ aralığında elastic net ceza fonksiyonu birinci türevi olmayan tekil bir fonksiyondur ve tüm $\alpha > 0$ için kesin konvektir (Zou ve Hastie, 2003, s. 303).

Şekil 3.7’de ridge, lasso ve elastic net yöntemlerinin HKT konturları bir arada gösterilmektedir. En dıştaki kesikli çizgi ridge ceza parametresinin, en içteki kesikli dörtgen lasso ceza parametresinin şeklini gösterirken; bu ikisinin ortasındaki düz çizgi ile gösterilen şekil de $\alpha = 0,5$ değeri için elastic net yönteminin ceza parametresinin kontur grafiğini göstermektedir. Ridge ve lasso cezalarını bir araya getiren elastic net probleminin kısıt alanı α parametresine bağlı olarak değişmektedir. $0 < \alpha < 1$ olduğu sürece elastic net cezasının kısıt alanı daima şekilde görüldüğü gibi lasso ve ridge cezalarının kısıt alanlarının arasında yer alacaktır. Şekil 3.6’da görülebilecek bir diğer önemli özellik ise elastic net ceza fonksiyonu $\alpha > 0$ için kesin konvektir ve $\alpha = 0$ için birinci türevi olmayan tekil bir fonksiyondur.



Şekil 3.6. Elastic Net tahmininin gösterimi (Zou ve Hastie, 2003).

3.4.1. Naive elastic net

Eşitlik (3.46)’i çözmek için verilen bir (Y, X) veri kümesi ve (λ_1, λ_2) parametre değerleri için (Y^*, X^*) yapay bir veri kümemiz olsun:

$$\mathbf{X}_{(n+p) \times p}^* = (1 + \lambda_2)^{-\frac{1}{2}} \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2} \mathbf{I} \end{pmatrix}, \mathbf{Y}_{(n+p)}^* = \begin{pmatrix} \mathbf{Y} \\ 0 \end{pmatrix} \quad (3.48)$$

$\beta^* = \sqrt{1 + \lambda_2} \beta$ ve $\gamma = \frac{\lambda_1}{\sqrt{1 + \lambda_2}}$ olduğunda naive elastic net fonksiyonu

$$L(\gamma, \beta^*) = \|\mathbf{Y}^* - \mathbf{X}^* \beta^*\|_2^2 + \gamma \|\beta^*\|_1 \quad (3.49)$$

biçiminde gösterilir. $\hat{\beta}^* = \arg \min_{\beta^*} L(\gamma, \beta^*)$ olduğuna göre eşitliği çözdüğümüzde katsayılar vektörünün tahmini

$$\hat{\beta}_{Naive} = \frac{1}{\sqrt{1 + \lambda_2}} \hat{\beta}^* \quad (3.50)$$

olur. (λ_1, λ_2) bilindiğinde ortogonal tabanda naive elastic net tahmincisi

$$\hat{\beta}_i(naive) = \frac{\left(\left| \hat{\beta}_i^{EKK} \right| - \frac{\lambda_1}{2} \right)_+}{1 + \lambda_2} \text{sgn} \left(\hat{\beta}_i^{EKK} \right) \quad (3.51)$$

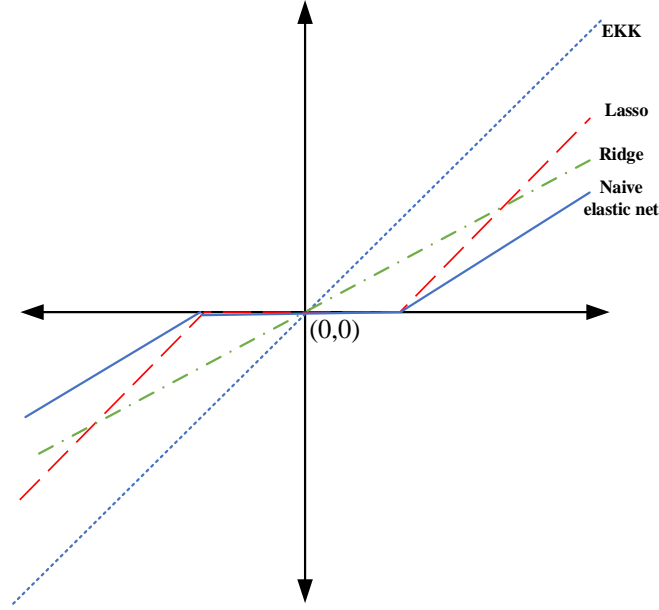
olur. Buna göre ridge regresyon tahmincisi $\hat{\beta}_{ridge} = \frac{\hat{\beta}_{EKK}}{1 + \lambda_2}$ ve lasso tahmincisi

$\hat{\beta}_{lasso} = \left(\left| \hat{\beta}_i^{EKK} \right| - \frac{\lambda_1}{2} \right)_+ \text{sgn} \left(\hat{\beta}_i^{EKK} \right)$ olacaktır. Naive elastic net yöntemi EKK katsayılarına

öncelikle yumuşak eşikleme uyguladıktan sonra katsayıları $1 + \lambda_2$ çarpanıyla düzenlemektedir (Zou ve Hastie, 2003, s. 7).

Ridge ve lasso regresyonun ceza parametrelerini birleştirerek her iki yöntemin özelliklerini bir araya getirirse de naive elastic net yöntemi ampirik çalışmalarda ridge ve lasso regresyon yöntemlerinden daha başarılı tahminler gerçekleştirmemektedir. Bunun sebebi iki aşamalı bir düzenleme yöntemine sahip olmasıdır. Naive elastic net yöntemi öncelikle λ_2 parametresini belirledikten sonra, yumuşak eşikleme operatörü yardımıyla lasso düzenlemesini gerçekleştirir. Düzenleyici yöntemler bir miktar yanlılığı artırarak varyansı azaltsalar da naive elastic net yöntemi, iki adımlı yöntemi gereği çift düzenleme gerçekleştirdiğinden dolayı varyansı azaltamamakla birlikte yanlılığı da lasso ve ridge regresyona göre daha fazladır (Zou ve Hastie, 2003, s. 8). Naive elastic net yönteminin regresyon katsayılarını düzenlediği görsel Şekil 3.7’de verilmiştir.

Görülebileceği üzere ortogonal durumda naive elastic net iki defa düzenelem yaptığı için katsayı tahminleri ridge ve lassoya göre sifıra daha yakındır.



Şekil 3.7. Ortogonal durumda naive elastic net (Zou ve Hastie, 2003)

3.4.2. Elastic net tahmincisi

Naive elastic net yönteminin bu kusurunun üstesinden gelmek için Zou ve Hastie (2003) naive elastic net tahmincisini yeniden ölçeklendirerek elastic net adını verdikleri tahminciyi önermişlerdir. Elastic net tahmincisi

$$\hat{\beta}_{EN} = \sqrt{1 + \lambda_2} \hat{\beta}^* \quad (3.52)$$

biçiminde tanımlanır. Naive elastic net tahmincisi $\hat{\beta}_{Naive} = \frac{1}{\sqrt{1 + \lambda_2}} \hat{\beta}^*$ olduğuna göre

elastic net tahmincisi naive elastic net tahmininin yeniden ölçeklendirilmiş bir formudur:

$$\hat{\beta}_{EN} = (1 + \lambda_2) \hat{\beta}_{Naive} \quad (3.53)$$

Bu dönüşüm sayesinde naive elastic net yönteminin avantajları korunurken çift düzenleme işlemi ortadan kaldırılır (Zou ve Hastie, 2003, s. 309).

$(1 + \lambda_2)$ çarpanının ölçekleme katsayısı olarak kullanılmasının sebebi ridge regresyonun çarpanlarına ayrıştırılmasında görülebilir. $\hat{\beta}_{ridge} = \mathbf{R}\mathbf{y} = (\mathbf{X}^T\mathbf{X} + \lambda_2\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$ tahmincisini ele aldığımızda \mathbf{R}^* , EKK şapka matrisi olmak üzere \mathbf{R} matrisi

$$\mathbf{R} = \frac{1}{1+\lambda_2} \mathbf{R}^* = \frac{1}{1+\lambda_2} \begin{pmatrix} 1 & \frac{\rho_{12}}{1+\lambda_2} & \dots & \frac{\rho_{1p}}{1+\lambda_2} \\ & 1 & \dots & \dots \\ & & \dots & \dots \\ & & & 1 & \frac{\rho_{p-1,p}}{1+\lambda_2} \\ & & & & 1 \end{pmatrix}^{-1} \mathbf{X}^T \quad (3.54)$$

denkleme eşit olur. Denklem (3.54)'te yer alan korelasyon matrisindeki kısmi korelasyonların $1/(1+\lambda_2)$ çarpanı ile düzenlendiği görülmektedir. Bu işlem sayesinde yüksek korelasyona sahip bağımsız değişkenlerin yer aldığı bir modelde ridge regresyon daha doğru tahminler üretebilmektedir. Ridge regresyonun gruplama etkisi özelliği olarak adlandırılan bu özellik ile lasso'nun seyreklik özelliği bir araya getirildiğinde $1/(1+\lambda_2)$ 'nin kullanıldığı düzenleyici adıma gerek kalmaz (Zou ve Hastie, 2003, s. 308)

Verilen (Y, X) veri kümesi ve (λ_1, λ_2) için elastic net regresyonunun tahmincisi

$$\hat{\beta}_{EN} = \arg \min_{\beta} \beta^T \left(\frac{\mathbf{X}^T \mathbf{X} + \lambda_2 \mathbf{I}}{1+\lambda_2} \right) \beta - 2\mathbf{Y}^T \mathbf{X} \beta + \lambda_1 \|\beta\|_1 \quad (3.55)$$

olur. Görülebileceği üzere $\lambda_2 = 0$ olduğunda lasso tahmincisi

$$\hat{\beta}_{Lasso} = \arg \min_{\beta} \beta^T X^T X \beta - 2Y^T X \beta + \lambda_1 \|\beta\|_1 \quad (3.56)$$

elastic netin özel bir durumu olarak elde edilir. Elastic net problemi için sağlanması gereken KKT koşulları aşağıdaki gibidir (R. J. Tibshirani, 2013, s. 1472):

$$\begin{aligned} X^T (y - X \hat{\beta}_{EN}) - \lambda_2 \hat{\beta}_{EN} &= \lambda_1 \gamma, \\ \gamma_i &\in \begin{cases} \left\{ \text{sign}(\hat{\beta}_i^{EN}) \right\}, & \hat{\beta}_i^{EN} \neq 0 \text{ ise} \\ [-1, 1] & , \hat{\beta}_i^{EN} = 0 \text{ ise} \end{cases} \end{aligned} \quad (3.57)$$

Elastic net tahmincisi ile lasso tahmincisi karşılaştırıldığında eşitlik (3.46)'nın lasso türü bir optimizasyon problemi olduğu görülür. Kısım 3.3.3.'te verilen döngüsel dik koordinat iniş algoritması (3.46)'daki minimizasyon problemine uygulandığında elastic net çözümü

$$\hat{\beta}_j \leftarrow \frac{S \left(\frac{1}{N} \sum_{i=1}^N r_i^{(j)} x_{ij}, \lambda_1 \right)}{1+\lambda_2} \quad (3.58)$$

biçiminde elde edilir. Eşitlikteki $S(x) = \text{sign}(x)(|x| - \lambda)_+$ yumuşak eşitleme operatörüdür (Friedman, Hastie ve Tibshirani, 2010, s. 5).

4. FİRMA PERFORMANS ANALİZİ

Bu bölümde düzenleştirici yöntemler olan ridge regresyon, lasso regresyon ve elastic-net yöntemleri ile sıradan değişken seçim yöntemleri olan tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemlerinin karşılaştırmalı analizleri yer almaktadır. Bu amaçla işletme finansal tablolarına dayalı olarak hesaplanan çeşitli finansal oranların firma performansına etkileri doğrusal regresyon modelleri ile uyumlandırılmış ve sonuçlar raporlanmıştır. İzleyen kesimde ilk olarak firma performansı ve finansal oranlara ilişkin literatür taramasına yer verilecek, araştırmanın evreni ve veri tanımlandıktan sonra ampirik bulgular sunulacaktır.

Finansal oranlar, işletmelerin finansal performansları hakkında bilgi edinmek amacıyla yatırımcılar tarafından yaygın bir kullanıma sahiptir. İşletmelerin bilanço, gelir tablosu ve nakit akış tablosundaki kalemlerden yararlanılarak hesaplanan finansal oranlar finansal tablolarda yer alan iki nicelik arasındaki matematiksel ilişkiyi gösterirler.

Barnes (1987, s. 450) finansal oranların iki temel amaç için kullanıldığını belirtmiştir:

- Farklı bilanço büyüklüklerine sahip firmaların likidite, kârlılık, borçluluk, etkinlik vb. durumlarını incelerken, işletme büyüklüğünün incelenen finansal değişkenler üzerindeki etkisini kontrol etmek
- Firmaların performanslarını sektör ortalamaları ile karşılaştırmak

Finansal oranlar esas olarak yatırımcılar tarafından bir firmanın mali durumu hakkında bilgi edinmek amacıyla kullanılırlar. Finansal oranlar sayesinde aynı sektörde faaliyet gösteren farklı bilanço büyüklüklerine sahip ve farklı sektörlerdeki firmalar birbirleriyle karşılaştırılabilirler. Ayrıca karar vericiler, bir firmanın sektöre göre nasıl bir finansal yapıya sahip olduğu hakkında da finansal oranlar yardımıyla bilgi sahibi olabilirler (Altman, 1968). Buna ek olarak işletmenin genel risk durumunun değerlendirilmesi ve işletmenin gelecekteki performansının tahmin edilmesi amacıyla da finansal oranlardan yararlanır. Öte yandan finansal oranlar borçlanmak isteyen firma ile borç veren finansal kuruluş arasında hazırlanan kredi sözleşmelerinde firmanın faaliyetlerini sınırlandırmak amacıyla birer gösterge olarak da kullanılırlar (Faello, 2015, s. 76).

Yukarıda sayılan farklı amaçlar için çeşitli faydalar sağlayan finansal oranlar birtakım kısıtlara da sahiptir. Bunların en başında veri girişinden kaynaklı insan hataları yer alır. İşletme tarafından finansal tablolarda hatalı olarak raporlanan bir kalem, ilgili finansal oranların da yanlış hesaplanmasına ve dolayısıyla akademik çalışmalarda da yanlış sonuçlara varılmasına yol açacaktır. Bir diğer kısıt ise işletmenin kullandığı muhasebe yönteminde zaman içinde değişikliğe gitmesidir. Böyle bir durumda işletmenin finansal bilgileri farklı dönemler arasında karşılaştırılmaz hale gelir. Yapılan birçok çalışmada finansal oran verilerinin normal dağılmadıkları ve değişen varyansa sahip oldukları görülmüştür. Öte yandan finansal oran verileri aykırı değerler barındırmaya son derece müsaittir. Sonuç olarak EKK gibi parametrik yöntemler ile finansal oranların analiz edilmesi için bir takım ön analizlerin yapılması gerekmektedir (Faello, 2015, s. 77).

Gelir tablosu tüm mali yılı kapsarken bilanço zamanda belirli bir ana ilişkin bilgi verdiği için bu iki tabloda yer alan kalemlerin finansal oranlar yardımıyla karşılaştırılması birtakım zorluklara yol açmaktadır. Bilanço kalemlerinin yıl boyunca ortalamaları hesaplandıktan sonra gelir tablosu kalemleri ile karşılaştırılması ve finansal oranların hesaplanması en ideal uygulamadır. Ancak firmanın harici kullanıcıları bu veriye erişemedikleri için finansal oranların hesaplanmasında genellikle yıl sonu bilanço verileri ya da birbirini takip eden iki yılın ortalama değerleri kullanılır. Bu yaklaşım yıl boyunca gerçekleşen değişimi bir ölçüde azaltsa da bilanço kalemlerinde meydana gelen mevsimsel ya da döngüsel değişimlere bağlı problemleri ortadan kaldırmaz (Gibson, 2000, s. 162).

Finansal oranlar sıklıkla likidite oranları, finansal yapı oranları, devir hızları ve kârlılık oranları biçiminde gruplandırılırlar. Likidite oranları bir işletmenin kısa vadeli borçlarını ve ani nakit çıkışlarını karşılayabilme gücünü değerlendirirler. Finansal yapı oranları ise işletmenin uzun dönemli borçlarını ödeyebilme gücü hakkında bilgi verirler. Faaliyet ya da etkinlik oranları, işletmenin satışlarını artırmak için varlıklarını ne kadar etkin kullandığını gösterirken son olarak kârlılık oranları ise işletmenin öz kaynaklarına, varlıklarına ya da satışlarına göre ne kadar kâr/zarar ürettiğini gösterir (Delen, Kuzey ve Uyar, 2013) (Ohlson, 1980). Firma performanslarının temsili için ise sıklıkla kârlılık oranlarından yararlanılmaktadır. En sık olarak kullanılan kârlılık oranları; yatırım sermayesi getirisi (ROI), öz kaynak getirisi (ROE), aktif getiri (ROA) ve gelir getirisi (ROR) (Rothaermel, 2018, s. 147).

Finansal oranlar, sađ kalım analizi ile mali başarısızlıđın tahmini (Aktař, 1997; Altman, 1968; Beaver, 1966; Deakin, 1972; Dikmen, 2007; Edmister, 1972; Gepp ve Kumar, 2015; Ohlson, 1980), kredi notunun belirlenmesi ve risk analizi (Backer ve Gosman, 1980; Golbayani, Florescu ve Chatterjee, 2020; Horrigan, 1966; Huang, Chen, Hsu, Chen ve Wu, 2004; Piotroski, 2000), deđerleme (Nissim ve Penman, 2001) ve firma performansını etkileyen faktörlerin ilişkişel analizlerinde de kullanılmaktadırlar.

Finansal oranların ilk olarak kullanılmaya başlanması 19. yüzyılın sonuna denk gelmektedir. A.B.D.'de ticari bankaların işletmelerin kısa vadeli borçlarını ödeme kabiliyetlerini anlamak amacıyla dönen varlıkların kısa vadeli borçlarına oranlanması sonucuyla elde edilen cari oranın hesaplanması finansal oranların kullanımının başlangıcı olarak kabul edilir (Horrigan, 1968).

İzleyen yıllarda finansal oranların hangi işletmelerin başarısız olacaklarının belirlenmesinde bir gösterge olarak kullanılıp kullanılmayacağına dair sorular ortaya atılmış ve finansal oranların öngörü kabiliyetleri incelenmiştir (Shimerda, 1978, s. 3). Ancak istatistiksel modellerin kullanıldığı ilk ampirik çalışmalar Beaver (1966) ve Altman'a (1968) aittir. Beaver (1966) çalışmasında işletmelerin başarısızlıklarını tahmin etmek amacıyla Bayeşçi ikili sınıflandırma yöntemi için 29 farklı finansal oranı kullanmış ve nakit/borç oranını işletme başarısızlığının tahmininde en önemli oran olarak belirlemiştir. Altman ise imalat sektöründeki işletmelerin başarısızlıklarını tahmin etmek için finansal oranları çoklu diskriminant analizi yöntemiyle incelemiş ve oluşturduğu diskriminant fonksiyonunda 5 finansal orana yer vermiştir: Çalışma sermayesi/toplam varlıklar, dağıtılmayan kârlar/toplam varlıklar, FVÖK/toplam varlıklar, öz kaynakların piyasa değeri/toplam borç, toplam satışlar/toplam varlıklar.

Performans, ulaşılmak istenen bir amacın etkinliğini ve etkililiğini ifade eden son derece dinamik bir kavramdır. Firma performansının bir deđişken olarak ele alınıp açıklandığı birçok nicel ve nitel çalışma olsa da firma performansının tanımı üzerinde tam bir uzlaşya varılamamıştır. Miller vd. (2013) çalışmalarında firma performansı ile ilgili olarak literatürde ciddi kavramsal ve ampirik sorunların var olduğuna işaret ederek inceledikleri 290 çalışmanın yalnızca %34'ünün firma performansı kavramının tanımına ilişkin olarak kendi içinde tutarlı olduklarını belirtmişlerdir. Firma performansının çok boyutluluğunu kabul eden çalışmalarda dahi çalışmanın kapsamına bađlı olarak firma

performansı tek boyutlu olarak temsil edilmekte ve genellikle tek bir gösterge ile ölçülmektedir (Santos ve Brito, 2012, s. 97).

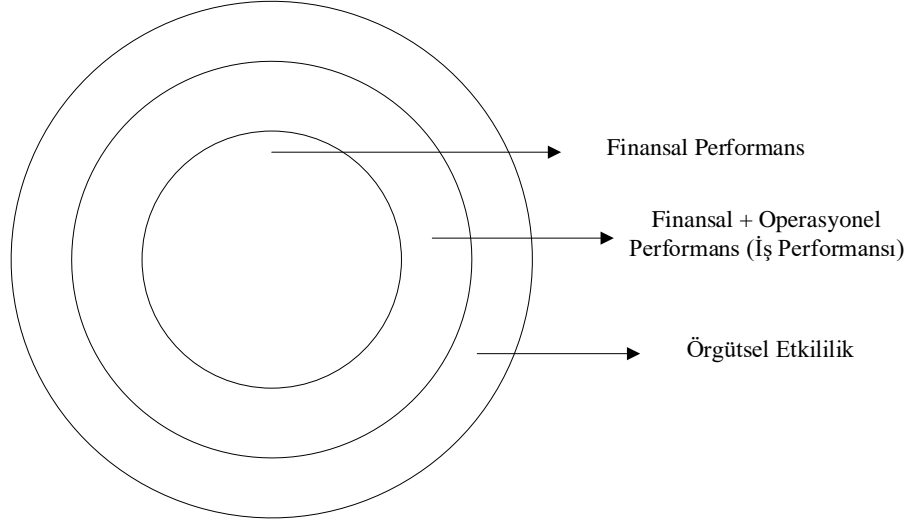
Venkatraman ve Ramanujam (1986, s. 802) birbirlerinden farklı araştırma sorularına sahip farklı alanlardaki çalışmaların işletme performansını da farklı biçimlerde ölçmelerinin tabii olduğunu belirtmişlerdir. Ayrıca Venkatraman ve Ramanujam'a göre çok disiplinli bir yaklaşım, performans ölçümünün nasıl yapılacağına ilişkin çeşitli disiplinlerdeki terminoloji ve varsayımların farklılıklarını vurgulamaktan öteye geçemeyeceği için performans ölçümünde kasıtlı olarak tek disiplinli yaklaşımlar benimsenmektedir.

Firma performansı genellikle kârlılık, büyüklük, müşteri memnuniyeti, piyasa değeri, hissedarların getirileri, ekonomik katma değer gibi birçok ölçütle tanımlanır ve ölçülür. Yatırımcılar, kredi verenler, paydaşlar ve diğer karar vericilerin hızlı ve kolay ulaşabildikleri bilgiler işletmenin finansal bilgileri olduğu için firma performansı finansal analiz yoluyla sıklıkla ölçülmektedir. Dolayısıyla muhasebe ve finans bakış açısından firma performansı kavramı ile işletmenin finansal performansı ifade edilmektedir (Herciu, 2017, s. 80). Rothaermel (2018) firma performansının üç geleneksel çerçevede ele alındığını belirtmiştir: Muhasebe kârlılığı, paydaş değer yaratımı ve ekonomik değer yaratımı. Muhasebe verisinden yararlanarak firma performansının değerlendirilmesi standart bir yönetim uygulamasıdır. Kârlılığın değerlendirilmesinde ise gelir tablosu ve bilançodaki kalemlerden elde edilen finansal veriden ve oranlardan yararlanır.

Venkatraman ve Ramanujam (1986, s. 803) ise çalışmalarında işletme performansının örgütsel etkililik tartışmasından farklı olarak ele alınıp alınmaması gerektiği sorusunu ortaya atmışlar ve işletme performansını örgütsel etkililik kavramının bir alt kümesi olarak tanımlamışlardır. İşletme performansının alanı Şekil 4.1.'deki gibi iç içe üç katmandan oluşan bir şema ile gösterilmiştir.

Firma performansı kavramı en dar anlamda firmanın ekonomik amaçlarına ulaşım ulaşmadığını gösteren finansal performans ile tanımlanır. Finansal performansı dikkate alan yaklaşımlar firma performansının göstergesi olarak finansal tablolardan elde edilen satış miktarındaki büyüme, kârlılık, hisse başına getiri gibi ölçütleri esas alırlar. Firma performansının daha geniş bir kavramsallaştırması ise işletmenin finansal performansına ek olarak operasyonel performansını birlikte ele alan yaklaşımdır. Buna göre finansal performans göstergelerine ek olarak ürün kalitesi, pazarlama, katma değer, etkin teknoloji

kullanımı, pazar payı gibi operasyonel göstergeler firma performansının ölçümünde kullanılır. En dış halkada yer alan örgütsel etkililik ise meşru faaliyetlere katılım, kaynak edinimi ve hedeflere ulaşılması gibi örgütün işleyişiyle ilgili diğer yönleri kapsamaktadır (Venkatraman ve Ramanujam, 1986, s. 804) (Santos ve Brito, 2012, s. 98).



Şekil 4.1. Firma performansının alanları (Venkatraman ve Ramanujam, 1986, s. 803)

Combs vd. (2005) organizasyon performansının boyutlarını inceledikleri derleme çalışmalarında organizasyon performansının ölçüldüğü 367 farklı çalışmanın 190'ında performans ölçütü olarak muhasebe getirilerinin kullanıldığını belirtmişlerdir. En sık kullanılan muhasebe getirisi ölçütleri ise sırasıyla her biri birer kârlılık oranı olan aktif getiri (ROA – Return on Assets) oranı, satış getiri (ROS – Return on Sales) oranı ve öz kaynak getiri (Return on Equity) oranıdır.

Kliestik vd. (2020) finansal oranlardan yararlanarak işletmelerin finansal performanslarını tahmin etmek amacıyla 1993-2018 arasında eski Sovyet Bloğu'ndan ayrılan Doğu Avrupa ülkelerinde yapılmış 400'den fazla model belirlemişler ve inceledikleri 180 modelde en fazla cari oran, toplam borç/toplam varlıklar, toplam satışlar/toplam varlıklar oranlarının kullanıldığını bulmuşlardır. En fazla kullanılan kârlılık oranları ise sırasıyla öz kaynak getiri oranı ve aktif getiri oranıdır. Ayrıca en fazla çalışmanın finansal başarısızlık üzerine yapıldığını belirlemişlerdir. En çok kullanılan yöntemler 101 çalışma ile çoklu diskriminant analizi ve 48 çalışma ile logit ve probit modellerdir.

Yousaf ve Dey (2022) işletme performansını en iyi temsil eden finansal oranı belirlemek için 2016-2020 arasındaki verileri kullanarak otomobil, inşaat ve imalat

sektöründe faaliyet gösteren 287 Çek firmasında yaptıkları çalışmalarında belirledikleri 16 adet finansal orandan yararlanmışlar ve makine öğrenmesi tabanlı karar ağacı algoritması (CHAID) kullanarak aktif getiri oranı, öz kaynak getir oranı ve adi hisse senedi sahiplerine ait öz kaynak oranı (*return on common equity*) arasından finansal performansı en iyi temsil eden ölçütün aktif getiri oranı olduğunu bulmuşlardır.

Bayraktaroğlu vd. (2019) 2003-2013 döneminde Türkiye’de faaliyet gösteren imalat firmalarında entelektüel sermaye ve firma performansı arasındaki ilişkiyi inceledikleri çalışmalarında firma performansını ölçmek amacıyla aktif getiri oranı, öz kaynak getir oranı ve aktif devir hızı oranlarını kullanmışlardır.

Akgün ve Karataş (2021) çalışma sermayesi yönetimi ile firma performansı arasındaki ilişkiyi araştırmak amacıyla 2003-2012 yılları arasında faaliyet gösteren AB-28 listesindeki 7812 firmayı en küçük kareler regresyon yöntemi ile incelemişlerdir. Çalışmada aktif getiri oranı, öz kaynak getir oranı ve faiz ve vergi öncesi kâr marjını (FVÖK) bağımlı değişkenler olarak belirlemişlerdir.

Yukarıda verilen ilişki analizlerde ve derleme çalışmalarında da açıklandığı üzere kârlılık oranları yaygın bir biçimde firma performansının birer ölçütü olarak kabul edilmektedir. Bununla birlikte firma performansının tahmin edilmesi amacıyla oluşturulan istatistiksel tahmin modellerinde hangi finansal oranların bağımsız değişkenler olarak yer alacağı belirlenmesi önem kazanmaktadır. Yapılan bazı çalışmalarda finansal oranlardan yararlanarak işletmenin farklı boyutları ile firma performansı arasındaki ilişki incelenmektedir. Örneğin; işletmelerin sermaye yapısı (Abor, 2005; Al-Zararee, Al-Sawalhah ve Samara, 2021; Avcı, 2016; Dao ve Ta, 2020; Salim ve Yadav, 2012), çalışma sermayesi (Akgün ve Memiş Karataş, 2021; Baños-Caballero, García-Teruel ve Martínez-Solano, 2014; Jaworski ve Czerwonka, 2022; Senan, Anagreh, Al-dalaien, Khaled ve Al-Homaidi, 2021), entelektüel sermaye (Bayraktaroğlu ve diğerleri, 2019; Yousaf, 2022) gibi faktörlerin firma performansı üzerindeki etkileri literatürde yaygın olarak araştırılmaktadır.

İşletmelerin tek bir boyutu ile firma performansı arasındaki ilişkiyi inceleyen bu tür çalışmalarda kullanılacak finansal oranların belirlenmesi görece daha kolay bir problem olarak karşımıza çıkmaktadır. Ancak işletmelerin birden fazla boyutunu (likidite, sermaye yapısı, çalışma sermayesi, operasyonel etkinlik gibi) birlikte ele alan çalışmalarda çok sayıda finansal oran arasından hangilerinin modelde yer alınacağı

belirlenmesi önemli bir sorundur. Firma performansı ile anlamlı bir ilişkisi olmayan ya da modeldeki diğer finansal oranlar ile ilişkili olan (çoklu bağıntı) finansal oranların modelde yer alması sınıflandırma ya da tahmin için gereksiz enformasyon sağlamakla birlikte hatalı tahminlerin yapılmasına yol açabilmektedirler. Ayrıca daha önce açıklandığı üzere firma performansı olarak hangi ölçütün kullanılacağına dair bir uzlaşmaya varılamamasına ek olarak firma performansının belirleyicileri olarak hangi finansal oranların modelde yer alacağına dair standart bir uygulama da mevcut değildir.

İstatistiksel tahmin modellerinde yer alacak finansal oranlar sıklıkla önceki literatür çalışmalarına veya uzmanlık bilgilerine dayalı olarak ya da keyfi biçimde belirlenmektedir. Bu yüzden finansal oranların yer aldığı modellerin tahmin edilmesi çok boyutlu bir istatistik problemidir. Buna bağlı olarak bağımsız değişken kümesinde yer alacak finansal oranların belirlenmesi bir değişken seçimi (*feature selection*) problemi olarak karşımıza çıkmaktadır. Tüm finansal oranların aynı anda modelde yer alması hem yüksek hesaplama maliyetine yol açacağı ve analitik olarak problemin çözümünü karmaşık hale getireceği için optimum sayıda ve en doğru finansal oranların belirlenmesi modelin tahmin doğruluğu için son derece önem arz etmektedir (Zaini ve Mahmuddin, 2019, s. 690).

Değişken seçimi problemi yalnızca firma performansının tahmini için geliştirilen tahmin modelleri için değil aynı zamanda işletme başarısızlığının tahmini için de geçerli bir problemidir. Örneğin, Zaini ve Mahmuddin (2019) öz kaynak getiri oranlarına bakarak işletmeleri yüksek performanslı ve düşük performanslı olarak iki gruba ayırmışlar ve farklı veri madenciliği yöntemlerinin (sinir ağları, kaba küme, diskriminant analizi, lojistik regresyon, karar tablosu, zincirleme minimal optimizasyon ve karar ağaçları) sınıflandırma performanslarını karşılaştırmalı olarak incelemiştir. Bunun için öncelikle bağımsız değişken olarak belirledikleri 24 finansal oran arasından farklı değişken seçim yöntemlerinden (en iyi ilk arama, tam arama, genetik arama, aç gözlü arama, doğrusal ileri doğru seçim, alt küme seçimi, rassal arama, sıralı arama, saçılım araması) yararlanarak modellerde yer alacak 7 adet finansal oran (dönen varlıklar/toplam varlıklar, satışlar/toplam varlıklar, gelir/ toplam varlıklar, brüt kâr marjı, öz kaynaklara serbest nakit akışı, vergi öncesi kâr/toplam varlıklar ve hisse başı getiri) belirlemiştir. Sonuç olarak belirlenen 7 değişkenli alt kümenin yer aldığı modellerin yüksek tahmin başarısına sahip oldukları ve az değişkenli modeller ile tüm değişken kümesinin yer aldığı

modellerin sınıflandırma başarıları arasında çok büyük farklılıkların olmadığı; böylece fazladan yer alan değişkenlerin tahmin başarısına katkılarının sınırlı olduğu sonucuna varmışlardır.

İşletme performansının başarılı olarak tahmin edilmesi için doğru finansal oranların belirlenmesinin önemine dikkat çeken Delen vd. (2013) çalışmalarında iki aşamalı bir yaklaşım izleyerek öncelikle açıklayıcı faktör analizi yardımıyla modelde yer alacak finansal oranları belirlemişler ve sonrasında karar ağaçları yardımıyla (CHAID, C5.0, QUEST ve C&RT) finansal oranların firma performansı üzerindeki etkilerini gösteren bir tahmin modeli kurmuşlardır. Delen vd.'ne göre en önemli değişkenler vergi öncesi kâr/öz kaynak oranı ile net kâr marjıdır.

Korkmaz ve Karaca (2014) BİST imalat sanayisindeki firmaların kârlılıklarının belirleyicilerini inceledikleri çalışmalarında hisse başına kazanç (HBK), ROE ve ROA oranlarını bağımlı değişken olarak seçerek üç farklı panel regresyon modeli belirlemişlerdir. Belirledikleri 12 finansal oran arasından her bir modelde yer alacak bağımsız değişkenleri ileriye doğru adımsal regresyon yöntemi ile seçmişlerdir. Çalışmada artan borçlanma oranının hisse başına kazanç, aktif getiri ve öz sermaye getirisini negatif yönde etkilediği ve işletmelerin duran varlıklar yerine sermayelerini artırmalarının kârlılıklarını olumlu etkilediği sonucuna varılmıştır.

Sun ve Hui (2006) çalışmaya dahil ettikleri 30 finansal oran arasından optimum finansal oran kümesini genetik algoritmalar yardımıyla belirlemişler ve karar ağacı algoritması yardımıyla finansal başarısızlık tahmin modeli kurmuşlardır

Xu vd. (2014) işletme başarısızlığı tahmin modeli için önceden belirlenen 27 finansal oran arasından optimum sayıda finansal oranı bağımsız değişken kümesi olarak belirlemek amacıyla yumuşak küme (soft set), kaba küme (rough set), asal bileşenler analizi başta olmak üzere farklı boyut azaltma yöntemlerini kullanmışlardır. Farklı yöntemlerle elde edilen finansal oran kümelerini destek vektör makineleri, lojistik regresyon ve yapay sinir ağları yöntemlerine uygulayarak performanslarını karşılaştırmalı olarak değerlendirmişlerdir. Çalışmada ilk kez önerilen yeni yumuşak küme (novel soft set) yaklaşımı ile elde edilen 9 finansal oranın yer aldığı aktif değişken kümesi kullanıldığında tahmin başarısının arttığı sonucuna ulaşılmıştır.

Tascon ve Castano (2017) ise işletme başarısızlığının tahmini için ortalama farkları ve medyan farkları yöntemlerine dayalı olarak modelde yer alacak finansal oranları

belirlemiş ve doğrusal, kuadratik, lojistik ve k-en yakın komşuluk diskriminant analizleri, logit ve probit modellerle işletme başarısızlığı tahmin modelleri kurmuşlardır. Tascon ve Castano'ya göre örneklemin yapısı yalnızca istatistiksel değişken seçimini değil aynı zamanda işletme başarısızlığının değerlendirilmesinde kullanılacak en uygun yöntemi de belirlemektedir. Doğru sınıflandırma yapabilmek için değişken seçim prosedürü ile sınıflandırma metodolojisi araştırmada yer alan değişkenlerin dağılımlarına bağlı olarak seçilmelidir. Parametrik yöntemler ortalama farkları seçim yöntemi ile belirlenen değişkenlerle daha iyi sınıflandırma gerçekleştirirken; parametrik olmayan yöntemler medyan seçimi ile daha başarılı sınıflandırma yapmaktadırlar.

Tablo 4.1'de firma performansını etkileyen faktörlerin belirlenmesi ve tahmin edilmesi amacıyla son yıllarda yapılmış bazı çalışmalara ait özet bilgiler verilmiştir. Verilen çalışmaların birçoğunda önerilen modellerde yer alan bağımsız değişkenlerin literatüre dayalı olarak yazarların sübjektif tercihleri doğrultusunda belirlendikleri görülebilir.

Tablo 4.1. *Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti*

Yazar	Performans Ölçütü	Ülke	Kapsadığı Yıllar	Sektör	Yöntem	Değişkenler	Bulgular	Değişkenlerin Seçimi
Ayoush, Toumeh, Shabaneh (2021)	ROA, ROE	Ürdün	2012-2018	Karışık	Çoklu Doğrusal Regresyon	Cari oran, asit test oranı, borçluluk oranı, kaldıraç oranı, faiz karşılama oranı,	Kârlılığı en çok etkileyen faktörler sırasıyla kaldıraç oranı, likidite oranı ve ödeme gücüdür.	Alanyazına dayalı
Kim vd. (2021)	ROE, ROS	Vietnam	2014-2019	Gıda	Panel Veri Analizi ve Blinder-Oaxaca Ayrıştırma Yöntemi	Asit test oranı, kaldıraç oranı, varlık devir hızı, firma büyüklüğü, satışlardaki büyüme, TÜFE	Varlık devir hızı ve satışlardaki büyüme ile finansal performans arasında anlamlı bir ilişki mevcuttur. Kaldıraç oranı satış getirisini olumsuz etkilemektedir.	Alanyazına dayalı
Senan vd. (2021)	ROA, ROE	Hindistan	2008-2018	Bankacılık	Çoklu Doğrusal Regresyon	Cari oran, toplam varlıklar, net kâr marjı, kullanılan sermaye getirisi, vergi sonrası kâr, politika faizi oranı	Net kâr marjı, vergi sonrası kâr, para politikası ile ROA arasında ilişki bulunmaktadır. Cari oran, bilanço büyüklüğü, net kâr marjı ile ROE arasında anlamlı ilişki mevcuttur.	Alanyazına dayalı
Akgün ve Karataş (2020)	ROA, ROE, EBITM	Avrupa Birliği	2003-2012	Karışık	Çoklu Doğrusal Regresyon	Nakit tutma seviyesi, nakit interaktif etkisi, brüt çalışma sermayesi oranı, net çalışma sermayesi oranı, nakit oranı, likidite oranı, firma büyüklüğü, cari oran, kaldıraç oranı	Nakit oranı ve likidite oranı ile ROA arasında pozitif yönlü bir ilişki bulunmaktadır. Çalışma sermayesi ile ROE arasında anlamlı bir ilişki bulunmamaktadır.	Alanyazına dayalı

Tablo 4.2. (Devam) *Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti*

Lan ve Cong (2020)	ROE, ROA, ROS	Vietnam	2014-2017	Karışık	Çoklu Doğrusal Regresyon	Sermaye yeterlik oranı, toplam varlıklar, likidite oranı, kaldıraç oranı, ödeme gücü oranı	İşletme büyüklüğü ve ödeme gücü oranı ile ROA ve ROS arasında pozitif yönlü anlamlı ilişki mevcut iken ROE ile ters yönlü ilişki vardır.	Alanyazına dayalı
Taysı (2020)	ROA	Türkiye	1998-2017	Karışık	Panel Regresyon	Aktif devir hızı, aktif büyüklük, aktifteki büyüme, net satışlar, stok devir hızı, çalışma sermayesi devir hızı, kaldıraç oranı	Aktif devir hızı, aktif büyüklük ve stok devir hızı ROA'yı pozitif; aktifteki büyüme, çalışma sermayesi devir hızı ve kaldıraç oranı negatif yönde etkilemektedir.	Alanyazına dayalı
Dizgil (2019)	ROA, FVÖK/Toplam Aktifler	Türkiye	2009-2017	İçecek ve Gıda	Panel Veri Analizi	Kısa vadeli borç oranı, uzun vadeli borç oranı, öz kaynak oranı, öz sermaye kârlılık oranı, net işletme sermayesi, alacakların ortalama tahsil süresi	Likidite ve faaliyet etkinliği ile kârlılık arasında anlamlı bir ilişki mevcuttur. Kısa ve uzun vadeli borç oranlarının artması kârlılığı negatif etkilemektedir.	Alanyazına dayalı
Melwani (2019)	ROA ve Tobin's Q	Hindistan	2007-2017	Otomobil	Çoklu Doğrusal Regresyon	17 finansal oran çalışmaya dahil edilmiştir.	Sermaye yapısı oranı, maddi duran varlıklar oranı ve faaliyet kârı otomobil işletmelerinin finansal performansını pozitif yönde etkilerken, işletme büyüklüğü, kaldıraç oranı ve AR&GE yoğunluğu negatif yönde etkilemektedir.	Alanyazına dayalı

Tablo 4.3. (Devam) *Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti*

Zaini ve Mahmuddin (2019)	ROE	Malezya	2001-2010	Karışık	Veri Madenciliği	24 finansal çalışmaya edilmiştir.	oran dahil	Dönen varlıklar/toplam varlıklar, satışlar/toplam varlıklar, gelir/toplam varlıklar, brüt kâr marjı, öz kaynaklara serbest nakit akışı, vergi öncesi kâr/toplam varlıklar ve hisse başı getiri değişkenleri tahmin modelinde yer almıştır.	En iyi ilk arama, tam arama, genetik arama, aç gözlü arama, doğrusal ileri doğru seçim, alt küme seçimi, rassal arama, sıralı arama, saçılım araması
Avcı (2016)	ROA, ROE	Türkiye	2003-2015	İmalat	Panel Veri Analizi	Kaldıraç oranı, kısa vadeli borç oranı, uzun vadeli borç oranı, toplam varlıklar	kısa vadeli borç oranı, uzun vadeli borç oranı, toplam varlıklar	Kısa ve uzun vadeli borç oranları kârlılığı negatif yönde işletme büyüklüğü ise pozitif yönde etkilemektedir.	Alanyazına dayalı
İslamoğlu ve Çelik (2015)	ROA, ROE	Türkiye	2011-2014	Kağıt Sanayi	Panel Regresyon	Toplam satışlar/toplam varlıklar, satış getiri oranı, asit test oranı, Pazar payı, dış ticaret açığı, faiz oranı	Toplam satışlar/toplam varlıklar oranı dışında tüm firma ve endüstriye ilişkin oranların performans üzerinde anlamlı bir etkisi bulunmaktadır. Dış ticaret açığı ile finansal performans arasında zayıf bir ilişki mevcutken faiz oranının performans üzerinde anlamlı bir etkisi bulunmamaktadır.	Alanyazına dayalı	

Tablo 4.4. (Devam) *Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti*

Korkmaz ve Karaca (2014)	HBK, ROE, ROA	Türkiye	2000-2011	İmalat	Panel Regresyon	Satış getiri oranı, satılan maliyet/stoklar, net satışlar/ticari alacaklar, cari oran, duran varlıklar/toplam varlıklar, maddi duran varlıklar/uzun vadeli yabancı kaynaklar, net satışlar/dönen varlıklar, borç oranı	Artan borçlanma oranı olumsuz yönde etkilerken, duran varlıklar yerine işletme sermayesinin artırılması kârlılığı olumlu yönde etkilemektedir.	İleriye doğru adimsal regresyon
Delen (2013)	ROE, ROA	Türkiye	2005-2011	Karışık	Karar Ağacı Algoritmaları	29 finansal orana çalışmaya edilmiştir.	VÖK/Öz kaynak oranı ve net kâr marjı ve borç oranı kârlılığın tahmin edilmesinde en önemli finansal oranlardır.	Açıklayıcı faktör analizi
Karadeniz ve İskenderoğlu (2011)	ROA	Türkiye	2002-2009	Turizm	Bütünleşik Regresyon	Kaldıraç oranı, kısa vadeli borç oranı, uzun vadeli borç oranı, toplam varlıklar, pazar payı, alacak devir hızı, stok devir hızı, net satışlar/toplam varlıklar, net işletme sermayesi/toplam varlıklar	Kaldıraç oranının aktif kârlılığı üzerinde negatif; ölçek, pazar payı, net işletme sermayesi devir hızı ve aktif devir hızının ise pozitif etkili olduğu sonucuna ulaşılmıştır.	Alanyazına dayalı

Tablo 4.5. (Devam) *Firma performansının finansal oranlarla tahmin edilmesine ilişkin literatür özeti*

Şahin (2011)	ROA, ROE, Kâr Marjı Oranı, Firma Değeri Oranı	Türkiye	2006-2010	Karışık	Panel Veri Analizi	16 finansal çalışmaya edilmiştir.	oran dahil	Öz sermaye karlılığı açısından en önemli faktörün maliyetler olduğu ve kaldıraç etkisinden dolayı borçluluğun artması firma performansını artırmaktadır.	Alanyazına dayalı
Albayrak ve Akbulut (2008)	ROA, ROE, Kâr Marjı oranları, Hisse Başına Kâr	Türkiye	2004-2006	Sanayi ve Hizmet	Çoklu Doğrusal Regresyon	Likidite, varlık kullanım etkinliği, sermaye yapısı, piyasa değeri ve işletme büyüklüğünü temsilen 18 farklı oran ve gösterge seçilmiştir.		Kârlılığı tanımlamada en önemli göstergeler likidite oranları, sermaye yapısı, işletme büyüklüğü ve stok devir hızıdır.	Alanyazına dayalı
Abor (2005)	ROE	Gana	1998-2002	Karışık	Panel Veri Analizi	Kısa vadeli borç/toplam sermaye, uzun vadeli borç/toplam sermaye, toplam borç/toplam sermaye, toplam satışlar, satışlardaki artış		KVB/T.Varlık oranı ile ROE arasında anlamlı pozitif ilişki. UVB/T.Varlık oranı ile ROE arasında negatif ilişki. Toplam Borç oranı ile ROE arasında pozitif ilişki.	Alanyazına dayalı

4.1. Araştırmanın Evreni ve Örneklemi

Araştırmada kullanılan finansal oranlar Türkiye Cumhuriyet Merkez Bankası'nın web sitesindeki yönergeye göre hesaplanmıştır. Buna göre finansal oranlar dört temel gruba ayrılmıştır: Likidite oranları, finansal yapı oranları, faaliyet oranları ve kârlılık oranları. Tablo 4.6'da Türkiye Cumhuriyet Merkez Bankası'nın (TCMB) önerdiği sekiz adet likidite oranı ve hesaplama formülleri yer almaktadır.

Tablo 4.6. Likidite oranları

Kısaltma	Oran	Formül
L1	Cari Oran	$\frac{\text{Dönen Varlıklar}}{\text{Kısa Vadeli Yabancı Kaynaklar}}$
L2	Asit-Test Oranı	$\frac{\text{Dönen Varlıklar} - (\text{Stoklar} + \text{Gelecek Aylara Ait Giderler} + \text{Diğer Dönen Varlıklar})}{\text{Kısa Vadeli Yabancı Kaynaklar}}$
L3	Nakit Oranı	$\frac{\text{Hazır Değerler} + \text{Menkul Kıymetler}}{\text{Kısa Vadeli Yabancı Kaynaklar}}$
L4	Stoklar/Dönen Varlıklar Oranı	$\frac{\text{Stoklar}}{\text{Dönen Varlıklar}}$
L5	Stoklar/Varlık (Aktif) Top. Oranı	$\frac{\text{Stoklar}}{\text{Varlık (Aktif) Toplamı}}$
L6	Stok Bağımlılık Oranı	$\frac{\text{Kısa Vadeli Yabancı Kaynaklar} - (\text{Hazır Değerler} + \text{Menkul Kıymetler})}{\text{Stoklar}}$
L7	Kısa Vadeli Alacaklar/Dönen Varlık Oranı	$\frac{\text{Kısa Vadeli Ticari Alacaklar} + \text{Diğer Kısa Vadeli Alacaklar}}{\text{Dönen Varlıklar}}$
L8	Kısa Vadeli Alacaklar/Varlık Toplamı Oranı	$\frac{\text{Kısa Vadeli Ticari Alacaklar} + \text{Diğer Kısa Vadeli Alacaklar}}{\text{Varlık (Aktif) Toplamı}}$

Kaynak: <http://www3.tcmb.gov.tr/sector/2017/Raporlar/oran.pdf>

İşletmelerin uzun dönem borç ödeyebilme kabiliyetlerini ölçmek amacıyla kullanılan finansal yapı oranları ve formülleri TCMB'nin önerdiği biçimde Tablo 4.7'de verilmiştir. İşletmelerin varlıklarının etkin kullanımları için hesaplanan faaliyet oranları ve formülleri Tablo 4.8'de yer almaktadır.

Bu oranlardan Tablo 4.6, Tablo 4.7, ve Tablo 4.8'de verilen likidite, finansal yapı ve faaliyet oranları araştırmada bağımsız değişkenler olarak kabul edilmiştir.

Tablo 4.7. Finansal yapı oranları

Kısaltma	Oran	Formül
F1	Kaldıraç Oranı	$\frac{\text{Kısa Vadeli Yabancı Kaynaklar} + \text{Uzun Vadeli Yabancı Kaynaklar}}{\text{Varlık Toplamı}}$
F2	Öz Kaynak/Varlık Toplamı Oranı	$\frac{\text{Öz Kaynaklar}}{\text{Varlık Toplamı}}$
F3	Öz Kaynak/Yabancı Kaynaklar Toplamı Oranı	$\frac{\text{Öz Kaynaklar}}{\text{Kısa Vadeli Yabancı Kaynaklar} + \text{Uzun Vadeli Yabancı Kaynaklar}}$
F4	Kısa Vadeli Yabancı Kaynaklar/Kaynak (Pasif) Toplamı Oranı	$\frac{\text{Kısa Vadeli Yabancı Kaynaklar}}{\text{Kaynak (Pasif) Toplamı}}$
F5	Uzun Vadeli Yabancı Kaynaklar/Kaynak (Pasif) Toplamı Oranı	$\frac{\text{Uzun Vadeli Yabancı Kaynaklar}}{\text{Kaynak (Pasif) Toplamı}}$
F6	Uzun Vadeli Yabancı Kaynaklar/Devamlı Sermaye Oranı	$\frac{\text{Uzun Vadeli Yabancı Kaynaklar}}{\text{Uzun Vadeli Yabancı Kaynaklar} + \text{Öz Kaynaklar}}$
F7	Maddi Duran Varlıklar/Öz Kaynaklar Oranı	$\frac{\text{Maddi Duran Varlıklar (Net)}}{\text{Öz Kaynaklar}}$
F8	Maddi Duran Varlıklar/Uzun Vadeli Yabancı Kaynaklar Oranı	$\frac{\text{Maddi Duran Varlıklar (Net)}}{\text{Uzun Vadeli Yabancı Kaynaklar}}$
F9	Duran Varlıklar/Yabancı Kaynaklar Toplamı Oranı	$\frac{\text{Duran Varlıklar}}{\text{Kısa Vadeli Yabancı Kaynaklar} + \text{Uzun Vadeli Yabancı Kaynaklar}}$
F10	Duran Varlıklar/Öz Kaynaklar Oranı	$\frac{\text{Duran Varlıklar}}{\text{Öz Kaynaklar}}$
F11	Duran Varlıklar/Devamlı Sermaye Oranı	$\frac{\text{Duran Varlıklar}}{\text{Uzun Vadeli Yabancı Kaynaklar} + \text{Öz Kaynaklar}}$
F12	Kısa Vadeli Yabancı Kaynaklar/Yabancı Kaynaklar Toplamı Oranı	$\frac{\text{Kısa Vadeli Yabancı Kaynaklar}}{\text{Kısa Vadeli Yabancı Kaynaklar} + \text{Uzun Vadeli Yabancı Kaynaklar}}$
F13	Banka Kredileri/Varlık Toplamı Oranı	$\frac{\text{Kısa Vadeli Banka Kredileri} + \text{Uzun Vadeli Kredi Anapara Taksit ve Faizleri} + \text{Uzun Vadeli Banka Kredileri}}{\text{Varlık Toplamı}}$
F14	Kısa Vadeli Banka Kredileri/Kısa Vadeli Yabancı Kaynaklar Oranı	$\frac{\text{Kısa Vadeli Banka Kredileri} + \text{Uzun Vadeli Kredi Anapara Taksit ve Faizleri}}{\text{Kısa Vadeli Yabancı Kaynaklar}}$
F15	Banka Kredileri/Yabancı Kaynaklar Toplamı	$\frac{\text{Kısa Vadeli Banka Kredileri} + \text{Uzun Vadeli Kredi Anapara Taksit ve Faizleri} + \text{Uzun Vadeli Banka Kredileri}}{\text{Kısa Vadeli Yabancı Kaynaklar} + \text{Uzun Vadeli Yabancı Kaynaklar}}$
F16	Dönen Varlıklar/Varlık Toplamı Oranı	$\frac{\text{Dönen Varlıklar}}{\text{Varlık Toplamı}}$
F17	Maddi Duran Varlıklar/Varlık Toplamı Oranı	$\frac{\text{Maddi Duran Varlıklar}}{\text{Varlık Toplamı}}$

Kaynak: https://www3.tcmb.gov.tr/sector/dosyalar/menu/ratios_tr.pdf

Tablo 4.8. Faaliyet oranları

Kısaltma	Oran	Formül
D1	Stok Devir Hızı	$\frac{\text{Satışların Maliyeti (Cari Yıl)}}{(\text{Önceki Yıl Stok.} + \text{Cari Yıl Stok.})/2}$
D2	Alacak Devir Hızı	$\frac{\text{Net Satışlar}}{\text{Kısa Vadeli Ticari Alacaklar} + \text{Uzun Vadeli Ticari Alacaklar}}$
D3	Çalışma Sermayesi Devir Hızı	$\frac{\text{Net Satışlar}}{\text{Dönen Varlıklar}}$
D4	Net Çalışma Sermayesi Devir Hızı	$\frac{\text{Net Satışlar}}{\text{Dönen Varlıklar} - \text{Kısa Vadeli Yabancı Kaynaklar}}$
D5	Maddi Duran Varlıklar Devir Hızı	$\frac{\text{Net Satışlar}}{\text{Maddi Duran Varlıklar (Net)}}$
D6	Duran Varlıklar Devir Hızı	$\frac{\text{Net Satışlar}}{\text{Duran Varlıklar}}$
D7	Öz Kaynak Devir Hızı	$\frac{\text{Net Satışlar}}{\text{Öz Kaynaklar}}$
D8	Aktif Devir Hızı	$\frac{\text{Net Satışlar}}{\text{Varlık (Aktif) Toplamı}}$

Kaynak: https://www3.tcmb.gov.tr/sector/dosyalar/menu/ratios_tr.pdf

Firma performansının ölçütü olarak literatürde sıklıkla kârlılık oranlarından yararlanıldığı daha önce açıklanmıştı. Bu çalışmada da firma performansı her biri birer kârlılık oranı olan öz kaynak kârlılık oranı ve aktif kârlılık oranı ile temsil edilmektedir. Böylece Tablo 4.9’da formülleri verilen bu iki oran araştırmanın bağımlı değişkenleri olarak belirlenmiştir. Önceki kısımlarda açıklanan doğrusal regresyon modelleri ile aktif kârlılığa ilişkin tahmin modelleri Bölüm 4.3’te ve öz kaynak kârlılığına ilişkin tahmin modelleri ise Bölüm 4.4’te incelenecektir.

Tablo 4.9. Kârlılık oranları

Kısaltma	Oran	Formül
ROE	Öz Kaynak Kârlılık Oranı	$\frac{\text{Net Kâr}}{\text{Öz Kaynaklar}}$
ROA	Aktif Kârlılık Oranı	$\frac{\text{Net Kâr}}{\text{Varlık (Aktif) Toplamı}}$

Kaynak: https://www3.tcmb.gov.tr/sector/dosyalar/menu/ratios_tr.pdf

Çalışmanın evrenini Borsa İstanbul'da işlem gören imalat sektöründeki firmalar oluşturmaktadır. Borsa İstanbul'da işlem gören imalat firmaları Kamu Aydınlatma Platformu'nda şu alt sektörlerle ayrılmıştır (<https://www.kap.org.tr/tr/Sektorler>):

- Ana Metal Sanayi
- Gıda, İçecek ve Tütün
- Kâğıt ve Kâğıt Ürünleri Basım
- Kimya İlaç Petrol Lastik ve Plastik Ürünler
- Metal Eşya Makine Elektrikli cihazlar ve Ulaşım Araçları
- Taş ve Toprağa Dayalı
- Tekstil, Giyim Eşyası ve Deri
- Orman Ürünleri ve Mobilya

Bu çalışmada Borsa İstanbul'da 2012-2019 yılları arasındaki 8 yıllık dönemde imalat sektöründe işlem gören firmalar örnekleme dâhil edilmiştir. İmalat sektörünün alt sektörlerindeki firmaların yıllara göre dağılımı Tablo 4.10'da yer almaktadır.

Tablo 4.10. İmalat firmalarının alt sektörlerle ve yıllara göre dağılımı

	2012	2013	2014	2015	2016	2017	2018	2019
Ana Metal	17	17	17	17	20	20	20	20
Gıda	25	25	25	25	26	27	27	28
Kâğıt	10	10	10	10	12	12	12	12
Kimya	30	30	30	30	30	33	33	33
Metal Eşya	30	30	30	31	33	33	33	33
Orman	3	4	4	4	5	5	5	5
Taş	19	19	19	19	19	19	19	19
Tekstil	21	21	21	21	22	22	22	22
TOPLAM	155	156	156	157	167	171	171	172

8 yıla ait toplam 1305 gözlemden 4'ü eksik veriye sahip oldukları için analizin dışında tutulmuşlardır. Öte yandan 29 gözlem ise negatif öz kaynaklara sahip oldukları için örnekleme dışında bırakılmışlardır. Formülü Tablo 4.9'da verilen öz kaynak kârlılık oranının pozitif olması söz konusu firmanın o dönem kâr ettiği, negatif olması ise firmanın o dönem zarar ettiği anlamına gelmektedir. Ancak bu yorum yalnızca öz kaynaklar grubu pozitif olduğunda geçerlidir. Şirketin geçmiş dönem ve mevcut dönem

zararının çok yüksek olması, aşırı borçlanma ve temettü ödemeleri gibi nedenlerle öz kaynaklar grubu negatif değere sahip olabilir. Bu durumda bir firmanın hem dönem net kârı hem de öz kaynakları negatif ise öz kaynak kârlılık oranı pozitif olarak elde edilir ve yalnızca orana bakıldığında firmanın söz konusu dönemde, yanlış bir yorumla, kâr ettiği sonucuna varılır. Veri kümesindeki öz kaynak grubu negatif olan işletmelerin net kâr kalemleri yanlış yorumlandığından ve aykırı değerlere sahip olduklarından dolayı öz kaynaklar kalemi negatif olan gözlemler analizin dışında bırakılmıştır. Sonuç olarak örneklem hacmi 1272 olarak belirlenmiştir.

4.2. Veri Ön Analizi

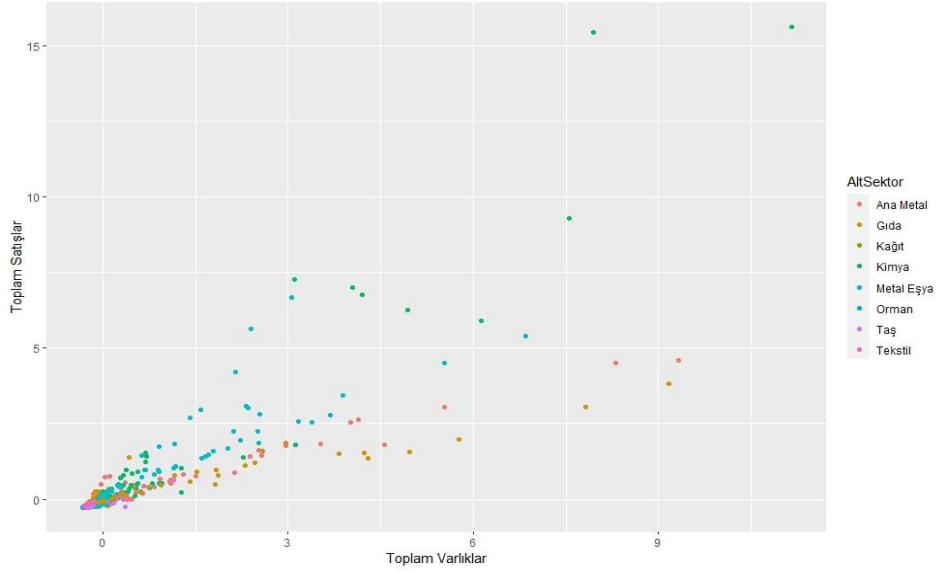
Küçük ve Orta Büyüklükteki İşletmelerin Tanımı, Nitelikleri ve Sınıflandırılması Hakkında Yönetmelik'e göre mikro, küçük, orta ve büyük işletmeler aşağıda gösterildiği biçimde sınıflandırılmaktadır (T.C. Resmi Gazete, 24 Haziran 2018, sayı: 30458):

a) Mikro işletme: On kişiden az yıllık çalışan istihdam eden ve yıllık net satış hasılatı veya mali bilançosundan herhangi biri 3 milyon Türk Lirasını aşmayan işletmeler.

b) Küçük işletme: Elli kişiden az yıllık çalışan istihdam eden ve yıllık net satış hasılatı veya mali bilançosundan herhangi biri 25 milyon Türk Lirasını aşmayan işletmeler.

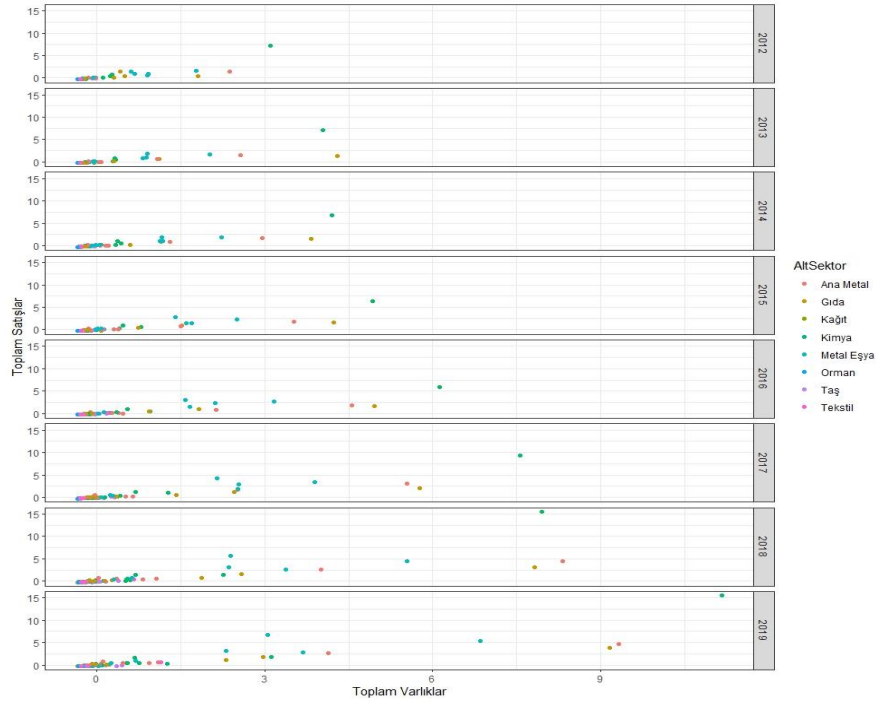
c) Orta büyüklükteki işletme: İki yüz elli kişiden az yıllık çalışan istihdam eden ve yıllık net satış hasılatı veya mali bilançosundan herhangi biri 125 milyon Türk Lirasını aşmayan işletmeler.”

İşletmelerdeki bilanço büyüklüğü ve yıllık net satış hasılatını göz önüne alarak yapılan sınıflandırmada 2019 yılında imalat sektöründe işlem gören firmaların 134'ü büyük, 34'ü orta büyüklükte, 4'ü küçük ölçekli işletmelerdir. Böylece örneklemdeki işletmelerin birçoğunun büyük işletmeler olduğu görülmektedir. Büyüklüğe ilişkin dağılımın daha kolay anlaşılabilmesi adına standartlaştırılmış toplam varlıklar ve toplam satışlar değişkenlerine göre işletmelerin dağılımı Şekil 4.2'deki saçılım grafiğinde verilmiştir. Grafikte görüldüğü üzere gözlemlerin birçoğu ortalamanın etrafında kümelenmiş olsa da bazı gözlemlerin toplam satışları ortalamadan 15 standart sapma ve toplam varlıkları ise 13 standart sapma kadar uzağında yer almaktadır. En büyük toplam satış ve toplam varlıklara sahip firmaların ağırlıklı olarak kimya, metal ve gıda sektörlerinde faaliyet gösterdikleri görülmektedir.



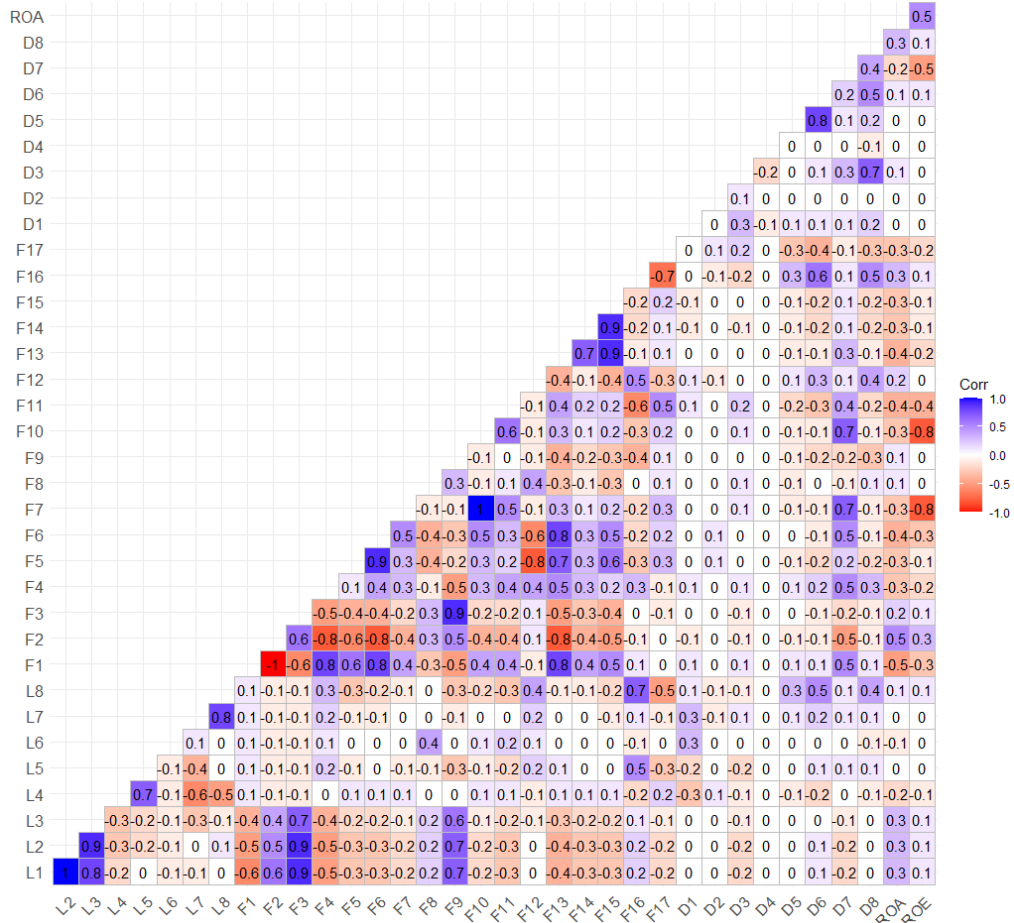
Şekil 4.2. İmalat firmalarının saçılım grafiği

Firmaların toplam varlıklar ve net satışlarını göz önüne alarak yıllara ve alt sektörlere göre dağılımı Şekil 4.3'te verilen nokta grafiğinde yer almaktadır. Grafikte en büyük toplam varlıklara sahip firmaların ana metal, gıda, kimya ve metal eşya alt sektörlerinde oldukları görülmektedir. Buna karşın kâğıt, orman, taş ve tekstil alt sektörlerine yer alan firmaların hem net satış hasılatı hem de bilanço büyüklükleri bakımından daha küçük ölçekli oldukları görülmektedir.



Şekil 4.3. Net satışlar ile toplam varlıkların yıllara ve sektörlere göre karşılaştırılması

Şekil 4.4'te bağımsız değişkenlerin arasındaki korelasyon grafiği verilmiştir. Bazı değişkenlerin arasında yüksek korelasyon olduğu görülmektedir. Modele eklenen 33 değişkenden F2 ve F5 değişkenleri sırasıyla F1 ve F6 değişkenleriyle negatif tam korelasyona sahip oldukları için modelden çıkarılmışlardır. Bunların dışında birçok değişken arasında yüksek korelasyonun mevcudiyeti kurulacak doğrusal regresyon modellerinde çoklu bağıntının gözlenebileceğine işaret etmektedir.



Şekil 4.4. Korelasyon grafiği

Maddi duran varlıklar/öz kaynaklar oranı (F7) ($r=-0,789$) ve duran varlıklar/öz kaynaklar oranı (F10) ($r=-0,753$), öz kaynak karlılığı oranı (ROE) ile en yüksek korelasyona sahip değişkenlerdir. Öte yandan kaldıraç oranı (F1) ($r=-0,453$) ve uzun vadeli yabancı kaynaklar/devamlı sermaye oranı (F6) ($r=-0,41$) aktif kârlılık oranı (ROA) ile en yüksek korelasyona sahiplerdir.

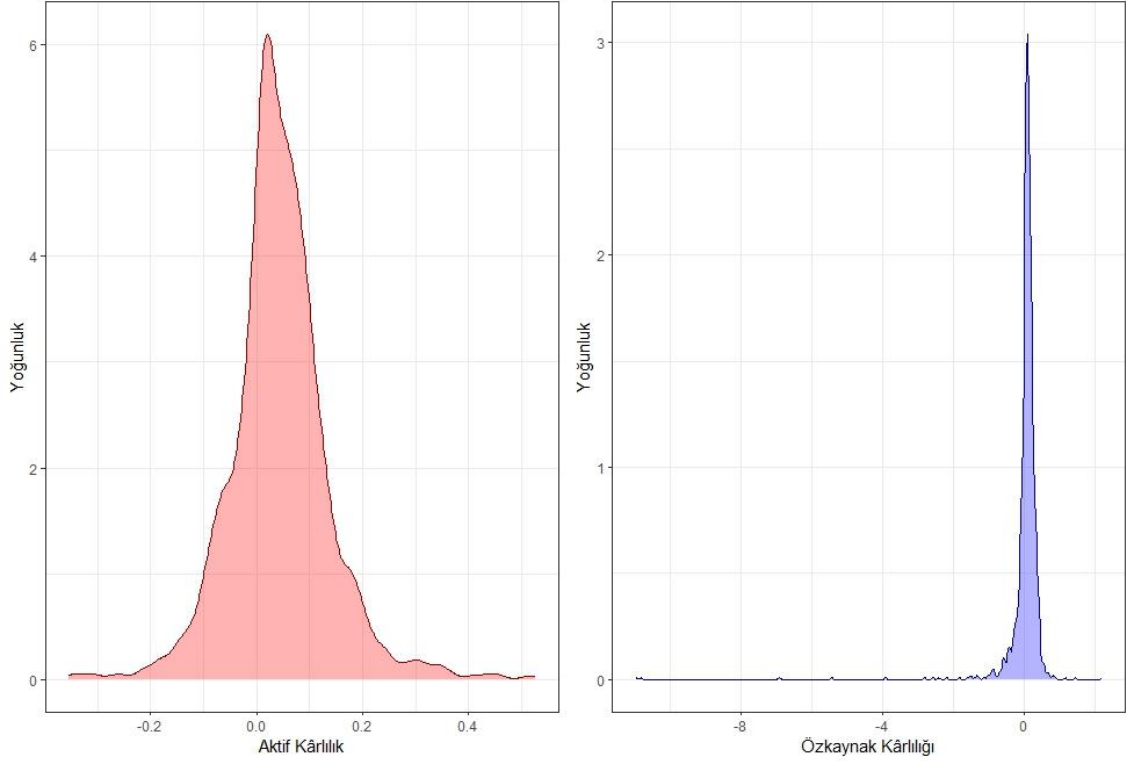
Analizde kullanılan değişkenlere ilişkin tanımlayıcı istatistikler Tablo 4.11'de verilmiştir. Likidite oranları arasında en fazla değişkenliğin stok/dönen varlık oranında (L4), finansal yapı oranları arasında ise en fazla değişkenliğin maddi duran varlıklar/uzun

vadeli yabancı kaynaklar oranında (F8) ve devir hızı oranları arasında ise en yüksek değişkenliğin net çalışma sermayesi (D4) devir hızında olduğu görülmektedir. Tablo incelendiğinde öz kaynak getirisinin ortalama %2 ve aktif karlılığın ise ortalama %4 civarında olduğu görülmektedir. Ayrıca öz kaynak kârlılığı -12,389 çarpıklık ve 204,372 basıklık değerleri ile sivri ve negatif asimetric bir dağılıma sahiptir. Aktif kârlılığın çarpıklık ve basıklık değerleri incelendiğinde normale yakın bir dağılıma sahip olduğu görülmektedir.

Tablo 4.11. *Finansal oranların tanımlayıcı istatistikleri*

Oran	Ortalama	Std. Sapma	Medyan	Minimum	Maksimum	Aralık	Çarpıklık	Basıklık
L1	2,020	2,067	1,482	0,032	43,864	43,832	8,896	145,31
L2	1,287	1,664	0,899	0,029	33,452	33,424	8,853	133,38
L3	0,410	1,009	0,141	0	14,951	14,951	7,282	73,328
L4	0,329	0,168	0,309	0,002	0,907	0,906	0,665	0,424
L5	0,173	0,109	0,156	0,000	0,678	0,677	1,372	2,895
L6	3,972	34,81	1,653	-18,075	987,731	1005,8	23,741	608,01
L7	0,437	0,184	0,434	0,016	0,982	0,966	0,242	-0,324
L8	0,240	0,142	0,221	0,005	0,878	0,874	0,999	1,581
F1	0,505	0,216	0,519	0,017	0,987	0,970	-0,037	-0,959
F3	1,725	2,824	0,928	0,013	56,399	56,386	8,963	132,31
F4	0,357	0,171	0,347	0,011	0,925	0,914	0,358	-0,487
F6	0,251	0,208	0,190	0,002	0,955	0,953	0,913	0,093
F7	1,028	2,070	0,646	0,011	41,172	41,160	12,756	214,11
F8	5,376	9,583	2,862	0,009	191,617	191,61	9,424	142,51
F9	1,277	1,607	0,895	0,011	29,525	29,514	7,980	100,47
F10	1,351	2,472	0,959	0,060	52,304	52,244	12,784	215,71
F11	0,761	0,472	0,743	0,028	8,383	8,356	6,955	100,10
F12	0,718	0,173	0,739	0,122	0,999	0,877	-0,578	-0,295
F13	0,247	0,188	0,241	0	0,846	0,846	0,373	-0,725
F14	0,367	0,247	0,374	0	0,958	0,958	0,095	-1,001
F15	0,433	0,260	0,478	0	0,979	0,979	-0,302	-1,037
F16	0,540	0,194	0,549	0,020	0,991	0,970	-0,074	-0,734
F17	0,340	0,179	0,324	0,002	0,877	0,875	0,370	-0,471
D1	6,938	17,92	4,212	0,002	450,482	450,48	17,639	385,53
D2	7,625	50,43	4,252	0,283	1761,042	1760,8	33,169	1146,4
D3	1,746	0,892	1,630	0,047	9,897	9,850	2,372	11,403
D4	2,915	209,8	3,035	-6210,4	2661,554	8871,9	-18,414	628,56
D5	5,615	18,16	2,585	0,071	456,243	456,17	17,202	372,15
D6	3,023	4,314	1,957	0,037	88,181	88,144	8,785	137,64
D7	2,564	3,108	1,694	0,091	50,247	50,156	7,001	82,042
D8	0,903	0,494	0,821	0,024	4,585	4,561	1,643	5,451
ROE	0,021	0,591	0,084	-10,962	2,187	13,148	-12,389	204,37
ROA	0,043	0,094	0,038	-0,354	0,525	0,879	0,415	3,302

Öz kaynak kârlılık oranının ve aktif kârlılık oranının yoğunluk grafikleri Şekil 4.5'te yer almaktadır. Görüldüğü üzere her iki dağılım da tek moda sahipleridir. Öz kaynak kârlılığı aktif kârlılığa göre çok daha sivri bir dağılıma ve negatif aykırı değerlere sahiptir.



Şekil 4.5. Aktif kârlılık ve öz kaynak kârlılığı yoğunluk grafikleri

İzleyen Bölüm 4.3'te aktif karlılık oranı (ROA) ve Bölüm 4.4'te öz kaynak karlılık oranı (ROE) bağımlı değişkenler olarak ele alınarak bu çalışmanın 2. ve 3. bölümlerinde ele alınan regresyon yöntemleri tahmin başarıları ve değişken seçim becerileri açılarından karşılaştırılacaklardır. Elde edilen modeller ve bu modellerin sonuçlarına ilişkin tüm hesaplamalar R programı (<https://www.r-project.org/>) ve bu program için geliştirilmiş çeşitli paketler yardımıyla gerçekleştirilmiştir. Çalışmada kullanılan paketlerden **glmnet** (J. H. Friedman ve diğerleri, 2010) lasso, ridge ve elastic net doğrusal regresyon modellerinin hesaplanmasında; **caret** (Kuhn, 2016) verinin ön işleme sürecinde, eğitim ve test kümelerine rassal olarak ayrılmasında ve çapraz geçerlilik hesaplamalarında; **ISLR** ve **leaps** (James ve diğerleri, 2013) ise tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemleriyle doğrusal regresyon modellerinin elde edilmesinde kullanılmışlardır.

Örneklemin eğitim ve test kümeleri olarak ayrılmasında farklı yaklaşımlar bulunmaktadır. Ancak genel uzlaşma verinin büyük bir kısmının eğitim verisi için ayrılması yönündedir. Verinin 2/3'ünün eğitim ve 1/3'ünün test verisi olarak ayrılması yaygın bir yaklaşımdır. Buna ek olarak %80/%20, %70/%30 ve %60/%40 biçiminde ayırmalara da yaygın olarak başvurulur (Raschka, 2018). Bu çalışmada toplam 1272 gözlemin %70'i eğitim kümesine ve %30'u test kümesine ayrılmıştır. Buna göre 892 gözlem eğitim verisi için 380 gözlem ise test verisi için ayrılmıştır.

EKK yönteminin aksine lasso ve elastic net yöntemleri değişkenlerin ölççeklerinden bağımsız olmadıkları için 0 ortalama ve 1 varyanslı olacak biçimde standartlaştırılmaları gerekmektedir. Bu konuyla ilgili açıklamalar ilgili bölümlerde yer almaktadır. Bu işlemin yalnızca lasso ve elastic net yöntemlerine uygulanması diğer modellerin katsayı tahminleri ile kıyaslanmalarını imkânsız kılacağından dolayı bu standartlaştırma işlemi her bir yöntem için analiz gerçekleştirilmeden önce eğitim kümesine uygulanmıştır. Ayrıca tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemleri sayesinde elde edilen değişkenler kümesinin belirlenmesi ve belirlenen değişkenlerin regresyon katsayılarının elde edilmesi için eğitim kümesine Bölüm 1.3.2.'de belirtilen 10 katlı çapraz geçerlilik yöntemi uygulanmıştır. Daha sonra modellerin performansları test kümesindeki tahminlerinin başarılarının karşılaştırılmasıyla değerlendirilmiştir.

4.3. Aktif Kârlılık Analizi

Bu bölümde Bölüm 2.3'te açıklanan sıradan değişken seçim yöntemleri olan tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemleri ile 3. bölümde açıklanan düzenleştirici yöntemlerin performanslarını karşılaştırmalı olarak analiz etmek amacıyla likidite, finansal yapı ve faaliyet etkinliğinin firma performansının üzerindeki etkileri incelenecektir. Bu amaçla firma performansının ölçütü olarak Tablo 4.9'da formülü (net kâr/toplam varlıklar) verilen aktif kârlılık oranı (ROA) bağımlı değişken olarak belirlenmiştir. Formülleri sırasıyla Tablo 4.6, Tablo 4.7 ve Tablo 4.8'de verilen likidite oranları, finansal yapı oranları ve devir hızları ise regresyon modellerindeki bağımsız değişkenlerdir.

4.3.1. En küçük kareler yöntemi

En küçük kareler yönteminin eğitim kümesine uygulanmasıyla elde edilen doğrusal regresyon modeline ilişkin sonuçlar Tablo 4.12'de yer almaktadır. Tablodaki sonuçlara göre L3, L4, L5, F8, F14, F16 ve D8 değişkenlerinin katsayılarının %5 anlamlılık

düzeyinde anlamlı oldukları görülmektedir. Buna ek olarak tablonun sağ sütununda çoklu bağıntıyı ölçen varyans büyütme çarpanları (VBÇ) yer almaktadır. Görüldüğü üzere 31 bağımsız değişkenin 20'sinin VBÇ değeri 10'dan büyüktür. Bu sonuca göre EKK modelinde çoklu bağıntının mevcut olduğu görülmektedir. F testi sonuçlarına göre ($F(31, 860) = 23,03, p < 0.001$) doğrusal regresyon modeli istatistiksel olarak anlamlıdır. Shapiro-Wilk normallik testi sonuçlarına göre ($W=0,960, p < 0,001$) hatalar normal dağılıma sahip değildir. Öte yandan Breusch-Pagan testi sonucuna göre ($BP(31) = 51,67, p < 0,01$) %1 anlamlılık düzeyinde sabit varyanslılığa sahip olduğu görülmektedir. Modelin belirlilik katsayısı ise %45,36'dır. Buna göre aktif kârlılık oranındaki değişimin %45,36'sı modelde yer alan finansal oranlar tarafından açıklanmaktadır.

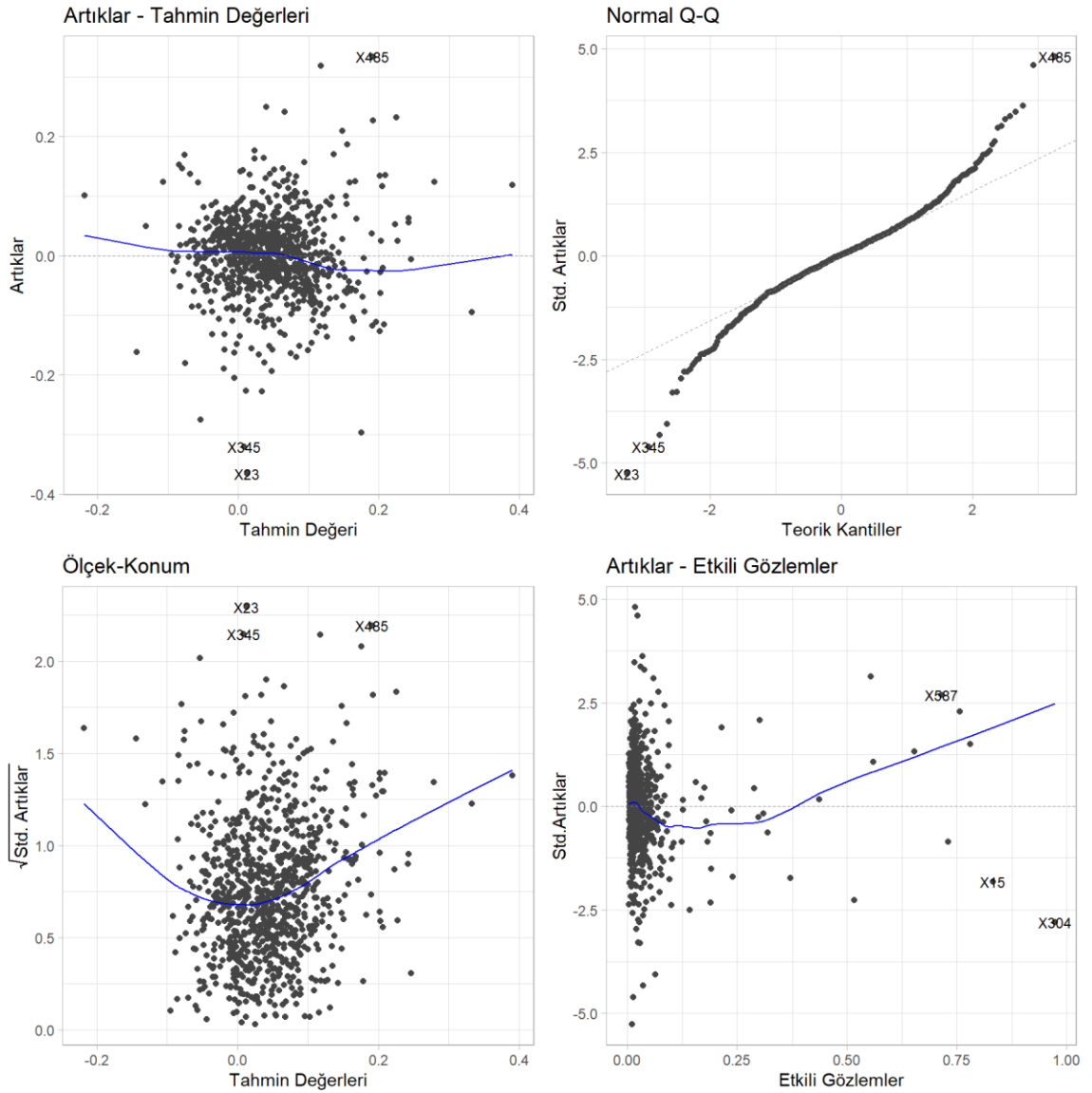
Tablo 4.12. *En küçük kareler aktif kârlılık doğrusal regresyon modeli*

	Katsayı Tahmini	VBÇ
Sabit	0,043***	66,2
L1	-0,035	53,4
L2	0,016	9,5
L3	0,023**	16,0
L4	0,035***	19,9
L5	-0,046***	2,0
L6	0,005	17,6
L7	0,016	29,0
L8	-0,022	110,2
F1	-0,039	157,5
F3	0,011	75,2
F4	-0,021	30,3
F6	0,005	90,1
F7	-0,019	2,3
F8	-0,008*	59,9
F9	-0,017	75,9
F10	0,020	5,0
F11	0,004	27,0
F12	0,009	20,9
F13	0,001	9,7
F14	-0,016*	18,4
F15	0,017	39,7
F16	0,057***	6,5
F17	-0,006	1,5
D1	-0,004	1,1
D2	-0,002	8,1
D3	0,006	1,2
D4	0,002	5,4
D5	-0,002	7,8

D6	-0,002	10,1
D7	-0,008	11,0
D8	0,021**	66,2

Anlamlılık düzeyleri: $p < 0.001$: '***', $p < 0.01$: '**', $p < 0.05$: '*'

EKK modelinin artıklarına ve etkili gözlemlerine ilişkin tanı grafikleri Şekil 4.6'da yer almaktadır. Şeklin sol üst köşesinde yer alan Artıklar – Uyum Değerleri grafiğinde doğruya yakın bir trend görülmekte ve artıklar sıfır doğrusunun etrafında rassal bir dağılım göstermektedir. Böylece sabit varyanslılığın mevcut olduğu söylenebilir. Sağ üstte verilen Normal QQ grafiğinde en düşük ve en yüksek kantillerde normallikten sapmaların olduğu görülmektedir. Ölçek-Konum ve Artıklar-Etkili Gözlemler grafikleri ise örnekleme aykırı ve etkili gözlemlerin var olduğunu işaret etmektedirler.



Şekil 4.6. EKK aktif kârlılık modelinin tanı grafikleri

10 katlı çapraz geçerlilik sonuçlarına göre tüm olası alt kümeler (TOAK), ileriye doğru seçim (İLERİ) ve geriye doğru seçim (GERİ) yöntemleri ile elde edilen doğrusal modellerin bulguları Tablo 4.13'te yer almaktadır. Tüm olası alt kümeler yöntemi ile en düşük çapraz geçerlilik hata kareleri ortalamasına sahip model 4 bağımsız değişkenli modeldir. Elde edilen sonuçlara göre L3, F1, F9 ve D8 değişkenleri modele dahil edilmiştir. Öte yandan ileriye doğru seçim modelinde 5 bağımsız değişken modele eklenmiştir ve dahil edilen değişkenler şunlardır: L5, L8, F1, F16 ve D8. Geriye doğru seçim yöntemine göre ise L3, L5, F4, F9, F12 ve D8 değişkenleri modele eklenmiş ve böylece 6 bağımsız değişkenli doğrusal regresyon modeli elde edilmiştir. Her üç modele seçilen tüm bağımsız değişkenler %5 anlamlılık düzeyinde anlamlıdır. Tüm modeller tarafından modele alınan tek değişken D8'dir. Öte yandan 22 değişken üç yöntem tarafından da aktif bağımsız değişkenler kümesine dahil edilmemişlerdir.

F-testi sonuçlarına göre (Tüm olası alt kümeler: $(F(31, 887) = 151.7, p < 0.001)$; ileriye doğru seçim: $(F(31, 886) = 119.4, p < 0.001)$; geriye doğru seçim: $(F(31, 885) = 103.6, p < 0.001)$) her üç model de istatistiksel olarak anlamlıdır.

Tablo 4.13. Tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim doğrusal regresyon modellerinin katsayı tahminleri

Değişkenler	TOAK	İLERİ	GERİ
Sabit	0,043***	0,043***	0,043***
L3	0,028***		0,028***
L5		-0,025***	-0,006***
L8		-0,022***	
F1	-0,053***	-0,044***	
F4			-0,056***
F9	-0,030***		-0,030***
F12			0,028***
F16		0,043***	
D8	0,021***	0,020***	0,023***

Anlamlılık düzeyleri: $p < 0.001$: '***', $p < 0.01$: '**', $p < 0.05$: '*'

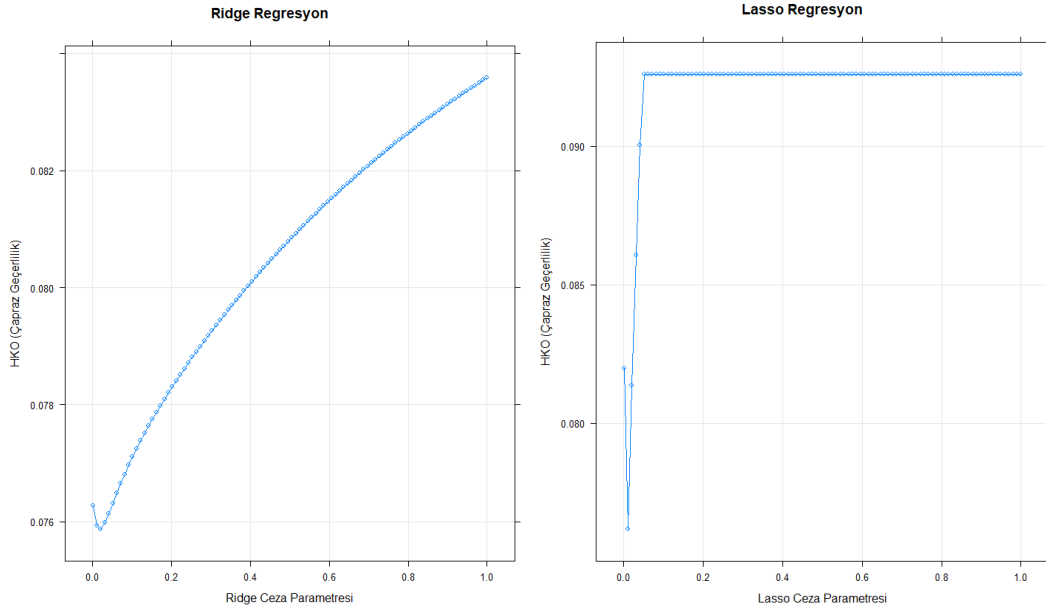
Tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemlerine ait değerlendirme ölçütleri Tablo 4.14'te verilmiştir. En yüksek belirlilik katsayısına sahip model bekleneceği üzere en çok değişkene sahip olan geriye doğru seçim modelidir. En düşük eğitim HKO'ya sahip model de benzer şekilde geriye doğru seçim modelidir. En düşük test hatasına sahip olan böylece en iyi genelleştirme özelliğine sahip model ise ileriye doğru seçim modelidir. Ancak tablodan görüleceği üzere üç yöntemin genelleştirme performansları arasında çok küçük farklar bulunmaktadır.

Tablo 4.14. Aktif kârlılık EKK modellerinin değerlendirilmesi

	TOAK	İLERİ	GERİ
Belirlilik Katsayısı	0,4062	0,4026	0,4125
Düzeltilmiş Belirlilik Katsayısı	0,4035	0,3992	0,4086
ÇG Eğitim HKO	0,0719	0,0724	0,0717
ÇG Test HKO	0,0807	0,0797	0,0802

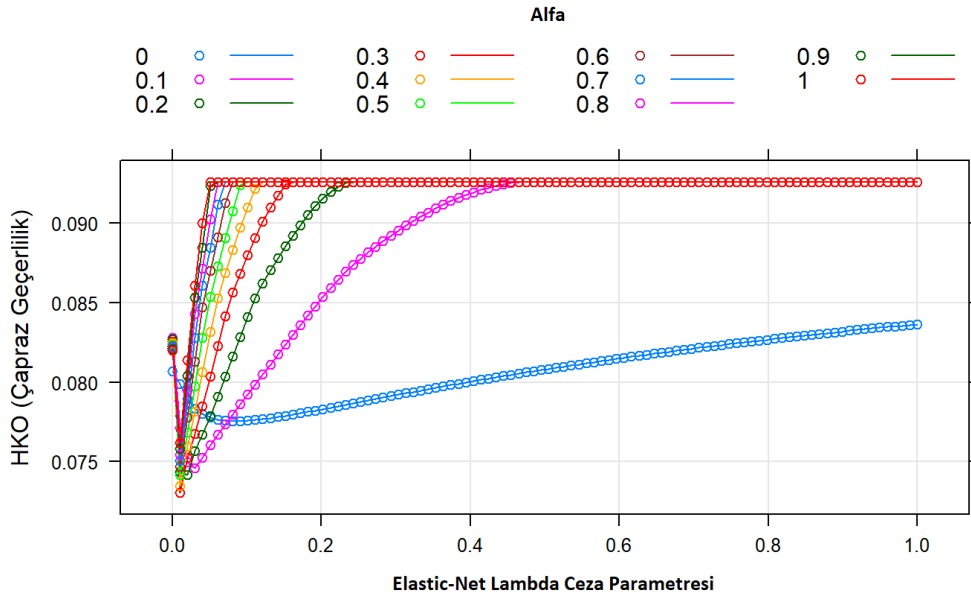
4.3.2. Düzenleştirici yöntemler analizi

Aktif kârlılık analizinde kullanmak üzere düzenleştirici yöntemlerden önceki bölümlerde açıklanan lasso regresyon, ridge regresyon ve elastic-net regresyon yöntemleri kullanılacaktır. Lasso regresyon ve ridge regresyon yöntemi ile doğrusal regresyon modelini elde edebilmek için öncelikle ceza parametresinin değerinin her iki yöntem için ayrı ayrı belirlenmesi gerekmektedir. Elastic-net yöntemi için ise s ceza parametresinin yanında α değerinin de belirlenmesi gerekmektedir. Söz konusu hiperparametreler 10 katlı çapraz geçerlilik yöntemi ile katsayı tahminleri gerçekleştirilmeden önce belirlenmiştir. Şekil 4.7’de çapraz geçerlilik yöntemi ile lasso ve ridge regresyon yöntemlerinin optimum s parametre değerlerinin belirlenmesine ilişkin grafikler yer almaktadır. Çapraz geçerlilik sonuçlarına göre $s_{lasso} = 0,0102$ ve $s_{ridge} = 0,0203$ olarak elde edilmiştir.



Şekil 4.7. Lasso ve ridge regresyon yöntemlerinin ÇG ile optimum ceza parametre değerleri

Elastic-net yönteminin hiperparametresi olan α 'nın belirlenebilmesi için $[0,1]$ arasında grid arama gerçekleştirilmiş ve bu aralıkta önceden belirlenen 10 farklı α değeri arasından en düşük çapraz geçerlilik hata kareleri ortalamasını veren s_{enet} ceza değeri bulunmuştur. Farklı α değerleri ve s_{enet} değerlerine göre çapraz geçerlilik grafiği Şekil 4.8'de yer almaktadır. Buna göre eğitim kümesinde gerçekleştirilen öğrenme işlemi sonucunda en düşük çapraz geçerlilik HKO değerini veren hiperparametre değerleri $\alpha = 0,3$ ve $s_{enet} = 0,0102$ olarak elde edilmiştir.



Şekil 4.8. Çapraz geçerlilik yöntemi ile elastic-net hiperparametrelerin değişim grafiği

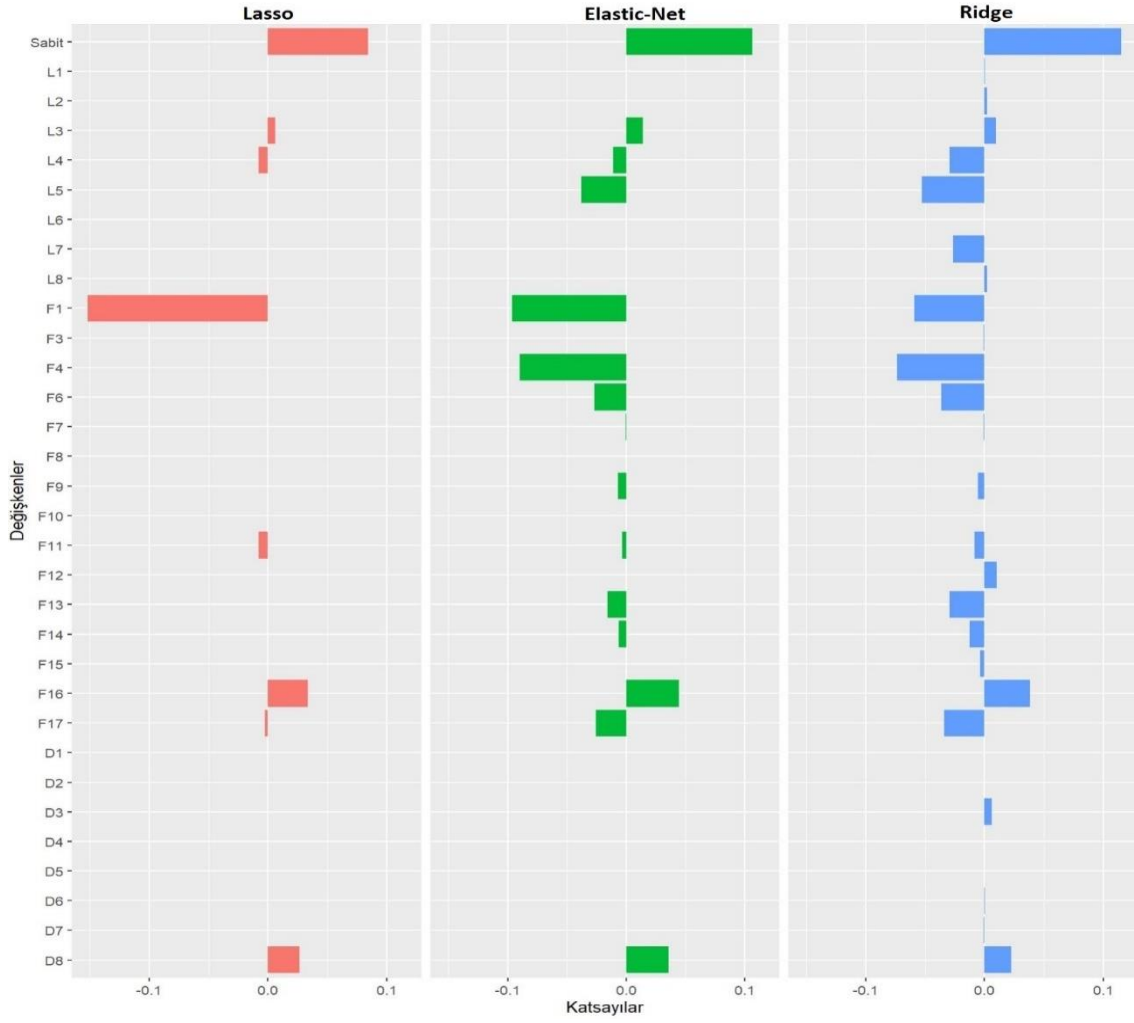
Tablo 4.15'te lasso, ridge ve elastic net doğrusal regresyon modellerine ilişkin katsayı tahminleri yer almaktadır. Görüldüğü üzere lasso ve elastic net yöntemleri seyrek (sparse) çözümlere sahip iken ridge regresyon seyrek bir çözüm üretememekte ve böylece değişken seçimini gerçekleştirememektedir. Lasso regresyon sonuçlarına göre L3, L4, F1, F11, F16, F17 ve D8 değişkenleri doğrusal regresyon modelinde yer almışken; elastic net yöntemi L3, L4, L5, L7, F1, F4, F6, F7, F8, F9, F11, F13, F14, F16 ve D8 değişkenlerinden oluşan bir doğrusal regresyon modelini tahminlemiştir. Böylece lasso regresyon 7 bağımsız değişkenli bir modelle 15 bağımsız değişkenli elastic net modeline göre daha seyrek bir model tahminlemiştir.

Lasso ve ridge regresyon ceza parametrelerinin her ikisini ağırlandırarak kullanan iki yöntem arasındaki ödünleşmeyi sağlayan elastic-net yönteminin α parametre değeri ridge regresyon modeline yaklaşırken değişken sayısının lasso yöntemine kıyasla daha fazla olması beklenen bir durumdur. Öte yandan α değeri sıfıra yaklaşırken elastic net yönteminin lasso modelindeki değişkenleri modelden çıkarmaması tutarlı bir davranış sergilediğini göstermektedir.

Tablo 4.15'te verilen katsayı tahminlerinin büyüklüklerinin karşılaştırma grafiği Şekil 4.9'da verilmiştir. Şekilde görüldüğü üzere ridge regresyon yöntemi değişken seçimi gerçekleştirememektedir. Öte yandan her üç yöntemde de sıfırdan farklı katsayılara sahip değişkenlerin katsayılarının işaretleri aynıdır. Ancak aynı değişken farklı yöntemler tarafından farklı büyüklüklerde sıfıra yaklaştırılmıştır. Örneğin, elastic-net yönteminde F1 değişkeni lasso regresyona kıyasla sıfıra daha yakın bir değer almışken; F16 değişkeni lasso regresyon yönteminde elastic-net'e göre sıfıra daha yakındır. Dolayısıyla bir yöntemin sıfırdan farklı katsayılara sahip değişkenlerin katsayılarında gerçekleştirdiği küçültme işlemi tüm değişkenler boyunca eşit ölçüde gerçekleştirilmemektedir. Bu durum Şekil 4.9, Şekil 4.11 ve Şekil 4.13'te ceza parametresine bağlı olarak katsayıların büyüklüklerinde gözlemlenen değişim grafiklerinde gözlemlenebilir.

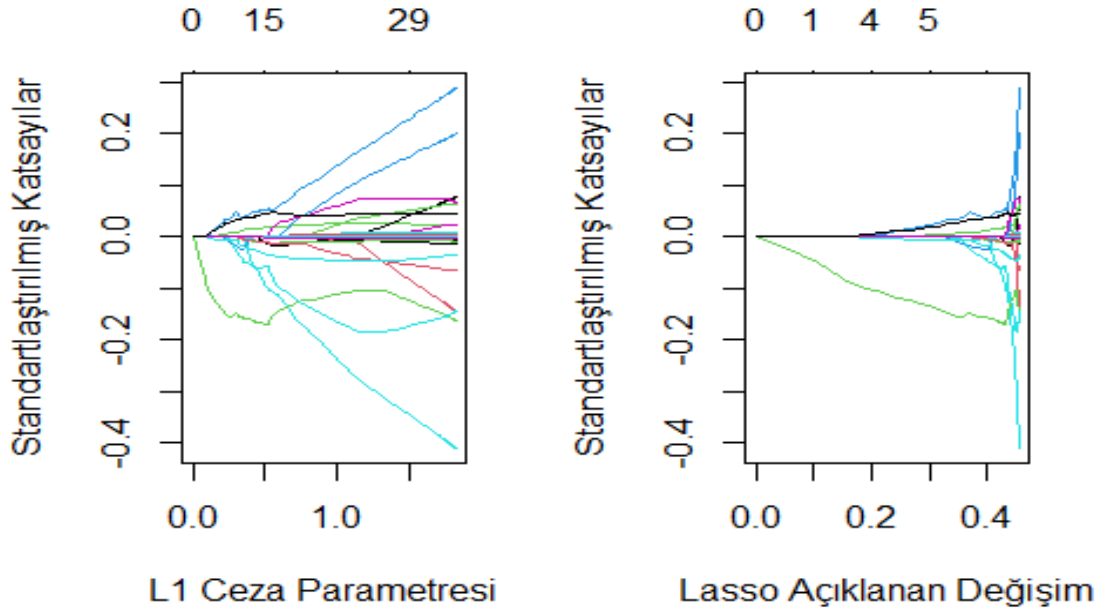
Tablo 4.15. *Lasso, elastic-net ve ridge regresyon modellerinin katsayı tahminleri*

	Lasso	Elastic Net	Ridge
Sabit	0,0845	0,1063	0,1108
L1	0	0	0,0007
L2	0	0	0,0021
L3	0,0061	0,0138	0,0085
L4	-0,0077	-0,0108	-0,0291
L5	0	-0,0377	-0,0419
L6	0	0	-0,0001
L7	0	-0,0001	-0,0230
L8	0	0	0,0023
F1	-0,1519	-0,0960	-0,0518
F3	0	0	-0,0005
F4	0	-0,0900	-0,0629
F6	0	-0,0266	-0,0345
F7	0	-0,0006	-0,0009
F8	0	-0,0001	-0,0002
F9	0	-0,0067	-0,0043
F10	0	0	-0,0005
F11	-0,0075	-0,0036	-0,0093
F12	0	0	0,0085
F13	0	-0,0155	-0,0297
F14	0	-0,0062	-0,0119
F15	0	0	-0,0064
F16	0,0339	0,0447	0,0335
F17	-0,0027	-0,0254	-0,0307
D1	0	0	-0,0001
D2	0	0	1x10-5
D3	0	0	0,0057
D4	0	0	4x10-6
D5	0	0	-0,0001
D6	0	0	0,0005
D7	0	0	-0,0007
D8	0,0269	0,0359	0,0202



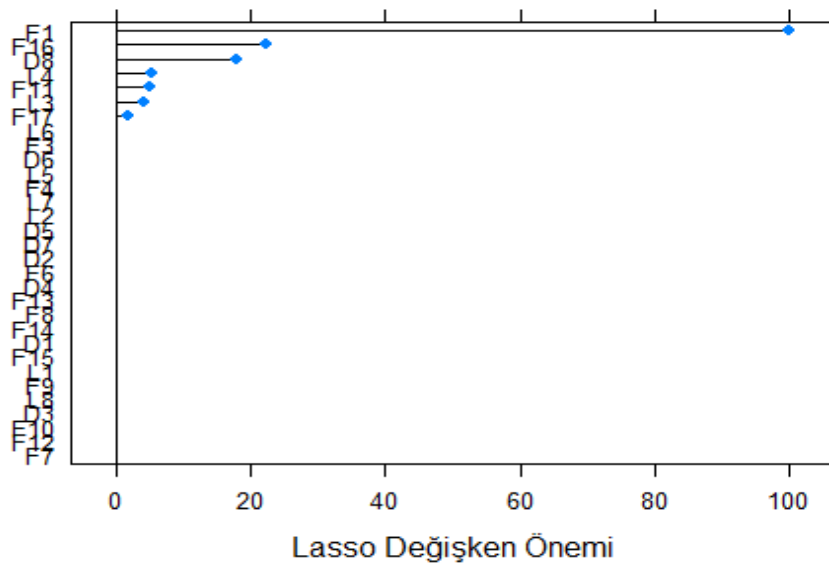
Şekil 4.9. Düzenleştirici yöntemlerin katsayı tahminlerinin çubuk grafikleri

Şekil 4.10.'da solda verilen grafikte ceza parametresi sıfıra yaklaştıkça her bir lasso değişken katsayılarında gerçekleşen değişim görülmektedir. Grafikte her bir renk farklı bir bağımsız değişkeni temsil etmektedir. Grafiğin en sağında EKK regresyon katsayılarının tahminleri yer alırken, ceza parametre değeri sıfıra yaklaştıkça bazı katsayı değerleri tam olarak sıfıra eşit olurken bazılarının değeri ise sıfıra yaklaşır. Grafiğin x ekseninde l_1 norm ceza parametreleri ve üstünde ise karşılık gelen ceza parametresi için modelde yer alan değişken sayısı görülmektedir. Şekil 4.10'un sağ tarafında ise değişkenler tarafından açıklanan değişim görülmektedir. Buna göre değişkenlerin katsayıları sıfıra yaklaştıkça aktif kârlılıktaki değişimin açıklanma yüzdesindeki değişim görülmektedir. Lasso regresyon modeli aktif kârlılıktaki değişimin %36'sını açıklamaktadır.



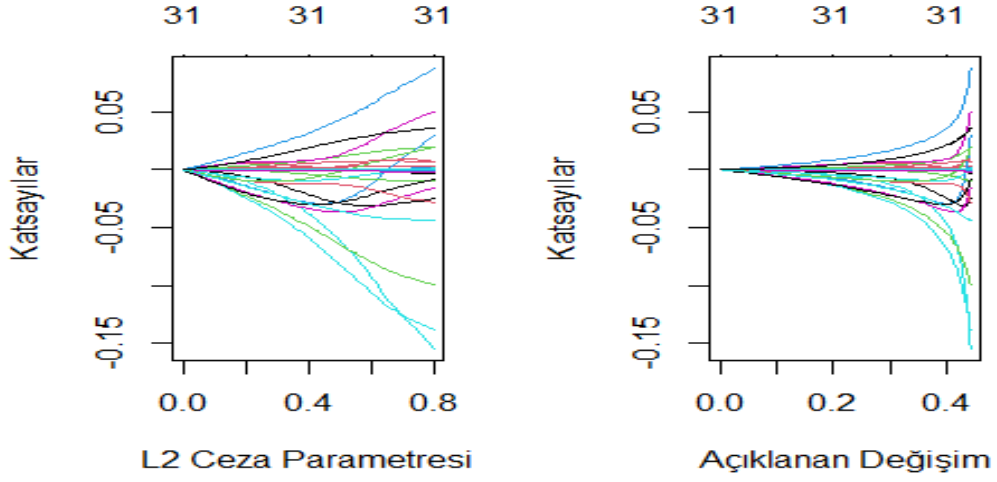
Şekil 4.10. Lasso katsayıların değişimi– Lasso açıklanan değişim

Şekil 4.11’de lasso modelinde yer alan değişkenlerin önemleri görülmektedir. Düzenleştirici yöntemler ile elde edilen katsayı tahminlerinden en yüksek mutlak değere sahip olan katsayı 100 kabul edilerek diğerleri buna oranlanır. Böylece değişkenlerin birbirlerine göre önemlerini veren değişken önemleri elde edilir. Grafikte görüldüğü üzere önem sırasına göre değişkenler F1, F16, D8, L4, F11, L3 ve F17’dir.



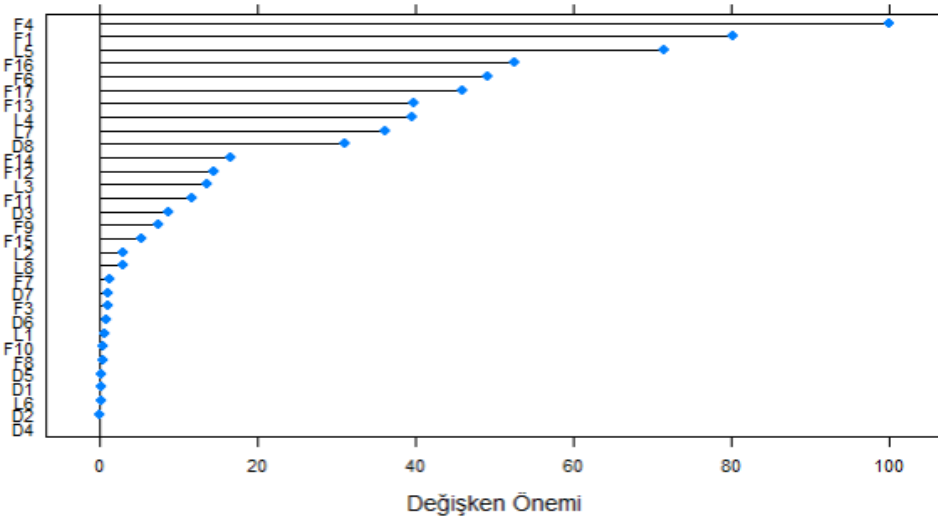
Şekil 4.11. Lasso regresyon değişkenlerin önem sıralaması

Şekil 4.12’de ridge regresyon yönteminde ceza parametresinin değişimine bağlı olarak katsayıların değişim yönü ve açıklanan değişim görülmektedir. Değişken seçimini gerçekleştirilmeyen ridge regresyonda, grafikte görülebileceği üzere tüm katsayılar sıfıra giderek yaklaşırsa da hiçbir zaman sıfıra eşit olmamaktadır. Düzenlenmiş katsayılarla sahip ridge regresyon yöntemi aktif kârlılıktaki değişimin %36,23’ünü açıklamaktadır.



Şekil 4.12. Ridge katsayıların değişimi– Ridge açıklanan değişim

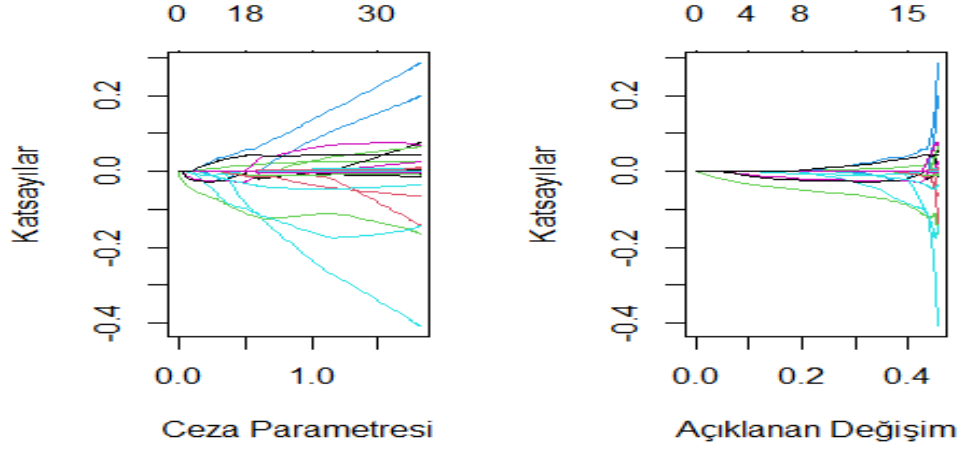
Ridge regresyon değişken önemi grafiğinde (Şekil 4.13) en önemli değişkenlerin finansal yapı oranları değişkenleri olduğu görülmektedir. Ridge regresyon modelinde en önemli beş değişken F4, F1, L5, F16 ve F6’dır.



Şekil 4.13. Ridge regresyon değişkenlerin önem sıralaması

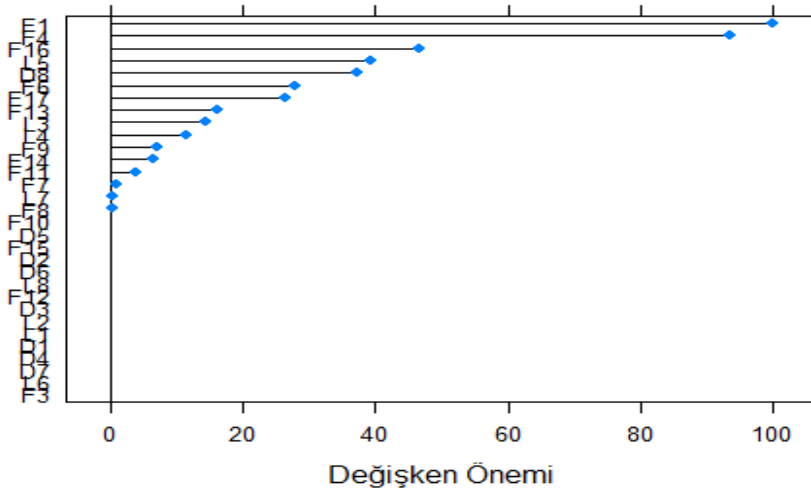
Şekil 4.14’te elastic-net yönteminde elde edilen katsayıların değişim grafiği ve açıklanan değişim grafiği gösterilmiştir. Ceza parametresinin değişmesine bağlı olarak

katsayıların takip ettiği yol da α hiperparametre değerine bağlı olarak elastic-net modeline 9 değişken daha eklenmiştir. Elastic-Net yönteminde açıklanan değişim %39,3'tür.



Şekil 4.14. Elastic-net katsayıların değişimi– Elastic-net açıklanan değişim

Elastic-net değişkenlerinin önem sıralamalarının gösterildiği Şekil 4.15'e göre elastic-net regresyon modelinde en önemli değişkenler sırasıyla F1, F4, F16, L5, D8, F6 ve F17'dir. Bu 7 değişken lasso regresyon modelinde de yer almaktadır. Elastic-net modelinde bu değişkenlerin önem sıralamaları değişmiş olsa da bu değişkenlerin elastic-net yöntemi tarafından en önemli değişkenler olarak belirlenip modele dahil edilmiş olmaları, elastic-net yönteminin α parametresine bağlı olarak katsayılarda gerçekleştirdiği değişimin lasso regresyon yöntemi ile tutarlı olduğunu göstermektedir.



Şekil 4.15. Elastic-net regresyon değişkenlerin önem sıralaması

Hastie vd. (2008, s. 91), lasso yöntemi ile elde edilen sıfırdan farklı regresyon katsayılarının tahminleri sıfıra doğru yanlı tahminler olduğundan dolayı; lasso regresyon yöntemince belirlenen sıfırdan farklı katsayılara sahip bağımsız değişkenlerin EKK yöntemi ile tekrardan uyumlandırmanın bu yanlılığı azaltacağını belirtmişlerdir. Böylece lasso yöntemi ile aktif değişkenler kümesi belirlenip kısıtlandırılmamış EKK katsayılarının iki aşamalı olarak elde edilmesi önerilmiştir. Bu amaçla lasso ve elastic-net yöntemlerince belirlenen bağımsız değişkenler EKK yöntemi ile uyumlandırılmış ve sonuçlar Tablo 4.16’da verilmiştir. F testi sonuçlarına göre hem Lasso-EKK modeli ($F(7, 884) = 79,16; p < 0.001$) hem de ElasticNet-EKK modeli ($F(16, 875) = 42,39; p < 0.001$) istatistiksel olarak anlamlıdır.

Tablo 4.16. *Lasso-EKK ve ElasticNet-EKK katsayı tahminleri*

	Lasso-EKK	ENet-EKK
Sabit	0,0429***	0,0429***
L3	0,0109***	0,0195***
L4	-0,0077**	0,0179*
L5	0	-0,0305***
L7	0	-0,0022
F1	-0,0404***	-0,0508*
F4	0	-0,0072
F6	0	0,0008
F7	0	-0,0037
F8	0	-0,0048
F9	0	-0,0226***
F11	-0,0061	0,0015
F13	0	0,0103
F14	0	-0,0083
F16	0,0085	0,0189**
F17	-0,0038	-0,0087*
D8	0,0207***	0,0210***

Anlamlılık düzeyleri: $p < 0.001$: ‘***’, $p < 0.01$: ‘**’, $p < 0.05$: ‘*’

Tablo 4.17’de lasso, ridge, elastic-net, lasso-EKK ve enet-EKK doğrusal regresyon modellerinin değerlendirme ölçütleri yer almaktadır. Lasso, ridge ve elastic-net modelleri incelendiğinde en düşük eğitim HKO ve en yüksek pseudo R^2 değerine sahip olan model elastic-net modelidir. Öte yandan en düşük test HKO değerine sahip model ise lasso modelidir. Böylece üç düzenleyici yöntem arasından en başarılı tahminleri

gerçekleştiren modelin lasso regresyon modeli olduğu sonucuna varılır. Öte yandan en küçük eğitim HKO ve test HKO değerlerine sahip modeller görüleceği üzere EKK yöntemi ile yeniden uyumlandırılmış lasso-EKK ve enet-EKK modelleridir. Sıfırdan farklı katsayılar üzerindeki kısıtlar kaldırılarak yeniden uyumlandırılan bu modeller ile hem lasso hem de elastic-net yöntemine göre eğitim ve test verisinde daha düşük HKO değerleri elde edilmiştir.

Tablo 4.17. *Aktif kârlılık düzenleyici modellerin değerlendirilmesi*

	Lasso	Ridge	Elastic-Net	Lasso-EKK	ElasticNet-EKK
Pseudo R²	0,3627	0,3515	0,3972	0,3853	0,4366
ÇG Eğitim HKO	0,0762	0,0764	0,0730	0,0739	0,0724
ÇG Test HKO	0,1021	0,1067	0,1066	0,0799	0,0800

4.3.3. Modellerin karşılaştırılması

Çalışmanın bu kısmında aktif kârlılık analizi için hesaplanan tüm doğrusal regresyon modellerin karşılaştırılmalı analizi yer almaktadır. Eğitim verisinden yararlanarak çapraz geçerlilik yöntemi yardımıyla hem sıradan değişken seçim yöntemleri ile hem de düzenleyici yöntemler ile elde edilen regresyon modellerinin genelleştirme performanslarını değerlendirmek amacıyla elde edilen modellerin her birinin test hatalarının hesaplanması gerekmektedir. Bu amaçla her bir model test verisindeki gözlemlere uyumlandırılarak HKO değerleri hesaplanmış ve ilgili sonuçlar Tablo 4.14 ve Tablo 4.17’de sunulmuştu. Düzenleyici yöntemler ile değişken seçim yöntemleri ile elde edilen regresyon modellerinin karşılaştırılması amacıyla ilgili sonuçlar bir arada Tablo 4.18’de verilmiştir.

Tablo 4.18’de en düşük eğitim hatasına sahip modeller tüm olası alt kümeler ve geriye doğru seçim yöntemleri ile elde edilmiştir. Öte yandan en yüksek eğitim hatasına sahip modeller ridge ve lasso regresyon modelleridir. En düşük test hatasına sahip model ise ileriye doğru seçim modelidir. Böylece test kümesindeki elde ettiği başarılı tahminlerle en iyi genelleştirme özelliğine sahip modelin ileriye doğru seçim modeli olduğu sonucuna varılır. Öte yandan Lasso-EKK ve ENet-EKK modellerinin ileriye doğru seçim modelinden sonraki en düşük test hatalarına sahiplerdir.

Lasso, ridge ve elastic-net modelleri ise hem en yüksek eğitim hatalarına hem de en yüksek test hatalarına sahiplerdir. Ancak Lasso-EKK ve ENet-EKK modellerinin eğitim

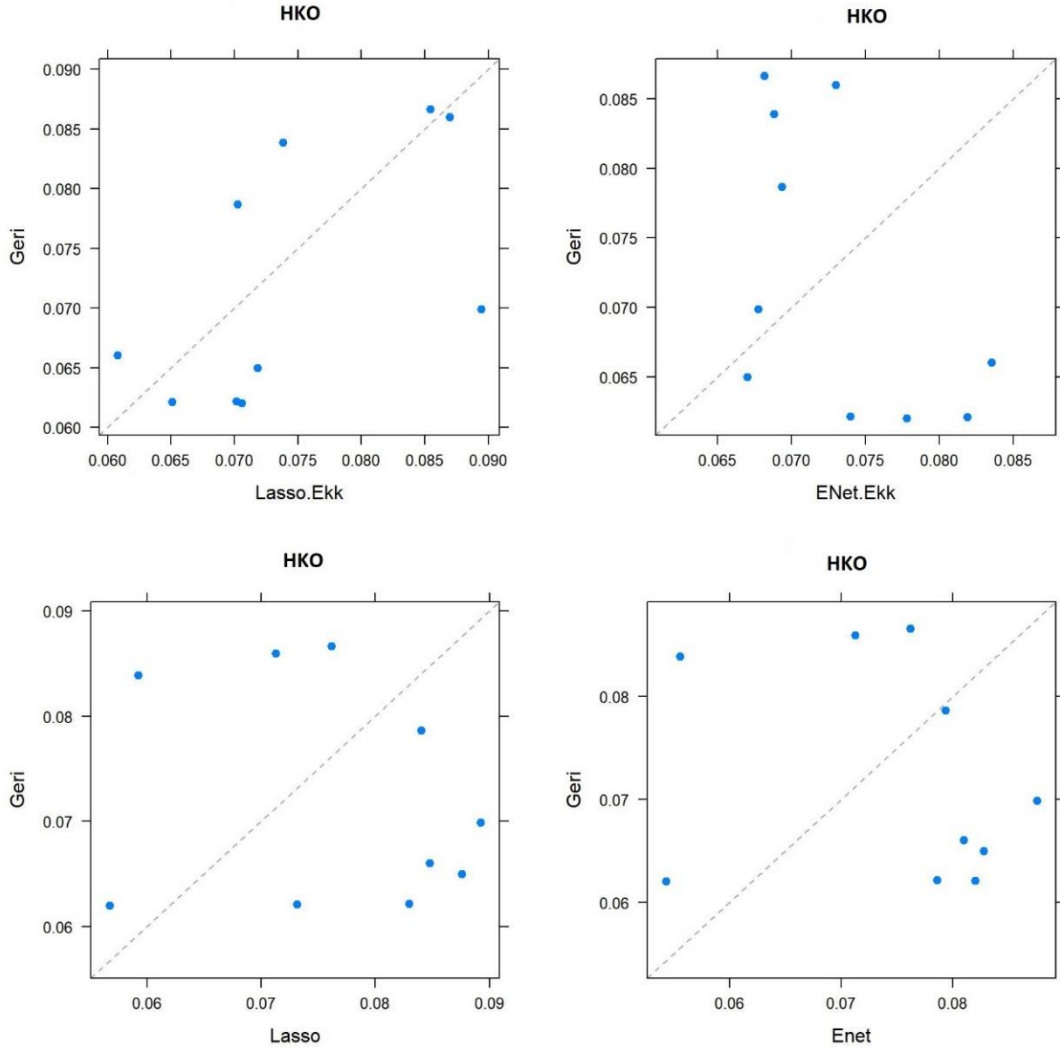
ve test hataları incelendiğinde lasso ve elastic-net yöntemleri ile belirlenen sıfırdan farklı katsayılara sahip bağımsız değişkenler EKK yöntemi ile tekrar aktif kârlılık değişkenine uyumlandırıldıklarında çok daha başarılı tahminler gerçekleştirdikleri görülmektedir. Lasso regresyonun test HKO değeri yaklaşık %27,8 ve elastic-net test HKO değeri ise %33,25 oranında azalmışlardır. Bu sonuca göre Lasso-EKK yöntemi en düşük ikinci ÇG test HKO değerine sahiptir.

Tablo 4.18. Doğrusal regresyon modellerin eğitim ve test hataları

	ÇG Eğitim HKO	ÇG Test HKO	Değişkenler
TOAK	0,0719	0,0807	L3, F1, F9, D8
ILERI	0,0724	0,0797	L5, L8, F1, F16, D8
GERI	0,0717	0,0802	L3, L5, F4, F9, F12, D8
Lasso	0,0762	0,1021	L3, L4, F1, F11, F16, F17, D8
Ridge	0,0764	0,1067	Tüm Değişkenler
Elastic-Net	0,0730	0,1066	L3, L4, L5, L7, F1, F4, F6, F7, F8, F9, F11, F13, F14, F16, F17, D8
Lasso-EKK	0,0739	0,0799	L3, L4, F1, F11, F16, F17, D8
ENet-EKK	0,0724	0,0800	L3, L4, L5, L7, F1, F4, F6, F7, F8, F9, F11, F13, F14, F16, F17, D8

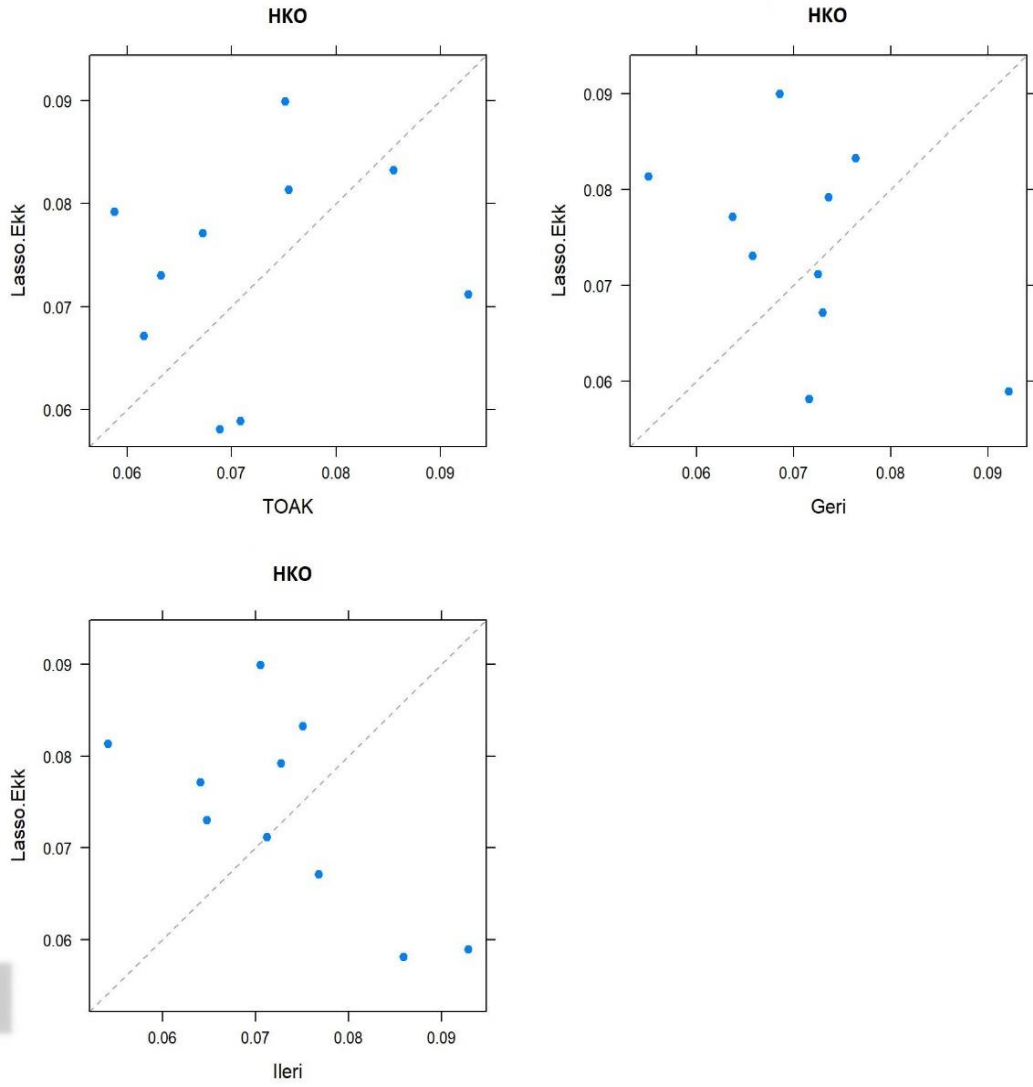
Tablo 4.18’de her bir regresyon modelinin 10 katlı çapraz geçerlilik yöntemi ile elde edilen çapraz geçerlilik HKO sonuçlarına yer verilmiştir. Şekil 4.16’da eğitim kümesinde gerçekleştirilen çapraz geçerlilik işlemi sırasında kullanılan 10 farklı çapraz geçerlilik katındaki HKO değerleri saçılım grafiklerinde sunulmuştur. Bu grafikte en düşük eğitim HKO değerine sahip geriye doğru seçim yöntemi ile düzenlileştirici yöntemler karşılaştırılmıştır. Ridge regresyon değişken seçimini gerçekleştirmediği için grafiğe dahil edilmemiştir. Şekil 4.16’da görüleceği üzere geriye doğru seçim modeli lasso ve elastic-net modellerine karşı net bir üstünlüğe sahip değildir. Böylece Tablo 4.18’de verilen eğitim seti ÇG HKO sonuçlarına göre geriye doğru seçim yönteminin bu üç yönteme göre daha başarılı tahminler elde etmesinin rassal bir sonuç olduğu neticesine

varılmaktadır.



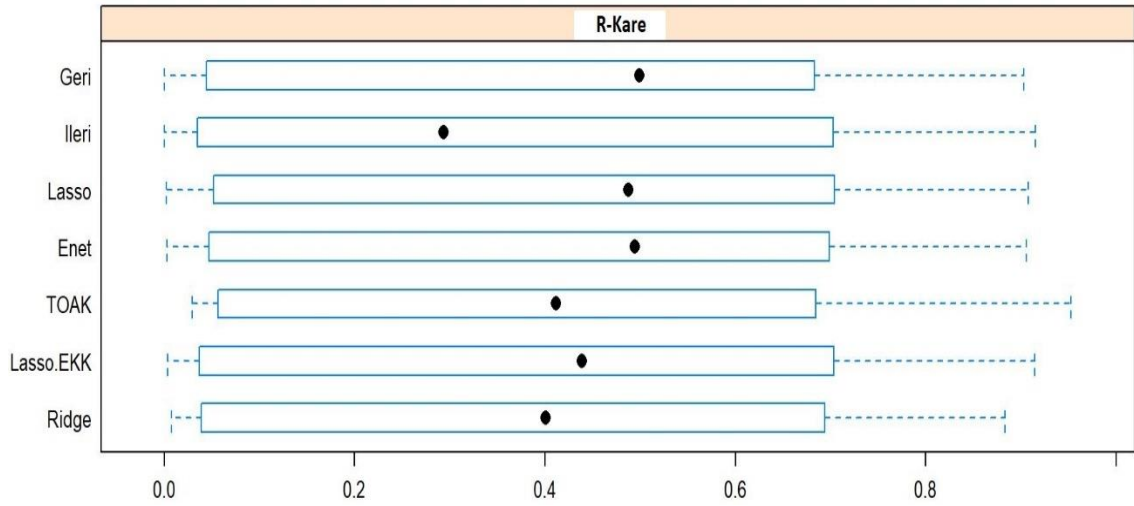
Şekil 4.16. Geriye doğru seçim yöntemi ÇG HKO sonuçlarının düzenleştirici yöntemler ÇG HKO sonuçları ile karşılaştırılması

Şekil 4.17’de ise Lasso-EKK modelinin ileriye doğru seçim, geriye doğru seçim ve tüm olası alt kümeler yöntemleri ile karşılaştırıldığı saçılım grafiği verilmiştir. Grafiklerde değişken seçim yöntemleri ile Lasso-EKK yönteminin ÇG HKO sonuçlarına göre birbirlerine karşı net bir üstünlüklerinin olmadığı görülmektedir. Dolayısıyla Lasso-EKK yöntemi ile diğer üç yöntemin Tablo 4.18’de verilen eğitim HKO sonuçları arasındaki farkların rassal nedenlerden kaynaklandığı sonucuna varılmaktadır.



Şekil 4.17. *Lasso-EKK ÇG HKO sonuçlarının değişik seçim yöntemleri ile karşılaştırılması*

Şekil 4.18’de yöntemlerin belirlilik katsayılarının kutu grafikleri incelendiğinde yöntemlerin kutu grafiklerinin örtüştüğü ve böylece elde edilen sonuçların birbirlerinden farklı olmadıkları görülebilir.



Şekil 4.18. Belirlilik katsayılarının kutu grafikleri

Önceki kısımlarda yapılan analizler veri kümesinin %70 eğitim ve %30 test biçiminde ayrılmasına dayanarak gerçekleştirilmiştir. Ancak literatürde 80/20, 60/40 gibi farklı eğitim/test kümesi ayrımları da mevcuttur. İdeal eğitim/test ayrımının hangi oranda olması gerektiği konusunda literatürde net bir uzlaşma olmaması sebebiyle ve modellerin tahmin edilmesinde kullanılan eğitim verisi kümesinde gerçekleşen artış ve azalışların yöntemlerin seçtiği değişkenler üzerindeki etkilerini gözlemleyebilmek adına önceki kısımda tahmin edilen tüm modeller 80/20 ve 60/40 eğitim/test ayrımına göre tekrar tahmin edilmiştir. 80/20 ayrımı sonrasında (1020 eğitim gözlemi ve 252 test gözlemi) eğitim verisinden yararlanarak çapraz geçerlilik yöntemi ile tahmin edilen modellerin katsayıları Tablo 4.19’da ve 60/40 ayrımı (764 eğitim ve 508 test gözlemi) ile elde edilen modellerin katsayı tahminleri ise Tablo 4.20’de yer almaktadır.

Tablo 4.19’da görüleceği üzere tüm modeller tarafından seçilen değişkenler F1 ve D8 değişkenleridir. Buna ek olarak F16 değişkeni tüm olası alt kümeler yöntemi haricinde ve L3 değişkeni de ileriye doğru seçim yöntemi dışında tüm modellerde yer almışlardır. Likidite oranlarından L2, L6; finansal yapı oranlarından F3, F8, F10, F12, F15 ve devir hızı oranlarından D1, D2, D5, D6, D7 değişkenleri değişkeni seçimi gerçekleştirilmeyen ridge regresyon modeli dışında herhangi bir modelde yer almayan değişkenlerdir. En seyrek seçimi gerçekleştiren yöntem 4 değişkenle hard-thresholding uygulayan tüm olası alt kümeler ve 5’er değişken seçen ileriye doğru seçim ile geriye doğru seçim yöntemleridir. Her ne kadar seyrek ve birbirlerine yakın sayıda değişkene sahip modeller seçmiş olsalar da bu üç yöntemin aktif bağımsız değişken kümeleri birbirlerinden farklılık göstermektedir.

Tablo 4.19. 80/20 eğitim/test ayrımı sonrası aktif kârlılık modellerinin katsayı tahminleri

	TOAK	İleri	Geri	Lasso	Enet	Ridge	EKK-Enet	EKK-Lasso
Sabit	0,0441***	0,0441***	0,0441***	0,0837	0,1086	0,1097	0,0441***	0,0441***
L1			-0,0293***			0,0007		
L2						0,0018		
L3	0,0271***		0,0316***	0,0047	0,0097	0,0072	0,0149***	0,0098***
L4				-0,0042	-0,0304	-0,0336	-0,0117***	-0,0068**
L5		-0,0231***				-0,0331		
L6						-0,0001		
L7					-0,0121	-0,0287	-0,0089**	
L8		-0,0249***				-0,0042		
F1	-0,0511***	-0,0430***	-0,0504***	-0,1477	-0,1017	-0,0509	-0,0465*	-0,0396***
F3						-0,0005		
F4					-0,0706	-0,0581	-0,0073	
F6					-0,0312	-0,0372	-0,0007	
F7					-0,0003	-0,0010	-0,0042	
F8						-0,0001		
F9	-0,0298***				-0,0043	-0,0041	-0,0204***	
F10						-0,0006		
F11				-0,0079	-0,0054	-0,0102	0,0012	-0,0055
F12						0,0125		
F13					-0,0050	-0,0271	0,0086	
F14					-0,0008	-0,0099	-0,0059	
F15						-0,0035		
F16		0,0442***	0,0217***	0,0351	0,0374	0,0359	0,0099	0,0091*
F17				-0,0061		-0,0327	0,0441	-0,0042
D1						-0,0001		
D2						-1x10 ⁻⁵		
D3						0,0062	0,0149	
D4						-4x10 ⁻⁶	-0,0117	
D5						-0,0001		
D6						0,0001		
D7						-0,0008	-0,0089*	
D8	0,0220***	0,0210***	0,0189***	0,0266	0,0336	0,0194	0,0206***	0,0211***

Anlamlılık düzeyleri: p<0.001: '***', p<0.01: '**', p<0.05: '*'

Tablo 4.20'de 60/40 eğitim/test ayrımına göre sekiz modele ait katsayı tahminleri yer almaktadır. L3, F4 ve D8 değişkenleri tüm modellerde yer alan değişkenlerdir. Tabloda gözlemlenebileceği üzere eğitim gözlem sayısı azaldığında tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemlerinin aktif bağımsız değişken kümeleri seyreklikten uzaklaşarak daha fazla sayıda değişkeni modele eklemektedirler. Öte yandan lasso ve elastic-net yöntemleri soft-thresholding ve hard-thresholding işlemlerini birlikte kullandıkları için seyreklik açısından ele alındıklarında eğitim

kümesindeki artış veya azalıştan daha az etkilenmekte ve sıradan değişken seçim yöntemlerine göre daha tutarlı bir davranış göstermektedirler.

Tablo 4.20. 60/40 eğitim/test ayrımı sonrası aktif kârlılık modellerinin katsayı tahminleri

60/40	TOAK	İleri	Geri	Lasso	Enet	Ridge	Enet-EKK	Lasso-EKK
Sabit	0,0421***	0,0421***	0,0421***	0,0609	0,0707	0,0744	0,0421***	0,0421***
L1						0,0013		
L2						0,0021		
L3	0,0351***	0,0332***	0,0312***	0,0102	0,0134	0,0052	0,0238***	0,0132***
L4	0,0282**				-0,0066	-0,0190	-0,0043	
L5	-0,0372***					-0,0066		
L6	0,0193**				3x10 ⁻⁵	-0,0003	-0,0044	
L7	0,0214*					-0,0101		
L8	-0,0241					0,0042		
F1		-0,0006		-0,1185	-0,1068	-0,0266	-0,0314	-0,0183***
F3	-0,0284***	-0,0287***				0,0002		
F4	-0,0571***	-0,0556***	-0,0489***	-0,0188	-0,0670	-0,0315	-0,0180	-0,0276***
F6					-0,0004	-0,0207	0,0011	
F7	-0,0595***		-0,0574***		-0,0005	-0,0014	-0,0038	
F8	-0,0083**	-0,0078				1x10 ⁻⁵		
F9			-0,0297***		-0,0027	-0,0015	-0,0239***	
F10	0,0554***		0,0508***			-0,0009		
F11				-0,0058	-0,0030	-0,0077	0,0031	0,0030
F12	0,0270***	0,0280***	0,0244***			0,0072		
F13						-0,0207		
F14						-0,0088		
F15						-0,0082		
F16	0,0516***	0,0084		0,0523	0,0576	0,0236	0,0071	0,0224***
F17		-0,0087*		-0,0140	-0,0272	-0,0237	-0,0094*	-0,0059
D1	-0,0253***	-0,0056*				-0,0001		
D2						1x10 ⁻⁵		
D3						0,0029		
D4						1x10 ⁻⁶		
D5						2x10 ⁻⁵		
D6	-0,0048					0,0003		
D7						-0,0007		
D8	0,0294***	0,0233***	0,0222***	0,0284	0,0351	0,0126	0,0209***	0,0234***

Anlamlılık düzeyleri: p<0.001: '***', p<0.01: '**', p<0.05: '*'

Tablo 4.21’de tüm yöntemlerin farklı oranlardaki eğitim/test ayrımlarına göre sahip oldukları eğitim hataları ve test hataları karşılaştırılmış ve buna ek olarak modellerde yer alan bağımsız değişken sayılarına yer verilmiştir. Tabloya göre en seyrek modeller geriye doğru seçim ve lasso regresyon modelidir. Modellerin eğitim ve test hataları

incelendiğinde farklı eğitim/test ayrımları için eğitim ve test hatalarında çok büyük farklılıklar görülmemektedir. En çok göze çarpan farklılık %60 eğitim verisiyle tahminlenen TOAK modelinin test hatasında görülmektedir. Önceki ayrımlarda daha düşük test hatalarına sahip TOAK yöntemi %60 eğitim verisi uygulandığında çok fazla sayıda değişkeni modele ekleyerek (model karmaşıklığının artması sonucunda) eğitim hatası azalmış ve aşırı uyum sorunu ile karşılaşmıştır. Tüm olası alt kümeler yöntemi 80/20 ve 70/30 ayırımına göre 4'er değişken seçmesine karşın 60/40 ayırımında seçtiği değişken sayısı 16'ya çıkmaktadır. Öte yandan ileriye doğru seçim yöntemi 60/40 ayırımında değişken sayısı diğer iki ayrıma göre tam iki katına çıkarak 10'a ulaşmıştır.

Tablo 4.21. 80/20, 70/30 ve 60/40 eğitim/test ayrımlarına ilişkin değerlendirme sonuçları

	80/20			70/30			60/40		
	Eğitim HKO	Test HKO	Değişken Sayısı	Eğitim HKO	Test HKO	Değişken Sayısı	Eğitim HKO	Test HKO	Değişken Sayısı
TOAK	0,0733	0,0802	4	0,0719	0,0807	4	0,0692	0,1337	16
ILERI	0,0730	0,0800	5	0,0724	0,0797	5	0,0727	0,0789	10
GERI	0,0733	0,0801	5	0,0717	0,0802	6	0,0712	0,0793	7
Lasso	0,0765	0,1063	7	0,0762	0,1021	7	0,0772	0,1055	7
Ridge	0,0773	0,1094	31	0,0764	0,1067	31	0,0797	0,1036	31
Elastic-Net	0,0748	0,1099	13	0,0730	0,1066	16	0,0756	0,1082	12
Lasso-EKK	0,0744	0,0791	7	0,0739	0,0799	7	0,0741	0,0793	7
Enet-EKK	0,0731	0,0784	13	0,0724	0,0800	16	0,0722	0,0797	12

Lasso, elastic-net ve ridge regresyon yöntemleri her üç ayrımda sıradan değişken seçim yöntemlerine göre daha yüksek test hatasına sahiplerdir. Elastic-net yöntemi 80/20 ayırımında 13, 70/30 ayırımında 16 ve 60/40 ayırımında 12 bağımsız değişkenli modeller vermiştir. Lasso yöntemi ise tüm modeller içinde üç farklı eğitim kümesinde de aynı sayıda ve 80/20 ve 70/30 ayrımlarında tam olarak aynı değişkenleri seçerek buna göre tahmin modelleri üretmiştir. Lasso regresyon yönteminde eğitim kümesi hacminden bağımsız olarak kararlı biçimde sıfırdan farklı katsayılar sahip bağımsız değişkenler L3, L4, F1, F11, F16, F17 ve D8'dir. Tablo 4.16'daki sonuçlara göre ise 60/40 ayırımı sonucunda elde edilen lasso modelinde ise L4 modelde yer almazken yerine F4 değişkeni dahil edilmiştir.

Firma performansının aktif kârlılık oranı ile ölçüldüğü ve bağımlı değişken olarak farklı regresyon yöntemleri ile modellendiği bu bölümde eğitim kümesindeki değişimlere en az duyarlı olan ve en kararlı değişken seçimine sahip yöntemin lasso-EKK regresyon modeli olduğu sonucuna varılmıştır. Lasso-EKK modeli her üç ayırmda da (80/20, 70/30 ve 60/40) en düşük test hatasına dolayısıyla en iyi genelleştirme performansına sahip modeller arasında ilk iki arasında yer almıştır. Böylece iyi uyumlu ve genelleştirme performansı yüksek tahmin modelleri üretmiştir. Buna ek olarak lasso yöntemi farklı eğitim kümeleri için 31 bağımsız değişkenden 7'sini seçerek seyrek sonuçlar üretmeyi başarmıştır. Buna göre aktif kârlılığın finansal oranlar ile açıklanması amacını gerçekleştirmek için en uygun modelin lasso-EKK modeli olduğu sonucuna varılmıştır. Diğerlerine kıyasla eğitim/test ayırımında ortada yer alan 70/30 ayırımı sonucunda elde edilen lasso-EKK doğrusal regresyon modelinin katsayı tahminleri ve katsayıların t değerleri Eşitlik (4.1)'deki gibidir:

$$\begin{aligned}
 ROA = & 0,043 + 0,011 L3 - 0,008 L4 - 0,040 F1 - 0,006 F11 + 0,009 F16 \\
 & (17,50) (3,97) \quad (-2,99) \quad (-13,30) \quad (-1,81) \quad (1,91) \\
 & - 0,004 F17 + 0,021 D8 \\
 & (-1,01) \quad (7,44)
 \end{aligned} \tag{4.1}$$

İstatistiksel modelde likidite oranlarından nakit oranı (L3) ve stoklar/dönen varlıklar oranı (L4); finansal yapı oranlarından kaldıraç oranı (F1), duran varlıklar/devamlı sermaye oranı (F11), dönen varlıklar/toplam varlıklar oranı (F16) ve maddi duran varlıklar/toplam varlıklar oranı (F17); faaliyet oranlarından ise D8 (aktif devir hızı) oranı modelde yer almıştır. Finansal yapı oranlarından duran varlıklar/devamlı sermaye oranı (F11), dönen varlıklar/toplam varlıklar oranı (F16) ve maddi duran varlıklar/toplam varlıklar oranı (F17) modelde yer almalarına rağmen aktif kârlılık ile %5 anlamlılık düzeyinde istatistiksel olarak anlamlı bir ilişkiye sahip değillerdir.

Doğrusal regresyon modelinde mutlakça en yüksek katsayıya (-0,040) ve böylece aktif kârlılık üzerinde en yüksek etkiye sahip finansal oran kaldıraç oranıdır. Buna göre imalat firmalarının toplam borçluluğunun sermaye yapısındaki oranı arttıkça aktif kârlılık önemli ölçüde azalmaktadır. Bir diğer yüksek katsayı değerine (0,021) sahip finansal oran aktif devir hızıdır (D8). Tablo 4.12, 4.14, 4.15 ve 4.16'dan görüleceği üzere aktif devir hızı tüm eğitim/test ayırımlarında analize dahil edilen tüm yöntemlerce modele dahil edilmiştir. Buna göre varlıkların etkin kullanımı ile aktif kârlılık arasında önemli bir ilişki olduğu sonucuna varılmaktadır. Nakit oranı (L3) incelendiğinde imalat işletmelerinin

dönen varlıkları içerisindeki nakit miktarını artırmaları aktif kârlılığı pozitif yönde etkilemektedir. Öte yandan imalat işletmelerinde stokların dönen varlıklar (L4) içerisindeki payının artması ise aktif kârlılığı negatif yönde etkilemektedir.

4.4. Öz Kaynak Kârlılık Analizi

Bu bölümde EKK yöntemi ile regresyon modeli tahmini gerçekleştiren sıradan değişken seçim yöntemleri ile düzenleyici yöntemlerin performanslarının karşılaştırılması amacıyla likidite oranları, finansal yapı oranları ve devir hızı oranlarının işletmenin kârlılığını ölçen bir diğer oran olan öz kaynak kârlılık oranı (ROE) üzerindeki etkileri incelenecektir. Regresyon modellerindeki bağımlı değişken olan öz kaynak kârlılık oranı Tablo 4.9’da gösterildiği üzere net kâr/öz kaynaklar biçiminde hesaplanmaktadır. Tablo 4.6, Tablo 4.7 ve Tablo 4.8’de verilen likidite, finansal yapı oranları ve devir hızları tahminlenecek regresyon yöntemlerindeki bağımsız değişkenlerdir.

4.4.1. En küçük kareler yöntemi

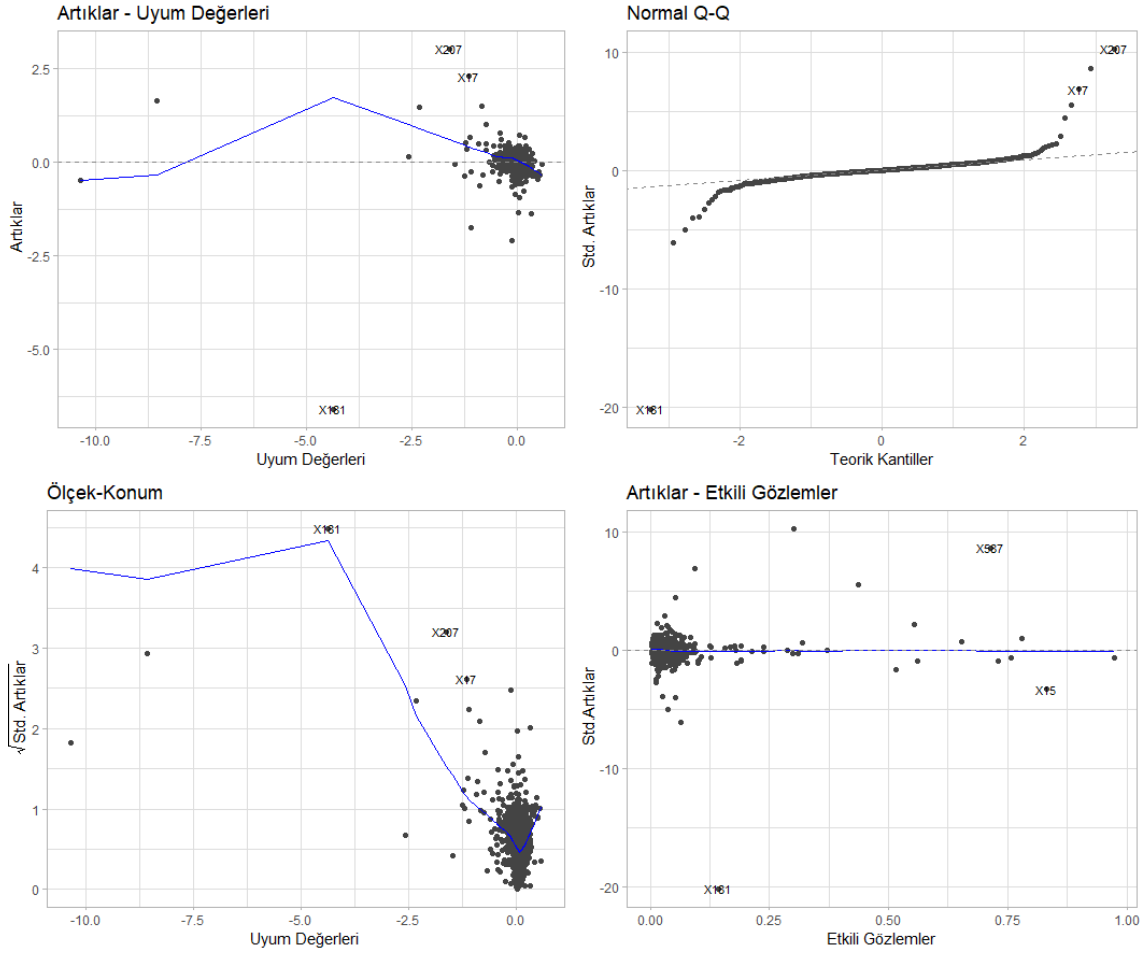
Tablo 4.6, Tablo 4.7 ve Tablo 4.8’de verilen tüm bağımsız değişkenlerin öz kaynak kârlılığına uyumlandırılması ile elde EKK doğrusal regresyon modelinin katsayı tahminleri Tablo 4.22’de verilmiştir. EKK modelinde yer alan değişkenlerden L4, L5, F1, F4, F6, F7 F8, F11, F12, F13, F16 ve D7 değişkenleri %5 anlamlılık düzeyinde istatistiksel olarak anlamlıdır. Tablo 4.22’nin sağ sütununda verilen varyans büyütme çarpanı değerleri incelendiğinde 19 değişkenin VBC değerlerinin 10’un üzerinde olduğu ve modelde çoklu bağıntının mevcut olduğu görülmektedir. F testi sonuçlarına göre ($F(31, 860) = 65,86, p < 0.001$) doğrusal regresyon modeli istatistiksel olarak anlamlı olduğu sonucuna varılmıştır. Hataların normal dağılıma sahip olup olmadığını test etmek amacıyla gerçekleştirilen Shapiro-Wilk normallik testi sonucunda ($W=0,784, p < 0,001$) hataların normal dağılıma sahip olmadıkları görülmektedir. Ayrıca Breusch-Pagan testi sonucuna göre ($BP(31) = 202,07, p < 0,001$) hataların değişen varyanslılığa sahip oldukları görülmektedir. EKK modelinin belirlilik katsayısı %48,25 ve düzeltilmiş belirlilik katsayısı ise %41,79 olarak elde edilmiştir.

Tablo 4.22. *En küçük kareler öz kaynak kârlılığı doğrusal regresyon modeli*

	Katsayı Tahminleri	Varyans Büyütme Çarpanı
Sabit	0,0117	66,2
L1	0,0261	53,4
L2	-0,0114	9,5
L3	0,0611	16,0
L4	0,1731***	19,9
L5	-0,1821***	2,0
L6	0,0213	17,6
L7	0,0754	29,0
L8	-0,0551	110,2
F1	0,7940***	157,5
F3	-0,0801	75,2
F4	-0,7453***	30,3
F6	-0,3426***	90,1
F7	-0,3648**	2,3
F8	-0,0368*	59,9
F9	0,0370	75,9
F10	-0,2271	5,0
F11	0,0772**	27,0
F12	0,1977**	20,9
F13	0,1479**	9,7
F14	-0,0501	18,4
F15	-0,0308	39,7
F16	0,1917*	6,5
F17	0,0344	1,5
D1	-0,0076	1,1
D2	-0,0211	8,1
D3	0,0629	1,2
D4	0,0080	5,4
D5	-0,0044	7,8
D6	-0,0062	10,1
D7	0,1036**	11,0
D8	-0,0407	66,2

Anlamlılık düzeyleri: $p < 0.001$: '***', $p < 0.01$: '**', $p < 0.05$: '*'

Şekil 4.19’da verilen tanı grafiklerinden “Artıklar-Uyum Değerleri” incelendiğinde artıkların sıfırın etrafında rassal dağılmadıkları görülmekte ve bu durum değişen varyanslılığın mevcut olduğunu işaret etmektedir. “Normal Q-Q” grafiğine göre artıklar normal dağılıma sahip değildir. “Ölçek-Konum” ve “Artıklar-Etkili Gözlemler” grafikleri ise veri kümesinde önemli etkili gözlemlerin olduğunu göstermektedir.



Şekil 4.19. EKK öz kaynak kârlılığı modelinin tanı grafikleri

892 gözlemden oluşan eğitim kümesinde uygulanan 10 katlı çapraz geçerlilik sonuçlarına göre tahminlenen tüm olası alt kümeler (TOAK), ileriye doğru seçim (İLERİ) ve geriye doğru seçim (GERİ) doğrusal regresyon modellerine ilişkin bulgular Tablo 4.23’te verilmiştir. Tüm olası alt kümeler yönteminde 13, ileriye doğru seçim yönteminde 8 bağımsız değişkenin modellere alınmasına karşın geriye doğru seçim yöntemi ile yalnızca F10 bağımsız değişkeni seçilmiştir. Tüm modellerde yer alan tüm bağımsız değişkenler %5 anlamlılık düzeyinde anlamlıdır. Geriye doğru seçim yöntemiyle yalnızca tek bir değişkenin seçilmesi açıklanacak.

F-testi sonuçlarına göre (Tüm olası alt kümeler: $(F(13, 878) = 154., p < 0.001)$; ileriye doğru seçim: $(F(8, 883) = 232,7, p < 0.001)$; geriye doğru seçim: $(F(1, 890) = 1584, p < 0.001)$) her üç model de istatistiksel olarak anlamlıdır.

Tablo 4.23. *Tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim doğrusal regresyon modellerinin katsayı tahminleri*

Değişkenler	TOAK	İleri	Geri
Sabit	0,0117	0,0117	0,0117
L4	0,1083***		
L5	-0,1401***		
F1	0,6602***	0,3207***	
F4	-0,6112***	-0,3622***	
F6	-0,2728***		
F7	-0,1740**	-0,1666*	
F10	-0,3638***	-0,4057***	-0.5102***
F11	0,1069***	0,0734***	
F12	0,1147**	0,1442***	
F13	0,1540**	0,0234***	
F14			
F15	-0,0935**		
F16	0,1365***		
D3	0,0584***	0,0553***	

Anlamlılık düzeyleri: $p < 0.001$: '***', $p < 0.01$: '**', $p < 0.05$: '*'

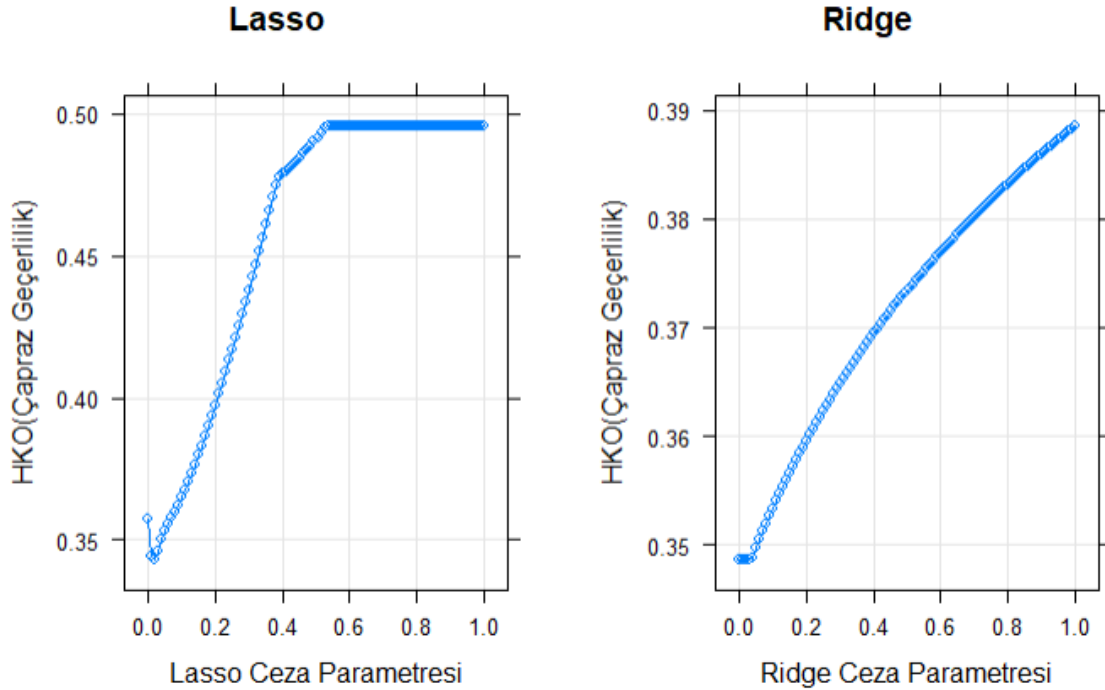
Tablo 4.23'te katsayı tahminleri verilen modellerin değerlendirme ölçütleri Tablo 4.24'te yer almaktadır. En düşük eğitim hatasına sahip model beklendiği üzere en çok değişkene modelde yer veren tüm olası alt kümeler modelidir. Tüm olası alt kümeler modeli aynı zamanda en düşük test hatasına yani en iyi genelleştirme performansına sahiptir. Belirlilik katsayıları incelendiğinde tek bir değişkene sahip olan geriye doğru seçim yönteminin belirlilik katsayısının %64,02 olduğu ve diğer iki yöntemin belirlilik katsayıları ile arasında son derece az bir farkın olduğu görülmektedir. Örnekleme yer alan etkili gözlemler, belirlilik katsayısının değerinin gerçekte olduğundan daha büyük olarak tahmin edilmesini sağlamış olabilir.

Tablo 4.24. Öz kaynak kârlılığı EKK modellerinin değerlendirilmesi

	TOAK	İLERİ	GERİ
Belirlilik Katsayısı	0,6954	0,6783	0,6402
Düzeltilmiş Belirlilik Katsayısı	0,6909	0,6754	0,6398
ÇG Eğitim HKO	0,3320	0,3576	0,3459
ÇG Test HKO	0,2971	0,2938	0,3115

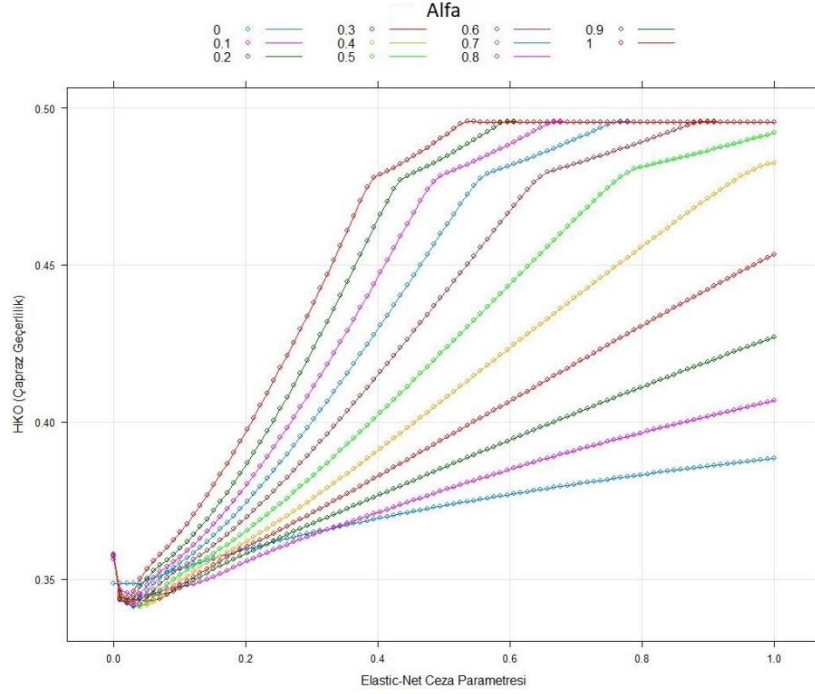
4.4.2. Düzenleştirici yöntemler analizi

Bu kısımda düzenleştirici yöntemlerden lasso, ridge ve elastic-net regresyon yöntemlerinin değişken seçimi ve tahmin performansları finansal oranların öz kaynak kârlılığı üzerindeki etkileri incelenerek değerlendirileceklerdir. Eğitim kümesinden yararlanarak çapraz geçerlilik sonuçlarına göre lasso, ridge ve elastic-net doğrusal regresyon modellerinin katsayılarını tahmin etmeden önce λ ceza parametresi ve elastic-net α parametreleri belirlenmiştir. Lasso ve ridge regresyon yöntemlerinin optimum ceza parametrelerinin belirlendiği çapraz geçerlilik grafikleri Şekil 4.21’de yer almaktadır. Grafiğin x ekseninde λ parametresinin duali olan s parametresi yer almaktadır. Buna göre lasso s kısıtı 0,0203 ve ridge regresyon s kısıtı 0,0304 olarak belirlenmiştir.



Şekil 4.20. Lasso ve ridge regresyon yöntemlerinin ÇG ile optimum ceza parametre değerleri

Elastic-net yönteminin ceza kısıtının $[0,1]$ aralığındaki 10 farklı α değeri için değişimleri Şekil 4.21’de yer almaktadır. Çapraz geçerlilik işlemi sonucunda $\alpha = 0.5$ için $s_{enet} = 0.0405$ değeri en düşük çapraz geçerlilik HKO değerini veren optimum ceza katsayısı olarak belirlenmiştir.



Şekil 4.21. Çapraz geçerlilik yöntemi ile elastic-net hiperparametrelerin değişim grafiği

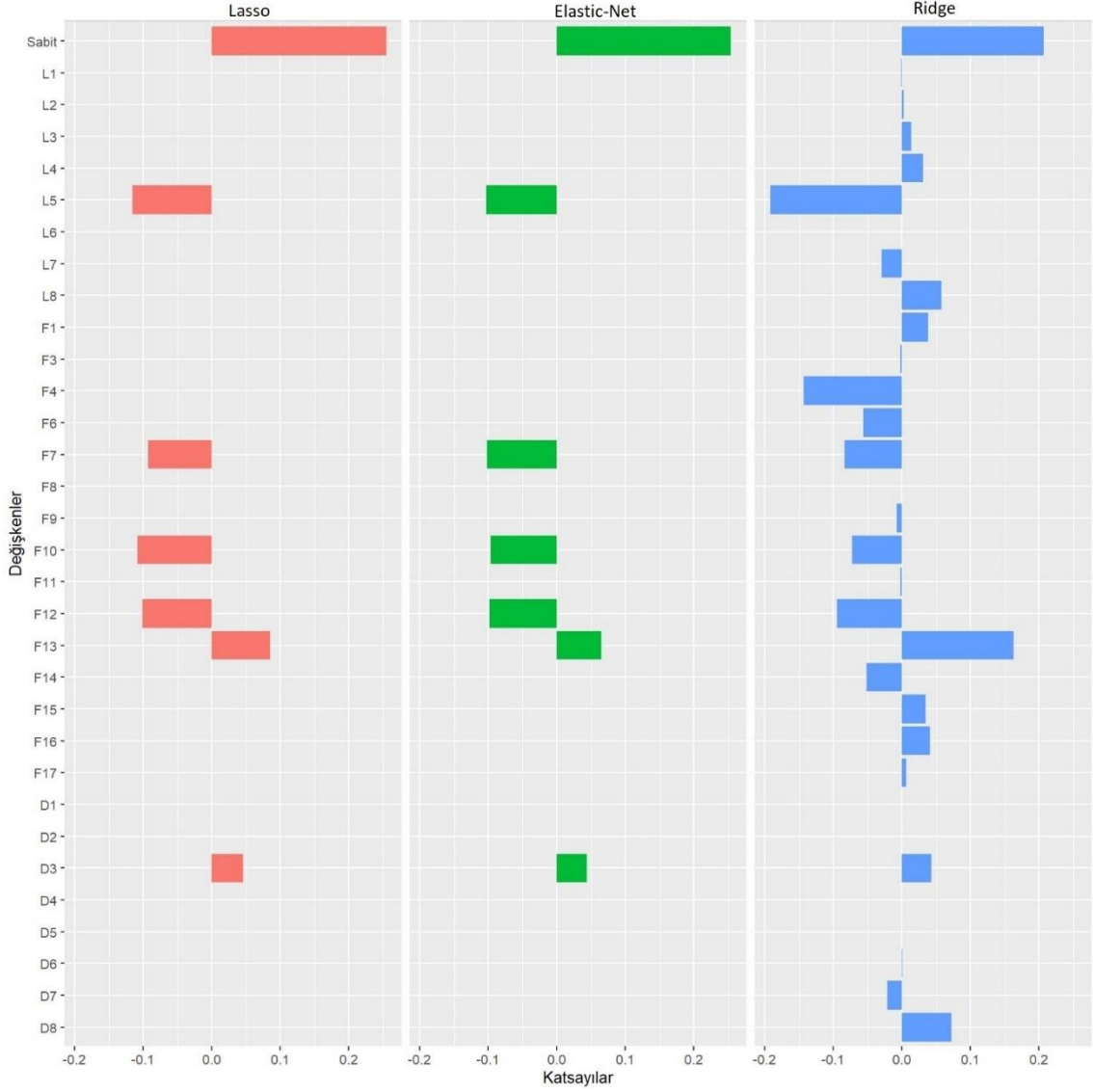
Öz kaynak kârlılığı için çapraz geçerlilik esasına göre oluşturulan lasso, ridge ve elastic-net doğrusal regresyon modellerinin katsayı tahminleri Tablo 4.25’te yer almaktadır. Seyrek çözüme sahip lasso regresyon ve elastic net regresyon modellerinin her ikisinde de L5, F7, F10, F12, F13 ve D3 değişkenleri sıfırdan farklı katsayılara sahiplerdir. Elastic-net yönteminin uygulamada $\alpha = 1$ değerine sahip olan ve dolayısıyla daha seyrek bağımsız değişken kümesine sahip lasso regresyondaki değişken sayısı ile uygulamada $\alpha = 0$ değerine sahip ve tüm bağımsız değişkenlerin yer aldığı ridge regresyon modelindeki değişken sayısı arasında bir aktif değişken kümesi oluşturması beklenir. Ancak elastic-net modeli lasso ve ridge regresyon modellerinin tam ortasında yer almasına rağmen lasso regresyon modeli ile aynı değişken kümesine sahiptir. Değişkenler arasındaki yüksek korelasyon değerleri ve veride yer alan etkili gözlemler elastic-netin kararsız davranmasına yol açtığı için bu sonuç elde edilmiştir. Öte yandan veride yer alan problemlere rağmen lasso ve elastic-net yöntemlerinin kendi aralarında tutarlı bir davranış sergiledikleri sonucuna varılmıştır.

Tablo 4.25. *Lasso, elastic-net ve ridge regresyon modellerinin katsayı tahminleri*

	Lasso	Elastic Net	Ridge
Sabit	0,2548	0,2539	0,0481
L1			-0,0054
L2			0,0045
L3			0,0273
L4			0,2523
L5	-0,1158	-0,1025	-0,5271
L6			0,0000
L7			-0,0166
L8			0,1256
F1			0,1589
F3			0,0019
F4			-0,4820
F6			-0,0669
F7	-0,0930	-0,1023	-0,1093
F8			-0,0005
F9			-0,0198
F10	-0,1082	-0,0968	-0,0971
F11			0,0741
F12	-0,1006	-0,0984	0,0173
F13	0,0849	0,0653	0,4556
F14			-0,1581
F15			-0,0010
F16			0,1312
F17			0,0582
D1			-0,0002
D2			-0,0001
D3	0,0458	0,0435	0,0526
D4			3×10^{-5}
D5			-0,0006
D6			-1×10^{-5}
D7			-0,0053
D8			0,0481

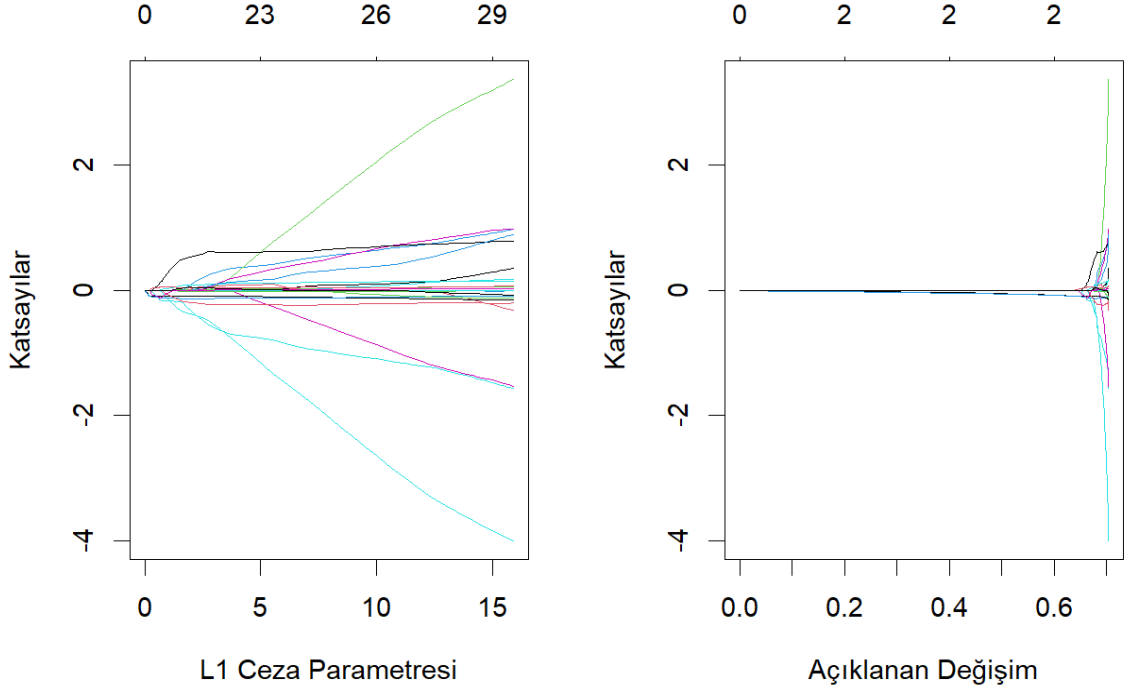
Tablo 4.25'te verilen katsayı tahminlerinin görsel olarak karşılaştırılması Şekil 4.22'de yer almaktadır. Lasso ve elastic-net yöntemlerinin katsayı tahminlerinin

birbirlerine oldukça yakın oldukları ve her üç yöntemde de yer alan değişkenlerin katsayı işaretlerinin aynı olması yöntemlerin tutarlı davranış sergilediklerini göstermektedir.



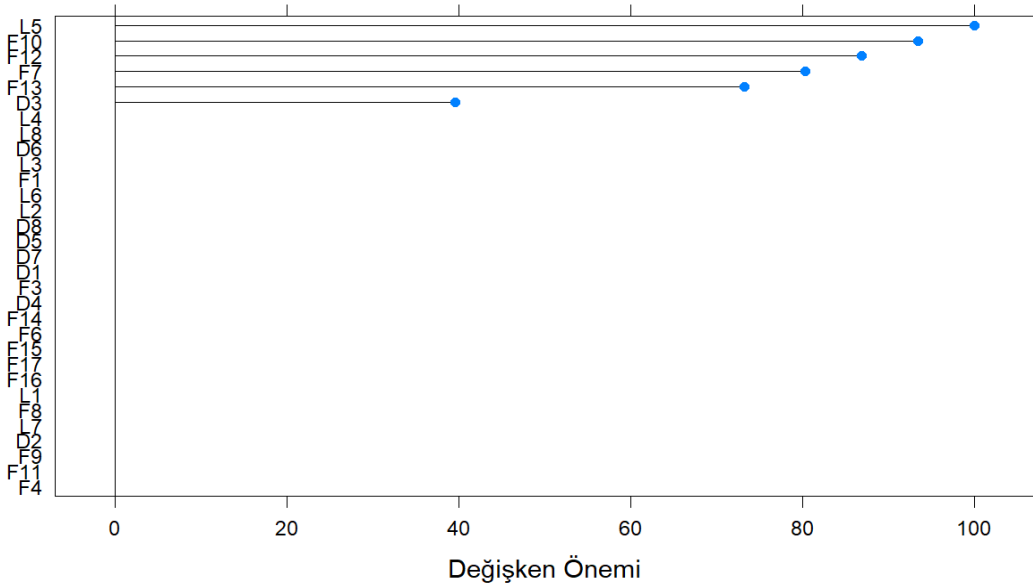
Şekil 4.22. Düzenleştirici yöntemlerin katsayı tahminlerinin çubuk grafikleri

Lasso regresyon modelinin çapraz geçerlilik yolu boyunca ceza parametresindeki değişime bağlı olarak katsayı tahminlerindeki dönüşüm grafiği ile açıklanan değişim grafiği Şekil 4.23'te yer almaktadır. Lasso regresyon analizine giren bağımsız değişkenlerin her biri farklı renkte bir eğri ile gösterilmiştir. Ceza kısıtı olan s değeri sıfıra yaklaştıkça katsayı tahminlerinin genel olarak sıfıra doğru yaklaştığı görülmektedir. Sağdaki grafiğin x ekseninde açıklanan değişim ve üst satırında ise her bir açıklanan değişime karşılık gelen değişken sayısı verilmiştir. Grafikten görüleceği üzere modelde yer alan 2 değişken öz kaynak kârlılığındaki toplam değişimin %60'ından fazlasını açıklamaktadır.



Şekil 4.23. Lasso katsayılarının değişimi– Lasso açıklanan değişim

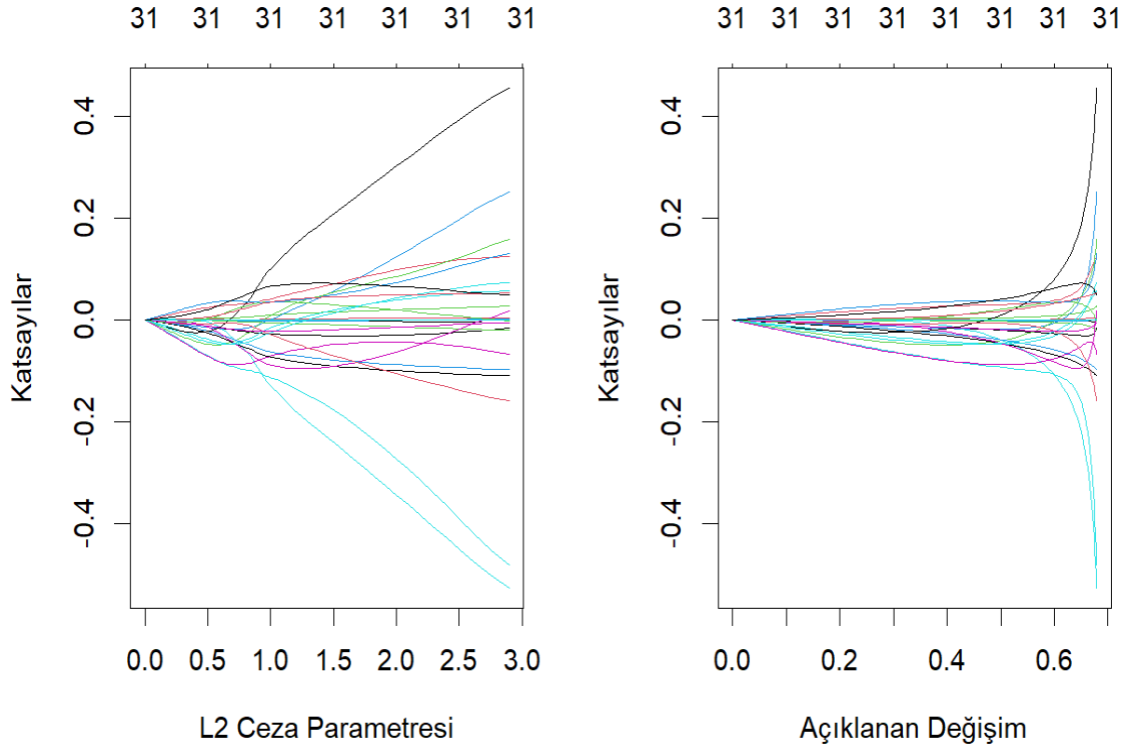
Lasso modelindeki sıfırdan farklı katsayıya sahip değişkenlerin göreceli önemleri Şekil 4.24’te verilmiştir. Değişken önemi sıralamasına göre en önemli iki değişken L5 ve F10 olarak belirlenmiştir. Bu iki değişken toplam değişimin %60’ından fazlasını açıklamaktadır.



Şekil 4.24. Lasso regresyon bağımsız değişkenlerinin önem sıralaması

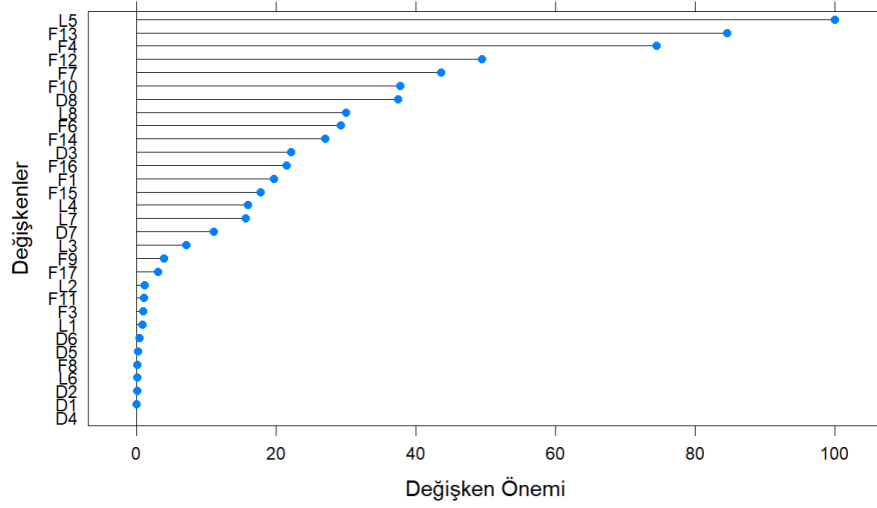
Ridge regresyon modelinde yer alan değişkenlerin katsayı tahminlerinin ceza katsayısının değerine bağlı olarak değişimleri ve açıklanan değişim grafikleri Şekil

4.25'te yer almaktadır. Ceza katsayısının değeri sıfıra yaklaştıkça regresyon katsayılarının tahmin değerlerinin de sıfıra yaklaştığı ve yalnızca ceza katsayısının değeri sıfıra eşit olduğunda katsayı tahminlerinin de sıfıra eşit olduğu görülmektedir. Düzenlenmiş katsayılara sahip doğrusal ridge regresyon modeli öz kaynak kârlılığındaki değişimin %49,51'ini açıklamaktadır.



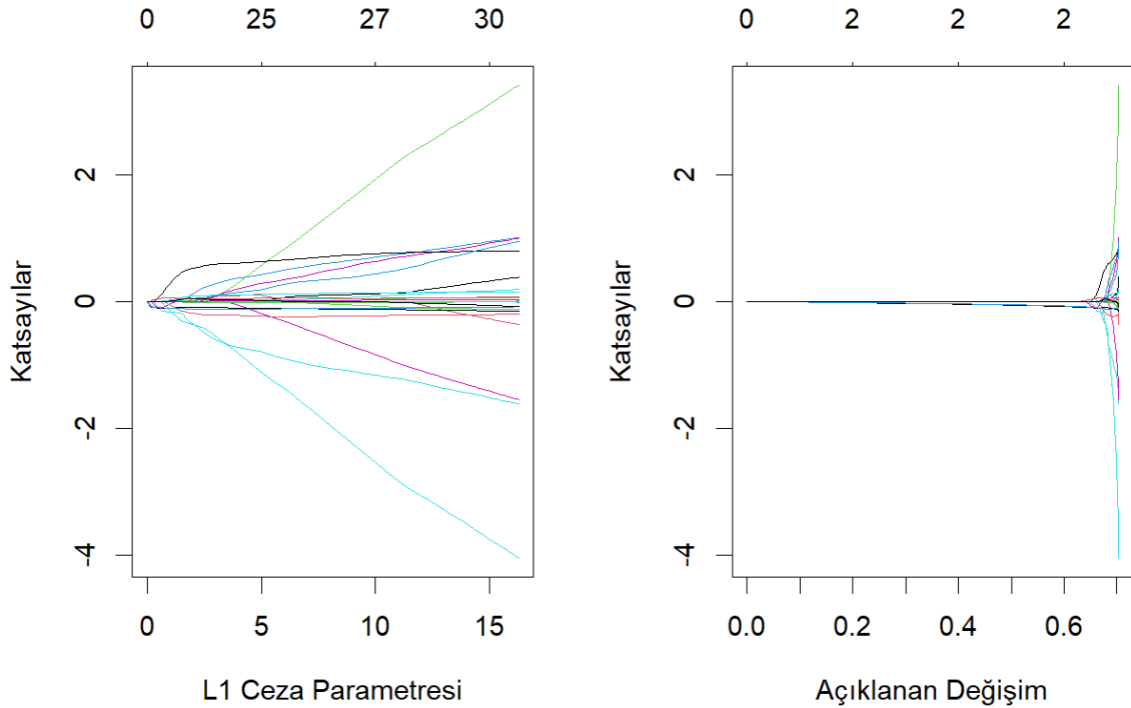
Şekil 4.25. Ridge regresyon katsayıların değişimi– Ridge açıklanan değişim

Ridge regresyon modelindeki en önemli değişkenlerin L5, F13 ve F4 değişkenlerinin olduğu Şekil 4.26'da verilen önem sıralaması grafiğinde görülmektedir. Lasso regresyon modelinde en önemli ikinci değişken olan F10 değişkeni ridge regresyon değişkenlerinin önem sıralamasında ikinci sırada yer almaktadır.



Şekil 4.26. Ridge regresyon bağımsız değişkenlerinin önem sıralaması

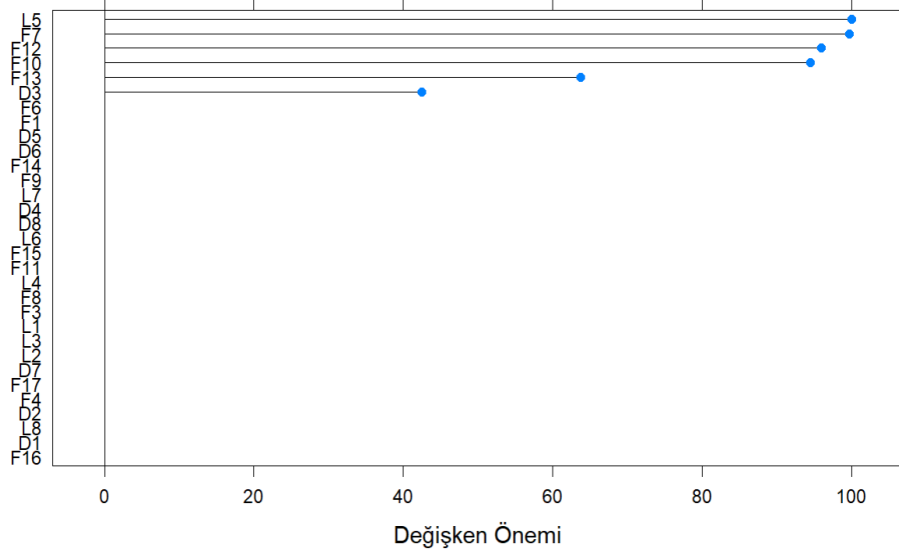
Şekil 4.27’de elastic-net yönteminde elde edilen katsayıların değişimi ve açıklanan değişime ilişkin grafikler gösterilmiştir. Elde edilen sonuçların Şekil 4.23’teki lasso regresyon sonuçları ile oldukça benzer oldukları görülebilir. Elastic-net yöntemine göre modelde yer alan 2 değişken öz kaynak kârlılığındaki toplam değişimin %60’ından fazlasını açıklamaktadır.



Şekil 4.27. Elastic-net katsayıların değişimi– Elastic-net açıklanan değişim

Şekil 4.28’de elastic-net doğrusal regresyon modelindeki değişimlerin önem sıraları yer almaktadır. Elastic-net yöntemi her ne kadar lasso ile aynı değişkenlere sahip olsa da

değişkenlerin önem sıralarında farklılık görülmektedir. Elastic-net modelinde en önemli iki değişken L5 ve F7 değişkenleri olarak belirlenmiştir.



Şekil 4.28. Elastic-net regresyon bağımsız değişkenlerinin önem sıralaması

Lasso regresyon ve elastic-net regresyon yöntemlerinin katsayılarındaki yanlılığı azaltmak amacıyla lasso ve elastic-net yöntemlerince belirlenen sıfırdan farklı katsayıya sahip değişkenler EKK yöntemi ile öz kaynak kârlılığı değişkenine tekrar uyumlandırılmıştır. Lasso ve elastic-net yöntemlerinin aktif bağımsız değişken kümeleri birebir aynı oldukları için tek bir EKK modeli belirlenmiş ve Lasso-EKK olarak adlandırılmıştır. Lasso-EKK modeline ilişkin regresyon katsayılarının tahminleri Tablo 4.26'da yer almaktadır. F testi sonuçlarına ($F(6, 885) = 79,16; p < 0.001$) göre Lasso-EKK modeli istatistiksel olarak anlamlıdır. Buna ek olarak tüm değişkenlerin regresyon katsayıları %5 anlamlılık düzeyinde istatistiksel anlamlılığa sahiptir.

Tablo 4.26. Lasso-EKK modelinin katsayı tahminleri

Değişkenler	Lasso-EKK
Sabit	0,0117
L5	-0,0330*
F7	-0,2011**
F10	-0,3409***
F12	-0,2889*
F13	0,4539**
D3	0,0623***

Anlamlılık düzeyleri: $p < 0.001$: '***', $p < 0.01$: '**', $p < 0.05$: '*'

Lasso, ridge, elastic-net ve lasso-EKK doğrusal regresyon modellerinin değerlendirme ölçütleri Tablo 4.27’de verilmiştir. Buna göre en düşük eğitim hatasına sahip model lasso regresyon ve lasso-EKK modelleridir. Öte yandan test hataları incelendiğinde lasso, ridge ve elastic-net yöntemlerinin genelleştirme performanslarının son derece düşük olduğu sonucuna varılmaktadır. En düşük test hatasına sahip model ise sıfırdan farklı katsayılar üzerindeki kısıtlar kaldırılarak yeniden uyumlandırılan lasso-EKK modelidir.

Tablo 4.27. *Düzenleştirici yöntemlerin değerlendirilmesi*

	Lasso	Ridge	Elastic-Net	Lasso-EKK
Pseudo R²	0,4269	0,4033	0,4264	0,4390
ÇG Eğitim HKO	0,3432	0,3607	0,3411	0,3341
ÇG Test HKO	0,6728	0,6304	0,6666	0,2907

4.4.3. Modellerin karşılaştırılması

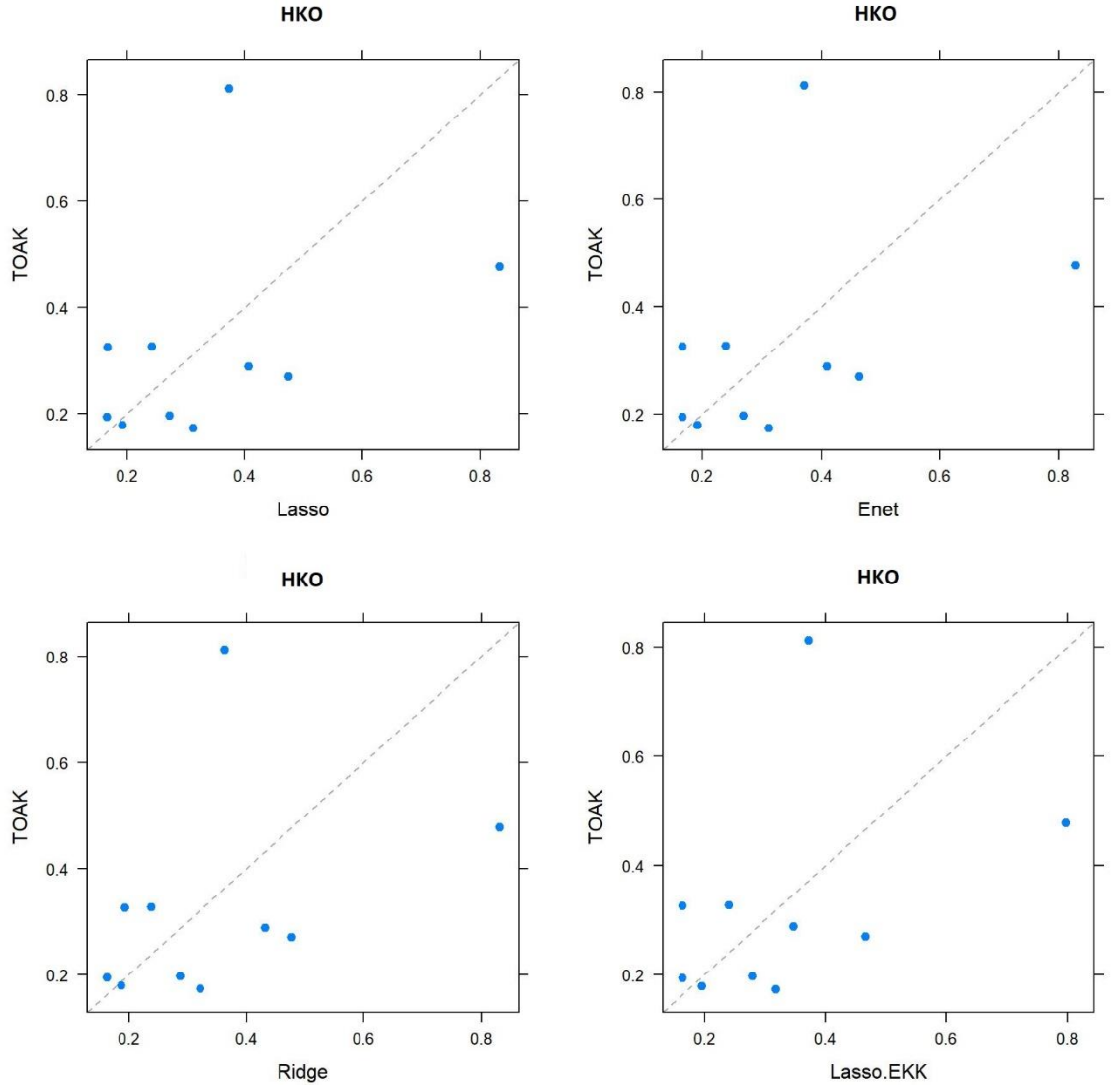
Öz kaynak kârlılık analizi için geliştirilmiş sıradan değişken seçim yöntemleri ve düzenleştirici yöntemlerin performanslarının karşılaştırılmaları yer almaktadır. Elde edilen tüm modellerin genelleştirme performanslarını karşılaştırmak amacıyla tüm modeller test verisine uyumlandırılmış ve her birinin test hataları hesaplanmıştır. Daha önce Tablo 4.24 ve Tablo 4.27’de verilen sonuçlar toplu halde Tablo 4.28’de verilmiştir. Tablo incelendiğinde modellerin eğitim hatalarının birbirlerine çok yakın değerler aldıkları görülmektedir. En düşük eğitim hatasına sahip olan modeller sırasıyla tüm olası alt kümeler, lasso regresyon ve lasso-EKK regresyon modelleridir. Modellerin test hataları incelendiğinde düzenleştirici yöntemler olan lasso, ridge ve elastic-net modellerinin test hatalarının son derece yüksek olduğu görülmektedir. Daha çok sayıda değişkene sahip tüm olası alt kümeler yönteminin eğitim hatası ile karşılaştırıldığında düzenleştirici yöntemlerin eğitim hataları arasında büyük farklılıkların olmadığı görülmektedir. Buna karşın düzenleştirici yöntemlerin tamamı test kümesinde gerçekleştirdiği tahminlerde son derece başarısız bir grafik sergilemişlerdir. Öte yandan lasso-EKK modeli ise en düşük test hatasına sahiptir. Bu sonuca göre lasso regresyon tarafından belirlenen değişkenler EKK ile yeniden uyumlandırıldıklarında azaltılan yanlılığın modelin tahmin başarısı üzerinde önemli bir etkiye sahip olduğu bulgusuna ulaşılmıştır.

Tablo 4.28. Doğrusal regresyon modellerin eğitim ve test hataları

	ÇG Eğitim HKO	ÇG Test HKO	Değişkenler
TOAK	0,3320	0,2971	L4, L5, F1, F4, F6, F7, F10, F11, F12, F13, F15, F16, D3
ILERI	0,3576	0,2938	F1, F4, F7, F10, F11, F12, F13, D3
GERI	0,3459	0,3115	F10
Lasso	0,3432	0,6728	L5, F7, F10, F12, F13, D3
Ridge	0,3607	0,6304	Tüm Değişkenler
Elastic-Net	0,3411	0,6666	L5, F7, F10, F12, F13, D3
Lasso-EKK	0,3443	0,2907	L5, F7, F10, F12, F13, D3

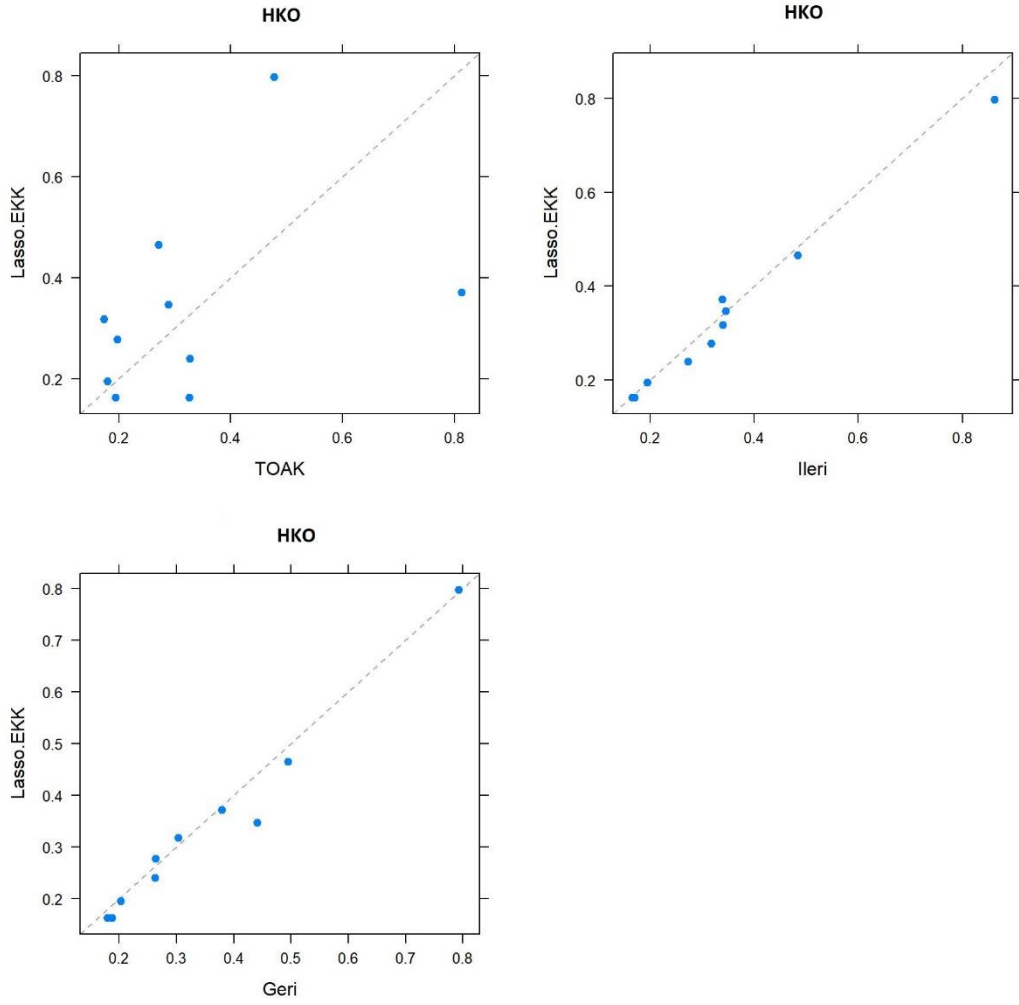
Toplam 31 bağımsız değişken arasından yalnızca tek bir değişkeni modele dahil eden geriye doğru seçim yönteminin elde ettiği bu sonuç etkili gözlemlerin ve çoklu bağıntının mevcut olduğu bir problemde başarılı bir değişken seçimi gerçekleştiremediğini göstermektedir. Sonuç olarak makul sayıda bir aktif değişken kümesi belirleyen lasso regresyon ve EKK yöntemine dayalı olarak tahmin edilen ve en düşük HKO değerini veren lasso-EKK modeli öz kaynak kârlılığı analizi sonucunda en başarılı model olarak belirlenmiştir.

Şekil 4.29’da en düşük eğitim hatasına sahip olan tüm olası alt kümeler modelinin 10 katlı çapraz geçerlilik sonuçları düzenleştirici yöntemlerin sonuçları ile karşılaştırılmıştır. Böylece bu yöntemlerin Tablo 4.28’de verilen çapraz geçerlilik HKO sonuçları arasındaki farklılıkların rassal nedenlerden kaynaklanıp kaynaklanmadığı incelenmesi amaçlanmıştır. Görüldüğü üzere tüm olası alt kümeler modeli ile düzenleştirici yöntemlerinin birbirlerine karşı bir üstünlüğe sahip değillerdir. Tablo 4.28’de verilen eğitim hataları arasındaki farklılığın rassal nedenlerden kaynaklandığı sonucuna varılabilir.



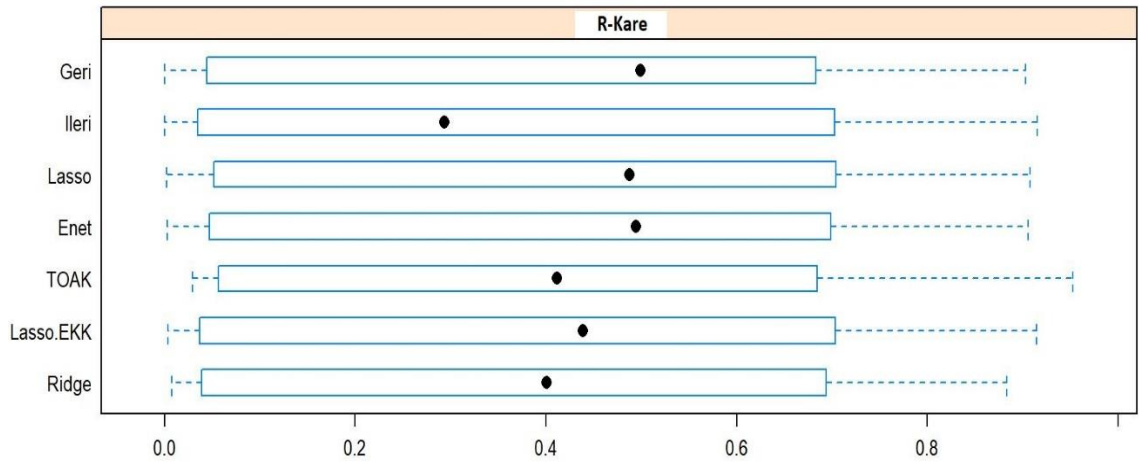
Şekil 4.29. Tüm olası alt kümeler ÇG HKO sonuçlarının düzenleştirici yöntemler ÇG HKO sonuçları ile karşılaştırılması

Lasso-EKK modelinin ileriye doğru seçim, geriye doğru seçim ve tüm olası alt kümeler yöntemleri ile karşılaştırıldığı saçılım grafikleri Şekil 4.30'da verilmiştir. Lasso-EKK yöntemi 10 ÇG katı sonuçları incelendiğinde, tüm olası alt kümeler modeli ile arasındaki farklılığın rassal nedenlerden kaynaklandığı ve ileriye doğru regresyon ve geriye doğru regresyon modellerinden nispeten daha iyi sonuçlara sahip olduğu görülebilir.



Şekil 4.30. Lasso-EKK ÇG HKO sonuçlarının değişken seçim yöntemleri ile karşılaştırılması

Şekil 4.31’de belirlilik katsayılarına ilişkin kutu grafikleri verilmiştir. Yöntemlerin eğitim kümesindeki değişkenliğin açıklama başarıları arasında anlamlı bir farklılığın olmadığı görülmektedir.



Şekil 4.31. Belirlilik katsayılarının kutu grafikleri

Öz kaynak kârlılığının analizi için oluşturulan doğrusal regresyon modellerinin tahminlerin toplam örneklemin %70'nin eğitim ve %30'unun test verisine ayrılması sonucunda gerçekleştirilmiştir. İzleyen kesimde örneklemin 80/20 ve 60/40 eğitim/test kümelerine ayrılması sonucunda modellerin tahminlerine ilişkin elde edilen bulgular incelenecektir. Buna göre 80/20 ayrımı sonrasında (1020 eğitim gözlemi ve 252 test gözlemi) eğitim verisinden yararlanarak çapraz geçerlilik yöntemi ile tahmin edilen modellerin katsayıları Tablo 4.29'da ve 60/40 ayrımı (764 eğitim ve 508 test gözlemi) ile elde edilen modellerin katsayı tahminleri ise Tablo 4.30'da yer almaktadır.

Tablo 4.29'daki sonuçlara göre örneklemin %80'i eğitim kümesine ayrıldıktan sonra ileriye doğru seçim ve geriye doğru seçim yöntemleri yalnızca F10 değişkeninin yer aldığı basit doğrusal regresyon modeli üretmişlerdir. 70/30 ayrımında aktif değişken kümesi 13 olan tüm olası alt kümeler yönteminin 80/20 ayrımında 6 değişkenli bir model tahmin ettiği görülmektedir. Tüm olası alt kümeler yöntemi hiçbir likidite oranına modelde yer vermemiştir. Bunun yanında L4, L5, F7, F13, F15, F16 ve D3 değişkenleri model dışında bırakılmış ve D8 değişkeni modele eklenmiştir.

Öte yandan düzenlileştirici yöntemler incelendiğinde lasso ve elastic-net yöntemlerinin aktif değişken kümeleri 70/30 ayrımında olduğu gibi aynıdır. Lasso ceza kısıtı $s_{lasso} = 0,0203$ iken elastic-net modeli için $\alpha = 0,4$ ve $s_{enet} = 0,0506$ parametreleri optimum olarak belirlenmiştir. Ceza kısıtı değerindeki farklılıktan dolayı iki modelin katsayı tahminleri arasında farklılıkla görülmektedir. Lasso yönteminde belirlenen sıfırdan farklı katsayıya sahip değişkenler EKK yöntemi ile tekrar uyumlandırıldıktan sonra elde edilen Lasso-EKK modelinin katsayı tahminleri tablonun en sağ sütununda yer almaktadır. Ridge regresyon modelinin ceza kısıtı $s_{ridge} = 0,1617$ olarak belirlenmiştir.

Tablo 4.29. 80/20 eğitim/test ayrımı sonrası öz kaynak kârlılığı modellerinin katsayı tahminleri

	TOAK	İleri	Geri	Lasso	Enet	Ridge	EKK-Lasso
Sabit	0,0200	0,0200	0,0200	0,2226	0,2206	0,1869	0,0200
L1						-0,0021	
L2						0,0018	
L3						0,0135	
L4						0,0224	
L5				-0,1146	-0,0924	-0,2420	-0,0319**
L6						-0,0001	
L7						-0,0531	
L8						0,0397	
F1	0,7009***					0,0536	
F3						-0,0011	
F4	-0,6067***					-0,1601	
F6	-0,2367***					-0,0967	
F7				-0,0708	-0,0967	-0,0913	-0,1407*
F8						-0,0003	
F9						-0,0082	
F10	-0,5095***	-0,4833***	-0,4833***	-0,1279	-0,1004	-0,0815	-0,3725***
F11	0,1062***					0,0152	
F12	0,1354***			-0,0561	-0,0537	-0,0692	-0,0213
F13				0,1044	0,0722	0,2323	0,0486***
F14						-0,0685	
F15						0,0420	
F16						0,0599	
F17						0,0317	
D1						-0,0001	
D2						-0,0001	
D3				0,0446	0,0421	0,0445	0,0633***
D4						3×10^{-5}	
D5						-0,0005	
D6						3×10^{-6}	
D7						-0,0175	
D8	0,0625***					0,0723	

Anlamlılık düzeyleri: $p < 0.001$: '***', $p < 0.01$: '**', $p < 0.05$: '*'

Tablo 4.30’da verilen örneklemin %60’ının eğitim ve %40’ının test kümesine ayrılması sonrasında elde edilen modellerin katsayı tahminleri incelendiğinde ileriye doğru seçim ve geriye doğru seçim modellerinin yalnızca F7 değişkeninin yer aldığı basit doğrusal regresyon modeli tahminledikleri görülmektedir. Tüm olası alt kümeler regresyon modelinde ise toplam 5 bağımsız değişken yer almaktadır.

Düzenleştirici yöntemlerden lasso regresyonun ceza kısıtı $s_{lasso} = 0,0304$, ridge regresyonun $s_{ridge} = 0,2223$ ve elastic-net regresyonun optimum parametreleri $\alpha = 0,4$ ve $s_{enet} = 0,0708$ olarak elde edilmiştir. Lasso regresyon modelindeki sıfırdan farklı katsayı tahminlerine sahip değişken kümesi 80/20 ve 70/30 ayrımlarında elde edilen kümeden farklılık göstermektedir. Buna göre L5 değişkeni modelden çıkarılarak 5 bağımsız değişkenli bir doğrusal regresyon modeli tahminlenmiştir. Lasso yöntemi ve elastic-net yöntemleri ile belirlenen bağımsız değişkenlerin EKK yöntemi ile tekrar uyumlandırılması sonucunda elde edilen Lasso-EKK ve Enet-EKK modellerinin katsayı tahminleri de Tablo 4.30'da yer almaktadır.

Tablo 4.31'de öz kaynak kârlılığının modellenmesi için kullanılan tüm modeller farklı eğitim/test ayrımlarına göre karşılaştırılmış ve modellerin eğitim ve test hataları ile değişken sayılarına yer verilmiştir. Sıradan değişken seçim yöntemlerini ele aldığımızda eğitim kümesinin hacmi değiştikçe tüm olası alt kümeler yönteminin aktif bağımsız değişken kümesinde önemli değişimler gözlemlenmektedir. 80/20 ayrımlarında 7, 70/30 ayrımlarında 13 ve 60/40 ayrımlarında 5 bağımsız değişkeni seçen tüm olası alt kümeler yöntemi eğitim kümesinde gerçekleşen değişimlerden büyük oranda etkilenmekte ve tutarsız sonuçlar vermektedir. Benzer şekilde ileriye doğru seçim yöntemi de 70/30 ayrımlarında 8 değişkenli bir model elde etmiş iken diğer iki ayrımda tek değişkenli regresyon modeli tahmin etmiştir.

Tablo 4.30'daki lasso ve elastic-net regresyonları ise değişken seçimi açısından en tutarlı yöntemlerdir. 80/20 ve 70/30 ayrımlarında aynı aktif değişken kümelerine sahip lasso ve elastic-net modelleri 60/40 ayrımlarında 5 değişkenli ve büyük ölçüde benzer bağımsız değişken kümelerine sahiptirler.

Eğitim hataları incelendiğinde tüm modellerin eğitim hatalarının birbirlerine yakın değerler aldıkları görülmektedir. Lasso, ridge ve elastic-net modelleri en yüksek test hatalarına ve böylece en düşük genelleştirme performansına sahip modellerdir. Test hataları incelendiğinde 80/20 ve 60/40 ayrımlarında Lasso-EKK yöntemi ve 70/30 ayrımlarında ise tüm olası alt kümeler yöntemi en iyi genelleştirme performanslarına sahiptirler. Ancak değişken sayıları incelendiğinde lasso-EKK yöntemi eğitim kümesindeki değişimlerden en az etkilenen yöntem olduğu sonucuna varılır. Geriye doğru seçim yöntemi her üç ayrımda da basit doğrusal regresyon modelleri verirken; ileri doğru seçim ve geriye doğru seçim yöntemleri eğitim kümesinin hacmine ve eğitim

kümesindeki gözlemlerdeki rassal değişimlere karşı son derece kararsız davranışlar sergilemişlerdir.

Tablo 4.30. 60/40 eğitim/test ayrımı sonrası öz kaynak kârlılığı modellerinin katsayı tahminleri

	TOAK	İleri	Geri	Lasso	Enet	Ridge	Lasso-EKK	Enet-EKK
Sabit	0,0191	0,0191	0,0191	0,1690	0,2419	-0,1307	0,0191	0,0191
L1						-0,0052		
L2						0,0051		
L3						0,0356		
L4						0,2579		
L5						-0,4438		
L6						0,0022		
L7						0,0245		
L8						0,1224		
F1	0,4993***					0,2266		
F3						0,0015		
F4	-0,5108***					-0,5760		
F6						0,0707		
F7	-0,5735***	-0,4887***	-0,4887***	-0,2426	-0,1359	-0,1491	-0,5383***	-0,4568***
F8						-0,0003		
F9						-0,0171		
F10					-0,0763	-0,0851		-0,0656
F11	0,1107***					0,1327		
F12	0,2162***			-0,0277	-0,0837	0,0482	-0,0147	-0,0344*
F13				0,1529	0,1124	0,5564	0,0726***	0,0660***
F14						-0,1258		
F15						-0,0761		
F16						0,2197		
F17				0,0959		0,0740	0,0537***	
D1						-0,0027		
D2						-0,0002		
D3				0,0305	0,0325	0,0624	0,0565***	0,0664***
D4						3×10^{-5}		
D5						-0,0003		
D6						-0,0013		
D7						-0,0181		
D8						0,0821		

Anlamlılık düzeyleri: p<0.001: '***', p<0.01: '**', p<0.05: '*'

Tablo 4.31. 80/20, 70/30 ve 60/40 eğitim/test ayrımlarına ilişkin değerlendirme sonuçları

	80/20			70/30			60/40		
	Eğitim HKO	Test HKO	Değişken Sayısı	Eğitim HKO	Test HKO	Değişken Sayısı	Eğitim HKO	Test HKO	Değişken Sayısı
TOAK	0,3463	0,3281	7	0,3320	0,2971	13	0,3526	0,3375	5
ILERI	0,3692	0,3359	1	0,3576	0,2938	8	0,3572	0,3234	1
GERI	0,3692	0,3359	1	0,3459	0,3115	1	0,3572	0,3234	1
Lasso	0,3518	0,6933	6	0,3432	0,6728	6	0,3483	0,6678	5
Ridge	0,3572	0,6742	31	0,3571	0,6743	31	0,3582	0,6424	31
Elastic-Net	0,3438	0,6847	6	0,3411	0,6666	6	0,3408	0,6567	5
Lasso-EKK	0,3524	0,3096	6	0,3431	0,3096	6	0,3565	0,3186	5
Enet-EKK	-	-	-	-	-	-	0,3519	0,5605	5

Lasso ve ridge regresyon modellerinin test HKO değerleri eğitim HKO değerlerinin yaklaşık iki katına eşittir. Test hatasının eğitim hatasından bu kadar yüksek olması iki sebeple açıklanabilir. Bunlardan ilki 1. Bölüm’de açıklanan ve Şekil 1.5’te gösterilen aşırı uyum sorunudur. Model karmaşıklığı (değişken sayısı) arttıkça yanlılığın artması ve varyansın azalması sonucunda test hatasında eğitim hatasına kıyasla orantısız bir artışın gözlemlenmesi beklenir. Öte yandan ridge regresyon 31 değişkene ve lasso regresyon 6 değişkene sahip olmasına rağmen test hataları birbirlerine oldukça yakın değerler almışlardır. Bununla birlikte lasso regresyondan daha fazla değişkene sahip bazı modellerin test hatalarının fazla yüksek olmadıkları hatta tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemleri ile tahmin edilen tüm modellerin test hatalarının teorik beklentinin tersine bir biçimde eğitim hatalarından daha düşük olarak elde edildikleri Tablo 4.31’de görülmektedir. Sonuç olarak elde edilen HKO değerlerindeki beklenmedik bu değerlerin mevcudiyetinin model karmaşıklığına bağlı olarak ortaya çıkan aşırı uyum ya da eksik uyum sorunlarından herhangi birisinin olmadığını söyleyebiliriz.

Bu beklenmedik sonuçların ikinci muhtemel açıklaması ise aykırı gözlemlerin yer aldığı veri kümesindeki gözlemlerin rassal olarak eğitim ve test kümeleri biçiminde ayrılmaları sonucunda zor uyumlanabilen (öğrenilebilen) aykırı gözlemlerin eğitim kümesinde, daha kolay tahminlenebilen gözlemlerin ise test kümesinde yer almasıdır.

Diğer bir ifade ile rassal olarak gerçekleştirilen eğitim/test ayrımlarında aykırı gözlemlerin çoğunun eğitim kümesinde yer alması sonucunda TOAK, ILERI ve GERI modelleri test HKO değerlerinden daha düşük eğitim HKO değerleri üretmişlerdir. Dolayısıyla bu modeller kötü uyumlanmış modellerdir. Lasso ve ridge regresyon modelleri ise teorik beklentiye uygun olarak eğitim hatalarından daha yüksek test hatalarına sahiplerdir. Ancak yanlış tahminlerde bulunmaları sebebiyle son derece başarısız genelleme performansı göstermişlerdir.

Öz kaynak kârlılığı oranının sıradan değişken seçim yöntemleri ve düzenlileştirici yöntemlere dayalı olarak incelendiği bu bölümde lasso regresyon ve elastic-net regresyon yöntemlerinin eğitim kümesindeki değişimlere karşı en kararlı ve seyrek değişken seçimi gerçekleştirildiği sonucuna varılmıştır. Buna göre öz kaynak kârlılığı doğrusal regresyon modeli ile tahmin etmek amacıyla lasso-EKK modeli en uygun model olarak belirlenmiştir. Lasso-EKK doğrusal regresyon modeli eşitlik 4.2’de verilmiştir.

$$\begin{aligned}
 ROE = & 0,117 - 0,033 L5 - 0,201 F7 - 0,341 F10 - 0,289 F12 \\
 & (0,95) \quad (-2,50) \quad (-3,03) \quad (-5,11) \quad (-2,10) \\
 & + 0,454 F13 + 0,062 D3 \\
 & (3,20) \quad (4,91)
 \end{aligned} \tag{4.2}$$

Lasso-EKK doğrusal regresyon modelinde likidite oranlarından L5 (stoklar/toplam varlıklar oranı); finansal yapı oranlarından F7 (maddi duran varlıklar/öz kaynaklar oranı), F10 (duran varlıklar/öz kaynaklar oranı), F12 (kısa vadeli yabancı kaynaklar/yabancı kaynaklar toplamı oranı) ve F13 (banka kredileri/toplam varlıklar oranı); devir hızlarından ise yalnızca D3 (çalışma sermayesi devir hızı) oranı yer almıştır. Modelde yer alan tüm katsayılar öz kaynak getirisi ile %5 düzeyinde anlamlı ilişkiye sahiplerdir.

Modelde yer alan finansal oranlar arasından mutlak değerce en yüksek değere sahip finansal oran banka kredileri/toplam varlıklar oranıdır (F13). Buna göre işletmelerin banka kredilerinin artması öz kaynak kârlılığı üzerinde pozitif bir etkiye sahiptir. Modelde yer alan diğer finansal yapı oranları olan maddi duran varlıklar/öz kaynaklar oranı (F7) ile duran varlıklar/öz kaynaklar oranının (F10) sahip oldukları negatif değerli katsayılar imalat işletmelerinin duran varlıklarının artmasının kârlılığı olumsuz etkilediğini göstermektedir. Ayrıca borçluluk içinde kısa vadeli borçların payının artması öz kaynak kârlılığını olumsuz etkilemektedir. Bununla beraber stokların toplam varlıklar içindeki payı (L5) arttıkça öz kaynak kârlılığı azalmaktadır. Faaliyet oranları arasından

modelde yalnızca alıřma sermayesi devir hızı oranı yer almıřtır. Buna gre imalat iřletmeleri alıřma sermayesini satıřlara dnřtrme oranını artırdıklarında z kaynak getirilerini artırmaktadırlar.

SONUÇ

Bu çalışmada doğrusal regresyon modeli için katsayı tahminlerini ve değişken seçimini eş zamanlı olarak gerçekleştiren düzenleyici yöntemler ile geleneksel değişken seçim yöntemlerinin gerçek bir işletme problemindeki performanslarının karşılaştırmalı olarak incelenmesi ve bu yöntemlerin güçlü ve zayıf yönlerinin belirlenmesi amaçlanmıştır. Bu amaçla öncelikle bir eğitim kümesinden elde edilen istatistiksel modellerin tahmin başarılarının bir test kümesinde değerlendirilmesi ilkesinin temelinde literatürde istatistiksel öğrenme alanına dayalı olduğu ve bu alanda kullanılan temel yaklaşımlar ile kavramlar ele alınmış, daha sonra doğrusal regresyon analizi, en küçük kareler yöntemi ve geleneksel değişken seçim yöntemleri olan tüm olası alt kümeler yöntemi, ileriye doğru seçim ve geriye doğru seçim yöntemlerinin özellikleri kısaca açıklanmıştır. Bunu takiben literatürde yaygın olarak kullanılan ve ceza parametrelerine sahip birer yanlı regresyon yöntemi olan düzenleyici yöntemlerden ridge, lasso ve elastic-net yöntemleri kısaca açıklanmıştır. Nihai olarak gerçek bir işletme problemi olan ve literatürde yaygın olarak araştırılan firma performansını etkileyen finansal oranların belirlenmesi, belirlenen finansal oranlar ile firma performansının tahmin edilmesi ve aralarındaki matematiksel ilişkinin yapısını ortaya koymak için ele alınan yöntemler kullanılmış ve elde edilen bulgular sunulmuştur.

Farklı sektörlerin (bankacılık, teknoloji, sigortacılık, hizmet gibi) finansal raporlarındaki kalemlerin birbirlerinden önemli ölçüde farklılıklar göstereceği göz önüne alınarak çalışmaya yalnızca hisseleri Borsa İstanbul'da işlem gören imalat işletmeleri dahil edilmiştir. 2012-2019 yılları arasında Borsa İstanbul'da işlem gören imalat firmalarının yıl sonu bilanço ve gelir tablolarından elde edilen veri kullanılarak TCMB'nin önerdiği formüller yardımıyla likidite, finansal yapı, faaliyet ve kârlılık oranları hesaplanmıştır. Firma performansının bir ölçütü olarak literatürde yaygın olarak kârlılık oranlarından yararlanıldığı açıklanmış ve birer kârlılık oranı olan aktif kârlılık oranı (ROA) ve öz kaynak kârlılık oranı (ROE) firma performansının ölçütü ve araştırmanın bağımlı değişkenleri olarak kabul edilmiştir. Araştırmanın bağımsız değişkenleri ise 33 adet likidite, finansal yapı ve faaliyet oranlarından oluşmaktadır.

Araştırmada sekiz yıla ait farklı büyüklüklerdeki (81'i küçük, 289'u orta, 902'si büyük ölçekli) toplam 1272 firmaya ait veriden yararlanılmıştır. Ön analizde gösterildiği üzere farklı büyüklükte ve farklı imalat alt sektörlerinde (ana metal, gıda, kâğıt, kimya,

metal eşya, orman, taş ve tekstil) faaliyet gösteren işletmelerin bir arada araştırmaya dahil edilmesi veri kümesinde birçok aykırı değer var olmasına ve işletme büyüklüğünü gösteren toplam satışlar ve toplam varlıkların normalden uzak dağılımlarına neden olmuştur. Ayrıca birçok finansal oranın pay veya payda değerlerinin ortak olması sonucunda beklendiği üzere bazı finansal oranlar yüksek korelasyona ve tahmin edilen regresyon modellerinde çoklu bağıntının olmasına yol açmıştır. İncelenen 33 bağımsız değişkenden 2'si diğer finansal oranlarla negatif tam korelasyona dahil oldukları için analizden çıkarılmıştır.

Bağımlı değişkenler olan aktif getiri ve öz kaynak getiri oranları incelendiğinde Borsa İstanbul'da faaliyet gösteren imalat işletmelerinin kârlılıklarının son derece düşük olduğu, aktif getiri oranının ortalamasının %4 ve öz kaynak getiri oranının ortalamasının ise %2 olduğu görülmektedir. Bununla beraber aktif getiri oranının normale daha yakın bir dağılıma sahip olduğu ancak öz kaynak getiri oranının normalden uzak bir dağılıma sahip olduğu görülmektedir.

Ele alınan regresyon yöntemleri ile modellerin katsayılarını tahmin etmek ve düzenleştireci yöntemlerin hiperparametre tahminlerini gerçekleştirmek amacıyla kullanılmak üzere verinin %70'i eğitim kümesine; elde edilen regresyon modellerinin tahmin başarılarını ölçmek amacıyla kullanılmak üzere verinin %30'u test kümesine ayrılmıştır. Böylece 892 gözlem eğitim verisinde ve 380 gözlem test verisinde yer almıştır.

İlk olarak 31 finansal oranın bağımsız değişken olarak modele dahil edildiği aktif kârlılık analizinde tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemleriyle değişken seçimleri gerçekleştirilmiş ve nihai bağımsız değişken kümeleri belirlendikten sonra değişkenler EKK yöntemi ile uyumlandırılmıştır. Sonuç olarak tüm olası alt kümeler yöntemi ile 4, ileriye doğru seçim yöntemi ile 5 ve geriye doğru seçim yöntemi ile 6 bağımsız değişkenli modeller elde edilmiştir. Elde edilen her üç model de istatistiksel olarak anlamlı modellerdir. Modellerin değişken kümeleri birbirlerinden farklılık göstermiş ve yalnızca D8 finansal oranı her üç modelde de yer almıştır. Modellerin eğitim ve test HKO değerleri birbirlerine oldukça yakın olarak elde edilmiştir. İleriye doğru seçim ile elde edilen doğrusal regresyon modeli en yüksek eğitim HKO'suna (0,724) sahip olsa da en düşük test HKO değerine (0,797) değerine sahip olduğu için en iyi genelleme performansına sahip model olarak belirlenmiştir.

Aktif kârlılık bağımlı değişkeni için doğrusal regresyon modelini tahmin etmek için lasso, ridge ve elastic-net tahmincileri kullanılmıştır. Bu amaçla öncelikle 10 katlı çapraz geçerlilik yöntemi ile ridge, lasso ve elastic-net yöntemleri için λ parametresinin duali olan s hiperparametre değerleri ($s_{lasso} = 0,0102$, $s_{ridge} = 0,0203$ ve $s_{enet} = 0,0102$) ve elastic-net için ayrıca optimum α hiperparametre değeri ($\alpha = 0,3$) belirlenmiştir. Ridge regresyon analize dahil edilen tüm değişkenlere nihai modelde yer verirken, lasso ve elastic-net yöntemleri sırasıyla 7 ve 15 değişkenli modeller tahmin etmiştir. s parametre değerleri birbirlerine eşit olan lasso ve elastic-net yöntemlerinin α parametre değerleri birbirlerinden farklı olduğu için (lasso için $\alpha = 0$ 'dır.) iki modelde yer alan değişken sayıları farklılık göstermektedir. Eşit s kısıt değerlerine sahip bu iki yöntemden elastic-net yönteminin α hiperparametre değeri 1'e yakın değer alması sonucunda teorik beklentiye uygun olarak lasso tahmincisi daha seyrek bir model tahmin etmiştir. Lasso, ridge ve elastic-net modellerinin tamamının test hata değerleri (sırasıyla 0.1021, 0.1067 ve 0.1066) ileriye doğru seçim yöntemine göre daha yüksek olarak elde edilmiştir. Sonuç olarak lasso, ridge ve elastic-net modelleri aktif kârlılığı tahmin başarısı ileriye doğru seçim modeline göre daha düşüktür.

Düzenleştirici yöntemlerin katsayı tahminleri sıfıra doğru yanlı tahminler olduğundan dolayı Hastie vd.(2008, s. 91), belirlenen sıfırdan farklı katsayılarla sahip bağımsız değişkenlerin EKK yöntemi ile tekrardan uyumlandırmanın bu yanlılığı azaltacağını belirtmişlerdir. Bu amaçla lasso ve elastic-net yöntemlerince belirlenen bağımsız değişkenler EKK yöntemince tekrar uyumlandırılmış ve Lasso-EKK ve ElasticNet-EKK olarak adlandırdığımız modeller elde edilmiştir. Elde edilen sonuçlara göre Lasso-EKK ve ElasticNet-EKK çapraz geçerlilik eğitim HKO ve test HKO değerlerinde önemli ölçüde azalma görülmüştür. Lasso test HKO değeri 0.1021'den 0,0799'a ve elastic-net test HKO değeri 0.1067'den 0.0800'e düşmüştür. Sonuç olarak lasso ve elastic-net yöntemleri ile elde edilen değişkenlerin EKK yansız tahmincisi ile tekrar uyumlandırılmasının tahmin başarısını artırdığı sonucuna varılmıştır.

Lasso-EKK 10 farklı çapraz geçerlilik katındaki eğitim HKO sonuçları tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemlerinin sonuçları ile karşılaştırılmış ve yöntemlerin birbirlerine kesin bir üstünlüklerinin olmadığı ve test HKO değerleri arasındaki farklılıkların rassal nedenlerden kaynaklandığı sonucuna ulaşılmıştır.

İdeal eğitim/test ayırımının hangi oranda olması gerektiği konusunda literatürde net bir uzlaşma olmaması sebebiyle ve modellerin tahmin edilmesinde kullanılan eğitim verisi kümesinde gerçekleşen artış ve azalışların yöntemlerin seçtiği değişkenler üzerindeki etkilerini gözlemleyebilmek adına tahmin edilen tüm modeller 80/20 ve 60/40 eğitim/test ayırımına göre tekrar tahmin edilmiştir. Elde edilen sonuçlara göre Lasso-EKK modeli eğitim kümesindeki değişime en az duyarlı ve en kararlı yöntem olarak belirlenmiştir. Ayrıca lasso-EKK modeli istikrarlı biçimde tüm veri ayırımlarında en düşük test hatasına sahip modeller arasında ilk iki arasında yer almıştır.

Sonuç olarak elde edilen aktif kârlılık modelinde nakit oranı (L3), stoklar/dönen varlıklar oranı (L4), kaldıraç oranı (F1), duran varlıklar/devamlı sermaye oranı (F11), dönen varlıklar/toplam varlıklar oranı (F16), maddi duran varlıklar/toplam varlıklar oranı (F17) ve D8 (aktif devir hızı) oranı modelde yer almıştır. Finansal yapı oranlarından duran varlıklar/devamlı sermaye oranı (F11), dönen varlıklar/toplam varlıklar oranı (F16) ve maddi duran varlıklar/toplam varlıklar oranı (F17) modelde yer almalarına rağmen aktif kârlılık ile %5 anlamlılık düzeyinde istatistiksel olarak anlamlı bir ilişkiye sahip değildir.

Aktif devir hızı tüm yöntemler tarafından seçilen tek finansal orandır. Kaldıraç oranı mutlak değerce en yüksek katsayı değerine (-0,04) sahiptir. Buradan toplam borçluluğun sermaye yapısındaki oranı arttıkça aktif kârlılığın azaldığı sonucuna varılmaktadır. Öte yandan nakit oranının ve aktif devir hızının artması imalat işletmelerinin aktif kârlılığını artırırken, stok/dönen varlıklar oranının artması aktif kârlılıklarını azaltmaktadır.

Yapılan analizlerden imalat işletmelerinin aktif kârlılıklarını artırmak için varlıklarını kullanarak satışlarını artırmalarının önemli olduğu sonucuna varılmıştır. Bununla beraber işletmeler dönen varlıkları içerisindeki nakit miktarını artırmaları ve stok miktarlarını azaltmaları aktif kârlılığı olumlu yönde etkilemektedir. Öte yandan işletmenin sermaye yapısı içerisinde toplam borcun payının artması aktif kârlılığı negatif yönde etkilemektedir.

Öz kaynak kârlılığı ile ilgili tanı grafikleri ve tanımlayıcı istatistikleri incelendiğinde öz kaynak kârlılığının normalden uzak bir dağılıma sahip olduğu ve örnekleme çok sayıda aykırı değer bulunduğunu tespit edilmiştir. Elde edilen bulgulara göre tüm olası alt kümeler yönteminde 13, ileriye doğru seçim yönteminde 8 bağımsız

değişken seçilirken, geriye doğru seçim yöntemi ile tek bir değişken seçilmiş ve basit doğrusal regresyon modeli tahmin edilmiştir. 31 adet finansal oran arasından yalnızca tek bir değişkeni modele dahil etmesi ve modelde tek bir değişkenin yer almasına rağmen belirlilik katsayısının %64 olması veride yer alan aykırı değerlerden kaynaklandığı sonucuna varılmaktadır. Ayrıca her üç yöntemin test hata değerleri teorik beklentinin aksine eğitim hata değerlerinden daha düşük olarak elde edilmiştir. Bu beklenmedik sonuç eğitim/test ayrımı sonucunda eğitim kümesinde regresyon yöntemi için öğrenmesi zor aykırı değerlerin yer alması ve test kümesinde göreceli olarak ortalamaya daha yakın gözlemlerin yer almasından kaynaklanmaktadır.

Lasso ($s=0,0203$), ridge ($s=0,0304$) ve elastic-net ($\alpha = 0,5$, $s=0,0405$) tahmincileri ile öz kaynak kârlılık modelleri tahmin edilmiştir. Ridge regresyon modelinde 31 finansal oranın tamamı yer alırken, lasso ve elastic-net yöntemlerinde eşit sayıda ve birebir aynı bağımsız değişkenler seçilmiştir. Lasso ve elastic-net modellerinin hiperparametre değerleri birbirlerinden farklı oldukları için tahmin edilen modellerin katsayı tahminleri birbirlerinden farklılık göstermiştir. Her üç modelin de test HKO değerleri eğitim HKO değerlerinden oldukça yüksektir. Elde edilen sonuçlar teorik beklentiye uygun olsalar da modellerin tahmin başarılarının son derece düşük olduğu sonucuna ulaşılmıştır. Bununla birlikte lasso ve elastic-net yöntemlerince belirlenen 6 değişkenin EKK tahmincileri ile yeniden uyumlandırılması sonucunda Lasso-EKK modeli tahmin edilmiş ve sonuç olarak test HKO değeri 0,6728'den 0,2907'ye düşmüştür. Sonuç olarak düzenleştirici yöntemler arasından en düşük test hatasına sahip model lasso-EKK modelidir.

Lasso-EKK modelinin eğitim kümesindeki 10 katlı çapraz geçerlilik sonuçları tüm olası alt kümeler, ileriye doğru seçim ve geriye doğru seçim yöntemlerinin çapraz geçerlilik HKO sonuçları ile karşılaştırılmış ve eğitim hatalarının aralarındaki farkın rassal nedenlerden kaynaklandığı ve lasso-EKK yönteminin sıradan değişken seçimi ile belirlenen modellere tahmin başarısı açısından bir üstünlüğe sahip olmadığı sonucuna ulaşılmıştır.

Aktif kârlılık analizinde olduğu gibi öz kaynak kârlılığı analizinde 70/30 eğitim/test ayrımının yanında 80/20 ve 60/40 ayrımlarına göre de doğrusal regresyon modelleri ele alınan yöntemlerce tahmin edilmiştir. Elde edilen sonuçlara göre sıradan değişken seçim yöntemlerinin eğitim kümesinin hacmi değiştikçe seçtikleri bağımsız değişken kümelerinin birbirlerinden farklılık gösterdiği görülmektedir. Buna karşın lasso ve

elastic-net yöntemleri ise en istikrarlı değişken seçimini gerçekleştirmişlerdir. Ayrıca lasso-EKK modeli 80/20 ve 60/40 ayrımlarında en düşük test HKO değerlerini veren model olmuştur. Tüm bu sonuçlar göz önüne alınarak öz kaynak kârlılığını en iyi tanımlayan modelin lasso-EKK modeli olduğu sonucuna varılmıştır.

Öz kaynak getiri oranının bağımlı değişken olarak kabul edildiği Lasso-EKK doğrusal regresyon modelinde stoklar/toplam varlıklar oranı, maddi duran varlıklar/öz kaynaklar oranı, duran varlıklar/öz kaynaklar oranı, kısa vadeli yabancı kaynaklar/yabancı kaynaklar toplamı oranı, banka kredileri/toplam varlıklar oranı ve çalışma sermayesi devir hızı yer almıştır. Modelde yer alan değişkenlerden banka kredileri/toplam varlıklar oranı ve çalışma sermayesi devir hızı değişkenleri öz kaynak getiri oranı ile pozitif yönlü bir ilişkiye sahip iken diğer finansal oranlar negatif yönlü bir ilişkiye sahiptir. Bu bulgulara göre imalat firmaları öz kaynak kârlılıklarını artırmak için stok ve maddi duran varlık miktarlarını azaltmaları gerekmektedir. Ayrıca imalat işletmeleri borç yüklerini kısa vadeli yabancı kaynaklar ile karşıladıkça öz kaynak kârlılıkları bundan olumsuz etkilemektedir. Bununla birlikte banka kredilerinin kullanımı ve çalışma sermayesinin etkin kullanımı öz kaynak kârlılığını artırmaktadır.

Elinizdeki bu çalışma diğer birçok çalışmada da olduğu gibi birtakım kısıtlara sahiptir. Öncelikle bu çalışmada ele alınan gerçek bir işletme problemine ilişkin bulgular ve yorumlar yalnızca Türkiye'deki imalat firmalarına ilişkindir. Araştırmada firmaların kamu aydınlatma platformunda yer alan finansal raporlarından elde edilen ikincil veri kullanılmıştır. Veri girişlerinde gerçekleştirilen hatalar finansal oranların hesaplanmasını ve bunların analiz edilmesini hiç şüphesiz olumsuz etkileyecektir. Bir diğer kısıt ise çalışmada yıl sonu bilanço verilerinin kullanılmış olmasıdır. Bilindiği üzere işletmelerin bilanço kalemleri günbegün farklılık göstermektedir. Bu nedenle yıl içerisinde sadece tek bir güne ait verinin kullanılması ile dönem içindeki değişimler göz ardı edilmiştir. Buna ek olarak çalışmada firma performansının ölçülmesinde yalnızca finansal veriden yararlanılması çalışmanın bir diğer kısıtıdır. Son olarak toplamsal ve doğrusal yapıya sahip doğrusal regresyon modellerinde var olan kısıtlar da bu araştırmanın kısıtları arasındadır. Örneğin aktif kârlılık analizinde borçluluk oranının aktif kârlılığı negatif etkilediği sonucuna ulaşılmıştır. Ancak literatürden bilindiği üzere borçluluk bir seviyeye kadar işletmelerin kârlılıklarını olumlu etkilerken bir noktadan sonra işletmeleri riskli hale getirdiği için olumsuz etkilemektedir. Dolayısıyla bu tür doğrusal olmayan ilişkiler kullanılan modellerin varsayımları gereği analizlerde belirlenememiştir.

Sonu olarak dzenliŒtirici yntemlerin deęiŐken seiminde sıradan deęiŐken seim yntemlerine gre daha kararlı davrandıkları gzlemlenmiŐtir. Bununla birlikte tahmin baŐarılarının sıradan deęiŐken seim yntemlerine gre bariz bir stnlę bulunmadığı sonucuna ulaŐılmıŐtır. Ancak yanlı tahmincilere sahip dzenliŒtirici yntemlerin belirledikleri deęiŐkenler EKK yntemi ile tekrar uyumlandırıldıklarında elde edilen modellerin tahmin baŐarılarının nemli lde arttığı sonucuna varılmıŐtır. alıŐmada ok boyutlu ve kt konumlanmış bir iŐletme problemi ele alınmıŐ ve hangi finansal oranların firma performansının belirleyicileri olduęuna karar verilmesinde dzenliŒtirici yntemlerin daha kararlı davrandıkları bulgusuna ulaŐılmıŐtır.

KAYNAKÇA

- Abor, J. (2005). The effect of capital structure on profitability: an empirical analysis of listed firms in Ghana. *The Journal of Risk Finance*, 6(5), 438-445. doi:10.1108/15265940510633505
- Abu-Mostafa, Y. S., Magdon-Ismail, M. ve Lin, H.-T. (2012). *Learning from data: A short course* (1. bs.). AMLbook.
- Akgün, A. I. ve Memiş Karataş, A. (2021). Investigating the relationship between working capital management and business performance: Evidence from the 2008 financial crisis of EU-28. *International Journal of Managerial Finance*, 17(4), 545-567.
- Akpınar, H. (2014). *Data: Veri madenciliği veri analizi* (1. bs.). İstanbul: Papatya Yayıncılık Eğitim.
- Aktaş, R. (1997). *Mali başarısızlık (işletme riski) tahmin modelleri*. Ankara: Türkiye İş Bankası Kültür Yayınları.
- Alpar, R. (2013). *Uygulamalı çok değişkenli istatistiksel yöntemler* (4. bs.). Ankara: Detay.
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589. doi:10.2307/2978933
- Al-Zararee, A., Al-Sawalhah, J. ve Samara, A. (2021). The relationship between the return on equity (ROE) and the capital structure of the Jordanian public shareholding industrial companies. *Academy of Accounting and Financial Studies Journal*, 25(1), 1-10.
- Avcı, E. (2016). Capital structure and firm performance: An application on manufacturing industry. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 38(1), 15-30.
- Aydın, D. (2014). *Uygulamalı regresyon analizi* (1. bs.). Ankara: Nobel.
- Backer, M. ve Gosman, M. L. (1980). The use of financial ratios in credit downgrade decisions. *Financial Management*, 9(1), 53-56. doi:10.2307/3665313
- Baños-Caballero, S., García-Teruel, P. J. ve Martínez-Solano, P. (2014). Working capital management, corporate performance, and financial constraints. *Journal of Business Research*, 67(3), 332-338. doi:10.1016/j.jbusres.2013.01.016

- Barnes, P. (1987). The analysis and use of financial ratios: A review article. *Journal of Business Finance & Accounting*, 14(4), 449-461. doi:10.1111/j.1468-5957.1987.tb00106.x
- Bayraktaroglu, A. E., Calisir, F. ve Baskak, M. (2019). Intellectual capital and firm performance: an extended VAIC model. *Journal of Intellectual Capital*, 20(3), 406-425. doi:10.1108/JIC-12-2017-0184
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71. doi:10.2307/2490171
- Berry, W. D. (1993). *Understanding regression assumptions. Series quantitative applications in the social sciences*. SAGE Publications, Inc. doi:10.4135/9781412986427
- Birkes, D. ve Dodge, Y. (1993). *Alternative methods of regression* (1. bs.). New York: Wiley.
- Bousquet, O., Boucheron, S. ve Lugosi, G. (2004). Introduction to statistical learning theory. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3176, 169-207. doi:10.1007/978-3-540-28650-9_8
- Boyd, S., Parikh, N., Chu, E., Peleato, B. ve Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1), 1-122. doi:10.1561/22000000016
- Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, 37(4), 373-384. doi:10.1080/00401706.1995.10484371
- Bühlmann, P. ve van de Geer, S. (2011). *Statistics for high-dimensional data*. Springer Series in Statistics (C. 1). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-20192-9
- Chatterjee, S. ve Simonoff, J. S. (2013). *Handbook of regression analysis*. John Wiley & Sons.
- Chen, X., Lin, Q., Kim, S., Carbonell, J. G. ve Xing, E. P. (2012). Smoothing proximal gradient method for general structured sparse regression. *The Annals of Applied Statistics*, 6(2), 719-752. doi:10.1214/11-AOAS514

- Cherkassky, V., Friedman, J. H. ve Wechsler, H. (2012). From statistics to neural networks: theory and pattern recognition applications (C. 136). Springer Science & Business Media.
- Cherkassky, V. ve Ma, Y. (2009). Another look at statistical learning theory and regularization. *Neural networks : the official journal of the International Neural Network Society*, 22(7), 958-969. doi:10.1016/J.NEUNET.2009.04.005
- Cherkassky, V. ve Mulier, F. M. (2007). *Learning from data: concepts, theory, and methods*. John Wiley & Sons.
- Combs, J. G., Russell Crook, T. ve Shook, C. L. (2005). The dimensionality of organizational performance and its implications for strategic management research. D. J. Ketchen ve D. D. Bergh (Ed.), *Research Methodology in Strategy and Management* içinde , Research Methodology in Strategy and Management (C. 2, ss. 259-286). Emerald Group Publishing Limited. doi:10.1016/S1479-8387(05)02011-4
- Cox, D. R. ve Snell, E. J. (1974). Choice of variables in observational studies. *Journal of Applied Statistics*, 23(1), 51-59. doi:10.2307/2347053
- Dao, B. T. T. ve Ta, T. D. N. (2020). A meta-analysis: capital structure and firm performance. *Journal of Economics and Development*, 22(1), 111-129. doi:10.1108/JED-12-2019-0072
- Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of accounting research*, 167-179.
- Delen, D., Kuzey, C. ve Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, 40(10), 3970-3983. doi:10.1016/j.eswa.2013.01.012
- Dikmen, B. (2007). *Finansal başarısızlık tahminlerinde matematiksel model uygulamaları*. Sermaye Piyasası Kurulu.
- Draper, N. R. ve Smith, H. (1998). Applied Regression Analysis. *Technometrics*, 47(3), 706. doi:10.1198/tech.2005.s303
- Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *The Journal of Financial and Quantitative Analysis*, 7(2), 1477-1493. doi:10.2307/2329929

- Efron, B., Hastie, T., Johnstone, I. ve Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407-451. doi:10.1214/009053604000000067
- Emmert-Streib, F. ve Dehmer, M. (2019). High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection. *Machine Learning and Knowledge Extraction 2019, Vol. 1, Pages 359-383*, 1(1), 359-383. doi:10.3390/MAKE1010021
- Faello, J. (2015). Understanding the limitations of financial ratios. *Academy of accounting and financial studies journal*, 19(3), 75.
- Frank, L. E. ve Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2), 109-135. doi:10.1080/00401706.1993.10485033
- Freund, R. J., Wilson, W. J. ve Sa, P. (2006). *Regression analysis* (2. bs.). Burlington: Elsevier.
- Friedman, J. H., Hastie, T. ve Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1-22. doi:10.18637/jss.v033.i01
- Friedman, J., Hastie, T., Höfling, H. ve Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2), 302-332. doi:10.1214/07-AOAS131
- Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7(3), 397. doi:10.2307/1390712
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*. doi:10.2307/2841583
- Gauraha, N. (2018). Introduction to the LASSO. *Resonance*, 23(4), 439-464. doi:10.1007/s12045-018-0635-x
- Genç, M. (2020). *Ridge, lasso and elastic net methods in high dimensional data*. (Yayımlanmamış doktora tezi). Çukurova Üniversitesi, Adana.
- Gepp, A. ve Kumar, K. (2015). Predicting financial distress: A comparison of survival analysis and decision tree techniques. *Procedia Computer Science*, 54, 396-404. doi:https://doi.org/10.1016/j.procs.2015.06.046

- Gibson, C. H. (2000). *Financial reporting and analysis : Using financial accounting information*. South-Western College Pub.
- Golbayani, P., Florescu, I. ve Chatterjee, R. (2020). A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. *The North American Journal of Economics and Finance*, 54, 101251. doi:<https://doi.org/10.1016/j.najef.2020.101251>
- Goodfellow, I., Bengio, Y. ve Courville, A. (2016). *Deep Learning* (1. bs.). MIT Press.
- Gruber, M. (1998). *Improving efficiency by shrinkage: The James-Stein and ridge regression estimators* (1. bs.). Routledge.
- Gunst, R. F. ve Webster, J. T. (2007). Regression analysis and problems of multicollinearity. *Communications in Statistics-Theory and Methods*, 4(3), 277-292. doi:10.1080/03610927308827246
- Hastie, T., Tibshirani, R. ve Friedman, J. (2008). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- Hastie, Trevor., Tibshirani, Robert. ve Wainwright, Martin. (2015). *Statistical learning with sparsity: The lasso and generalizations*. CRC Press.
- Herciu, M. (2017). Drivers of firm performance: exploring quantitative and qualitative approaches. *Studies in Business and Economics*, 12(1), 79-84. doi:10.1515/SBE-2017-0006
- Hocking, R. R. ve Leslie, R. N. (1967). Selection of the best subset in regression analysis. *Technometrics*, 9(4), 531-540.
- Hoerl, A. E. ve Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67. doi:10.1080/00401706.1970.10488634
- Hoerl, A. E. ve Kennard, R. W. (1970b). Ridge Regression: Applications to nonorthogonal problems. *Technometrics*, 12(1), 69-82. doi:10.1080/00401706.1970.10488635
- Hoerl, R. W. (2020). Ridge regression: A historical context. *Technometrics*, 62(4), 420-425. doi:10.1080/00401706.2020.1742207
- Horrigan, J. O. (1966). The determination of long-term credit standing with financial ratios. *Journal of Accounting Research*, 4, 44-62. doi:10.2307/2490168

- Horrigan, J. O. (1968). A short history of financial ratio analysis. *The Accounting Review*, 43(2), 284-294. <http://www.jstor.org/stable/243765> adresinden erişildi.
- Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H. ve Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, 37(4), 543-558. doi:[https://doi.org/10.1016/S0167-9236\(03\)00086-1](https://doi.org/10.1016/S0167-9236(03)00086-1)
- İmir, E. (1986). *Çoklu Bağıntılı Doğrusal Modellerde Ridge Regresyon Yöntemiyle Parametre Kestirimi*. (Yayımlanmamış doktora tezi). Eskişehir.
- James, G., Witten, D., Hastie, T. ve Tibshirani, R. (2013). *An introduction to statistical learning*. Springer Texts in Statistics (C. 103). doi:10.1007/978-1-4614-7138-7
- Jaworski, J. ve Czerwonka, L. (2022). Profitability and working capital management: evidence from the Warsaw Stock Exchange. *Journal of Business Economics and Management*, 23, 180-198. doi:10.3846/jbem.2022.15087
- Jones, Z. M., Desmarais, B. ve Zorn, C. (2017). *Interpretable statistical learning*. ProQuest Dissertations and Theses. (Yayımlanmamış doktora tezi). Ann Arbor.
- Kecman, V. (2001). *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. MIT press.
- Kliestik, T., Valaskova, K., Lazaroiu, G., Kovacova, M. ve Vrbka, J. (2020). Remaining financially healthy and competitive: The role of financial predictors. *Journal of Competitiveness*, 12(1), 74.
- Korkmaz, Ö. ve Karaca, S. S. (2014). Üretim işletmelerinde firma karlılığının finansal belirleyicileri ve BİST imalat sanayi uygulaması. *Ege Akademik Bakış*, 14(1), 21-29.
- Kroese, D. P., Botev, Z. I., Taimre, T. ve Vaisman, R. (2019). *Data science and machine learning: mathematical and statistical methods*. Chapman and Hall/CRC.
- Kuhn, M. (2016). caret: Classification and Regression Training. R package version 6.0-71. <https://CRAN.R-project.org/package=caret> adresinden erişildi.
- Kutner, M. H., Nachtsheim, C. J., Neter, J. ve Li, W. (2005). *Applied linear statistical models* (C. 5). McGraw-Hill Irwin Boston.

- Marill, K. A. (2004). Advanced Statistics: Linear Regression, Part I: Simple Linear Regression. *Academic Emergency Medicine*, 11(1), 87-93. doi:10.1197/j.aem.2003.09.005
- McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 93-100. doi:10.1002/WICS.14
- Miller, C. C., Washburn, N. T. ve Glick, W. H. (2013). Perspective—The myth of firm performance. *Organization Science*, 24(3), 948-964.
- Mitchell, T. M. (1997). *Machine learning*. McGraw Hill.
- Montgomery, D. C., Peck, El. A. ve Vining, G. G. (2013). *Doğrusal Regresyon Analizine Giriş*. (M. A. Erar, Ed.) (5. bs.). Ankara: Nobel.
- Nissim, D. ve Penman, S. H. (2001). Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies*, 6(1), 109-154. doi:10.1023/A:1011338221623
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109. doi:10.2307/2490395
- Pardoe, I. (2012). *Applied regression modeling* (2. bs.). Hoboken: Wiley.
- Park, C. ve Yoon, Y. J. (2011). Bridge regression: Adaptivity and group selection. *Journal of Statistical Planning and Inference*, 141(11), 3506-3519. doi:10.1016/J.JSPI.2011.05.004
- Piotroski, J. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, 38, 1-41.
- Raschka, S. (2015). *Python machine learning*. Packt publishing ltd.
- Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*.
- Refaeilzadeh, P., Tang, L. ve Liu, H. (2007). On comparison of feature selection algorithms. *Proceedings of AAAI workshop on evaluation methods for machine learning II* içinde (C. 3, s. 5). AAAI Press Vancouver.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408. doi:10.1037/H0042519
- Rothaermel, F. T. (2018). *Strategic management* (4. bs.). New York: McGraw-Hill.

- Salim, M. ve Yadav, R. (2012). Capital structure and firm performance: Evidence from Malaysian listed companies. *Procedia - Social and Behavioral Sciences*, 65, 156-166. doi:10.1016/j.sbspro.2012.11.105
- Santos, J. ve Brito, L. (2012). Toward a subjective measurement model for firm performance. *BAR - Brazilian Administration Review*, 9, 95-117. doi:10.1590/S1807-76922012000500007
- Seber, G. A. F. ve Lee, A. J. (2003). *Linear regression analysis*. Wiley Series in Probability and Statistics. Wiley. doi:10.1002/9780471722199
- Senan, N., Anagreh, S., Al-dalaien, B., Khaled, S. ve Al-Homaidi, E. (2021). Working capital management and banks' performance: Evidence from India. *Journal of Asian Finance Economics and Business*, 8, 747-758. doi:10.13106/jafeb.2021.vol8.no6.0747
- Shimerda, T. A. (1978). *Financail ratios as predictors of profitability*. (Yayımlanmamış doktora tezi). The University of Nebraska, Lincoln.
- Sun, J. ve Hui, X. (2006). An application of decision tree and genetic algorithms for financial ratios' dynamic selection and financial distress prediction. *2006 International Conference on Machine Learning and Cybernetics* içinde (ss. 2413-2418). doi:10.1109/ICMLC.2006.258771
- Şıklar, E. (2000). *Regresyon analizine giriş*. Eskişehir: Anadolu Üniversitesi.
- Tascón, M. T. ve Castano, F. J. (2017). Selection of variables in small business failure analysis: Mean selection vs. median selection. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 24, 54-88.
- Tezel Özturan, A. (2019). *Optimizasyon ve Matlab uygulamaları*. Ankara: Nobel Yayın Dağıtım.
- Tibshirani, R. J. (2013). The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7, 1456-1490. doi:10.1214/13-EJS815
- Tibshirani, Robert. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- Tibshirani, Ryan ve Wasserman, L. (2015). Stein's unbiased risk estimate. *Course notes from "Statistical Machine Learning*, 1-12.

- Tseng, P. (2001). Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization. *Journal of Optimization Theory and Applications*, 109(3), 475-494.
- van Wieringen, W. N. (2015). Lecture notes on ridge regression. <https://arxiv.org/abs/1509.09169v7> adresinden erişildi.
- Vapnik, V. (1991). Principles of Risk Minimization for Learning Theory. J. Moody, S. Hanson ve R. P. Lippmann (Ed.), *Advances in Neural Information Processing Systems* içinde (C. 4). Morgan-Kaufmann.
- Vapnik, V. N. (Vladimir N. (1998). *Statistical learning theory*. Wiley.
- Vapnik, Vladimir. (2000). *The nature of statistical learning theory. The Nature of Statistical Learning Theory*. Springer New York. doi:10.1007/978-1-4757-3264-1
- Venkatraman, N. ve Ramanujam, V. (1986). Measurement of business performance in strategy research: A comparison of approaches. *The Academy of Management Review*, 11(4), 801-814.
- Xu, W., Xiao, Z., Dang, X., Yang, D. ve Yang, X. (2014). Financial ratio selection for business failure prediction using soft set theory. *Knowledge-Based Systems*, 63, 59-67. doi:<https://doi.org/10.1016/j.knosys.2014.03.007>
- Yousaf, M. (2022). Intellectual capital and firm performance: evidence from certified firms from the EFQM excellence model. *Total Quality Management & Business Excellence*, 33(13-14), 1472-1488. doi:10.1080/14783363.2021.1972800
- Yousaf, M. ve Dey, S. K. (2022). Best proxy to determine firm performance using financial ratios: A CHAID approach. *Review of Economic Perspectives*.
- Zaini, B. J. ve Mahmuddin, M. (2019). Classifying firms' performance using data mining approaches. *International Journal Supply Chain Management*, 8(1), 690.
- Zou, H. ve Hastie, T. (2003). Regression shrinkage and selection via the elastic net with applications to microarrays. *JR Stat Soc Ser B*, 67, 301-320.
- Zou, H., Hastie, T. ve Tibshirani, R. (2007). On the “degrees of freedom” of the lasso. *Ann. Statist.*, 35(5), 2173-2192. doi:10.1214/009053607000000127