



T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İSTATİSTİK ANA BİLİM DALI

**ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON
EĞRİLERİNDE ALTERNATİF BİLGİ KRİTERLERİ İLE
MODEL SEÇİMİ**

Doktora Tezi

Meryem BEKAR ADIGÜZEL

Danışman
Prof. Dr. Mehmet Ali CENGİZ

SAMSUN
2021

T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İSTATİSTİK ANA BİLİM DALI

**ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON
EĞRİLERİNDE ALTERNATİF BİLGİ KRİTERLERİ İLE
MODEL SEÇİMİ**

Doktora Tezi

MERYEM BEKAR ADIGÜZEL

Danışman
Prof. Dr. Mehmet Ali CENGİZ

SAMSUN
2021

TEZ KABUL VE ONAYI

Meryem BEKAR ADIGÜZEL tarafından, Prof. Dr. Mehmet Ali CENGİZ danışmanlığında hazırlanan “Çok Değişkenli Uyarlanabilir Regresyon Eğrilerinde Alternatif Bilgi Kriterleri İle Model Seçimi” başlıklı bu çalışma, jürimiz tarafından 26.04.2021 tarihinde yapılan sınav sonucunda oy birliği / oy çokluğu ile başarılı bulunarak doktora tezi olarak kabul edilmiştir.

	Unvanı Adı Soyadı Üniversitesi Ana Bilim/Ana Sanat Dalı	İmza	Sonuç
Başkan	Prof. Dr. XXXX XXXX Ondokuz Mayıs Üniversitesi Anabilim Dalı		<input type="checkbox"/>
			Kabul
			<input type="checkbox"/>
			Ret
Üye (Danışman)	Dr. Öğr. Üyesi XXXX XXXX Ondokuz Mayıs Üniversitesi Anabilim Dalı		<input type="checkbox"/>
			Kabul
			<input type="checkbox"/>
			Ret
Üye	Doç. Dr. XXXX XXXX Ondokuz Mayıs Üniversitesi Anabilim Dalı		<input type="checkbox"/>
			Kabul
			<input type="checkbox"/>
			Ret
Üye	Prof. Dr. XXXX XXXX Ondokuz Mayıs Üniversitesi Anabilim Dalı		<input checked="" type="checkbox"/>
			Kabul
			<input checked="" type="checkbox"/>
			Ret
Üye	Prof. Dr. XXXX XXXX Ondokuz Mayıs Üniversitesi Anabilim Dalı		<input type="checkbox"/>
			Kabul
			<input type="checkbox"/>
			Ret

Bu tez, Enstitü Yönetim Kurulunca belirlenen ve yukarıda adları yazılı jüri üyeleri tarafından uygun görülmüştür.

ONAY

... / ... / ...

Prof. Dr. Ali BOLAT
Enstitü Müdürü

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Hazırladığım doktora tezinin bütün aşamalarında bilimsel etiğe ve akademik kurallara riayet ettiğimi, çalışmada doğrudan ve dolaylı olarak kullandığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin Kaynaklar'da gösterilenden oluştuğunu, her unsurun enstitü yazım klavuzuna uygun yazıldığını ve TÜBİTAK Araştırma ve Yayın Etiği Kurulu Yönetmeliği'nin 3. bölüm 9. maddesinde belirtilen durumlara aykırı davranılmadığını taahhüt ve beyan ederim.

İmza

26/ 04 / 2021

Meryem BEKAR ADIGÜZEL

TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI

Tez Başlığı : ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON EĞRİLERİNDE ALTERNATİF BİLGİ KRİTERLERİ İLE MODEL SEÇİMİ

Yukarıda başlığı belirtilen tez çalışması için şahsım tarafından 08.03.2021 tarihinde intihal tespit programından alınmış olan özgünlük raporu sonucunda;

Benzerlik oranı : % 14 Bir öge seçin.

Tek kaynak oranı : % 2 çıkmıştır.

İmza

26 / 04 / 2021

Prof. Dr. Mehmet Ali CENGİZ

ÖZET

ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON EĞRİLERİNDE ALTERNATİF BİLGİ KRİTERLERİ İLE MODEL SEÇİMİ

Meryem BEKAR ADIGÜZEL

Ondokuz Mayıs Üniversitesi

Lisansüstü Eğitim Enstitüsü

İstatistik Ana Bilim Dalı

Doktora, Nisan/2021

Danışman: Prof. Dr. Mehmet Ali CENGİZ

Çok değişkenli uyarlanabilir regresyon eğrileri tekniği, yüksek boyutlu verilerden kaynaklanan doğrusal olmama probleminin çözümündeki başarısı, bağımsız değişkenler ile bağımlı değişken arasında herhangi bir varsayım gerektirmemesi ve bağımlı değişkeni tanımlamak için bağımsız değişkenlerin eklemeli ve etkileşimsel katkılarına yer vermesi bakımından elverişli bir yöntemdir. MARS tekniği, model seçiminde genelleştirilmiş çapraz geçerlilik kriterini (GCV) kullanmaktadır. Ancak GCV kriteri ise algoritmasında kullandığı düzeltme parametresinin keyfi değer alması ve bu kriter ile yüksek boyutlu modeller elde edilmesi nedeniyle eleştirilmektedir.

Bu çalışmada, MARS tekniğinin model seçiminde kullandığı GCV yerine alternatif bilgi kriterleri kullanılarak (AIC, SBC, ICOMP(IFIM)_PEU), MARS tekniği için keyfi seçim eleştirisinin ortadan kaldırılması ve bağımlı değişkeni en iyi açıklayan, en fazla bilgiyi içeren, en yalın modelin bulunması amaçlanmıştır.

MARS'ta kullanılan bilgi kriterlerinin (GCV, AIC, SBC, ICOMP(IFIM)_PEU) model seçimindeki başarılarını test edebilmek için öncelikle bağımlı değişken üzerinde katkısı olan ve olmayan değişkenlerden oluşan veri uzayı ile simülasyon çalışması gerçekleştirilmiştir. Simülasyon çalışması sonucunda bağımlı değişkene katkısı olmayan değişkenlerin regresyon modeline dâhil edilmemiş olması kriterlerin model seçimindeki başarısını göstermiştir. Gerçek veri seti olarak ise Türkiye'de faaliyet gösteren 18 bankanın 2005-2019 yıllarındaki verilerinden faydalanılarak kredilerin takibe düşme nedenleri incelenmiştir. Elde edilen sonuçlar, ICOMP kriterinin diğer kriterlere kıyasla, daha yalın sonuçlar verdiğini ortaya koymaktadır.

Anahtar Sözcükler: Çok değişkenli uyarlanabilir regresyon eğrileri (MARS), akaike bilgi kriteri (AIC), schwarz bayes bilgi kriteri (SBC), bilgi karmaşıklığı kriteri ICOMP(IFIM)_PEU, banka, takipteki krediler.

ABSTRACT

MODEL SELECTION IN MULTIVARIATE ADAPTIVE REGRESSIONS SPLINES WITH ALTERNATIVE INFORMATION CRITERIA

Meryem BEKAR ADIGÜZEL

Ondokuz Mayıs University

Institute of Graduate Studies

Department of Statistics

Doctoral Dissertation, April, 2021

Supervisor: Prof. Dr. Mehmet Ali CENGİZ

Multivariate adaptive regression splines (MARS) technique is a convenient method due to its success in solving nonlinearity problem caused by high dimensional data, not to require any assumption between independent variables and dependent variable and using additive and interactive contributions of independent variables to define dependent variable. The MARS technique uses generalized cross validation (GCV) criterion to select the model. However, GCV criterion is criticized because of the arbitrary value in smoothing parameter which is used in its algorithm and obtaining high dimensional models by using this criterion.

This study aims to lift the arbitrary value criticism and try to find the best model which is the most informative and best explainer of the dependent variable in MARS technique by using alternative information criterion (AIC, SBC, ICOMP(IFIM)_PEU) instead of GCV.

In order to test the success of information criteria (GCV, AIC, SBC, ICOMP(IFIM)_PEU) which are used to select the model in MARS technique, first of all a simulation study is carried out with the data set composed of data which have or not have contribution on the dependent variable. As a result of the simulation study not including the variables which do not have contribution on the dependent variable to the regression model, it shows the success of the criteria in choosing the model. Moreover, the causes of non performing loans are being analyzed using the data set between the years 2005 – 2019 of 18 banks operating in Turkey. The results show that ICOMP (IFIM)_PEU criterion gives the best results compared to the other criteria.

Keywords: Multivariate adaptive regression splines (MARS), akaike information criterion (AIC), schwarz bayesian information criterion (SBC), information complexity criterion (ICOMP(IFIM)_PEU), bank, non performing loans.

ÖNSÖZ VE TEŞEKKÜR

Engin bilgi ve tecrübelerinden faydalandığım, çalışmamın her aşamasında desteğini hissettiğim danışman hocam Sayın Prof. Dr. Mehmet Ali CENGİZ'e, zor zamanlarımda hep yanımda olan, yönlendirmeleriyle çalışmalarımı kolaylaştıran, iyiki var dediğim canım abim Doç. Dr. Murat BEKAR'a, hayatıma tarifsiz değer katan, bu sürecin güzelliklerini de zorluklarını da birlikte yaşadığım sevgili eşim Rıza ADIGÜZEL'e ve küçücük bedenine sonsuz huzur ve mutluluk sığdıran biricik oğlumuz Burak ADIGÜZEL'e sonsuz teşekkürler.

Meryem BEKAR ADIGÜZEL

İÇİNDEKİLER

1. GİRİŞ	1
2. LİTERATÜR ARAŞTIRMASI	3
3. LİTERATÜR DEĞERLENDİRMESİ	15
4. ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON EĞRİLERİ	16
4.1 MARS Modeli.....	17
4.2 Temel Fonksiyon.....	19
4.3 Düğüm Noktası	19
4.4 Genelleştirilmiş Çapraz Geçerlilik (GCV) Kriteri	20
5. MODEL SEÇİMİ ve BİLGİ KRİTERLERİ	22
5.1 Entropi.....	23
5.2 Kullback-Leibler Ölçüsü.....	24
5.3 Fisher Bilgi Matrisi	25
5.3.1 Dublikasyon Matrisi ve Moore-Penrose Tersisi	26
5.3.2 Kronecker Çarpım.....	26
5.4 Akaike Bilgi Kriteri (AIC).....	27
5.5 Schwarz Bayes Bilgi Kriteri (SBC)	28
5.6 Tutarlı Akaike Bilgi Kriteri (CAIC)	28
5.7 Fisher Bilgi Matrisine Dayalı Tutarlı Akaike Bilgi Kriteri (CAICF)	29
5.8 Bilgi Karmaşıklığı Kriteri (ICOMP).....	29
5.8.1 Bilgi Ölçümü ve Bilgi Karmaşıklığı	29
5.8.2 ICOMP (IFIM)_PEU Bilgi Kriteri.....	33
6. SİMÜLASYON ÇALIŞMASI	35
7. GERÇEK VERİ SETİ UYGULAMASI	44
7.1 MARS_GCV Model Seçimi	57
7.2 MARS_AIC Model Seçimi	64
7.3 MARS_SBC Model Seçimi	66
7.4 MARS_ICOMP(IFIM)_PEU Model Seçimi	68
8. SONUÇ	74
9. KAYNAKLAR	76
EK 1. Kredi Veri Seti.....	85

KISALTMALAR

MARS	: Çok Değişkenli Uyarlanabilir Regresyon Eğrileri
BF	: Temel Fonksiyon
GCV	: Genelleştirilmiş Çapraz Geçerlilik
KL	: Kullback-Leibler Uzaklığı
FIM	: Fisher Bilgi Matrisi
IFIM	: Ters Fisher Bilgi Matrisi
AIC	: Akaike Bilgi Kriteri
SBC	: Schwartz Bilgi Kriteri
CAIC	: Tutarlı Akaike Bilgi Kriteri
CAICF	: Fisher Bilgi Matrisinin Kullanıldığı Tutarlı Akaike Bilgi Kriteri
ICOMP	: Bilgi Karmaşıklığı Kriteri
CICOMP	: Tutarlı Bilgi Karmaşıklığı Kriteri
ICOMP(IFIM)_PEU	: Sonraki Beklenen Faydaya Bir Yaklaşım Olarak Ters-Fisher Bilgi Matrisine Dayalı Bilgi Karmaşıklığı Kriteri
ANOVA	: Varyans Analizi
EKK	: En Küçük Kareler
RSS	: Artık Kareler Toplamı
TCMB	: Türkiye Cumhuriyeti Merkez Bankası
GSYİH	: Gayri Safi Yurtiçi Hasıla
BKZ	: Bakımız

ŞEKİLLER DİZİNİ

Şekil 4.1: (a) İleri Doğru Seçim Modeli, (b) Geriye Doğru Eleme Modeli

Şekil 4.2: Temel Fonksiyonlar, $(x - 0,5)_+$ ve $(0,5 - x)_+$

Şekil 4.3: MARS ile belirlenen düğüm değerleri aracılığıyla oluşturulan üç düğümlü parçalı doğrusal regresyon modeli (solda gerçek veri)

Şekil 5.1: Büyük patlama olayı üzerinden entropi kavramı

Şekil 6.1: Doğrusal regresyon modeli

Şekil 6.2: Karmaşık yapıdaki regresyon modeli

Şekil 6.3: Kesikli doğrusal hale getirilmiş MARS regresyon modeli

Şekil 6.4: X değişkeninden BF'lerin elde edilmesi

Şekil 7.1: MARS Modelleri Akış Diyagramı

Şekil 7.2: TSÜE Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği

Şekil 7.3: Faiz Oranları Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği

Şekil 7.4: İthalat Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği

Şekil 7.5: İşsizlik Oranı Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği

Şekil 7.6: Enflasyon Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği

Şekil 7.7: Döviz Kuru Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği

Şekil 7.8: İhracat Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği

TABLULAR DİZİNİ

Tablo 6.1: 10 deęişken ve 100 veriyi içeren matris

Tablo 6.2: Bir deęişkendeki düęüm noktalarının ve temel fonksiyonun oluşması

Tablo 6.3: GCV kriteri kullanılarak elde edilen ANOVA ayrışımı

Tablo 6.4: ICOMP(IFIM)_PEU kriteri kullanılarak elde edilen ANOVA ayrışımı

Tablo 7.1: Çalışmada kullanılan veri seti

Tablo 7.2: Çalışmada kullanılan mevduat ve katılım bankaları

Tablo 7.3: Gerçek veri deęişkenlerine ait düęüm deęerleri ve sıra sayıları

Tablo 7.4: 12 adet ana etkiye sahip BF'nin oluşturulması

Tablo 7.5: 24 adet BF'ye ilişkin 2'li etkileşimlerin oluşturulması

Tablo 7.6: GCV kriteri kullanılarak elde edilen ANOVA ayrışımı

Tablo 7.7: AIC kriteri kullanılarak elde edilen ANOVA ayrışımı

Tablo 7.8: SBC kriteri kullanılarak elde edilen ANOVA ayrışımı

Tablo 7.9: ICOMP(IFIM)_PEU kriteri kullanılarak elde edilen ANOVA ayrışımı

1. GİRİŞ

Değişkenler arasındaki ilişkinin modellenmesinde klasik yöntemler sıkça kullanılmaktadır. İstatistiksel bakış açısına göre, kurulan modellerde değişkenler arasındaki ilişkinin gerçek yapısının daha iyi anlaşılabilmesi için kullanılan gözlemlerin değişken sayısından anlamlı bir şekilde fazla olması istenir. Ancak bunun tersi durumlar da gözlenebilmektedir. Örneğin sağlık taraması ile ilgili bir kişiden (gözlem, n) kan değerleri, tansiyon, ağırlık, boy gibi onlarca değişken (boyut, p) ölçümü gerekebilmektedir. Yani çalışma için ulaşılabilen gözlemler onlarla veya yüzlerle ifade edilirken, tek bir gözlem için onlarca hatta yüzlerce değişkene, boyuta sahip olunabilmektedir ($n \ll p$). Klasik yöntemler ‘büyük boyut, küçük gözlem – bip , small n ’ denilen bu tarz verilerle başa çıkabilecek şekilde tasarlanmış değıllerdir. Böyle durumlarda klasik metotlara alternatif olarak sunulan yöntemler gündeme gelmektedir. Bu yöntemlerden birisi de “Çok Değişkenli Uyarlanabilir Regresyon Eğrileri” (Multivariate Adaptive Regression Splines / MARS) tekniğidir. MARS, değişkenler arasındaki olası ilişkiyi belirlemekte “düzleştirme uzanımları” (smoothing splines) tekniğini kullanmaktadır (Friedman, 1991). Kullanılan algoritma ile değişkenler ve\veya değişkenlerin birbirleriyle olan kombinasyonlarından oluşturulan temel fonksiyonlardan meydana gelen modellerde değişkenlerin ve etkileşimlerinin tek tek yorumlanabilmesi çalışmalarda büyük bir kolaylık sağlamaktadır.

MARS tekniği ileri doğru seçim ve geriye doğru eleme adımlarından oluşur. İleri doğru seçim adımında değişken ve\veya değişken kombinasyonlarından oluşan temel fonksiyonlar ile maksimum karmaşıklığıdaki model elde edilir. Geriye doğru eleme adımında ise ileriye doğru seçim adımında oluşturulan modelin karmaşıklığı, modele daha az katkıda bulunan temel fonksiyonların elenmesiyle giderilir. Bu işlem için Craven ve Wahba (1979) tarafından geliştirilen GCV (Genelleştirilmiş Çapraz Geçerlilik) kriteri kullanılmaktadır. En iyi ve en yalın model için en küçük GCV değerini veren model tercih edilmektedir. Ancak literatürde MARS algoritmasında model seçimi için ‘en iyi’ kriterin GCV kriteri olup olmadığı yönünde sorular

mevcuttur (Koc, Bozdogan, 2014). Bunun üzerine GCV kriteri yerine alternatif bilgi kriterleri kullanarak verilere en iyi uyumu sađlayan ve en yalın MARS modelinin bulunması amaçlanmaktadır.

Bu çalıřmada ilk olarak MARS tekniđi ve GCV kriteri tanıtılacak ve Entropi kavramından Kullback-Leibler ölçüsüne kadar uzanan bir gelişim içinde kronolojik sırası ile, Akaike bilgi kriterinden (AIC) Bilgi karmaşıklığı kriterine (ICOMP(IFIM)_PEU) kadar olan model seçim kriterlerine yer verilecektir. Akabinde bir simülasyon çalıřması ile seçilen en iyi alt modeller için model seçim kriterlerinin performansları deđerlendirilecek ve çalıřmaya gerçek bir veri seti üzerinde devam edilecektir.

2. LİTERATÜR ARAŞTIRMASI

MARS ile ilgili çalışmalar:

Charron ve Annoni (2021), çalışmasında geleneksel medya olarak adlandırılan profesyonel radyo, televizyon, gazeteler ile sosyal medya grubunda yer alan YouTube, Facebook, Twitter gibi iki medya ayağının vatandaşların siyasi tutum ve davranışlarını nasıl etkilediğini incelemiştir. Yani bir kişinin haber kaynağının siyasi yolsuzluk üzerinde etkisi var mı? Sorusuna cevap aranmıştır. Araştırmacı sosyal medya kullanıcılarının geleneksel medya kullanıcılarına kıyasla daha fazla siyasi yolsuzluk algısına sahip olduğunu, sosyal medya kullanıcıları arasında daha yüksek kutuplaşma olduğunu ve hükümet yanlıları ile muhalif siyasi görüşe sahip olanlar arasında yolsuzluk algısı farkının sosyal medya kullanıcılarında daha fazla görüldüğü yönündeki hipotezlerini savunmaktadır. Çalışma için Avrupa Devlet Kalitesi Endeksi tarafından hazırlanan anket kullanılmıştır. Ankete Avrupa'nın 21 ülkesinden 18 yaş ve üstü 77.966 kişi katılmıştır. Modelleme hem parametrik (ikili ortalamalar t-testi) hem de parametrik olmayan (büyük veri setleri için hesaplama kolaylığı sağlayan, etkileşim etkilerini modele dahil eden ve doğrusal regresyon varsayımları barındırmayan MARS) istatistik teknikleri kullanılmıştır. Elde edilen sonuçlar sosyal medya kullanıcılarının siyasi yolsuzluk algısının daha baskın olduğu yönündedir. Bunun nedeni olarak da sosyal medyanın bağımsızlığının yerel medyaya göre daha fazla olması şeklinde açıklanmıştır.

Serrano vd. (2020), cinsiyetin kas ve iskelet sistemi bozuklukları üzerindeki etkisini incelemek için çok değişkenli uyarlanabilir regresyon eğrileri (MARS) tekniğini kullanmışlardır. Amaçları kas ve iskelet bozukluklarına neden olan faktörler ve bu faktörlerin cinsiyetlere göre farklılıklarını belirlemektir. Çalışma için 2015 yılında 35 ülkede 18 ile 89 yaşları arasındaki 21.578 kadın ve 21.922 erkek işçiyle görüşülmüştür. Bunun için kadın ve erkekler için iki farklı MARS modeli oluşturulmuştur. Her iki model de rahatsızlığın oluşumunu başarılı bir şekilde tahmin ettiği görülmüştür. erkekler fiziksel risk faktörlerine (ağır eşyalar taşımak, yorucu pozisyonlarda bulunmak) diğer bazı çalışma koşullarına (gürültüye maruz kalmak, düşük sıcaklıklarda çalışmak, vardiyalı çalışmak) daha savunmasız bulunurken, kadınlar ise psikososyal risk faktörlerinden (yoğun çalışmak, işyerinde yaşanan sıkıntılar) ve çalışma saatleri dışında gerçekleştirilen faaliyetlerden (iş dışında ev

işleri, yemek, çocuklarla ilgilenmek) daha fazla etkilendiği görülmüştür. Sonuçlar çalışan nüfusun kas ve iskelet bozukluklarına müdahalede başarılı olunabilmesi için cinsiyet faktörünün önemini de ortaya koymuştur

Nacar (2020), çalışmasında iklim değişikliğinin Doğu Karadeniz Havzası sıcaklık ve yağış parametreleri üzerindeki olası etkisini incelemiştir. Havza içinde ve yakınlarında bulunan 30 yıllık kesintisiz gözlem verisine sahip 12 meteoroloji istasyonu belirlemiştir. Bu istasyonların her biri için Çok Değişkenli Uyarlanabilir Regresyon Eğrileri Yöntemi kullanılarak ölçek indirgeme modelleri kurulmuştur. 12 adet küresel ölçekli iklim parametresi girdi değişkeni olarak, aylık toplam yağış ve ortalama sıcaklık değerleri ise model çıktıları olarak kullanılmıştır. Yapılan çalışmalar sonucunda havzanın güneyi için iyimser senaryoya göre ortalama 2,5 C, kötümser senaryoya göre ortalama 3,5 C, kuzey bölümünde ise iyimser senaryoya göre ortalama 2,0 C, kötümser senaryoya göre ortalama 3,0 C sıcaklık artışı beklendiği sonucuna ulaşılmıştır. Ayrıca yağış değerlerinde havzanın güney bölümünde düşüş beklenirken, batı ve kıyı bölümlerinde özellikle ilkbahar aylarında önemli artış beklendiği sonucuna varılmıştır.

Lokuge vd. (2018), yüksek oranda CO₂ salınımı sonucunda elde edilen Portland Çimentolu betonlara karşı çevre dostu ve yenilikçi bir ürün olarak bilinen geopolimer betonunu alternatif olarak sunmuştur. Bu çalışmada MARS modeli kullanılarak geopolimer betonu için uçucu kül, kalsiyum, iri ve ince agrega, katalitik sıvı, cüruf karışım maddeleri kullanımı ve basınç dayanımları değerlendirilmiştir.

Ghasemzadeh ve Ahmed (2018), çalışmalarında olumsuz hava koşullarında ve kaygan zeminlerde sürücülerin şerit tutma kabiliyetleriyle ilgili çalışmaların çoğunluğunun sürüş simülatörleri gibi kontrollü ortamlarda gerçekleştirildiğini bu nedenle doğal ortam koşullarında olumsuz hava koşulları dikkate alınarak sürücü şeridi tutma kabiliyetinin araştırıldığı çalışmaların bulunmadığını ileri sürmüşlerdir. Bu çalışma, 19-89 yaş arası 141 sürücü ile naturalistik sürüş verileri kullanılarak gerçekleştirilmiştir. Sürücü şeridi tutma kabiliyetini etkileyen faktörlerin belirlenmesinde MARS ve lojistik regresyon teknikleri kullanılmıştır. Sonuçlar, şiddetli yağmurun şerit tutma yeteneğinde standart sapmayı önemli ölçüde artırdığını ortaya koymuştur. Trafik koşullarının, sürücü yaşının, deneyimin ve hız sınırının sürücü şerit tutma yeteneği üzerinde önemli etkileri olduğu da bulunmuştur. Bu

çalışmanın ilginç bir bulgusu da, sürücülerin daha yüksek hız sınırı olan yollarda daha iyi şerit tutma kabiliyeti göstermiş olduklarıdır.

Kuter vd. (2018), MODIS (Moderate Resolution Imaging Spectrometer) verilerini kullanarak Avrupa Alpleri'nde 500m uzamsal çözünürlükteki fraksiyonel kar örtüsünü (FSC) tahmin etmek için MARS modelini, Yapay Sinir Ağı (ANN) modelini ve Modis kar fraksiyonu ürününü (MOD10) kullanmıştır. MARS ve ANN modellerinin FSC verisi $R=0,93$ ile aynı değerlere sahip çıkmıştır. MOD10 FSC verisi $R=0,88$ ile biraz daha zayıf bir performans sergilemiştir. Çalışmada MARS modelinin ANN ile aynı performansa sahip olduğu ancak model oluşturmada hesaplama açısından daha verimli olduğu savunulmuştur.

Heddam ve Kişi (2018), en küçük kareler destek vektörü (LSSVM), MARS, M5 model ağacı (M5T) tekniklerini Amerika Birleşik Devletlerindeki üç istasyona ait su kalitesi verileri ile çözünmüş oksijen (DO) konsantrasyonunun tahmininde kullanmışlardır. Model girdileri olarak günlük olarak ölçülen su sıcaklığı, pH, özel iletkenlik kullanılmıştır. Elde edilen sonuçlara göre DO konsantrasyonunun üç model kullanılarak başarılı bir şekilde tahmin edilebildiği ve üçü arasından en iyi modelin bir istasyondan diğerine farklılık gösterdiği şeklindedir.

Roy vd. (2018), çalışmalarında yoğun enerji kullanımına sahip sektörlerden biri olan inşaat sektöründe, enerji verimli binaların ısıtma yükünü ve soğutma yükünü yani enerji tüketimini tahmin etmek için MARS ve Aşırı Öğrenme Makinesi'nin (ELM) hibrid modeli yöntemlerini kullanmışlardır. Edinilen sonuçlar hem MARS hem de ELM tekniğinin enerji tüketimi tahmininde kullanılabileceğini göstermekle beraber MARS, ELM'ye göre daha hızlı hesaplama yapabilmektedir.

Özmen vd. (2018), "Konut kullanıcıları, özellikle kış mevsiminde önemli miktarda gaz talep eden tüketicilerdir. Öncelikle alan ısıtması için ve de yemek, sıcak su ihtiyaçlarında kullanılan doğal gaz kesintiye uğramamalıdır. Sınırlı bir kapasiteye sahip olan dağıtım merkezlerinin tüm yıl boyunca uygun planlama ve tahminleme yapmaları gerekli hale gelmiştir" diyen Özmen ve arkadaşları 2009-2012 dönemlerinde Ankara'daki konut kullanıcılarının doğal gaz tüketimlerini tahmin etmek için MARS ve CMARS (Konik Çok Değişkenli Uyarlanabilir Regresyon Eğrileri) modellerini kullanmışlardır. Gaz tüketim verileri Ankara'nın yerel dağıtım şirketi olan Başkentgaz'dan, günlük minimum, maksimum ve ortalama sıcaklık

değerleri ise meteorolojiden elde edilmiştir. Bu çalışmanın sonuçları her iki yöntemden benzer sonuçlar elde edildiğini ve bu yöntemlerin doğal gaz endüstrisi için iki yeni rekabetçi yaklaşım olarak sunulabileceğini göstermektedir.

Keshtegar vd. (2018), Türkiye'nin Doğu Akdeniz Bölgesi'nde yer alan Adana ve Antakya istasyonlarının aylık güneş radyasyonunun tahmini için Kriging (yüksek dereceli doğrusal olmayan modelleri tahmin etmek için kullanılan bir yöntemdir), yanıt yüzey metodu (RSM), çok değişkenli uyarlanabilir regresyon eğrileri (MARS) ve model ağacı (M5) tekniklerini kullanmışlardır. Girdi değişkenleri olarak; maksimum sıcaklık (Tmax), minimum sıcaklık (Tmin), güneş saatleri (Hs), rüzgar hızı (Ws) ve bağıl nem (RH) kullanılmıştır. MARS modeli Adana istasyonunda en iyi tahmini sağlarken, Kriging modeli Antakya istasyonunda en iyi tahminde bulunmuştur.

Al-Musaylh vd. (2018), Avustralya'nın Queensland Eyaleti için MARS, SVR (destek vektör regresyonu) ve ARIMA modellerini kullanarak kısa vadeli (0.5 saat, 1 saat ve 24 saatlik öngörmeler kullanılmıştır) elektrik talebi tahmininde bulunmuşlardır. 0.5 ve 1 saatlik elektrik tahminlerinde MARS tekniği diğer teknikleri geride bırakırken bir günlük tahmin değerlerinde SVR tekniği üstünlük sağlamıştır. Bu nedenle, MARS ve SVR modelleri ARIMA modeline kıyasla Queensland Eyaletinde kısa süreli elektrik talebi için daha uygun olarak düşünülebilir. Çalışmanın gerçek zamanlı elektrik talebi araştırmaları için de yararlı olacağı düşüncesi belirtilmiştir.

Mehdizadeh vd. (2017), çalışmalarında İran gibi kurak ve yarı kurak bölgelerde su kaynaklarının optimal yönetiminin öneminden bahsedip İran'da bitkilerin su tüketimleri ve buharlaşma ile birlikte toplam su kayıpları aylık olarak incelenmiştir. Bunun için MARS, SVM (destek vektör makinesi), GEP (gen ekspresyonu programlama) ve ampirik denklemler kullanılmış ve performansları değerlendirilmiştir. MARS ve SVM yöntemlerinin genellikle GEP ve ampirik denklemlerden daha iyi performans gösterdiği bulunmuştur.

Bozağaç vd. (2016), MARS kullanarak zaman-alan koordinatlarının bir fonksiyonu olan dinamik sistem simülasyonu üzerinde çalışmışlardır. Eksik veri girişleri nedeniyle eksik çıktı işlemlerini ele almak için iki aşamalı bir metamodel planı geliştirilmiş ve bu metodoloji dinamik bir radar simülasyon modelinde

uygulanmıştır. Ortaya çıkan metamodelin tahmin performansı, dört farklı örnekleme tekniği ve 16 örnek büyüklüğü ile test edilmiş, sonuçlar MARS'ın özellikle uzman görüşü hazır olmadığında, dinamik simülasyonların metamodellenmesi için etkili bir yöntem olduğunu göstermiştir.

Yüksel (2016), Türkiye'deki bankaların takipteki krediler oranını belirleyen faktörleri MARS tekniği kullanarak tespit etmiştir. Çalışmasında 5 adedi makroekonomik ve 8 adedi bankalara özgü olmak üzere 13 adet bağımsız değişkene ait 1988-2014 dönem aralığındaki yıllık verileri kullanmıştır. Elde edilen bulgular döviz kurunun, faiz oranlarının ve büyüme oranlarının kredi takibini etkilediği yönündedir.

Shahraiyini vd. (2015), daha iyi hava kirliliği yönetimi için İran Tahran'da karbon monoksitin (CO) dağılım modellemesi ile ilgilenmişlerdir. Bu çalışmada, 2008 yılında Tahran'da saatlik ve günlük olarak 16 izleme istasyonu tarafından ölçülen CO verileri kullanılmıştır. MARS, Cokriging tekniği, Ters Mesafe Ağırlıklandırma, İnce Plaka Pürüzsüz Spline ve Kriging teknikleri kullanılmış ayrıca teknikler, ortalama mutlak hata yüzdesi (MAPE) değerine göre de karşılaştırılmıştır. Cokriging tekniği, Ters Mesafe Ağırlıklandırma, İnce Plaka Pürüzsüz Spline ve Kriging tekniklerinden daha az hata yüzdesi vermesine rağmen MARS en az hata yüzdesi sergilemiştir. Buna ek olarak, MARS modelleme prosedürü kolaydır. Bu nedenle MARS, mekansal dağılım modellemesi için uygun bir yöntem olarak sunulmaya hak kazanmıştır. Sonuçlar en kirli bölgelerin; Tahran'ın merkezi, doğu ve güneydoğu kısımları olduğunu ve yıllık ortalama CO konsantrasyonunun 4,2 ile 4,6 ppm arasında olduğunu göstermiştir.

Fernandez vd. (2014), İspanya'nın kuzeybatısındaki bir rezervuarda ötrofikasyona (su ekosisteminde çeşitli nedenlerle besin maddelerinin büyük oranda artması sonucu plankton ve alg varlığının aşırı şekilde artması) karşı risk önleme modellemesi üzerinde çalışmışlardır. Ötrofikasyon klorofil konsantrasyonunu öngörücü olarak kullanılmıştır. Ötrofikasyonun tatlı su çevreleri için tehlike arz ettiği bilinmektedir. Bu duruma önlem alabilmek amacıyla ötrofikasyonu meydana getiren sebepler araştırılmış bunun için MARS tekniği kullanılmış elde edilen korelasyon katsayısı 0,99 bulunmuştur.

Migueis vd. (2013), müşteri ilişkilerinden kaynaklanan kar, şirketler açısından büyük öneme sahiptir düşüncesiyle müşteri yıpranmasını araştırmışlardır. Bu amaçla MARS ve Lojistik regresyon yöntemlerini kullanarak müşteri memnuniyeti-memnuniyetsizliği üzerine modelleme geliştirmişlerdir. Ampirik sonuçlar, MARS'ın Lojistik regresyondan daha iyi performans sergilediği ancak adım adım özellik seçimli Lojistik regresyon ile yürütüldüğünde MARS'ın üstünlüğünü kaybettiği görülmüştür.

Chang vd. (2012), çalışmalarında otoyol kazaları üzerindeki otoyol geometrisi özellikleri, trafik faktörleri ve çevre koşulları gibi davranış dışı faktörlerin etkilerini araştırmak için parametrik olmayan çok değişkenli uyarlanabilir regresyon eğrileri (MARS) modelleme tekniğini kullanmışlardır. Tayvan'daki National Freeway yoluna ait 2007-2008 kaza verileri analiz için elde edilmiştir. Sonuçlar, yatay hizalama, dikey hizalama, günlük ortalama trafik hacmi, ağır vasıta ve yıllık yağışların otoyol kazaları üzerinde önemli etkileri olduğunu göstermektedir.

Andres vd. (2011), iflas tahminlerinde sıkça kullanılan hibrit modele ek olarak bulanık kümeleme ve MARS tekniklerini bir araya getiren bir hibrit sistem önermişlerdir. Çalışmada 59.396 iflas ettirilmemiş İspanyol şirketten oluşan ve 2007 yılında iflas etmiş 138 sıkıntılı firmadan elde edilen gerçek veriler kullanılmıştır. Her iki model de, kullanılan bilgilerin şirket mali tablolarından alındığı durumlarda, teorik avantajlarından dolayı, iflas tahmini için uygun bulunmuştur. Sonuçlar, hibrit modelin, hem doğru sınıflandırma yüzdesi hem de kredi verme kararları tarafından üretilen kar açısından diğer tekniklerden daha iyi performans sergilediğini göstermiştir.

Shahsavani vd. (2010), model çıktısının duyarlılık analizi için MARS modelleme tekniğini kullanmışlardır. Eldeki orijinal model verilerinde herhangi bir değişiklik, çalışmayı baştan yapmayı gerektirebilir ki bu da ek zaman ve maliyet gerektirmektedir. Duyarlılık analizi, modelde herhangi bir değişiklik durumunda orijinal model üzerinden yeni çözüme ulaşmayı hedefler. Bu çalışmada, MARS yöntemi doğrusal olmayan test fonksiyonunun davranışını taklit etmek için kullanılmış ve duyarlılık endekslerinin tahmininde düşük maliyet sağladığı görülmüştür.

Bilgi kriterleri ile ilgili çalışmalar:

Aydoğan Ünal (2020), 2018 Türkiye seçmen tercihlerinin oluşturduğu optimal küme sayısı ve kümelerin içerdiği illeri iki aşamalı kümeleme analiziyle belirlemiştir. İlk aşamada Akaike Bilgi Kriteri (AIC) kullanarak çeşitli kümeleme çözümlerinin performansı karşılaştırılmış, ikinci aşamada en iyi performans gösteren kümeleme seçeneği belirlenmiştir. Bu yöntemin sağladığı avantajların başında küme sayısına dair ön bilginin olmadığı durumlarda, optimal küme sayısının veri setinin analizi neticesinde belirlenebildiğine vurgu yapan Ünal, bireysel olarak seçmenler seçimden seçime tercihlerini değiştirseler bile illerin uzun yıllar boyunca aynı kümede kaldıkları sonucuna ulaşmıştır. Bu durumu ise, bireylerin rasyonel hesaplamalar yapmak yerine, hazır bilgi kaynağı olarak gördükleri yakın çevrelerine güvenmeleri ve karar verirken bu yollara başvurmalarının etkisi olduğu şeklinde açıklamıştır.

Karaman (2019), cep telefonu kullanımının, uyku düzeni üzerine etkisinin olup olmadığını, var ise hangi değişkenlerce açıklandığını incelemiştir. Çalışmada 491 kişiye uygulanan anket verileri analiz edilmiştir. Veri seti tamamı kategorik 1 bağımlı, 25 bağımsız değişkenden oluşmaktadır. Çalışmanın temel savı, en az değişkenle en uygun modeli saptayabilmektir. Bunun için Hampursam Bozdoğan (1988) tarafından geliştirilen model yalınlığı ve uyum iyiliğinin yanısıra modelin karmaşıklığını da dikkate alan ICOMP (Information Complexity) ve onun özel hali ICOMP (IFIM) belirlenmiştir. Final modelde ICOMP (IFIM) sonucuna göre 5 bağımsız değişken yer almıştır. Final modelde elde edilen değişkenler üzerinden cep telefonu bağımlılığının uyku kalitesi üzerindeki olumsuz etkisi saptanmıştır.

Güney (2018), gelişen teknoloji ile bir istatistiksel problemin çözümü için daha karmaşık modellerin ortaya çıktığını bu durumun ise alternatif model sayısında artışa neden olduğundan bahsetmiştir. Bu modellerden hangisinin veriye en uygun, en basit model olduğunu belirlemek için ICOMP kriterini kullanan Güney, ayrıca dağılım varsayımının sağlanmaması, verinin aykırı değer içermesi gibi durumlar söz konusu olduğunda güvenilir sonuçlar elde etmek için robust model seçim kriterlerine ihtiyaç duyulduğunu ifade etmiştir. Yaygın kullanılan kriterlerin robust formları ile ilgili zengin literatür olmasına rağmen, Bilgi Karmaşıklığı (ICOMP) kriterinin robust formlarına ilişkin literatürde az sayıda çalışma vardır. Bu çalışmada farklı birtakım robust istatistiksel yöntemlere dayalı robust ICOMP kriteri kullanılmıştır.

Pamukçu (2017), istatistiksel çalışmalarda temel bileşenler analizi, faktör analizi, sınıflama ve kümeleme analizleri, regresyon katsayılarının çıkarımı ve tahmini gibi klasik çok değişkenli tekniklerin birçoğunun verinin kovaryans matrisi ve/veya onun tersinin tahminini gerektiren çalışmalar olduğunu söylemiştir. Değişken sayısının (p) örnek sayısından (n) fazla olduğu durumlarda örnek varyans-kovaryans matrisinin dejenere olmasından ve tersinin hesaplanamayacağından bahsetmiştir. Çalışmalarında, yüksek boyutlu veri setlerindeki kovaryans probleminin üstesinden gelmek için hibrit kovaryans tahmin edicisini (HCE) geliştirmiştir. Simülasyon çalışması ile farklı p/n oranlarına sahip veri setleri HCE ile analiz edilmiş, uygun kovaryans yapısının seçimi AIC CAIC ve ICOMP bilgi kriterleri ile yapılmıştır ayrıca sonuçlar klasik regresyon analizi yöntemi ile karşılaştırılmıştır.

Beak vd. (2016), verilerin normal dağılım sergilemediği durumlarda diskriminant analizinde (DA) değişken alt kümesinin seçilmesi için etkin ve kullanımı kolay yeni bir hibrit yaklaşım önerilmiştir. Önerilen yaklaşım, diskriminant analizi için çekirdek yoğunluğu tahminini (kernel density estimation KDE-DA) ve bilgi karmaşıklığı kriteri (ICOMP) ile genetik algoritmayı (GA) birleştirir. ICOMP, KDA-DA için değişkenlerin en iyi alt kümesinin bulunmasında önemli bir rol oynamıştır. Önerilen yaklaşım diğer model seçme yaklaşımları olan lineer diskriminant analizi (linear discriminant analysis, LDA), kuadratik diskriminant analizi (quadratic discriminant analysis, QDA), ve k-en yakın komşu diskriminant analizi (k-nearest neighbor discriminant analysis, k-NNDA) ile kıyaslanmış ve daha üstün performans gösterdiği gözlenmiştir.

Pamukçu (2015), çalışmasında yapısal olarak aşırı derecede küçük örneklem problemine sahip veri setleri için en temel problem olan boyut indirgeme için HDR (Hybrid Dimension Reduction) yaklaşımını ve aykırı gözlem tespiti için HOD (Hybrid Outlier Detection) yaklaşımını ve model seçim kriterleri olarak ise AIC, CAIC ve bilgi karmaşıklığı kriteri (CICOMP) kullanılmıştır. Uygulama için altı tane mikrodizilim gen ifade verileri tercih edilmiştir. Bunlar; lösemi, kolon, prostat, lenfoma, SRBCT ve beyinden oluşmaktadır. 423 hasta üzerinden toplamda 23.535 gen ifadesi incelenmiştir. Sonuçlar HDR yaklaşımı ile seçilen boyutların başarısının ortaya koymuştur. HOD yaklaşımı aykırı gözlemlerin tespiti için literatüre farklı bir bakış açısı kazandırmıştır. Model seçiminde ise AIC ve CAIC tamamen aynı

performansı sergilemiş ancak CICOMP, AIC tipi kriterlere göre olağan üstü başarı sergilemiştir.

Koc ve Bozdoğan (2014), çalışmalarında yüksek boyutlu veri modellerinde, parametrik olmayan regresyon tekniği olarak kullanılan çok değişkenli uyarlanabilir regresyon eğrileri (MARS) tekniğini ele almışlardır. MARS tekniği model seçimi için geliştirilmiş çapraz geçerlilik (GCV) kriterini kullanır. Koc ve Bozdoğan GCV kriteri yerine model karmaşıklığını da hesaba katan ve GCV kriterine kıyasla daha yalın bir model seçiminde bulunabilen bilgi karmaşıklığı kriterini (ICOMP) kullanmışlardır. Uygulama kısmında 252 erkeğin (yaş, ağırlık, boy, boyun çevresi, göğüs çevresi, karın çevresi, kalça çevresi, uyluk çevresi, diz çevresi, ayak bileği çevresi, biceps çevresi, bilek çevresi, ön kol çevresi) vücut ölçüm gözlemleri bağımsız değişken; vücut yağı ise bağımlı değişken olarak alınmış ve obezite üzerine bir çalışma yapılmıştır. Elde edilen sonuçlarda GCV kriteri doğru modeli 15 temel fonksiyon kullanarak elde ederken, ICOMP kriteri ise 6 temel fonksiyon kullanarak elde etmiştir.

Deniz vd. (2011), çalışmalarında mevcut problemlerin üstesinden gelmek için yapısal eşitlik modellerinde (YEM) kategorik, iki kategorili (dikotom) ve karışık veri kümelerinin analizi yerine yeni ve sayısal olarak hesaplanabilir alternatif bir yaklaşım sunmaktadırlar. Sunulan yaklaşım Gifi sistemine dayanmaktadır. Gifi sistemi, gözlenen kategorik değişkenleri ölçmek için en uygun ölçekleme metodunu kullanmaktadır. Gifi sistemi kategorik değişkenlerin ölçek özelliklerini bozmadan kategorik verileri sürekli verilere dönüştürür. Dönüşümden sonra, dönüştürülmüş veri seti YEM kullanarak multinominal dağılım varsayımına dayanarak analiz edilmiştir. Ayrıca AIC, CAIC ve ICOMP gibi bilgi-teorik model seçim kriterleri tanıtılmış ve yapısal eşitlik modelinde uyum ölçütü olarak geliştirilmişlerdir. Ölçütlerin minimum değerine sahip model, aday modellerin portföyü arasında en uygun model olarak seçilmiştir. Veri setinde Gifi dönüşümü kullanarak yaklaşımlarının çok yönlülüğünü ve esnekliğini yaşam kalitesini ölçen bir çalışma üzerinde gösterilmiştir.

Deniz (2007), yapısal eşitlik modelleri analizinde yer alan akaike bilgi kriteri (AIC) ve tutarlı akaike bilgi kriteri (CAIC) ile bunlara alternatif olarak Bozdoğan tarafından önerilen (ICOMP) kriterinin değerlendirilmesini ve karşılaştırılmasını amaçlamıştır. Veri kümesi için bir içecek şirketinin reklam ve marka iletişim

etkinliklerini sürekli olarak ölçmek ve bu yolla iletişim yatırımlarının verimlilikleri ile ilgili değerlendirme yapmak amacıyla toplanmış 135 bilgidir. Verilerde ürünün, serinletici, tadı harika, yemeklerle iyi giden, beni canlandıran, kendimi iyi hissettiren, kaliteli ürünleri olan marka ve ödenen paraya değer olmak üzere yedi özellik dikkate alınmıştır. Sonuç olarak ICOMP, ICOMP(IFIM), ICOMP(IFIM)_PEU, YEM’de kullanılmakta olan AIC ve CAIC kriterleri kadar performans gösterdiği ICOMP(IFIM)_PEU kriterinin araştırmanın kapsamına göre fazla uyumlu modellerle çalışmanın sorun olmadığı durumlarda kullanılabilir olduğu, ICOMP ve ICOMP(IFIM) kriterleri ile YEM’i değerlendirmenin doğru olmayacağı söylenebilir.

Clark ve Troskie (2006), çalışmalarında bir regresyon simülasyonu ile ICOMP, AIC ve BIC model seçim kriterlerinin, doğrusal bağlantı durumu (collinearity), örneklem büyüklüğü ve artık varyans düzeyleri altındaki davranışlarını incelemiştir. Tasarım matrisindeki değişim büyük olduğunda yani tasarım matrisindeki doğrusallık seviyeleri arttıkça tüm bilgi kriterleri için doğru model seçim yüzdeleri azalırken Kullback-Leibler uzaklığını kullanan ICOMP kriterinin daha doğru karar verdiği söylenmiştir ayrıca örnek büyüklüğü arttıkça ICOMP ‘un AIC ve BIC’e göre daha sık tercih edildiğinden bahsedilmiştir.

Capron vd. (2005), temel bileşenler regresyon modelinin öngörü kabiliyetinin değişken sayısına büyük ölçüde bağlı olduğunu, optimal olmayan bir karmaşıklığın regresyon katsayılarının yüksek varyansı nedeniyle doğru olmayan tahmin edicilere sahip modele yol açacağından bahsetmişlerdir. Tek çıkışlı çapraz doğrulama ve monte carlo çapraz doğrulama gibi popüler çapraz doğrulama yöntemleri, özellikle verilerde atipik örnek mevcutsa ve büyük veri kümesi üzerinde çalışılıyorsa, her zaman doğru sayıda değişken tutmaz. Bahsedilen sebepler nedeniyle temel bileşenler regresyon modelini seçmek için ICOMP kriteri kullanılmıştır. Elde edilen sonuçlar bu bilgi kriterinin en az çapraz geçerlilik kriterleri kadar iyi davrandığını ve genellikle model seçimi ve performansı açısından daha iyi performans sergilediğini göstermişlerdir.

Bozdoğan ve Bearse (2003), vektör otoregresyonlar (vector autoregressions, VAR) olarak bilinen çoklu zaman serilerinde etkili gözlemleri saptamak için yeni bir bilgi tabanlı yaklaşım geliştirmişlerdir. Yaklaşım iki aşamadan oluşmaktadır. İlk olarak uygunluk fonksiyonu için Bozdoğan’ın bilgi karmaşıklığı (ICOMP) kriteri ile

genetik algoritma kullanılmıştır. İkinci aşamada ise etkili gözlemleri tespit etmek için GA tarafından seçilen vektör otoregresyonlar alt kümesi üzerinde ICOMP kullanılmıştır. Çalışmada amaç hem model uyumsuzluğunu hem de model karmaşıklığını hesaba katan bir model seçim kriteri ile çok değişkenli zaman serilerinde etkili gözlemlerin sezgisel, pratik ve titiz iki boyutlu grafiksel gösterimini vermektir.

Bearse ve Bozdoğan (2002), çok değişkenli regresyon modellerinde en iyi tahmin alt kümesini seçmek için bir teknik geliştirmişlerdir. Geliştirilen teknik genetik algoritma (GA) ile bilgi karmaşıklığı kriterinin bütünleşmesine dayanmaktadır. Yaklaşımlarının amacı, genetik algoritma ile karmaşıklık kriterinin teorik-bilgi ölçüsüne dayanan yeni istatistiksel modelleme prosedürlerini bütünleştirmektedir. ICOMP ile bütünleşen bir optimizasyon algoritması olan GA, istatistiksel model seçiminde tek başına sergilediği performanstan çok daha etkili performans sergilemiştir. Çalışmalarında 30 gözlemlili bir veri kümesine sahip Japon pirinç şarabının tat ve koku dokuları incelenmiştir. Bağımsız değişken olarak, pH, asidite, doku ölçüsü, indirgenmiş şeker, toplam şeker, azot, alkol alınmıştır. Sadece GA tabanlı yaklaşımla elde edilen sonuçlar GA-ICOMP ile elde edilen sonuçlara göre çok daha kompleks yapıya sahiptir. ICOMP bilgi kriterinin daha yalın bir alt model seçtiği gözlenmiştir.

Pan (2001), Genel doğrusal modellerin ilişkili veriler için geliştirilen bir uzantısı olarak görülen genelleştirilmiş tahmin denklemleri (GTD) üzerinde çalışmıştır. Bu yöntem normal dağılım koşulunun sağlanmadığı veriler için daha etkin ve yansız regresyon tahminleri üretmeyi amaçlar. Yöntemin üstünlüğü, varyans homojenliğine ihtiyaç duyulmaması ve hata terimleri arasında ilişki olduğu durumlarda da kullanılabilmesinden kaynaklanmaktadır. Bahsedilen ilişkiyi gösteren matris korelasyon matrisidir. Korelasyon yapısının seçimi için oldukça az kriter kullanılabilen ancak kullanımı oldukça yaygın olan Akaike Bilgi Kriteri kullanılamamaktadır. Bunun nedeni AIC kriterinin en çok olabilirlik tahminine dayanmasıdır. Pan çalışmasında AIC de kullanılan en çok olabilirlik tahminini yarı olabilirlik (quasi-likelihood) tahmini olarak GTD için geliştirmiştir. Son olarak geliştirilen tekniğin performansı, gerçek veri seti kullanılarak diyabetik risk unsurlarının belirlenmesinde gösterilmiştir.

Bozdođan (2000), alıřmasında Akaike'nin bilgi ltn (AIC) incelemiřtir. Daha sonra model seim kriteri iin gncel geliřmeleri sunarak, entropi, ters-Fisher ve Kullback-Leibler kavramlarından bahsetmiřtir. Ayrıca Monte Carlo simlasyon rnekleri ve gerek verilerle AIC ve varyantlarının ICOMP ile performanslarını kıyaslamıřtır.

Bozdođan ve Haughton (1998), alıřmalarında oklu regresyon bađlamında  ařamadan oluřan bir makale sunmuřlardır. Birincisi; model karmařıklık ls ve model uyumsuzluđunu birleřtiren model seim kriteri olan Bozdođan'ın geliřtirdiđi ICOMP kriteri tanıtılmıřtır. İkinci ařamada, gerek veri zerinde ICOMP kriteri; Akaike bilgi kriteri (AIC) ve Schwarz bilgi kriterinin (BIC) performansları ve asimptotik tutarlılıkları incelenmiřtir. Son olarak ise gerek veri yerine simlasyon alıřması ile model seim kriterlerinin davranıřları incelenmiřtir. Elde edilen bulgular ICOMP kriterinin daha iyi bir model tahmininde bulunduđu ynnde olmuřtur.

3. LİTERATÜR DEĞERLENDİRMESİ

Mevcut literatür incelendiğinde çok değişkenli uyarlanabilir regresyon eğrileri tekniği, yüksek boyutlu verilerden kaynaklanan doğrusal olmama probleminin çözümündeki başarısı, bağımsız değişkenler ile bağımlı değişken arasında belirli bir ilişki öngörmemesi ve bağımlı değişkeni tanımlamak için bağımsız değişkenlerin eklemeli ve etkileşimsel katkılarına yer vermesi bakımından elverişli bir yöntemi olarak önerilmektedir. Ancak MARS tekniği, model seçiminde kullandığı GCV kriteri algoritmasında kullandığı düzleştirme parametresi (d)'nin keyfi değer alması ve bu teknik ile yüksek boyutlu modeller (aşırı uyum gösteren modeller (overfitting)) elde edilmesi nedeniyle eleştirilmektedir (Barron ve Xiao 1991; Koc ve Bozdoğan 2014). Model seçimi denildiğinde 'bağımlı değişkeni en iyi açıklayan, bağımsız değişkenlerin oluşturduğu, en fazla bilgiyi içeren en yalın modelin bulunmasıdır' tanımından hareketle ve Ockham'ın usturası veya basitlik yasası olarak da ifade edilen 'tam olarak aynı tahminleri yapan iki rakip teoriniz olduğunda, daha basit olan daha iyidir' tanımından yola çıkıldığında elde edilen modelin karmaşıklığının dikkate alınması oldukça önemlidir. Bu noktada aşırı uyum (overfitting) ve uyum eksikliği (lack of fit) dengeleyen bir model elde edebilmek amacıyla parametre cezalandırma yöntemlerini kullanan bilgi kriterleri devreye girmektedir. Konuyla ilgili literatür taraması yapıldığında AIC, SBC gibi bilgi kriterleri ile oldukça fazla çalışma yapıldığı söylenebilir. Bu kriterler modellemede parametre sayısını cezalandırarak daha yalın modelleri seçebilmektedirler. Ancak bahsi geçen bilgi kriterleri değişkenler arasındaki ilişkileri göz ardı ederek model seçimi yapmaları hususunda eleştirilmişlerdir (Koc ve Bozdoğan, 2014). Bozdoğan bu noktadan hareketle model seçiminde sadece uyum iyiliği ve model yalınlığını değil, bunlarla birlikte modelin karmaşıklığını da dikkate alan ICOMP kriterini, son hali ICOMP (IFIM)_PEU'yu geliştirmiştir .

İlgili literatür incelendiğinde modelleme yöntemi olarak MARS veya bilgi kriterleri kullanılarak çok sayıda çalışma yapıldığı görülmektedir. Biz ise çalışmamızda kendine özgü yeniliklerle başarılı bir modelleme yöntemi olan MARS tekniğinin, kullandığı GCV kriteri nedeniyle, eleştirilen yönlerini ortadan kaldırarak yani modelleme için MARS tekniğini alternatif bilgi kriterleri (AIC, SBC ve ICOMP (IFIM)_PEU) ile güncelleyerek hem MARS tekniğinin başarısından faydalanmayı, hem de daha yalın bir model elde etmeyi öneriyoruz.

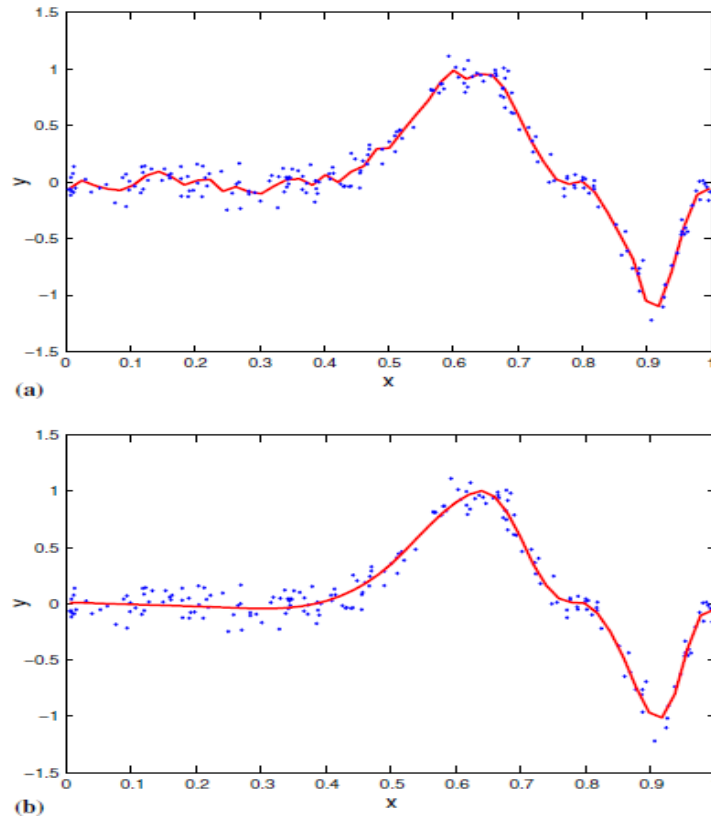
4. ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON EĞRİLERİ

Mars tekniği ilk olarak 1991’de istatistikçi J. H. Friedman tarafından geliştirilmiş, parametrik olmayan bir regresyon yöntemidir. MARS; regresyon yapısını, veri uzayını küçük doğrusal regresyon denklemlerine (temel fonksiyonlara) bölerek oluşturmaktadır. MARS prosedürü adım adım seçme (Stepwise Selection) yöntemine benzemekle birlikte MARS yönteminde modele eklenen değişkenlerin yanısıra değişkenlerin dönüşümleri ve etkileşimlerinden de söz edilmektedir. Bu bağlamda MARS yöntemi için adım adım seçim yönteminin geliştirilmiş halidir denebilir (Kayhan, 2010). Model oluşumu ileriye doğru seçim ve geriye doğru eleme olmak üzere iki adımda gerçekleşir (Kolyshkina vd, 2004).

İleriye Doğru Seçim: Model kurulumuna $B_0(x) = 1$ sabit fonksiyonu ile başlanır. Ardından mümkün olan tüm temel fonksiyonlar oluşturulur. Çok büyük bir model bulununcaya kadar yani modelin karmaşıklığı maksimum seviyeye ulaşıncaya kadar eklenen temel fonksiyonlarla model geliştirilir. Bu aşamada, hemen hemen tüm seçim prosedürlerinde sıklıkla kullanılan Akaike Bilgi Kriteri (AIC) ve Determinasyon (açıklayıcılık) Katsayısı (R^2), arasından tercihimizi AIC yönünde kullandık. Bunun nedeni, modele alınan açıklayıcı değişken sayısı arttıkça R^2 ’nin yükseliyor olması; AIC değerinin ise model seçiminde olabilirlik tahminini baz alarak uyum hassasiyeti ile bu uyumu sağlamak için kullandığı parametre sayısı arasında bir denge oluşturmaya odaklanmasıdır. Temel fonksiyonlar oluşturulurken tüm değişkenler ve değişkenlerin birleşimleri tek tek ele alınır. En güçlü bağıntıyı gösteren değişken veya değişken bileşimi ile işe başlanır. Bu bağıntının AIC değeri hesaplanır. Sonraki aşamada, ikinci derecede güçlü bağıntıya sahip bağımsız değişken de modele katılır. Oluşan yeni denklemde elde edilen AIC, daha öncekine göre anlamlı olarak yüksek bulunursa ikinci bağımsız değişken de modele eklenir. Eğer ikinci bağımsız değişkenin olduğu model de AIC de anlamlı bir yükseliş olmaz ise ikinci bağımsız değişken modele alınmaz. Bu şekilde devam edilerek öncelikle tüm bağımsız değişkenler ve değişkenler arası etkileşimler modele eklenerek en karmaşık model elde edilir (Xu, 2004).

Geriyeye Doğru Eleme: Geriyeye doğru eleme algoritmasında ise amaç ileriye doğru adımda oluşturulmuş olan modelin karmaşıklığını azaltmaya çalışmaktır. Bu amaçla ileriye doğru adım algoritmasının devamı olarak aşağıdaki sıra takip edilir.

İlk olarak; modele giren bütün temel fonksiyonlar (sabit temel fonksiyon hariç) kümesi araştırılır. Burada uyum iyiliği kriteri olan Genelleştirilmiş Çapraz Geçerlilik (Generalized Cross Validation: GCV) kriterine en az katkısı olan temel fonksiyonlar elenir. Böylelikle tek tek sonuca en az etkisi olan yani artıkların hata karelerinde (residual squared error) artışa neden olan temel fonksiyonlar (sabit temel fonksiyon hariç) her adımda elenir. Bu işlem, en iyi alt model bulununcaya yani genel model için olan GCV kriteri en küçük değerine ulaşıncaya kadar devam eder (Xu, 2004). Aşağıda verilen (a) şekli ileri doğru seçim adımında oluşturulan maksimum karmaşıklığa sahip model ve (b) şekli de geriyeye doğru eleme adımı sonunda karmaşıklığı azaltılmış model olarak düşünülebilir.



Şekil 4.1: (a) İleri Doğru Seçim Modeli, (b) Geriyeye Doğru Eleme Modeli (Koc ve Bozdoğan, 2014)

4.1 MARS Modeli

$$Y = \beta_0 + \sum_{m=1}^M \beta_m B_m(x) + \varepsilon_i \quad (4.1)$$

Burada;

M : Temel fonksiyon sayısını,

x : Bağımsız değişkeni,

$B_m(x)$: m . Temel fonksiyonu,

β_m : m . Temel fonksiyon katsayısını veya sabiti ($m=0$),

β_0 : Modeldeki sabit terimi ifade etmektedir.

Temel Fonksiyon ise;

$$B_m(x) = \prod_{k=1}^{K_m} [S_{km} \cdot (x_{v(k,m)} - t_{k,m})]_+ \quad (m = 1, 2, \dots, M) \quad (4.2)$$

Buradaki;

K_m : Toplam kesikli doğrusal fonksiyonların sayısı,

$S_{k,m}$: 1 ya da -1 değerlerini alıp adım fonksiyonunun sağ ya da sol tarafını temsil eder, $S_{k,m} \in \{\pm 1\}$.

$x_{v(k,m)}$: Bağımsız değişken vektörünü,

$t_{k,m}$: Düğüm değerlerini göstermektedir.

Temel fonksiyon Eş. 4.1'de yerine yazılırsa;

$$Y = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{K_m} [S_{k,m} \cdot (x_{v(k,m)} - t_{k,m})]_+ + \varepsilon_i \quad (4.3)$$

eşitliğindeki MARS modeli elde edilmiş olur (Friedman, 1991). Modelin yapısını daha iyi anlayabilmek için Eş. 4.3, Eş. 4.4'deki gibi de ifade edilebilir.

$$Y = \beta_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i, x_j) + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k) + \dots \quad (4.4)$$

Eşitlikteki ilk terim sabit terimdir. İkinci terim $\{\sum_{K_m=1} f_i(x_i) = \sum_{K_m=1} \beta_m B_m(x_i)\}$, tek değişken içeren tüm temel fonksiyonların toplamını ifade etmektedir. Üçüncü terim $\{\sum_{K_m=2} f_{ij}(x_i, x_j) = \sum_{K_m=2} \beta_m B_m(x_i, x_j)\}$ ise değişken ikilileri ve varsa bu değişkenlerin etkileşimlerini gösteren tüm temel fonksiyonlar olarak tanımlanmaktadır. Dördüncü terim, üç değişkenli etkileşimleri temsil eden temel

fonksiyonları içermektedir ve modelin yapısı bu şekilde devam etmektedir (Ünal, 2009).

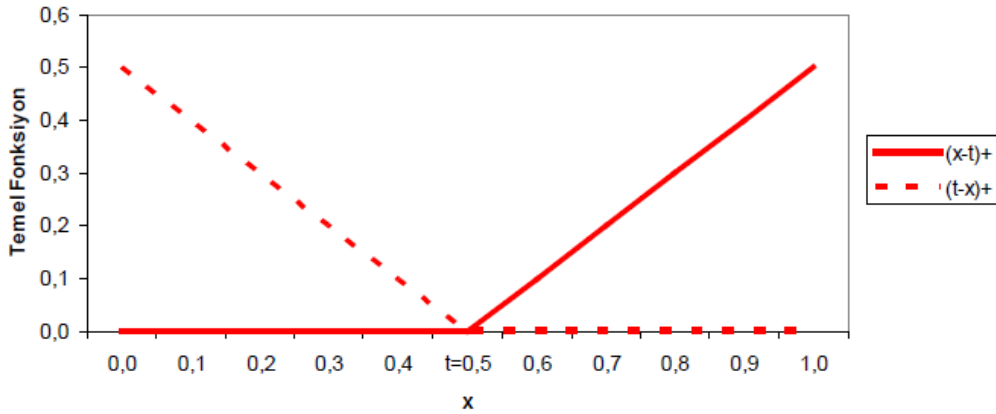
4.2 Temel Fonksiyon

MARS ile oluşturulan yapısal model, $(x - t)_+$ ve $(t - x)_+$ formunda gösterilen parçalı doğrusal temel fonksiyon açılımını kullanır. “+” alt indisi pozitif kısmı belirtip, istenen koşulun sağlanamadığı durumda temel fonksiyonun sıfır sonucunu alacağını belirtir ve aşağıdaki gibi tanımlanır (Friedman, 1991).

$$(x - t)_+ = \begin{cases} x - t, & x > t \\ 0, & \text{diğer durum} \end{cases} \quad (4.5)$$

$$(t - x)_+ = \begin{cases} t - x, & x < t \\ 0, & \text{diğer durum} \end{cases} \quad (4.6)$$

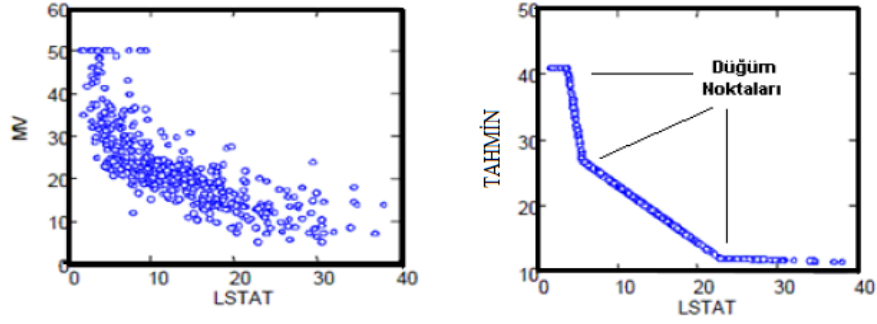
Burada “t” düğüm (knot) değeri olup, her fonksiyon “t” değerinde parçalı doğrusaldır. $(x - t)_+$ ve $(t - x)_+$ temel fonksiyonları aynı zamanda, sırasıyla “t” düğümünün sağ ve sol bölgelerini ifade edip, yansıtılmış bir çift (ayna görüntü) olarak da adlandırılmaktadır (Sevimli, 2009). Şekil 4.2’de $t=0,5$ için $(x - t)_+$ ve $(t - x)_+$ temel fonksiyonları gösterilmiştir.



Şekil 4.2: Temel Fonksiyonlar, $(x - 0,5)_+$ ve $(0,5 - x)_+$ (Ünal, 2009)

4.3 Düğüm Noktası

Aynı bağımsız değişken üzerinde ilişkinin şeklinin değiştiği bağımsız değişken değerine düğüm denir. Bir başka ifadeyle bağımsız değişkenin, tanımlı bulunduğu aralıklarda, doğrunun eğimini değiştirmeyen en son değerine düğüm değeri denir. Aşağıda gerçek bir veri ve bu veriden oluşturulan üç düğüm noktası olan bir MARS kesiti gösterilmiştir.



Şekil 4.3: MARS İle Belirlenen Düğüm Değerleri Aracılığıyla Oluşturulan Üç Düğümlü Parçalı Doğrusal Regresyon Modeli (Solda Gerçek Veri) (Chou vd.,2004)

4.4 Genelleştirilmiş Çapraz Geçerlilik (GCV) Kriteri

MARS modelinin oluşum sürecinde ikinci adım olan geriye doğru adım algoritmasının, ilk adımda oluşturulan maksimum modelin karmaşıklığının azaltıldığı aşama olduğundan bahsedilmiştir. Bu aşamada Craven ve Wahba (1979) tarafından geliştirilen Genelleştirilmiş Çapraz Geçerlilik (GCV) ölçütü kullanılarak, modele en az katkısı olan temel fonksiyonlar tek tek elenerek modelden çıkarılır.

$$GCV(M) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{f}_M(x_i))^2}{\left(1 - \frac{P(M)^*}{n}\right)^2} \quad (4.7)$$

$$P(M) = iz(\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T) + 1$$

$$P(M)^* = P(M) + dM \quad (4.8)$$

Burada;

y_i : i. gözlemlenen bağımlı değişken vektörü

\hat{f}_M : Temel fonksiyonlara dayanan MARS tahmin modeli

x_i : i. gözlemlenen bağımsız değişken vektörü; $x_i = (x_{i1}, \dots, x_{ip})^T, i = 1, \dots, n$

n: Veri setindeki gözlem sayısı

M: Modele giren temel fonksiyon sayısı

B: Modele ait BF'lerin oluşturduğu matris

P(M): Modele giren temel fonksiyonların sayısı ile artan ceza ölçüsü

d: Düzleştirme parametresi (smoothing parameter)

d 'nin büyük deęeri daha az sayıda temel fonksiyon oluřacaęı, dolayısıyla daha düz model tahmini anlamına gelmektedir. Model düzleřtirildikçe kaybolan deęerler ceza olarak adlandırılır. Yapılan alıřmalar d için en iyi deęerin $2 \leq d \leq 4$ aralıęında olduęunu göstermiřtir (Friedman, 1991). Bylelikle Eř. 4.7'daki modelin pay kısmı, M tane temel fonksiyon ($\hat{f}_M(x_i)$) modelinin uyumsuzluęunu lerken payda kısmı ise, model karmařıklıęını belirtir. En uygun MARS modeli ise en kük GCV lmne sahip olan deęerdir.

5. MODEL SEÇİMİ ve BİLGİ KRİTERLERİ

Model seçimi, açıklayıcı değişkenlerden hangisi veya hangilerinin yanıt değişkeni üzerinde etkili olduğunu ortaya koyarak mevcut modeller alt kümesi arasından en iyi modelin bulunmasıdır. Açıklayıcı değişken sayısının çok olduğu durumlarda en iyi modelin belirlenmesi en büyük problemlerden biri haline gelmektedir. Ayrıca en çok karşılaşılan problemlerden bir diğeri de çoklu doğrusal bağlantı (multicollinearity) durumudur.

Günümüzde birçok sahada, onlar veya yüzlerle ifade edilen gözlem (n) elde edilirken, tek bir gözlem fazlaca özelliğe-boyuta (p) sahip olabilmektedir ($n \ll p$). Örneğin gözlem olarak bir insan ele alalım. Bu gözlemin kan değerleri, kilosu, boyu, tansiyonu, ... gibi özelliklerinin ölçülmesi gerekebilmektedir. Özellikle açıklayıcı değişkenlere ilişkin alt küme sayısının milyonları bulabildiği durumlarda ($p = 20$ için $2^{20} = 1048576$ adet alt küme kombinasyonunun olması gibi) model seçimi için nümerik tekniklere gereksinim duyulur. Model seçim ile ilgili birçok çalışmaya rastlamak mümkündür. Bunlar arasında klasik model seçim metotları ve bilgi kriterlerine dayanan model seçim metotları yer almaktadır.

Klasik seçim metotları genellikle hipotez testlerinin kullanımına dayanmaktadır ve açıklayıcı değişkenlerin sonuç modelinde yer alıp almayacağına karar vermek için keyfi bir güven düzeyi seçilir. Literatürde klasik model seçiminde kullanılan güven düzeyinin temelsiz olduğu yönünde eleştiriler mevcuttur (Linhart ve Zucchini, 1986).

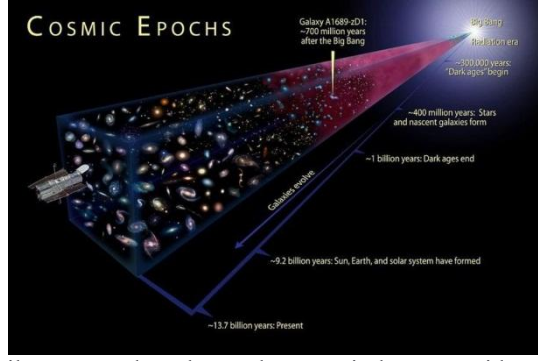
İstatistik paketlerinin birçoğunda ileriye doğru seçim (Forward Selection), geriye doğru eleme (Backward Elimination) ve adımsal teknikler (Stepwise) gibi klasik model seçim prosedürleri mevcuttur. Ancak bu seçim metotları değişkenler arasındaki bağımlılık yapısıyla ilgilenmezler. Ayrıca klasik tekniklerin birçoğu verinin kovaryans matrisinin ve/veya tersinin tahminini gerektirir. Bağımsız değişkenlere ait kovaryans matrisinin determinanı 0 ile 1 arasında değer alır. Determinant 1'e eşit olduğunda değişkenler arasında doğrusal bağımlılığın olmadığı, 0 olduğunda ise çoklu doğrusal bağlantının olduğu ifade edilir. Bu durumda kovaryans matrisi dejenere olur ve tersi hesaplanamaz. Bu, klasik metotlar açısından karşılaşılabilecek en önemli zorluklardan biridir (Pamukçu, 2017). Ayrıca kovaryans

matrisinin maksimum olabilirlik tahmini, p/n büyük olduğu zaman yansız ve pozitif tanımlı olmasına rağmen, doğru bir tahmin edicisi değildir (Stein, 1975).

Bahsi geçen kovaryans problemlerinin üstesinden gelmek için bilgi kriterlerine dayanan model seçim prosedürleri klasik yaklaşımlara alternatif oluşturmaktadır. Bilgi kriterleri, bir modelin hem uygunluğunu hem de bu uygunluğu sağlamak için gerekli olan modelin karmaşıklığını aynı anda dikkate alarak, beklenen entropiyi (verilerde beklenen belirsizliğe benzer) tahmin etmek için kullanılır. Model karmaşıklığı, hem parametre sayısını hem de parametreler arasındaki etkileşimi (korelasyon) içerir. Bu tür kriterlerin temel fikri, gerçek model ve en uygun model arasındaki Kullback-Leibler uzaklığını minimize etmektir (Chaloner ve Verdinelli, 1995). Bu kriterler model seçiminde uyum iyiliğini ve model yalınlığını dikkate almaktadırlar. Ockham'ın Usturası olarak da bilinen “her şeyin birbirine eşit olduğu bir ortamda, en basit açıklama doğruya en yakın olandır” prensibi ile modelin aşırı uyum (overfitting) ve uyum eksikliğini (lack-of-fit) dengeleyen en yüksek bilgiye sahip en yalın model elde edilmelidir. Bu kriterlerin yapısının daha iyi anlaşılabilmesi için öncelikle bilgi teorisinde önemli yere sahip olan Entropi, Kullback-Leibler ve Fisher Bilgi Matrisi kavramları incelenecektir.

5.1 Entropi

Evrendeki her şey düzenli (olasılığı daha az) bir durumdan düzensizliğe (olasılığı daha yüksek bir duruma) doğru evrilmektedir. Örneğin büyük patlama (big-bang) olayında enerji bir çekirdek içindeyken patlama sonucu enerji sürekli yayılmaktadır (Şekil 5.1), enerjinin düzensizliği artmaktadır. Yani entropi sürekli artan bir olgudur.



Şekil 5.1: Büyük patlama olayı üzerinden entropi kavramı
<https://www.forbes.com/sites/startswithabang/2017/04/15/>

Bilgi kuramında Shannon (1948) tarafından tanımlanan entropi kavramı, bir rastlantı değişkeni için belirsizlik ölçüsü olarak tanımlanmaktadır. N tane muhtemel değerlere sahip bir X olayını ele alalım. X 'in kesikli değerleri x_i ($i = 1, 2, \dots, n$) olsun. x_i 'nin her bir değeri, olasılık şartını sağlayan $p(x_i)$ olasılıklarına sahip olsunlar. Yani; $0 \leq p(x_i) \leq 1$ ve $\sum_{i=1}^n p(x_i) = 1$ olsun. $p(x_i)$ olasılığı ile $X = x_i$ olayı gözlemlendikten sonra elde edilen bilgi miktarı,

$$I(x_i) = \log\left(\frac{1}{p(x_i)}\right) = -\log p(x_i) \quad (5.1)$$

olarak tanımlanır. x_i değeri tarafından tanımlanan bilgi miktarı, olayın meydana gelme olasılığının tersi ile ilişkilidir. Yani $p(x_i)$ olasılığının düşük olması daha yüksek bilgi içermesi anlamı taşımaktadır (Thomaz, 2004).

X olayının entropisi ise bilginin ortalaması veya beklenen değeri olarak tanımlanır ve $H(X)$ ile gösterilir.

$$H(X) = E\{I(X)\} = \sum_{i=1}^n p(x_i)I(x_i) = -\sum_{i=1}^n p(x_i)\log p(x_i) \quad (5.2)$$

benzer şekilde, sürekli X olayı için entropi aşağıdaki gibidir.

$$H(X) = E\{I(X)\} = \int_{-\infty}^{\infty} p(x_i)I(x_i) dx = -\int_{-\infty}^{\infty} p(x_i) \log p(x_i) dx \quad (5.3)$$

5.2 Kullback-Leibler Ölçüsü

Model ve gerçek dağılım arasındaki uzaklığın bir ölçümü varsa, iyi bir tahmin bu uzaklığı minimum yapmalıdır. Bu iki dağılım arasındaki uzaklığın ölçülmesi için

kullanılan Kullback-Leibler (KL) ölçüsü, ilk defa Kullback ve Leibler (1951) tarafından tanıtılmıştır. Mesafe anlamına gelen ‘distance’ kelimesinin ilk harfi olan D_{KL} ile gösterilir. $D_{KL}(p \parallel q)$ gösterimi ile p dağılımı kullanılarak q dağılımının tahmin edilmesi halindeki bilgi kaybı ifade edilir.

$p(x)$ model dağılımını ve $q(x)$ gerçek dağılımı ifade etmek üzere, bu iki dağılım arasındaki KL uzaklığı, kesikli ve sürekli durumlar için sırasıyla aşağıdaki gibidir.

$$D_{KL}(p \parallel q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \quad (5.4)$$

$$D_{KL}(p \parallel q) = \int p(x) \log \frac{p(x)}{q(x)} dx \quad (5.5)$$

5.3 Fisher Bilgi Matrisi

Herhangi bir kestiricinin minimum varyans sınırını bulabilmek için Fisher Bilgi Matrisine ihtiyaç duyulur (Tiku ve Suresh, 1992). n bağımsız bileşenden oluşan $X = (x_1, x_2, \dots, x_n)$ rassal değişkeninin olasılık yoğunluk fonksiyonu $\prod_{i=1}^n f(x_i; Q) = f(X; Q)$ olmak üzere $L(Q; X) = f(X; Q)$ ve $\log L(Q; X)$ fonksiyonları sırasıyla X rassal değişkeninin olabilirlik ve log-olabilirlik fonksiyonları olarak tanımlansın. X rassal değişkeni için Fisher bilgi matrisi (Fisher Information Matrix – FIM) aşağıdaki gibi tanımlanır.

$$F(Q) = -E \left(\frac{d^2}{dQ^2} \log L(Q; X) \right) \quad (5.6)$$

Eşitlik (5.6) dan yola çıkarak kovaryans matrisinin tahmini ‘ki bu Ters Fisher Bilgi Matrisi’ne (Inverse Fisher Information Matrix - IFIM) eşdeğerdir’ aşağıdaki gibidir.

$$IFIM = F^{-1} = \left\{ -E \left(\frac{d^2}{dQ^2} \log L(Q; X) \right) \right\}^{-1} \quad (5.7)$$

Q parametresinin tahmini \hat{Q} ile hesaplanan kovaryans matrisi $\hat{\Sigma}$ ’nın kullanılmasıyla, IFIM aşağıdaki gibi elde edilir.

$$\hat{F}^{-1} = \begin{bmatrix} \hat{\Sigma}(\hat{Q}) & 0 \\ 0 & \left(\frac{2}{n} D_{p+q}^+ [\hat{\Sigma}(\hat{Q}) \otimes \hat{\Sigma}(\hat{Q})] D_{p+q}^+ \right) \end{bmatrix} \quad (5.8)$$

Burada D ile duplikasyon matrisi, D^+ ile $(D_p' D_p)^{-1} D_p'$ şeklinde tanımlanan Moore-Penrose tersi ve \otimes ile kronecker çarpım ifade edilmektedir (Williams ve Bozdoğan, 1995).

5.3.1 Dublikasyon Matrisi ve Moore-Penrose Tersisi

A $n \times n$ karesel bir matris ve $V(A)$, $Vec(A)$ 'dan elde edilen $\frac{1}{2}p(p+1) \times 1$ boyutlu vektör. $V(A)$, diyagonal elemanların üzerindeki elemanların çıkarılması ile elde edilir. Örneğin 3×3 lük bir A matris için;

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \begin{aligned} Vec(A) &= (a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, a_{13}, a_{23}, a_{33})' \\ V(A) &= (a_{11}, a_{21}, a_{31}, a_{22}, a_{32}, a_{33})' \end{aligned}$$

Böylece simetrik A matrisi için $V(A)$ da sadece farklı elemanlar bulunur. $V(A)$ 'yı $Vec(A)$ 'ya dönüştüren $p^2 \times \frac{1}{2}p(p+1)$ boyutlu, sıfır ve birim elemanlardan oluşan eşsiz bir matris vardır. Bu matrise dublikasyon matrisi denir ve D ile gösterilir. (Magnus ve Neudecker, 1980; 1988).

$$DV(A) = Vec(A) \quad (A = A') \quad (5.9)$$

Benzer şekilde $Vec(A)$ 'yı $V(A)$ 'ya dönüştüren matris de dublikasyon matrisinin Moore-Penrose tersi olarak tanımlanır ve D^+ ile gösterilir.

$$D^+ = (D' D)^{-1} D' \quad (5.10)$$

$$D^+ Vec(A) = V(A) \quad (A = A') \quad (5.11)$$

5.3.2 Kronecker Çarpım

$A = [a_{ij}]$ ve $B = [b_{ij}]$ matrisleri sırasıyla $m \times n$ ve $p \times q$ boyutlu iki matris olsun.

$$[A \otimes B]_{m \times p \times n \times q} = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} \quad (5.12)$$

çarpımına kronecker çarpım denir. Örneğin $[A]_{1 \times 3} = [0 \ 1 \ 2]$ ve $[B]_{2 \times 2} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ matris olmak üzere $[A \otimes B]_{2 \times 6} = [0B \ 1B \ 2B] = \begin{bmatrix} 0 & 0 & 1 & 2 & 2 & 4 \\ 0 & 0 & 3 & 4 & 6 & 8 \end{bmatrix}$ şeklinde hesaplanır.

5.4 Akaike Bilgi Kriteri (AIC)

İstatistiksel model değerlendirme sürecinde eldeki veriye en iyi uyumu gösteren değişkenlerin seçiminde modelleme hatası ve yan-varyans kavramlarının oluşturduğu kestirim hatası (Q parametre vektörünün kestiriminden kaynaklanan hata) olmak üzere iki tip hata dikkate alınır. Bu iki tip hata toplam riski (R) oluşturur.

Modelleme Riski, değişken seçim kriterleri kullanılırken genellikle modelin önceden belirlenen bir parametrik dağılıma uyduğu varsayılarak model kestirilir. Ancak veri analizi süreci boyunca sıklıkla, eldeki veriler bilinen dağılımlara uyum göstermez. Bu gibi durumlarda modelin dağılımının doğru belirlenebilmesi anlamında hatalı modelleme riski ortaya çıkmaktadır.

Kestirim Riski ise gerçek parametre vektörünün belirlenen model üzerinden sınırlandırılmış bir parametre uzayı için kestirildiği durumlarda ortaya çıkar. Burada risk, kestirimde ‘varyans’ bileşeni olarak adlandırılır. Gerçek parametre vektörünün, modelin sınırlandırılmış parametre uzayı tarafından içerilmediği durumda ‘yan’ oluşur. Başka bir ifade ile varyans, modelin kabul edilen parametre uzayının büyüklüğüne karşılık getirilen ceza olarak yorumlanırken; yan, modelin gerçek parametre vektörü ile indirgenmiş parametre uzayı arasındaki uzaklıktır.

Model değerlendirme sürecinde, değişkenlerin seçimindeki amaç toplam riskin en küçüklenmesidir. Bu bağlamda değişken seçim kriterleri en çok olabilirlik kestirimi altında modelin toplam riskinin kestiricileridir (Bozdoğan, 2000).

Akaike tarafından 1973 yılında önerilen AIC, Q parametre vektörünün en çok olabilirlik tahmininin olabilirlik fonksiyonunun logaritmasının -2 katının yansız kestiricisidir. AIC, modelin uyum eksikliğini değerlendirilmesi ve parametre sayısının cezalandırılması esasına dayalı bir kriterdir (Akaike, 1973).

$$AIC = -2\log L(\hat{Q}) + 2k \quad (5.13)$$

Eşitlikteki ilk terim¹ uyum eksikliğini, yanlılığı ikinci terim ise birinci terimdeki yanlılığı telafi etmenin bir ölçümü olarak ceza terimini² ifade etmektedir.

¹ Eldeki bilgiyle olabilirlik fonksiyonu oluşturup modelin uyumunu değerlendirdiğimiz aşamadır. Mevcut bilgi ne kadar iyiyse $2\log L(\hat{Q})$ değeri okadar yüksek çıkacak başındaki ‘-’ ise kriterin değerini okadar küçültecektir. Bilgi kriteri için en küçük değere sahip model en iyi model seçildiği için gözlem sayısı arttıkça eldeki bilgi artacak ve daha küçük değerli kriter elde edilecektir.

² Bir modelde değişken sayısı arttıkça dolayısıyla parametre sayısı arttıkça o model o kadar iyi açıklama yüzdesine sahip olur, varyansı küçülür. Bu da ilgisi olmasa dahi parametre ekleyerek yapay

k , modeldeki bağımsız parametre sayısını göstermek üzere; Q , k boyutlu bilinmeyen parametre vektörü \hat{Q} ise Q 'nın en çok olabilirlik tahmin edicisidir.

En küçük AIC değerine sahip model en iyi model olarak belirlenir. Karşılaştırılan modeller için AIC değerinin enküçüklenmesi ile bir anlamda kestirilen model ile gerçek model arasındaki KL uzaklığı enküçüklenmiş olur (Akaike, 1973).

5.5 Schwarz Bayes Bilgi Kriteri (SBC)

Bayesci bilgi kriteri olarak da bilinen SBC, AIC'in bir türevi olarak Schwartz (1978) tarafından önerilmiştir. Bayes kuramını temel alan SBC'nin ceza terimi AIC'e göre daha büyüktür. Dolayısıyla SBC ile seçilecek olan modelin, AIC ile seçilecek modelden daha küçük veya en azından eşit boyutta olması beklenir. Eşitlikte yer alan örneklem genişliğine (n) dayalı $k\log(n)$ terimi ceza terimini artırmanın etkisidir.

$$SBC = -2\log L(\hat{Q}) + k\log(n) \quad (5.14)$$

5.6 Tutarlı Akaike Bilgi Kriteri (CAIC)

AIC formülünde örneklem büyüklüğü arttıkça ilk terim artar fakat ceza terimi sabit kalır. Bunun anlamı ceza teriminin yanlılığı telafi etmedeki etkisinin, 'n' örneklem büyüklüğü arttıkça yetersiz kalmasıdır. Bu nedenle Akaike bilgi kriteri tutarlılığının asimptotik durumda şüpheli olduğu (Bozdoğan, 1987) ve ceza teriminin önündeki 2 rakamının rastgele seçilmiş olması (Rissanen, 1978) yönündenden eleştirilmektedir.

AIC'e göre daha tutarlı bir kriter olarak Bozdoğan (1987) tarafından önerilen CAIC'de parametre sayısına ilişkin ceza terimi, AIC'in ceza terimine göre oldukça fazladır. Bu özelliği ile CAIC, daha az parametrelili model seçilmesini sağlar.

$$CAIC = -2\log L(\hat{Q}) + k[\log(n) + 1] \quad (5.15)$$

CAIC'e göre en iyi model, en küçük CAIC değerini veren modeldir. CAIC formülü SBC'ye benzemekle beraber, CAIC'in ceza terimi SBC'nin ceza teriminden

olarak modeli iyileştirme kandırmacasıdır. Bu işlem 'model seçimi' tanımındaki en az parametreyle en iyi modelin bulunması kavramıyla örtüşmemektedir. Bu nedenle bilgi kriterleri gereksiz parametre sayısı artışını engellemek adına parametre sayısının artırdıkça kriteri büyütecek şekilde bu cezalandırmayı koyar. Böylece sahte olarak modele parametre eklendiğinde ceza terimi bilgi kriterinin değerini büyütür ve o modelin seçimini zorlaştırır (bilgi kriteri için en küçük değere sahip olan model en iyi model seçildiği için). Bu durum da parametrelerin cezalandırılması olarak ifade edilir.

bilinmeyen parametre sayısı kadar daha fazladır. Bu da CAIC ile seçilen modelin SBC'ye kıyasla daha basit bir model olacağı anlamına gelmektedir.

5.7 Fisher Bilgi Matrisine Dayalı Tutarlı Akaike Bilgi Kriteri (CAICF)

CAICF, ceza terimine fisher bilgi matrisinin eklenmesiyle, CAIC'e farklı bir yaklaşım getirilmiştir (Bozdoğan, 1987).

$$\begin{aligned} \text{CAICF} &= -2\log L(\hat{Q}) + k[\log(n) + 2] + \log|F(\hat{Q})| \\ &= \text{AIC} + k\log(n) + \log|F(\hat{Q})| \end{aligned} \quad (5.16)$$

CAIC ve CAICF'nin kullanılmasıyla aşırı uyum (overfitting) gösteren model elde edilmesi olasılığı azaltılarak daha yalın bir model elde edilmesi sağlanır. Sonuç olarak örneklem büyüklüğünün fazla olduğu durumlarda CAIC veya CAICF'nin kullanılması AIC'in kullanılmasına tercih edilebilir.

Buraya kadar bahsi geçen bilgi kriterlerinin en büyük eksiklikleri değişkenler arasındaki ilişkilerin göz önünde bulundurulmamasıdır. Bilindiği üzere bağımsız değişkenler arasında doğrusal (yada doğrusala yakın) ilişki olması çoklu doğrusal bağlantı (multicollinearity) durumu meydana gelmekte dolayısıyla model tahminlerinde sapmalara neden olmaktadır. Bu noktadan hareketle Bozdoğan (1988), sadece uyum iyiliği ve model yanlılığını değil aynı zamanda modelin karmaşıklığını da göz önüne alarak ICOMP kriterini geliştirmiştir.

5.8 Bilgi Karmaşıklığı Kriteri (ICOMP)

ICOMP (Information Complexity) Bozdoğan tarafından 1988 yılında model seçimi için oluşturulmuş bir kriterdir. ICOMP bilgi kriterine geçmeden önce konunun daha iyi anlaşılabilmesi için ICOMP'u oluşturan bilgi ve karmaşıklık-komplekslik kavramlarından bahsedilmesinde fayda görülmektedir.

5.8.1 Bilgi Ölçümü ve Bilgi Karmaşıklığı

Van Endem (1971), p değişkenli bir Normal dağılım için karmaşıklığın tanımını kovaryans matrisi (Σ) aracılığı ile yapmıştır. Karmaşıklık diğer bir ifade ile komplekslik, bileşenler arasındaki bağımlılığın veya etkileşimin ölçüsü olarak tanımlanmaktadır. $X = X_1, X_2, \dots, X_p$ rasgele değişkeninin $f(X) = f(X_1, X_2, \dots, X_p)$

ortak yoğunluk fonksiyonu ve $f_j(X_j)$, $j = (1, 2, \dots, p)$ marjinal yoğunluk fonksiyonu olmak üzere değişkenler arasındaki bağımlılığın ölçüsü:

$$I(X) = I(X_1, X_2, \dots, X_p) = E_f \left[\log \frac{f(X_1, X_2, \dots, X_p)}{f_1(X_1) \cdot f_2(X_2) \dots f_p(X_p)} \right] \quad (5.17)$$

veya

$$I(X) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(X_1, X_2, \dots, X_p) \log \frac{f(X_1, X_2, \dots, X_p)}{f_1(X_1) \cdot f_2(X_2) \dots f_p(X_p)} dX_1 \dots dX_p \quad (5.18)$$

şeklinde tanımlanır. Aynı zamanda $I(X)$ bilgi miktarı veya Kullback-Leibler (KL) ölçüsü olarak da ifade edilmektedir (Kullback, 1968; Harris, 1978; Theil ve Fiebig, 1984). Her X_1, X_2, \dots, X_p rasgele değişkeni istatistiksel olarak bağımsız ise $f(X_1, X_2, \dots, X_p) = f_1(X_1) \cdot f_2(X_2) \dots f_p(X_p)$ olacağı için $I(X) = 0$ olacaktır. Bu durumda değişkenler arasında bir etkileşimden bahsedilemez. $I(X) \neq 0$ ise değişkenler arasında etkileşim vardır yorumu yapılabilir.

KL ölçüsü, entropi kavramı ile ilişkilidir. $X = (X_1, X_2, \dots, X_p)$ rassal değişkeninin olasılık yoğunluk fonksiyonu $f_x(X)$ olmak üzere X olayının entropisi $H(X) = -E\{\log f_x(X)\} = -\sum f_x(X) \log f_x(X)$ 'dir. Bu ifade X değişkeninin $f_x(X)$ dağılımına ne kadar uyduğu bilgisini vermektedir. $H(X_j)$, $j = 1, \dots, p$ marjinal entropiyi $H(X)$, $X = (X_1, X_2, \dots, X_p)$ ortak entropiyi göstermek üzere, entropi kavramını kullanarak bilgiyi aşağıdaki gibi gösterebiliriz.

$$I(X) = I(X_1, X_2, \dots, X_p) = \sum_{j=1}^p H(X_j) - H(X) \quad (5.19)$$

İspat:

$$\begin{aligned} I(X) &= E \left[\log \frac{f(X_1, X_2, \dots, X_p)}{f_1(X_1) \cdot f_2(X_2) \dots f_p(X_p)} \right] \\ &= \sum f(X_1, X_2, \dots, X_p) \log \frac{f(X_1, X_2, \dots, X_p)}{f_1(X_1) \cdot f_2(X_2) \dots f_p(X_p)} \\ &= \sum f(X_1, X_2, \dots, X_p) [\underbrace{\log f(X_1, X_2, \dots, X_p)}_{f(X)} - \underbrace{\log f_1(X_1) \cdot f_2(X_2) \dots f_p(X_p)}_{\log f_j(X_j)}] \\ &= f(X) \log f(X) - \sum f(X) \log f_j(X_j) \\ &= -H(X) - (-\sum H(X_j)) = \sum_{j=1}^p H(X_j) - H(X) \end{aligned}$$

Yapı içinde ne kadar çok bağımlılık varsa marjinal entropilerin toplamı da o kadar büyük olacaktır.

Çalışmamızda kullanılmak üzere çok değişkenli normal dağılımın bilgi ölçüsüne geçmeden önce, p değişkenli normal dağılım fonksiyonun

$$f(X) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (X - \mu)' \Sigma^{-1} (X - \mu) \right\}; X \sim N_p(\mu, \Sigma)$$

olduğunu hatırlayalım. Burada; $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$, $-\infty < \mu_j < \infty, j = 1, 2, \dots, p$ ve Σ pozitif tanımlı olmak üzere, çok değişkenli dağılımın komplekslik bilgi ölçüsünü tanımlamak için $f(X) = f(X_1, X_2, \dots, X_p)$ 'yi çok değişkenli normal dağılım fonksiyonu olarak ele alalım. $I(X) = \sum_{j=1}^p H(X_j) - H(X)$ bilgisinden yola çıkarak $\mu = 0$ için çok değişkenli normal dağılımın $H(X) = H(X_1, X_2, \dots, X_p)$ ortak entropisi ve $H(X_j)$ marjinal entropisi aşağıdaki gibi elde edilir (Bozdoğan, 2004).

$$\begin{aligned} H(X) &= -E\{\log f(X)\} = -\int f(X) \log f(X) dx \\ &= \int f(X) \left[\frac{p}{2} \log(2\pi) |\Sigma| + \frac{1}{2} (X - \mu)' \Sigma^{-1} (X - \mu) \right] dx \\ &= \frac{p}{2} \log(2\pi) |\Sigma| + \frac{1}{2} \text{tr} \left[\int f(X) \Sigma^{-1} (X - \mu)(X - \mu)' dx \right]. \end{aligned}$$

$E[(X - \mu)(X - \mu)'] = \Sigma$ eşitliği kullanılarak

$$H(X) = \frac{p}{2} [\log(2\pi) + 1] + \frac{1}{2} \log |\Sigma| \quad (5.20)$$

$$H(X_j) = -E\{\log f(X_j)\} = -\int f(X_j) \log f(X_j) dx_j$$

$$H(X_j) = \frac{1}{2} [\log(2\pi) + 1] + \frac{1}{2} \log |\sigma_{jj}|, \quad j = 1, \dots, p \quad (5.21)$$

$$\begin{aligned} I(X) &= \sum_{j=1}^p \frac{1}{2} [\log(2\pi) + 1] + \frac{1}{2} \log |\sigma_{jj}| - \frac{p}{2} [\log(2\pi) + 1] - \frac{1}{2} \log |\Sigma| \\ &= \sum_{j=1}^p \frac{1}{2} \log |\sigma_{jj}| - \frac{1}{2} \log |\Sigma| \end{aligned} \quad (5.22)$$

$$C_0(\Sigma) = \sum_{j=1}^p \frac{1}{2} \log |\sigma_{jj}| - \frac{1}{2} \log |\Sigma| \quad (5.23)$$

σ_{jj} , Σ 'nin j . köşegen elemanı; p , Σ 'nin boyutu ve $C_0(\Sigma)$ ise Van Emden tarafından tanımlanan kovaryans matrisi için komplekslik ölçüsüdür. Eğer değişkenler bağımsız ise yani Σ bir diyagonal matris ise $C_0(\Sigma) = 0$ olur. Eğer

değişkenlerden bir tanesi diğer değişkenlerin lineer bir fonksiyonu olarak yazılabiliyorsa $|\Sigma| = 0$ olacağı için $C_0(\Sigma) = \infty$ olur (Van Endem, 1971).

$C_0(\Sigma)$, X_1, X_2, \dots, X_p rassal değişkenlerin ortak ve marjinal dağılımlarına bağlıdır. Ayrıca $C_0(\Sigma)$ 'nin ilk terimi ortonormal dönüşüme uğrayabilir. $C_0(\Sigma)$, Σ matrisinin komplekslik miktarının etkili bir ölçüsü olmadığı yönünde eleştiriler mevcuttur (Bozdoğan, 1988). Bunun üzerine Bozdoğan $X = X_1, X_2, \dots, X_p$ değişkenlerini en büyükleme işlemi ile $C_0(\Sigma)$ 'in herhangi bir ortonormal dönüşüm altında maksimumu hesaplanmış ve çok değişkenli normal dağılımlı doğrusal veya doğrusal olmayan modelin kovaryans matrisi için maksimum komplekslik ölçüsü olan $C_1(\Sigma)$ 'i tanımlamıştır. T herhangi bir ortonormal dönüşüm olmak üzere $C_1(\Sigma)$, eşitlik 5.24'deki gibi tanımlanır (Bozdoğan, 2004).

$$\begin{aligned} C_1(\Sigma) &= \max_T C_0(\Sigma) = \max_T \{H(X_1) + H(X_2) + \dots + H(X_p) - H(X)\} \\ &= \max_T \left\{ \frac{1}{2} \sum_{j=1}^p \log(\sigma_{jj}) - \frac{1}{2} \log|\Sigma| \right\} \\ &= \frac{p}{2} \log \left[\frac{\text{tr}(\Sigma)}{p} \right] - \frac{1}{2} \log|\Sigma| \end{aligned} \quad (5.24)$$

İspat: (Van Emden, 1971) (Rissanen, 1978) Σ 'nin ortonormal dönüşümünü bulmalıyız:

$$\sum_{j=1}^p \log(\sigma_{jj}) + \dots + \log(\sigma_{pp}) \quad (5.25)$$

σ_{jj} , $\text{Cov}T(X) = \Sigma \{T(X) = T(X_1, X_2, \dots, X_p)\}$ matrisinin diyagonal elemanlarıdır. Ortonormal dönüşümler $\text{tr}(\Sigma) = \sigma_{11} + \dots + \sigma_{pp}$ de değişiklik meydana getirmezler. $\max \sum_{j=1}^p \log(\sigma_{jj})$ işlemini gerçekleştirmek için $\sigma_{11}, \dots, \sigma_{pp}$ 'nin aritmetik ve geometrik ortalaması kullanılarak $\left(\prod_{j=1}^p \sigma_{jj}\right)^{\frac{1}{p}} \leq \frac{1}{p} \left(\sum_{j=1}^p \sigma_{jj}\right)$ eşitliği göz önünde bulundurulursa

$$\begin{aligned} \max \sum_{j=1}^p \log(\sigma_{jj}) &= \max \log \prod_{j=1}^p \sigma_{jj} \\ &= p \log \text{tr}(\Sigma) - p \log p = p \log \left[\frac{\text{tr}(\Sigma)}{p} \right] \end{aligned} \quad (5.26)$$

elde edilir. Bu değer $C_0(\Sigma)$ 'de yerine yazılırsa

$$C_1(\Sigma) = \frac{p}{2} \log \left[\frac{\text{tr}(\Sigma)}{p} \right] - \frac{1}{2} \log|\Sigma| \quad (5.27)$$

çok deęişkenli normal daęılımının kovaryans matrisinin (Σ) karmaşıklığının teorik maksimum bilgi ölçüsünü buluruz. $C_1(\Sigma)$, $C_0(\Sigma)$ için bir üst sınırdır ve hem varyanslar arasındaki eşitsizliği hem de Σ 'de kovaryansların katkısını ölçer. Böyle bir ölçü, model içindeki yüksek dereceden korelasyonları, benzerlikleri, farklılıkları, model yapısını tanımlamada, deęerlendirmede ve model seçiminde çok önemlidir. $C_1(\Sigma)$ varyanslarla ilişkili koordinat sisteminden baęımsızdır. Örneęin, herhangi bir σ_{jj} 'nin sıfır olması durumunda $C_0(\Sigma)$ " $\infty - \infty$ " belirsiz deęerini alırken, $C_1(\Sigma)$ matematiksel anlam ifade eden " ∞ " sonsuz deęerini alır. Ayrıca $C_1(\Sigma)$ skaler çarpım ve ortonormal dönüşümler altında deęişmezdir (Bozdoğan, 1990). 5.27 eşitliğinde gerekli düzenlemeler yapılırsa,

$$C_1(\Sigma) = \frac{1}{2} \log \left[\frac{\text{tr}(\Sigma)}{|\Sigma|} \right] \quad (5.28)$$

elde edilir. Eşitlikteki $|\Sigma|$ genelleştirilmiş varyansı, $\frac{\text{tr}(\Sigma)}{p}$ ise toplam varyans ortalaması olduęu için kompleksliği ifade eder. Eęer Σ 'nın özdeęerleri $\lambda_1, \lambda_2, \dots, \lambda_p$ olarak alınırsa $\frac{\text{tr}(\Sigma)}{p} = \frac{1}{p} \sum_{j=1}^p \lambda_j = \bar{\lambda}_a$ özdeęerlerin aritmetik ortalaması ve $|\Sigma|^{\frac{1}{p}} = \prod_{j=1}^p \lambda_j^{\frac{1}{p}} = \bar{\lambda}_g$ özdeęerlerin geometrik ortalaması olarak yazılabilir. Bu durumda 5.28'deki komplekslik

$$C_1(\Sigma) = \frac{p}{2} \log \frac{\bar{\lambda}_a}{\bar{\lambda}_g} \quad (5.29)$$

olarak da ifade edilebilir.

$C_1(\Sigma)$ 'in büyük deęer alması deęişkenler arasındaki etkileşimin yüksek olduęunu, küçük deęer alması ise etkileşimin zayıf olduęunu gösterir. Varyans kovaryans matrisinin (Σ) birim matrise eşit olduęu durumlarda komplekslik '0' olacaktır. Yani en az kompleksliğe sahip olan matris birim matristir. Dięer durumlarda $C_1(\Sigma) > 0$ 'dır (Bozdoğan, 2004).

5.8.2 ICOMP (IFIM)_PEU Bilgi Kriteri

ICOMP'un gelişimi kısmen AIC kısmen de bilgi karmaşıklığı kavramlarına dayanmaktadır. ICOMP kriteri, serbest parametre sayısı yerine modelin kovaryans karmaşıklığını cezalandırmaktadır. Amacı ise bir modelin karmaşıklığı ile uyum iyilięi arasındaki en uygun dengeyi sağlamaktır. ICOMP

$$\text{ICOMP} = -2\log L(\hat{Q}) + 2C_1(\Sigma) \quad (5.30)$$

şeklinde tanımlanır. Eşitlikteki ilk ifade uyum eksikliğini ikinci ifade ise karmaşıklığı ifade etmektedir. $L(\hat{Q})$, maksimize edilmiş olabilirlik fonksiyonu, \hat{Q} , Q parametre vektörünün maksimum olabilirlik tahmini, $C_1(\Sigma)$, 5.29 nolu eşitlikte verilen komplekslik ölçüsü ve Σ ise modelin parametre vektörlerinin kovaryans matrisidir. Kovaryans tahmini için eşitlik 5.8'de verilen ters Fisher bilgi matrisi kullanılırsa,

$$\text{ICOMP(IFIM)} = -2\log L(\hat{Q}) + 2C_1(\hat{F}^{-1}) \quad (5.31)$$

denklemini elde edilir. Burada $C_1(\hat{F}^{-1})$ aşağıdaki şekilde hesaplanır.

$$C_1(\hat{F}^{-1}) = (M + 1) \log \left[\frac{\text{tr} \hat{\sigma}^2 (B' B)^{-1} + \frac{2\hat{\sigma}^4}{n}}{M + 1} \right] - \log |\hat{\sigma}^2 (B' B)^{-1}| \log \frac{2\hat{\sigma}^4}{n} \quad (5.32)$$

$$\text{Cov}(\hat{\beta}, \hat{\sigma}^2) = \hat{F}^{-1} = \begin{bmatrix} \hat{\sigma}^2 (B' B)^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix} \quad (5.33)$$

Eşitlik 5.32'deki köşeli parantezin pay kısmının eşitlik 5.33'de verilen kovaryans matrisinin izi olduğuna dikkat edilmelidir.

Son olarak ICOMP(IFIM) kriterine farklı bir yaklaşım olarak sonsal beklenen faydayı (Posterior Expected Utility_{PEU}) en büyükmek amacıyla $\text{ICOMP(IFIM)}_{\text{PEU}}$ elde edilmiştir.

$$\text{ICOMP(IFIM)}_{\text{PEU}} = -2\log L(\hat{Q}) + k(1 + \log(n)) + 2C_1(\hat{F}^{-1}) \quad (5.34)$$

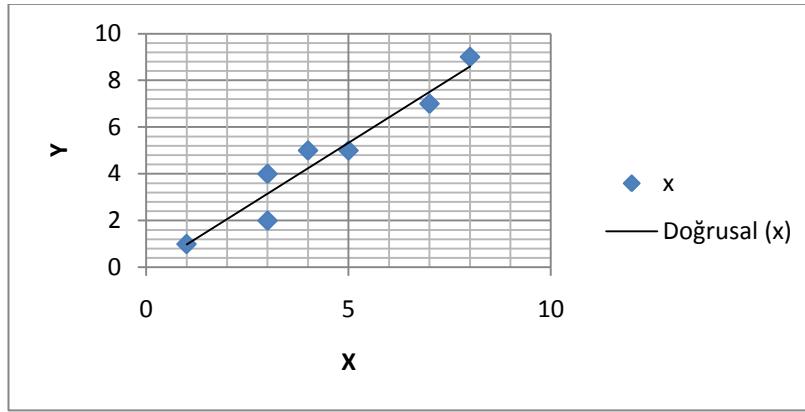
ICOMP(IFIM) kriterinden farklı olarak $\text{ICOMP(IFIM)}_{\text{PEU}}$ kriterine serbest parametre sayısı $k(1 + \log(n))$ eklenmiştir. $\text{ICOMP(IFIM)}_{\text{PEU}}$ aynı zamanda Bozdoğan tarafından geliştirilen $\text{CAIC} = -2\log L(\hat{Q}) + k \log(n) + k$ kriteri ile $2C_1(\hat{F}^{-1})$ 'den elde edildiği için Tutarlı ICOMP (CICOMP) olarak da adlandırılmaktadır (Pamukçu, Bozdoğan, 2015).

6. SİMÜLASYON ÇALIŞMASI

Bir veya birden çok açıklayıcı değişken ile açıklanan değişken arasındaki istatistiksel ilişkinin modellenmesi regresyon analizi ile incelenir. Regresyon analizinde amaç, açıklayıcı değişkenlerden açıklanan değişkeni tahmin edebilen en iyi modelin oluşturulması veya açıklanan değişkenin hangi açıklayıcı değişkenlerden daha çok etkilendiğinin tespit edilmesidir. Açıklayıcı değişken sayısının az olduğu durumlarda regresyon denkleminin çözümlenmesi ve yorumlanması açıklayıcı değişken sayısının çok olduğu durumlara göre oldukça kolaydır.

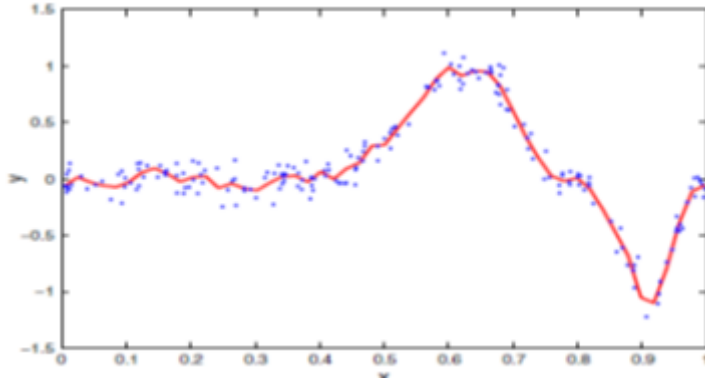
Örneğin; $y_i = 1, 3, 3, 5, 4, 8, 7$ $x_i = 1, 2, 4, 5, 5, 9, 7$ veri setimiz için regresyon denklemi ve grafiği aşağıdaki gibidir.

$$y_i = 0.40 + 0.86x_i \text{ (EKK yöntemiyle elde edildi.)}$$

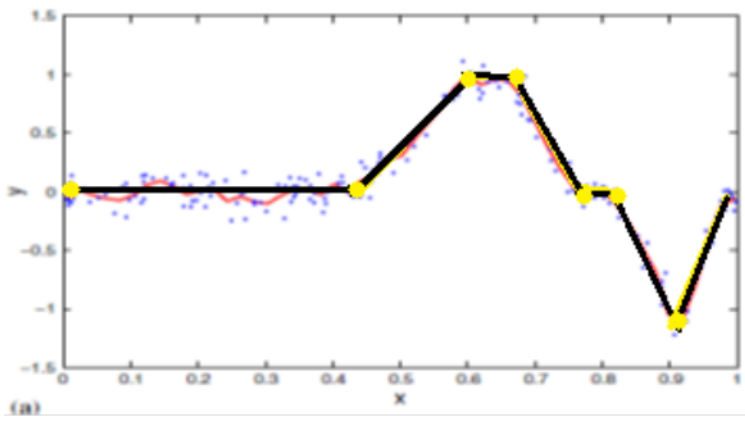


Şekil 6.1: Doğrusal regresyon modeli

Değişken ve veri sayısı arttıkça regresyon denklemini elde etmek ve yorumlamak daha zor hale gelir. Aşağıdaki gibi regresyon yapısına sahip bir modeli daha anlaşılır kılmak için MARS yöntemi kullanılır. MARS bu şekildeki modeli doğrusal denklemlere (temel fonksiyonlara (BF)) bölerek kuadratik veya kübik yapıdan kurtarıp parçalı doğrusal hale getirir.



Şekil 6.2: Karmaşık yapıdaki regresyon modeli (Koc ve Bozdoğan, 2014)



Şekil 6.3: Kesikli doğrusal hale getirilmiş MARS regresyon modeli

Şekil 6.3 te 7 temel fonksiyon (BF) ve 7 adet (sarı ile gösterilen) düğüm değeri oluşmuştur. Uygulamaya, RStudio programı kullanılarak, simülasyon çalışması ile başlanmıştır. İlk olarak her biri 100 gözlem içeren 10 açıklayıcı değişken üretilmiştir.

Tablo 6.1: 10 değişken ve 100 veriyi içeren matris

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & \cdots & x_{1,10} \\ x_{2,1} & x_{2,2} & \cdots & \cdots & x_{2,10} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ x_{100,1} & x_{100,2} & \cdots & \cdots & x_{100,10} \end{bmatrix}$$

Sonrasında her bir değişken için düğüm değeri 3 olarak atanmıştır (Friedman, 1991). Yani her bir 33 veride düğüm değeri oluşacaktır. Böylece her değişken için 3 BF oluşmuştur.

Tablo 6.2: Bir değişkenden düğüm noktalarının ve temel fonksiyonun oluşması

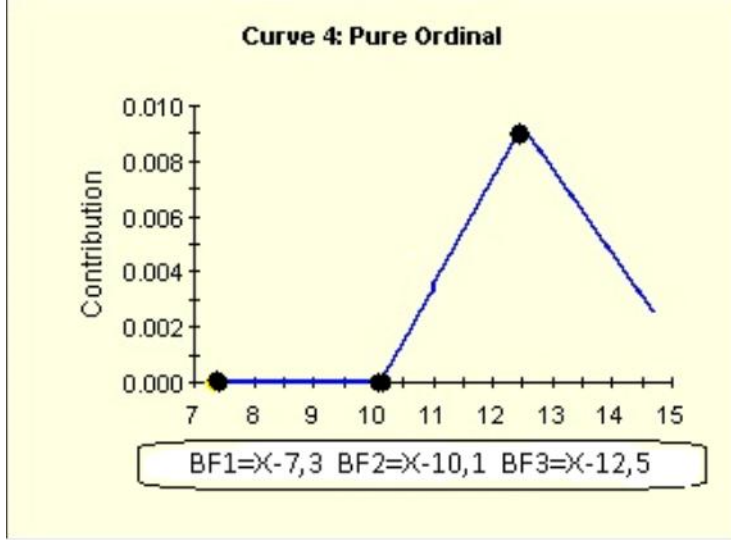
x	
1	}
2	
⋮	
33	}
34	
⋮	
66	}
67	
⋮	
100	

Modele dahil edilecek maksimum BF sayısı 25 olarak belirlendi. Oluşturulan 10 değişkenin ilk 5'i (x_1, x_2, x_3, x_4, x_5) modele katkısı olan değişkenler olarak, son 5'i ise katkısı olmayan değişkenler olarak seçilmiştir. Bu seçim, kriterlerin modeli oluşturma başarısını gözlemlenebilir kılmıştır. Her bir değişken için 3 düğüm değeri elde edilmişti, böylece toplamda 30 düğüm değeri sayısı ayna (mirror) fonksiyonlarıyla birlikte 60'a çıkmıştır.

Her bir değişken için 0-1 arasında uniform dağılımdan 100 tane sayı üretilmiştir. Bunun yanı sıra standart normal dağılım kullanılarak hata değişken vektörü üretilmiştir ($\varepsilon \sim N(0,1)$). Son olarak $y = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + 0.5\varepsilon$ başlangıç denklemi ile y vektörü oluşturuldu (Koç ve Bozdoğan, 2014). Bu şekilde x matrisi ve y ve ε (epsilon) vektörleri elde edildi.

X değişkenlerinden BF'lerin elde edilmesi

Temel fonksiyonlar, x değişkeni ile düğüm değerinin farkı (x-t) ve ayna görüntüsü (t-x) şeklinde ifade edilir. Örneğin X değişkenine ait BF'lerin oluşumu aşağıdaki şekil yardımıyla gösterebiliriz. X değişkeni için 3 adet düğüm değeri elde edilmiş varsayımıyla temel fonksiyonlar aşağıdaki gibi elde edilir. Temel fonksiyonları oluşturduktan sonra, X'ler yerine artık BF'ler ile çalışmaya devam edilir



Şekil 6.4: X değişkeninden BF'lerin elde edilmesi

Her değişkenden 3 adet olmak üzere 30 adet BF elde edildi ve ayna fonksiyonlarıyla birlikte 60 adet BF oluşmuş oldu.

Şu ana kadar etkileşimsiz değişkenler üzerinden işlem yapıldı. MARS'ın diğer avantajı, değişkenlerin etkileşimlerinin de modele katkısını dikkate almasıdır. Etkileşimler de oluşturulduğunda (BF1*BF2, ... BF1*BF60, BF2*BF3, ... BF2*BF60, BF3*BF4, ... BF3*BF60, ... BF58*BF60, BF59*BF60) toplamda 1770 ikili etkileşim içeren BF, 60 adet etkileşimsiz BF ve y vektörü olmak üzere elimizde 1831 değişken oluşmuş oldu. Bu şekilde çok büyük bir veri seti (veri setini çok daha fazla büyüteceği için 3'lü ve daha büyük etkileşimler çalışma açısından pratik kabul edilmiyor (Koç ve Bozdoğan, 2014)) olduğundan, veri setimiz artık büyük veriye (bigdata) dönüşmüş oldu. Bu kadar büyük değişkenli bir regresyon yapısının incelenmesi mümkün olmadığından temel fonksiyon sayısına kısıt getirmemiz gerekiyor. Bu amaçla, modele dahil edilecek maksimum temel fonksiyonlar 25 sayısı ile kısıtlandı. Sonuç olarak diyebiliriz ki, 1830 BF (etkileşimsiz, etkileşimli ve ayna fonksiyonları) içinden MARS yöntemi ileri doğru seçim ve geriye doğru eleme adımlarından sonra, en fazla 25 temel fonksiyon içeren final modeli oluşturulabilecek.

İleriye doğru seçim adımı (forward selection)

İleriye doğru seçim aşamasında, hemen hemen tüm seçim prosedürlerinde sıklıkla kullanılan Akaike Bilgi Kriteri (AIC) ve Determinasyon (açıklayıcılık) Katsayısı (R^2), arasından tercihimizi AIC yönünde kullandık. Bunun nedeni, R^2 'nin

değişkenler arasında ilişki olması durumunda (örneğin çalışmamızda kullandığımız enflasyon ve faiz oranı arasında ilişkiden söz edebiliriz) yapay (kovaryans _özilişki_ hatası) olarak yükseliyor olması (Heo vd. 2008); AIC değerinin ise model seçiminde olabilirlik tahminini baz alarak uyum hassasiyeti ile bu uyumu sağlamak için kullandığı parametre sayısı arasında bir denge oluşturmaya odaklanmasıdır.

İlk olarak $Y = \widehat{\beta}_0$ için AIC değerini hesaplıyoruz. $Y = \widehat{\beta}_0 + \widehat{\beta}_1 BF_1$ denklemi için de AIC değeri hesaplanıyor ve bir önceki modelde (sadece sabit içeren model) elde edilen değerden daha küçük bir değer elde edilmiş ise BF_1 modele dahil ediliyor. Aksi durumda ise BF_1 'i modele almayı işlemeye devam ediyoruz. Sonrasında BF_2 'nin modele alınıp alınmayacağını aynı şekilde test ediyoruz. $Y = \widehat{\beta}_0 + \widehat{\beta}_1 BF_1 + \widehat{\beta}_2 BF_2$ (BF_1 'in modele dahil olduğu varsayılırsa) modelinin AIC değerinin $Y = \widehat{\beta}_0 + \widehat{\beta}_1 BF_1$ modelinin AIC değerinden büyük çıktığı varsayılırsa, BF_2 modele dahil edilmeden bir sonraki temel fonksiyonun modele alınıp alınmayacağına bakılır. Oluşturulan tüm BF'ler için aynı işlem yapılır ve maksimum karmaşıklığıdaki model elde edilmiş olur.

Geriyeye doğru eleme adımı (backward elimination)

Maksimum büyüklükteki modeli daha yalın hale getirmek için modele katkısı diğer BF'lere göre daha az olan temel fonksiyonlar her adımda elenir. Eleme için MARS yöntemi GCV kriterini kullanmaktayken (Friedman, 1991) biz ise bu kriterin yanında AIC, SBC ve ICOMP(IFIM)_PEU kriterlerini de kullanarak, daha doğru ve daha yalın modeli veren kriteri belirlemek istiyoruz. Örneğin ileriye doğru seçim aşamasında elde edilen modelimiz $Y = \widehat{\beta}_0 + \widehat{\beta}_1 BF_1 + \dots + \widehat{\beta}_a BF_b BF_c$ şeklinde olsun. Bu model için GCV, AIC, SBC, ICOMP(IFIM)_PEU değerleri hesaplanır. Sonra en sondaki temel fonksiyon olan $\widehat{\beta}_a BF_b BF_c$ modelden çıkarılır ve tekrar kriter değerleri hesaplanır. Eğer BF çıkarıldığında kriter değerinde azalma varsa, azalan kriter değerlerine sahip modelden o BF çıkarılır, aksi durumda ise BF'ye dokunulmaz. Bu şekilde budamalar yapılarak en küçük kriter değerine ulaşılır ve final modeli elde edilir. Bu çalışmada GCV kriterinin

$$\left(\text{GCV}(M) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{f}_M(x_i))^2}{\left(1 - \frac{P(M)^*}{n}\right)^2} \right)$$

pay kısmı olan hata kareler toplamı, $\left(e^{\log\text{-likelihood} * \frac{-2}{n} * n} \right)$ olarak alınmıştır. Rstudio programında hazır olan komut aynı sonucu vermekle birlikte işlemlerde kolaylık sağlamaktadır.

Aşağıda sırasıyla GCV, AIC, SBC ve ICOMP(IFIM)_PEU kriterleri kullanılarak elde edilen modeller ve GCV kriteri kullanılarak elde edilen ANOVA ayrışımı tablosu verilmiştir.

GCV kriteri ile elde edilen MARS modeli

$$Y = 2.41 + 2.31*BF4 + 3.29*BF7 + 9.79*BF11 + 3.06*BF13 - 0.33*BF31 + 3.88*BF38 + 2.36*BF39 + 2.34*BF1*BF5 - 11.44*BF1*BF6 + 4.3*BF1*BF34$$

$$“BF4= \max(0, x2-0.179)”$$

$$“BF7= \max(0, x3-0.6578)”$$

$$“BF11= \max(0, x4-0.0158)”$$

$$“BF13= \max(0, x5-0.0255)”$$

$$“BF31= \max(0, 0.4922-x1)”$$

$$“BF38= \max(0, 0.6367-x3)”$$

$$“BF39= \max(0, 0.1594-x3)”$$

$$“BF1*BF5= \max(0, x1-0.4922*x2-0.674)”$$

$$“BF1*BF6= \max(0, x1-0.4922*x2-0.0832)”$$

$$“BF1*BF34= \max(0, x1-0.4922*0.179-x2)”$$

GCV kriteri kullanılarak elde edilen MARS modeli 10 temel fonksiyon içermekte. Bunların 3 tanesi etkileşim fonksiyonları olarak modele dahil olmuştur. Etkileşim gösteren değişkenler x1-x2'dir. Modelimiz x6,x7,x8,x9,x10 değişkenlerini içermediği için GCV kriterinin model tahmininde başarısının yüksek olduğunu söyleyebiliriz.

Tablo 6.3:GCV kriteri kullanılarak elde edilen ANOVA ayrışımı

Fonksiyon	St. Sapma	GCV	#Temel Fonksiyon	#Serbestlik Derecesi	Değişkenler
1	3,8909	15,6059	1	2,5	x1
2	4,4838	20,7241	1	2,5	x2
3	4,5239	25,0146	3	7,5	x3
4	3,9784	16,3159	1	2,5	x4
5	4,4386	20,3087	1	2,5	x5
6	4,0955	20,5012	3	7,5	x1x2

Yukarıda MARS regresyonun GCV kriterine bağlı olarak elde edilmiş anlamlı değişkenlerinin ANOVA ayrışımı sonuçları verilmiştir. Bu tabloda final modele giren her bir değişkenin bağımlı değişken üzerine olan genel etkileri görülmektedir. ANOVA tablosundaki standart sapmaların büyüklüğü o değişkenin bağımlı değişken üzerinde olan genel etkisinin fazla olduğunu gösterir. Üçüncü sütunda yer alan GCV değerleri, söz konusu değişkene ait temel fonksiyonları

modelden çıkardığımızda tahminlerde meydana gelecek kaybı göstermektedir. Örneğin x_3 değişkeni toplam 3 temel fonksiyon ile modele girmiş ve standart sapması incelendiğinde bağımlı değişken üzerinde en etkili değişken olduğu görülmüştür. Modele katkı sağlama yönünden x_3 değişkenini sırasıyla x_2, x_5 ve x_1x_2 değişkenleri izlemektedir. x_1 ve x_4 değişkenlerinin modele diğer değişkenlere kıyasla en az katkıyı sağladıkları söylenebilir.

AIC kriteri ile elde edilen MARS modeli

$$Y = -2.65 + 2.12*BF1 -2.64*BF2 + 12.16*BF4 + 6.45*BF5 + 0.16*BF7 + 8.1*BF11 + 5.54*BF13 - 0.89*BF20 - 8.85*BF31 - 0.76*BF34 - 8.03*BF37 + 7.55*BF38 - 14.39*BF39 + 7.94*BF46 - 2.61*BF48 + 14.03*BF49 + 8.76*BF1*BF5 + 1.6*BF1*BF6 + 3.42*BF1*BF10 - 6.32*BF1*BF28 - 2.4*BF1*BF34$$

$$"BF1= \max(0, x1-0.4922)"$$

$$"BF2= \max(0, x1-0.6096)"$$

$$"BF4= \max(0, x2-0.179)"$$

$$"BF5= \max(0, x2-0.674)"$$

$$"BF7= \max(0, x3-0.6578)"$$

$$"BF11= \max(0, x4-0.0158)"$$

$$"BF13= \max(0, x5-0.0255)"$$

$$"BF20= \max(0, x7-0.1337)"$$

$$"BF31= \max(0, 0.4922-x1)"$$

$$"BF34= \max(0, 0.179-x2)"$$

$$"BF37= \max(0, 0.6578-x3)"$$

$$"BF38= \max(0, 0.6367-x3)"$$

$$"BF39= \max(0, 0.1594-x3)"$$

$$"BF46= \max(0, 0.9516-x6)"$$

$$"BF48= \max(0, 0.035-x6)"$$

$$"BF49= \max(0, 0.7125-x7)"$$

$$"BF1*BF5= \max(0, x1-0.4922*x2-0.674)"$$

$$"BF1*BF6= \max(0, x1-0.4922*x2-0.0832)"$$

$$"BF1*BF10= \max(0, x1-0.4922*x4-0.5186)"$$

$$"BF1*BF28= \max(0, x1-0.4922*x10-0.7223)"$$

$$"BF1*BF34= \max(0, x1-0.4922*0.179-x2)"$$

AIC kriteri kullanılarak elde edilen MARS modeli 21 temel fonksiyon içermekte. Etkileşim gösteren değişkenler $x1-x2, x1-x4$ ve $x1-x10$ 'dur. Modelimizin $x6, x7$ ve $x10$ değişkenlerini içermesi doğru modelin seçilmesi yönünden istenmeyen bir durumdur. Ayrıca GCV kriterinden elde edilen modele kıyasla oldukça fazla BF içeren bir model oluştuğunu görüyoruz. Doğru değişken seçimiyle birlikte en yalın modelin tercihi göz önünde bulundurulduğunda, GCV kriteri AIC kriterine göre daha başarılı olmuştur diyebiliriz.

SBC kriteri ile elde edilen MARS modeli

$$Y = -0.46 + 3.07*BF4 + 3.06*BF7 + 5.82*BF11 - 0.11*BF13 - 7.15*BF31 - 0.38*BF34 + 0.97*BF37 + 3.82*BF38 - 3.12*BF39 + 3.14*BF1*BF5 + 8.08*BF1*BF6 + 6.92*BF1*BF34$$

$$"BF4= \max(0, x2-0.179)"$$

$$"BF7= \max(0, x3-0.6578)"$$

$$"BF11= \max(0, x4-0.0158)"$$

$$"BF13= \max(0, x5-0.0255)"$$

$$"BF31= \max(0, 0.4922-x1)"$$

$$"BF34= \max(0, 0.179-x2)"$$

$$"BF37= \max(0, 0.6578-x3)"$$

$$"BF38= \max(0, 0.6367-x3)"$$

$$"BF39= \max(0, 0.1594-x3)"$$

$$"BF1*BF5= \max(0, x1-0.4922*x2-0.674)"$$

$$"BF1*BF6= \max(0, x1-0.4922*x2-0.0832)"$$

$$"BF1*BF34= \max(0, x1-0.4922*0.179-x2)"$$

SBC kriteri ile 12 BF içeren model elde edilmiştir. x1-x2 bu kriterde de etkileşim gösteren değişkenler olarak belirlenmiştir. Modele katkısı olmayan değişkenleri içermemesi kriterin model seçiminde başarılı olduğunu gösteriyor. Temel fonksiyon sayısı yönünden GCV kriterinden çok fazla olmamakla birlikte GCV'nin daha sade model seçiminde bulunduğunu söyleyebiliriz.

ICOMP(IFIM)_PEU kriteri ile elde edilen MARS modeli

$$Y = 1.97 + 10.01*BF4 - 2.48*BF5 - 1.41*BF7 + 9.17*BF11 - 4.78*BF13 + 1.58*BF31 + 4.92*BF38 + 2.73*BF39 + 0.42*BF1*BF34$$

$$"BF4= \max(0, x2-0.179)"$$

$$"BF5= \max(0, x2-0.674)"$$

$$"BF7= \max(0, x3-0.6578)"$$

$$"BF11= \max(0, x4-0.0158)"$$

$$"BF13= \max(0, x5-0.0255)"$$

$$"BF31= \max(0, 0.4922-x1)"$$

$$"BF38= \max(0, 0.6367-x3)"$$

$$"BF39= \max(0, 0.1594-x3)"$$

$$"BF1*BF34= \max(0, x1-0.4922*0.179-x2)"$$

ICOMP(IFIM)_PEU kriteri 9 BF içermesiyle ve içerdiği değişkenlerin modele katkısı olan değişkenlerden oluşması sebebiyle, ICOMP(IFIM)_PEU'nun model seçiminde en başarılı kriter olduğunu söyleyebiliriz.

Tablo 6.4:ICOMP(IFIM)_PEU kriteri kullanılarak elde edilen ANOVA ayrışımı

Fonksiyon	St.Sapma	ICOMP	#Temel Fonksiyon	#Serbestlik Derecesi	Değişkenler
1	3,890933	560,2431	1	2,5	x1
2	4,462392	599,0788	2	5	x2
3	4,523992	613,3811	3	7,5	x3
4	3,97846	566,4229	1	2,5	x4
5	4,438636	590,2337	1	2,5	x5
6	4,325727	589,1087	1	2,5	x1x2

Yukarıda MARS regresyonun ICOMP(IFIM)_PEU kriterine bağlı olarak elde edilmiş anlamlı değişkenlerinin ANOVA ayrışımı sonuçları verilmiştir. x_3 değişkeni toplam 3 temel fonksiyon ile modele girmiş ve standart sapması incelendiğinde 4,52 skoru ile bağımlı değişken üzerinde en etkili değişken olduğu görülmüştür. Modele katkı sağlama yönünden x_3 değişkenini sırasıyla x_2, x_5 ve x_1x_2 değişkenleri izlemektedir. Beklenildiği üzere, değişkenlerin model üzerindeki etkileri bakımından ICOMP ile GCV kriteri kullanılarak elde edilen ANOVA ayrışım tablolarında benzer sonuçlar elde edilmiştir.

7. GERÇEK VERİ SETİ UYGULAMASI

Modern liberal dünyada ekonominin birçok belirleyeni vardır. Liberal ekonomik modelin merkezinde ise sermaye yer alır. Tüm sistem sermaye birikimi ve bu sermayenin yatırıma dönüştürülmesi için kurgulanmıştır. Bu sermayenin birikmesi ve yatırıma dönüşmesi için bankalar çok önemli bir rol üstlenmektedir. Sermayenin yatırımcılara ve ihtiyaç sahiplerine güvenli bir şekilde ulaştırılmasını ve paranın hem ulusal hem de küresel düzlemde yayılmasına aracılık eden bu kurumların ayakta kalması ve sürdürülebilir bir altyapıya sahip olması büyük önem taşımaktadır. Bunun için sağlanması gereken koşulların başında bankaların vermiş oldukları kredilerin³ geri ödemesinin önemli bir oranda gerçekleşmiş olması gelmektedir. Aksi takdirde kredilerin takibe dönüş oranı artacak ve bu doğal olarak bankaların performansını olumsuz etkileyecektir. Makro düzeyde ise bu durum ülkenin genel ekonomik tablosuna negatif olarak yansıtacaktır. Kredilerin takibe dönüş oranını etkileme potansiyeline sahip birçok gösterge vardır Bu göstergeler arasından farklı seçimler yapmak mümkündür. Bununla birlikte bu çalışmada, ilgili literatür çalışmaları da dikkate alınarak, kredilerin takibe dönüş oranı üzerinde önemli bir etkiye sahip oldukları varsayılan sanayi üretim endeksi, işsizlik oranı, döviz kuru, enflasyon, faiz oranı, gayrisafi yurt içi hâsıla, ithalat ve ihracat faktörleri bağımsız değişkenler olarak seçilmiştir. Takibe dönüş oranını ve onunla ilişki içerisinde olduğu varsayılan yukarıdaki göstergeleri kısaca tanımlamak gerekirse:

³ Kredi belirli miktardaki satın alma gücünün, belirli bir süre için, faiz gibi bir bedel karşılığında, gerçek ya da tüzel kişilere verilmesini ifade eder. Aynı zamanda fon fazlası olanlar ile fon ihtiyacı olanlar arasındaki iletişimi sağlayan önemli bir finansal araçtır banka kredileri. 1 Kasım 2005 tarihinde Resmi Gazete’de yayımlanarak yürürlüğe giren 5411 Sayılı Bankacılık Kanunu’nun 48. Maddesince kredi şöyle tanımlanmaktadır:

“Bankalarca verilen nakdi krediler ile teminat mektupları, kontrgarantiler, kefaletler, aval, ciro, kabul gibi gayrinakdi krediler ve bu niteliği haiz taahhütler, satın alınan tahvil ve benzeri sermaye piyasası araçları, tevdiatta bulunmak süretiyle ya da herhangi bir şekil ve surette verilen ödünçler, varlıkların vadeli satışından doğan alacaklar, vadesi geçmiş nakdi krediler, tahakkuk etmekle birlikte tahsil edilmemiş faizler, gayrinakdi kredilerin nakde tahvil olan bedelleri, ters repo işlemlerinden alacaklar, vadeli işlem ve opsiyon sözleşmeleri ile benzeri diğer sözleşmeler nedeniyle üstlenilen riskler, ortaklık payları ve Kurulca kredi olarak kabul edilen işlemler izlendikleri hesaba bakılmaksızın bu Kanun Uygulamasında kredi sayılır.” (BDDK, 5411 Sayılı Bankacılık Kanunu, 1 Kasım 2005 tarih ve 25983 mükerrer sayılı Resmi Gazete)

Takibe Dönüşüm Oranı: Takibe dönüşüm oranı bankacılık sektörünün aktif kalitesinin⁴ en önemli göstergelerinden biridir. Bankaların kullandığı kredilerde geri dönmeyenlerinin oranının artması, banka aktif kalitesini bozar ve karlılığını da olumsuz etkiler. Yine takibe dönüşüm oranı kredi hacmi üzerinde de etkilidir. Yapılan bazı akademik çalışmalarda kredi hacmini etkileyen en önemli faktör olarak takibe dönüşüm oranı bulunmuştur. Bu çalışmalarda takibe dönüşüm oranı arttıkça kredi hacminin azaldığı tespit edilmiştir. Dolayısıyla takibe dönüşüm oranlarındaki artışın hem banka karlılığını, hem de kredi hacmini baskılayarak ekonomik büyümeyi ve istihdamı olumsuz etkilediği söylenebilir. Anapara veya faizin ya da her ikisinin ödenmesi gereken tarihten itibaren 90 gün geçmesine rağmen tahsil edilemeyen kredi ve alacaklar takip hesabına aktarılırlar (Finansal Göz, 2018).

Sanayi Üretim Endeksi: Bir ülkenin gelişmesinde aynı zamanda uluslararası rekabette ülkeye güç katan en önemli sektörlerden olan Sanayi Sektörü milli gelir büyümesinin ve ekonominin en önemli itici güçlerindedir. İstihdama önemli bir katkı sağlayan sanayi sektörü büyümeye de ciddi bir ivme kazandırmaktadır. Sanayi üretim endeksi, ekonomide meydana gelen gelişmelerin ve uygulanan ekonomik politikaların, kısa dönemde olumlu veya olumsuz etkilerinin ölçülebilmesi için aylık olarak hesaplanan bir endekstir (TCMB Terimler Sözlüğü, 2020). Sanayi sektörünün en büyük alt sektörü konumundaki üretim (imalat) sanayi sektörü; özellikle Türkiye'nin içinde bulunduğu gelişmekte olan ülkelerde ihracatın, istihdamın ve inavasyonların önemli bir kaynağı olarak ekonomik büyümeye katkı sağlamaktadır. İmalat sanayi işletmeleri banka kredilerini sabit varlık ve çalışma sermayesi finansmanında yoğun kullanmaktadır. Analizler; imalat sanayi sektöründe üretim ve banka kredilerinin eşbütünleşik olduğunu, aralarında uzun dönemde pozitif bir ilişki olduğunu ve üretimde banka kredilerine doğru bir nedenselliğin varlığını göstermektedir⁵.

⁴ Bankaların başarı göstergelerinden biri olarak kabul gören fakörlerden biri de aktif kalitesidir. Aktif kalitesinin büyüklüğü, bankaların büyüklüğünü belirlemekte ve sahip olunan varlıkların etkin yönetimi için aktif kalitesinin yüksek olması beklenmektedir (Dursun, Bozkır, 2018).

⁵ Takibe Dönüşüm Oranı ve Sanayi Üretim Endeksi arasındaki ilişki için bkz.(Investaz, 2018; Yüksel, 2016; Demirci, 2016; Curak vd., 2013; Bhattarai, 2016; Chaibi ve Ftiti, 2015; Macitve Keçeli, 2012; Yağcılar ve Demir, 2015; Castro, 2013).

İşsizlik Oranı: İşi olmayan, iş arayan ve bu konuda girişimde bulunmuş olan kişiler işsiz olarak tanımlanmaktadır. Çalışma çağında olan nüfustan çalışma arzusu göstererek işgücü piyasasına dahil olanlar işgücüne kabul edilmektedir. İş aradığı halde henüz iş bulamayanlar, diğer bir ifadeyle işsiz konumda olanlar ile çalışanların toplamı işgücünü oluşturmaktadır. İşsizlerin işgücüne oranına ise işsizlik oranı denilmektedir (TCMB Terimler Sözlüğü, 2020). Ekonomilerde, mevcut işgücünün ekonomik faaliyetler içerisine katılıp çalışması veya çalıştırılması istihdam olarak ifade edilmektedir. Bir ülkenin istihdam düzeyi, o ülkenin kalkınmışlık seviyesi, doğal kaynakları, nüfusu, eğitimi, teknolojik yenilikleri ve ekonomik koşullarına bağlıdır. Eğer ekonomilerde üretim faktörlerinin (doğal kaynaklar, sermaye, emek ve girişimci) hepsi çalışıp üretime katılıyor ise burada tam istihdam durumundan söz edilebilir. Dolayısıyla işsizliğin azalması istihdamın artması, ekonomik koşulların iyileşmesi anlamına gelecektir. Bu durumda işsizlik ile kredi kullanımı ve kredilerin vadesinde geri ödenmesi arasında ters yönlü bir ilişki söz edebiliriz ⁶.

Döviz Kuru: Ulusal bir paranın, diğer bir ulusal para cinsinden ifadesidir (TCMB Terimler Sözlüğü, 2020). Döviz kuru bir ülkenin göreceli ekonomik seviyesinin en önemli belirleyicilerinden biridir. Döviz kuru arttıkça kredilerin takibe dönüşüm oranı artmaktadır. Bu durum firmaların genelinin yapısal olarak döviz açığı bulunması ve pozitif döviz şoklarında firmaların ödeme güçlüğüne düşmesi ile açıklanabilir. Günümüzde sürdürülebilir kalkınma ve istikrarlı bir ekonominin tesisi için döviz kuru adeta bir gösterge niteliğindedir. Döviz kurunda meydana gelebilecek değişiklikler yerli ve yabancı firmaların yatırım kararları ile tüketicilerin harcamaları üzerinde etkili olmaktadır⁷. Örneğin 1 dolar almak için ‘günümüzü baz alacak olursak’ 7,84 Türk lirası ödememiz gerekecek. Yani düşük değerli para birimi bir ülkenin ithalatını daha pahalı hale getirir. İşletmelerin ticari ilişkilerdeki gelirleri düşecek ve almış oldukları kredileri ödemekte zorlanacaklardır.

⁶ Takibe Dönüşüm Oranı ve İşsizlik arasındaki ilişki için bkz. (Avcı,2020; Konstantakis vd., 2016; Bhattarai, 2016; Chaibi ve Ftiti, 2015; Milenis, 2012; Yüksel, 2016; Castro, 2013).

⁷ Takibe Dönüşüm Oranı ve Döviz Kuru arasındaki ilişki için bkz. (Altınöz, 2018; Castro, 2013; Bhattarai, 2016; Tanınmış Yücememiş ve Sözer, 2011; Chaibi ve Ftiti, 2015; Yüksel, 2016; Macit ve Keçeli, 2012).

Enflasyon: :Fiyatlar genel seviyesindeki sürekli artış eğilimidir. Enflasyon, tüketici ve üretici fiyat gelişmelerini gösteren ve TÜİK tarafından yayımlanan resmi endeksler ile ölçülmektedir (TCMB Terimler Sözlüğü, 2020). İktisat literatüründe belirlendiği şekilde banka kredileri ile enflasyon arasında karşılıklı nedensellik ilişkisi vardır⁸. Bireysel kredilerdeki artışın enflasyona yol açtığı bilinmektedir. Bu kişilerin ellerinde daha fazla para bulunmasıyla birlikte daha fazla tüketime yönelmeleri ve dolayısıyla da fiyatlar genel seviyesinde artışa neden olacağı gerçeğiyle açıklanabilir. Tersisi durumda ise enflasyondaki artış faiz oranlarını doğru orantılı bir şekilde etkileyecek bu durumda da faiz oranlarındaki yükseliş kredi kullanımını azaltacak veya satın alınan kredilerin geri ödenmesini zorlaştıracaktır.

Faiz Oranı: Üretim faktörlerinden sermayenin elde ettiği getiridir. Diğer bir ifade ile paranın kullanım bedelidir (TCMB Terimler Sözlüğü, 2020). Bankalar açısından bilançonun hem aktif hem de pasif tarafını etkileyen faiz oranının bankanın sürekliliği ve karlılığı açısından önemli bir unsur olduğu bilinmektedir. Kaynak yaratabilmek için gider kalemi olarak görülen faiz oranı, kredi satışında gelir kalemi olarak yer almaktadır. Bankaları diğer ticari işletmelerden ayıran temel unsur, içinde buldukları dönemden ziyade gelecek dönemle ilgili beklentileridir. Bu açıdan bakıldığında alınan veya satılan krediler önem arz etmektedir. Faiz oranlarında oluşabilecek herhangi bir değişim toplanacak veya satılacak krediler üzerinde etkili olurken, aynı zamanda bu durum bankayı finansal olarak da riskli hale getirebilmektedir. Borç almak isteyenlerin, yüksek faiz oranları ile elde ettikleri kredileri geri ödemekte zorluk çekmektedirler⁹

GSYİH: Gayri safi yurtiçi hasıla bir ülkede belli bir dönemde o ülkenin yurttışları tarafından üretilen mal ve hizmetlerin toplamının parasal karşılığıdır. Başka bir ifade ile GSYİH, bir ülkenin sınırları içinde bir dönemde üretilen katma değerdir (TCMB Terimler Sözlüğü, 2020). Yani GSYİH iktisadi büyümenin bir nevi Pazar büyüklüğünün göstergesidir. Yapılan çalışmalar iktisadi büyümenin mevduatın

⁸ Takibe Dömuşüm Oranı ve Enflasyon arasındaki ilişki için bkz. (Altınöz, 2018; Yüksel, 2016; Castro, 2013; Bhattarai, 2016; Chaibi ve Ftiti, 2015; Macit ve Keçeli, 2012; Mileris, 2012; Yağcılar ve Demir, 2015).

⁹ Takibe Dömuşüm Oranı ve Faiz Oranı arasındaki ilişki için bkz.(Altınöz, 2018; Bayrakdaroğlu ve Aydın, 2017; Yüksel, 2016; Castro, 2013; Bhattarai, 2016; Chaibi ve Ftiti, 2015; Yağcılar ve Demir, 2015).

krediye dönüş oranını ivmelediği yönündedir¹⁰. Ekonomik büyümenin finansal gelişmeyi de beraberinde getirdiği bilinmektedir. Bir yandan ekonominin iyiye gitmesi kullanılan kredileri artıracak ticari kazanımlar sayesinde de geri ödemede sorun yaşanmayacaktır yorumu yapılabilirken diğer yandan gelişmekte olan bir ülke için iktisadi büyümenin yeterince yüksek ve istikrarlı seyretmemesi tasarruf açığı sebebiyle mevduatın krediye dönüş oranında olumsuzluklara yol açabilmektedir. Ekonomik açıdan olumsuz bir durum kredilerin geri ödenmesinde de olumsuz sonuçlar doğurabilecektir.

İthalat: Bir ülkede yerleşik kişi ve kurumların diğer ülkelerden mal almasıdır (TCMB Terimler Sözlüğü, 2020). Kredi ve ithalat arasındaki dolaylı etki kendisini tüketim ve harcamalar yoluyla göstermektedir. Kişiler harcamalarını gelirlerinden fazla yaptıklarında bu açığı kredi yoluyla kapatmaya çalışırlar. Kredi kullanımının arttığı dönemlerde ithalat oranlarında da artış olduğu gözlenmektedir¹¹.

İhracat: Bir ülkede yerleşik kişi veya kurumların diğer ülkelere mal satmasıdır (TCMB Terimler Sözlüğü, 2020). İhracat faaliyetleri dünya genelinde olduğu gibi ülkemizde de ekonomik büyümeye önemli katkı sağlamaktadır. İhracattaki başarı ülkedeki üretim sistemiyle doğru orantılıdır. Üretim varsa ihracat gerçekleşebilecektir. Sanayi üretim endeksinde de bahsedildiği gibi İmalat sanayi işletmeleri banka kredilerini sabit varlık ve çalışma sermayesi finansmanında yoğun kullanmaktadırlar. Dolayısıyla ihracat ve kredi kullanımı pozitif yönlü ilişki içerisindedirler¹².

¹⁰ Takibe Dömuşüm Oranı ve GSYİH arasındaki ilişki için bkz.(Investaz, 2018; Altınöz, 2018; Yüksel, 2016; Demirci, 2016; Curak vd., 2013; Bhattarai, 2016; Chaibi ve Ftiti, 2015; Kavcıoğlu, 2012; Macitve Keçeli, 2012; Yağcılar ve Demir, 2015; Castro, 2013).

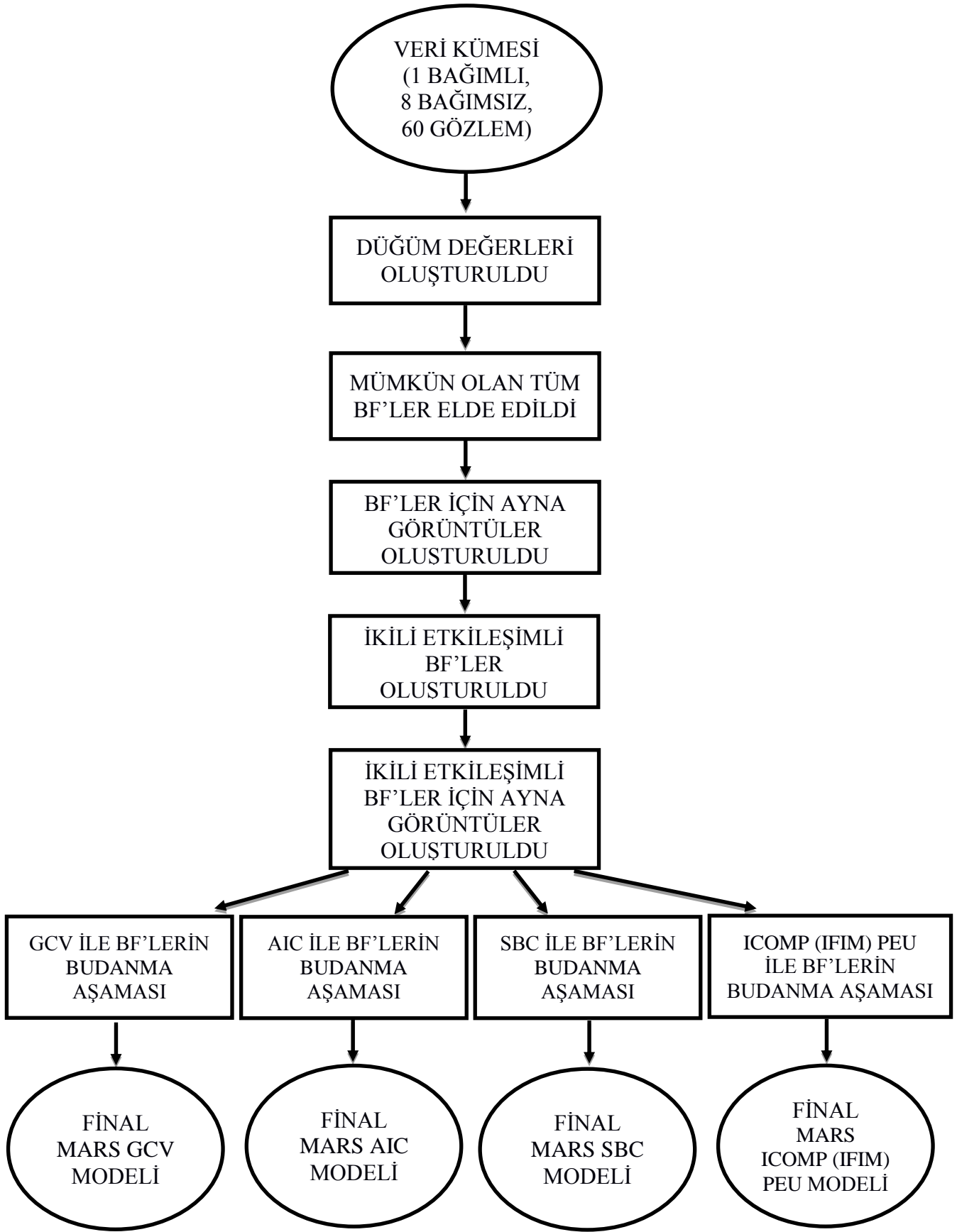
¹¹ Takibe Dömuşüm Oranı ve İthalat arasındaki ilişki için bkz.(Demirezen, 2015; Yurdakul, 2014; Kavcıoğlu, 2012; Şahbaz ve İnkaya, 2010).

¹² Takibe Dömuşüm Oranı ve İhracat arasındaki ilişki için bkz.(Investaz, 2018; Yüksel, 2016; Demirci, 2016; Curak vd., 2013; Bhattarai, 2016; Chaibi ve Ftiti, 2015; Kavcıoğlu, 2012; Macitve Keçeli, 2012; Yağcılar ve Demir, 2015; Castro, 2013).

Tablo 7.1: Çalışmada kullanılan ver seti

DEĞİŞKEN ADI	MODELE ALINMA ŞEKLİ	DÖNEM	GÖZLEM	ÖLÇÜMÜ	KAYNAK
Kredilerin Takibe Düşüş Oranı	Bağımlı Değişken (Y)	2005-2019	60	Katılım bankaları ve mevduat bankalarının kullandırmış olduğu kredileri içerisinde takibe düşen kredilerin oranını göstermektedir.	TKBB, TBB ve Banka Faaliyet Raporları
Sanayi Üretim Endeksi	Bağımsız Değişken (X1)	2005-2019	60	Yıllık toplam sanayi üretim endeksini göstermektedir.	TCMB (Elektronik Veri Dağıtım Sistemi)
İşsizlik	Bağımsız Değişken (X2)	2005-2019	60	Toplam işgücü üzerinden yüzde olarak işsizlik oranını göstermektedir.	TCMB (Elektronik Veri Dağıtım Sistemi)
Döviz Kuru	Bağımsız Değişken (X3)	2005-2019	60	Yıllık döviz kurlarını göstermektedir.	TCMB (Elektronik Veri Dağıtım Sistemi)
Enflasyon	Bağımsız Değişken (X4)	2005-2019	60	Tüketici fiyatlarında bir önceki döneme göre meydana gelen fiyat artışlarını yüzde olarak göstermektedir.	TCMB (Elektronik Veri Dağıtım Sistemi)
Faiz Oranı	Bağımsız Değişken (X5)	2005-2019	60	Bankalarca açılan ticari kredilere uygulanan ağırlıklı ortalama faiz oranlarını göstermektedir.	TCMB (Elektronik Veri Dağıtım Sistemi)

GSYİH	Bağımsız Değişken (X6)	2005-2019	60	Ülkenin yıllık büyüme oranı değerlerini göstermektedir.	TCMB (Elektronik Veri Dağıtım Sistemi)
İthalat	Bağımsız Değişken (X7)	2005-2019	60	Yıllık toplam ithalat hacmini göstermektedir.	TCMB (Elektronik Veri Dağıtım Sistemi)
İhracat	Bağımsız Değişken (X8)	2005-2019	60	Yıllık toplam ihracat hacmini göstermektedir.	TCMB (Elektronik Veri Dağıtım Sistemi)



Şekil 7.1: MARS Modelleri Akış Diyagramı

Bu çalışmada Türkiye’de faaliyet gösteren 12’si mevduat 6’sı katılım olmak üzere 18 bankanın 2005-2019 yıllarındaki (her bir yıl için 4’er çeyreklik olmak üzere) verilerinden faydalanılarak takibe dönüş oranının yukarıda adları sayılan değişkenler ile ilişkisi ortaya konacaktır. Veri seti oluşturulurken takibe dönüşüm oranı bağımlı değişken, toplam sanayi üretim endeksi, işsizlik, döviz kuru, enflasyon, faiz oranları, ithalat, ihracat ise bağımsız değişkenler olarak alınmıştır.

Tablo 7.2: Çalışmada kullanılan mevduat ve katılım bankaları

Mevduat Bankaları	Mevduat Bankaları	Katılım Bankaları
Akbank	QNB Finansbank	Albaraka Türk Katılım B.
Denizbank	Şekerbank	Emlak Katılım B.
Garanti B.	Türk Ekonomi B. (TEB)	Kuveyt Türk Katılım B.
Halkbank	Vakıfbank	Türkiye Finans Katılım B.
ING B.	Ziraat B.	Vakıf Katılım
İş B.	Yapı Kredi	Ziraat Katılım

Uygulama RStudio programında yapılmıştır. Verilere ilişkin düğüm değerleri MARS programından elde edilmiştir. Mars programı bir arama prosedürü algoritması ile düğüm değerlerini belirler (Oğuz, 2014). Kırılma noktası mevcut olan değişkenleri bünyesinde tutup diğer gözlem değerlerini veri setinden çıkararak 2. aşama olan GCV kriteri ile model seçimine başlar. GSYİH değişkeni için herhangi bir düğüm değeri oluşmadığı ve MARS tarafından bu değişkenin veri setinden çıkarıldığı görülmüştür. Bu bilgi ışığında amacı literatürdeki GCV’ye yönelik eleştirilere bir çözüm sunabilmek için GCV yerine alternatif bilgi kriterlerini kullandığımız çalışmamızda, GSYİH ‘yı veri setimize dahil etmedik.

$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$ ($i=1,2,\dots,n$) gibi çoklu regresyon modelinin β katsayılar vektörünün EKK tahmin edicisi $\hat{\beta} = (X'X)^{-1}X'Y$ formülü ile elde edilmiştir. Ancak elde edilen GSYİH, İhracat ve İthalat verilerinin sayısal değerlerinin çok büyük olması nedeniyle bu verilerin doğal logaritmaları alınmış ve analizde dönüştürülmüş veriler kullanılmıştır. Çünkü logaritması alınmamış veriler kullanıldığında tahmin edicinin $(X'X)^{-1}$ kısmı hesaplandığında (1/çok büyük sayı) oluştuğu için β katsayıları aşırı küçük değerler olarak elde edilmiştir. Regresyon denkleminin yorumlanması açısından da zorluk yaratan bu durum değerlerin logaritması alınarak çözülmüştür.

MARS programından elde ettiğimiz düğüm değerlerini kullanarak 4 farklı kriter için model seçimi gerçekleştirilmiştir. Her bir değişken için elde edilen düğüm değerleri ve düğüm değerlerine karşılık gelen sıra sayıları aşağıdaki gibidir.

Tablo 7.3: Gerçek veri değişkenlerine ait düğüm değerleri ve sıra sayıları

Değişken Adı	Düğüm Değeri		Ln Alınmamış Değeri		Sıra Değeri	
İşsizlik Oranı	7,3	10,1	-	-	31	16
Enflasyon	3,99	-	-	-	36	-
Faiz Oranı	8,53	-	-	-	36	-
Toplam Sanayi Üretim Endeksi	53,3	82,99	-	-	60	30
Döviz Kuru	1,2	2,26	-	-	48	21
İhracat (Ln)	16,71	17,33	18.214.793	33.797.332	3	28
İthalat (Ln)	17,14	17,86	27.822.542	57.234.288	1	50

Bu düğüm değerleri ile 12 adet normal ve 12 adet ayna görüntü olmak üzere 24 adet “temel fonksiyon” (Base Function) (BF) üretilmiştir.

Tablo 7.4: 12 adet ana etkiye sahip BF'nin oluşturulması

BF1	BF2	BF3	...	BF12
işsizlik(1)-7,3	işsizlik (1)-10,1	enflasyon(1)-3,99	...	ithalat(Ln)(1)-17,86
işsizlik(2)-7,3	işsizlik (2)-10,1	enflasyon(2)-3,99	...	ithalat(Ln)(2)-17,86
⋮	⋮	⋮	⋮	⋮
işsizlik(60)-7,3	işsizlik (60)-10,1	enflasyon(60)-3,99	...	ithalat(Ln)(60)-17,86

Ayna görüntüleri için de [düğüm değeri-X(i)] olacak şekilde BF13,...,BF24 temel fonksiyonları elde edilmiştir. Burada unutulmaması gereken nokta; $(x - t)_+$ ve $(t - x)_+$ formunda gösterilen parçalı doğrusal temel fonksiyon açılımında “+” alt indisi pozitif kısmı belirtip, istenen koşulun sağlanmadığı durumda temel fonksiyonun sıfır sonucunu alacağıdır. Bu sayede 60 satır ve 24 sütuna sahip BF matrisimiz oluşmuş oldu. Değişkenlerin etkileşimlerinin de modele katkısını inceleyebilmek için ikili etkileşimler oluşturulacaktır.

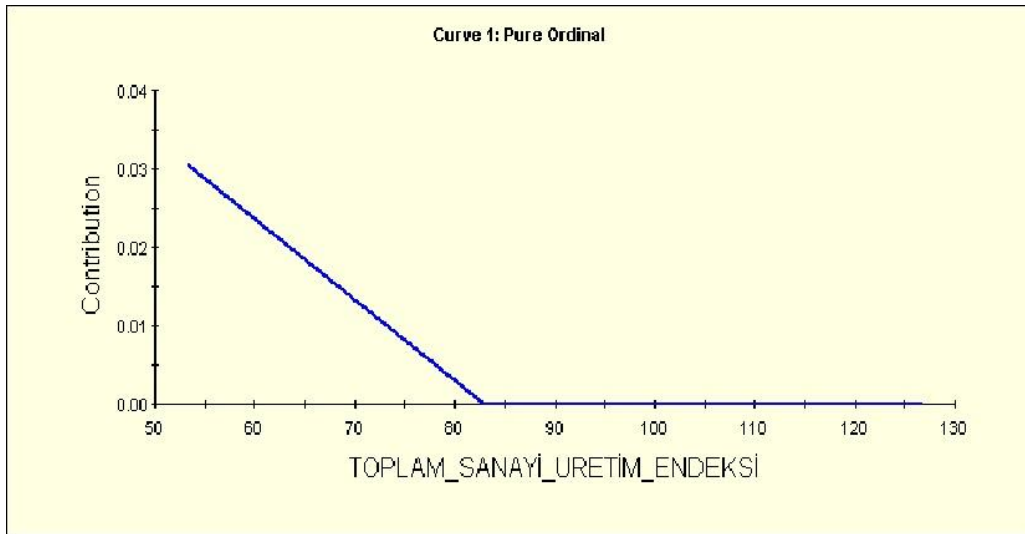
Tablo 7.5: 24 adet BF'ye ilişkin 2'li etkileşimlerin oluşturulması

BF1*BF2	BF1*BF3	BF1*BF4	...	BF1*BF24	→	23 tane 2'li
	BF2*BF3	BF2*BF4	...	BF2*BF24	→	22 tane 2'li
			⋮			⋮
				BF23*BF24	→	1 tane 2'li etkileşim
					=	Toplamda 276 adet

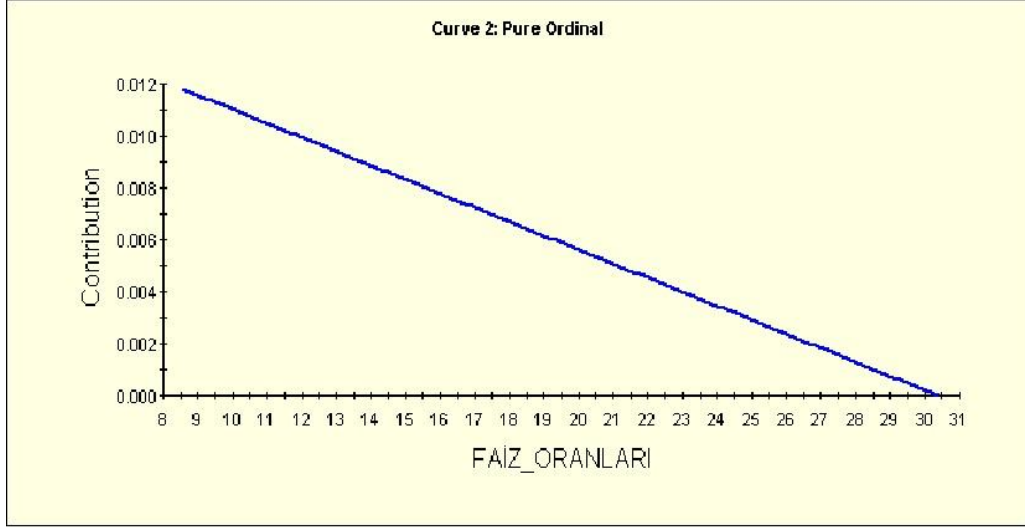
Bu BF'lere ilişkin ikili etkileşimler de modele dahil edilerek 24 BF ana etki, 276 BF ikili etkileşimli olmak üzere 300 açıklayıcı değişken oluşturuldu.

Bu şekilde çok büyük bir veri seti (veri setini çok daha fazla büyüteceği için 3'lü ve daha büyük etkileşimler çalışma açısından pratik kabul edilmiyor (Koç ve Bozdoğan, 2014)) olduğundan, veri setimiz artık büyük veriye (bigdata) dönüşmüş oldu. Bu kadar büyük değişkenli bir regresyon yapısının incelenmesi mümkün olmadığından temel fonksiyon sayısına kısıt getirmemiz gerekiyor. Bu amaçla, modele dahil edilecek maksimum temel fonksiyonlar 25 sayısı ile kısıtlandı. Sonuç olarak diyebiliriz ki, 300 BF (ana etki (etkileşimsiz), etkileşimli ve ayna fonksiyonları) içinden MARS yöntemi ile doğru seçim ve geriye doğru eleme adımlarından sonra, en fazla 25 temel fonksiyon içeren final modeli oluşturulabilecek.

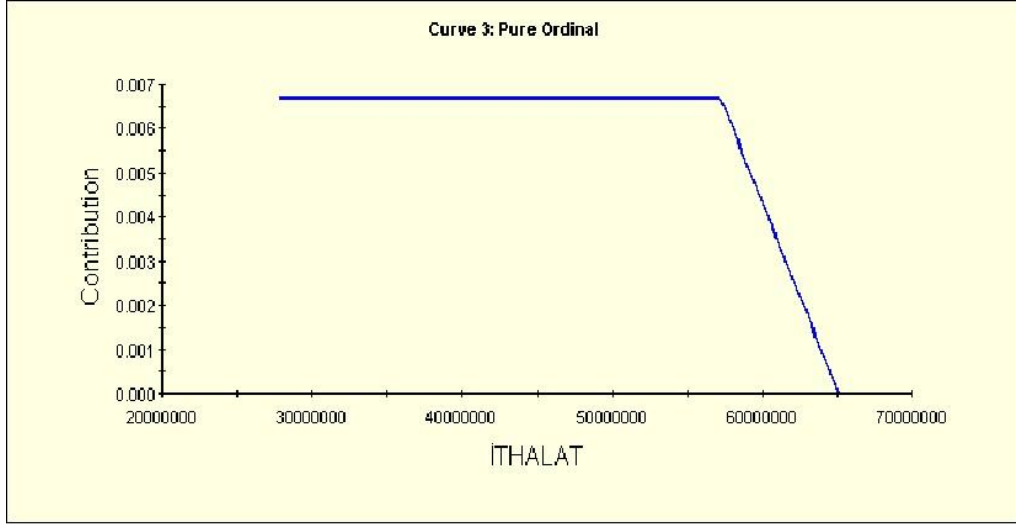
Bunlara ek olarak, MARS GCV final modelinde kalan değişkenlerin düğüm değerlerinin belirlenip dönüşümler aracılığıyla doğrusala çevrilmiş olan hallerinin, bağımlı değişken üzerindeki ilişkilerini gösteren iki boyutlu parçalı doğrusal grafikleri aşağıda gösterilmiştir.



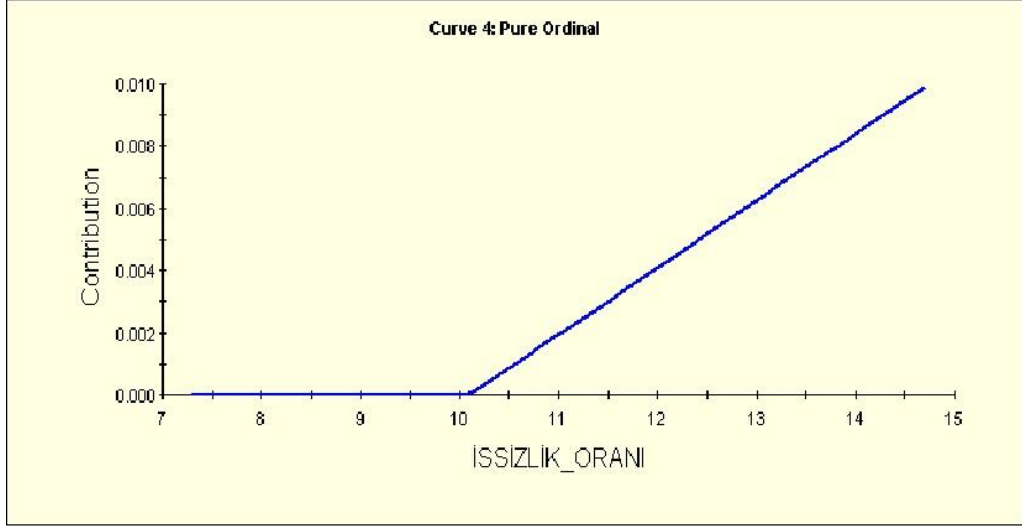
Şekil 7.2: Toplam Sanayi Üretim Endeksi Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği [$\max(0, \text{Toplam_Sanayi_Üretim_Endeksi} - 53, 3)$; $\max(0, \text{Toplam_Sanayi_Üretim_Endeksi} - 82, 99)$]



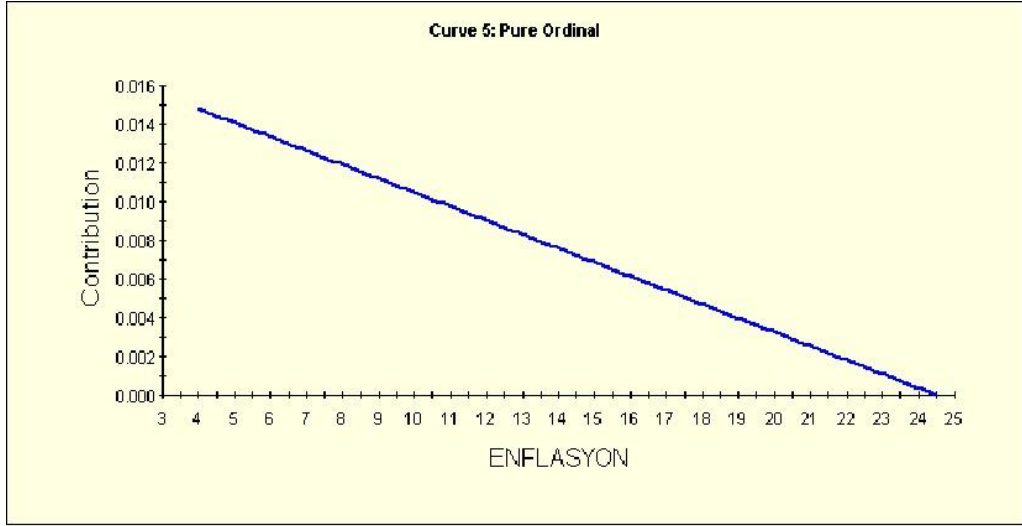
Şekil 7.3: Faiz Oranları Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği [$\max(0, \text{Faiz Oranları}-8,5)$]



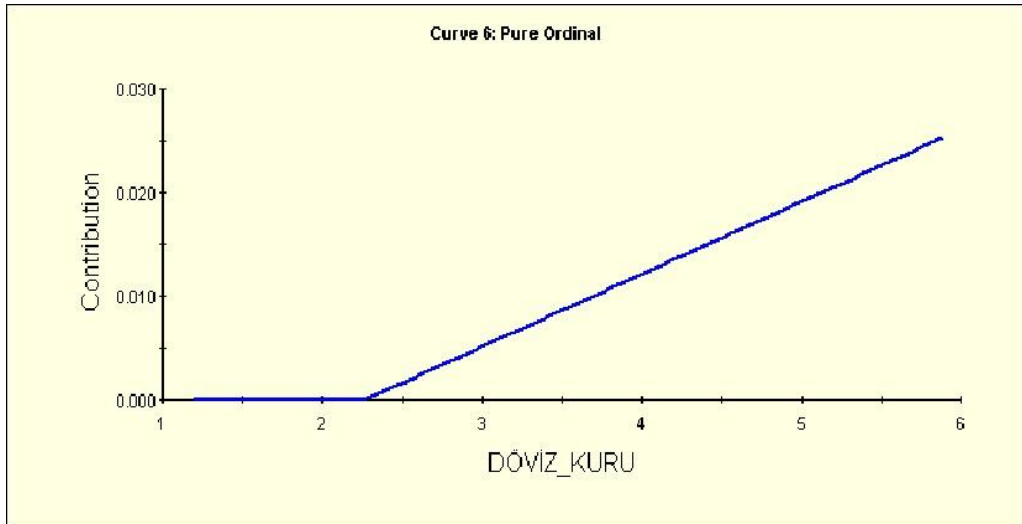
Şekil 7.4: İthalat Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği [$\max(0, \text{İthalat}-27.822.542)$; $\max(0, \text{İthalat}-65.085.408)$]



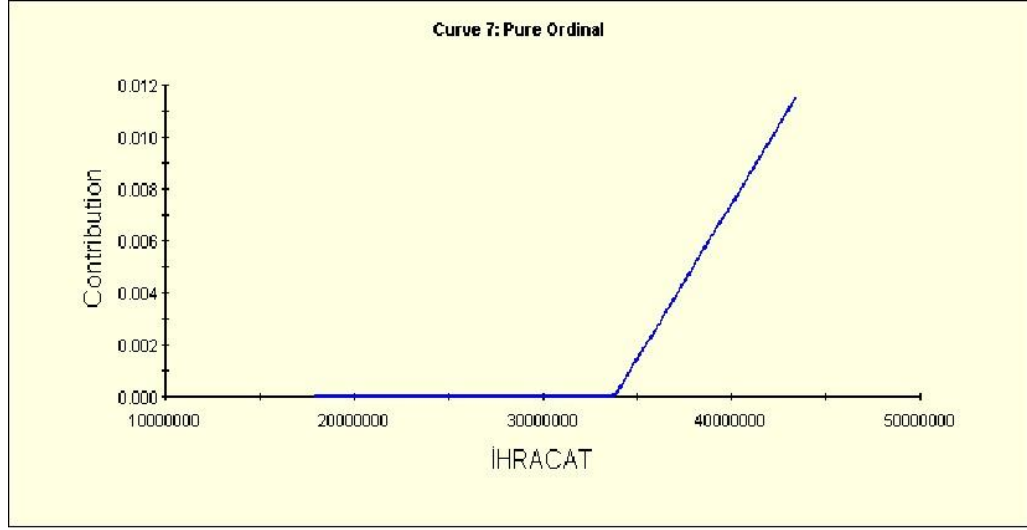
Şekil 7.5: İşsizlik Oranı Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği [$\max(0, \text{İşsizlik_Oranı}-7,3)$; $\max(0, \text{İşsizlik_Oranı}-10,1)$]



Şekil 7.6: Enflasyon Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği [$\max(0, \text{Enflasyon}-3,99)$]



Şekil 7.7: Döviz Kuru Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği [max (0, Döviz_Kuru-1,2); max (0, Döviz_Kuru-2,26)]



Şekil 7.8: İhracat Değişkeninin Dönüşüm İle Doğrusallaştırılmış İki Boyutlu Grafiği [max (0, İhracat-18.214.793); max (0, İhracat-33.797.332)]

7.1 MARS_GCV Model Seçimi

GCV kriterinin RStudio programına yazdırılması:

$$*GCV(M) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{f}_M(x_i))^2}{\left(1 - \frac{P(M)^*}{n}\right)^2}$$

$$P(M) = \text{iz}(\mathbf{B}(\mathbf{B}^T)^{-1}\mathbf{B}^T) + 1$$

$$P(M)^* = P(M) + dM$$

```
critGCV <- function(loglik,k,betahat,n,Xm)
```

```
{
```

```
  PM=(sum(diag((Xm%*%ginv(t(Xm)%*%Xm)%*%t(Xm)))))+1+(3*(k-1)))
```

```
  GCV_crit=(1/n)*(exp(loglik*(-2/n))*n)/((1-PM/n)^2)
```

```
  return(GCV_crit)
```

```
}
```

GCV formülünde verilen \mathbf{B} BF lerden oluşan matrisi ifade etmektedir. Aynı zamanda yeni X değerlerimiz anlamına da geldiği için X_m olarak kodlandı. Friedman tarafından 'd' için en iyi değer $2 \leq d \leq 4$ aralığında olduğu ifade edildiği için $d=3$ olarak alınmıştır. Kod ile ifade edilen k değeri modelimizdeki parametre

sayısını ifade etmektedir. Örneğin modelimizin $Y = \beta_0 + \beta_1 BF1 + \beta_5 BF5$ şeklinde oluştuğunu varsayalım. Bu durumda k değeri 3'e eşit olacaktır ($\beta_0, \beta_1, \beta_5$). M ise modeldeki BF ların sayısını temsil eder. Bu durumda örneğimiz için M=2 olacaktır (BF1, BF5). Aynı zamanda M=k-1 eşitliğinden söz edebiliriz. Bu sebeple formüldeki M değeri kod kısmında (k-1) ile ifade edildi. GCV formülünün pay kısmı ($\sum_{i=1}^n (y_i - \hat{f}_M(x_i))^2$) olan hata kareler toplamı ($e^{\log\text{-likelihood} * \frac{-2}{n} * n}$) olarak alınmıştır. Rstudio programında hazır olan komut aynı sonucu vermekle birlikte işlemde kolaylık sağlamaktadır. Bilindiği gibi MARS tekniği ileri doğru seçim ve geriye doğru eleme adımlarından oluşmaktadır. GCV kriteri kullanılarak elde edilecek MARS modeli için bu adımların kodları ise aşağıdaki gibi oluşturulmuştur.

```
fast_forward(datatabig, crit=aic,maxf = limit)%>%
```

```
multi_backward(crit=critGCV)->mmGCV
```

İleri doğru seçim aşamasında, hemen hemen tüm seçim prosedürlerinde ortak olarak kullanılan Akaike Bilgi Kriteri (AIC) ve Determinasyon (açıklayıcılık) Katsayısı (R^2), arasından tercihimizi AIC yönünde kullandık. Bunun nedeni, R^2 'nin değişkenler arasında ilişki olması durumunda (örneğin çalışmamızda kullandığımız enflasyon ve faiz oranı arasında ilişkiden söz edebiliriz) yapay (kovaryans _özilişki_ hatası) olarak yükseliyor olması (Heo vd. 2008); AIC değerinin ise model seçiminde olabilirlik tahminini baz alarak uyum hassasiyeti ile bu uyumu sağlamak için kullandığı parametre sayısı arasında bir denge oluşturmaya odaklanmasıdır.

Geriye doğru eleme adımı ise GCV formülü kullanılarak hesaplatılmıştır. mmGCV kodu ise GCV ile oluşan multi modeli yani mümkün tüm temel fonksiyonları (BF) oluştur anlamına gelmektedir.

GCV kriteri ile elde edilen MARS modeli

$$\begin{aligned}
 Y = & 0.0007971 + 0.0007471*BF1 - 0.0015716*BF2 + 0.0005623*BF3 - \\
 & 0.0001953*BF4 + 0.0014005*BF5 - 0.0021203*BF6 + 0.0002353*BF8 + \\
 & 0.0000713*BF9 - 0.0001196*BF11 + 0.0007151*BF13 + 0.0000815*BF15 - \\
 & 0.0008827*BF18 + 0.0005647*BF20 - 0.000047*BF21 + 0.0003499*BF1*BF4 - \\
 & 0.0000503*BF1*BF5 - 0.000032*BF1*BF7 + 0.0006535*BF1*BF8 - \\
 & 0.0001343*BF1*BF10 - 0.0002163*BF1*BF12 - 0.0001099*BF1*BF17 +
 \end{aligned}$$

$$0.0007017*BF1*BF18 - 0.0007195*BF1*BF20 - 0.0002283*BF1*BF21 + 0.0006874*BF1*BF22$$

GCV için BF gösterimleri

- [1] "BF1=ISSIZLIK_ORANI-7.3"
- [2] "BF2=ISSIZLIK_ORANI-10.1"
- [3] "BF3=ENFLASYON-3.99"
- [4] "BF4=FAIZ_ORANLARI-8.5392"
- [5] "BF5=TOPLAM SANAYI_URETIM_ENDEKSI-82.9933"
- [6] "BF6=TOPLAM SANAYI_URETIM_ENDEKSI-53.3"
- [7] "BF8=DOVIZ_KURU-2.26"
- [8] "BF9=lnIHRACAT-16.71"
- [9] "BF11=lnITHALAT-17.8627"
- [10] "BF13=7.3-ISSIZLIK_ORANI"
- [11] "BF15=3.99-ENFLASYON"
- [12] "BF18=53.3-TOPLAM SANAYI_URETIM_ENDEKSI"
- [13] "BF20=2.26-DOVIZ_KURU"
- [14] "BF21=16.71-lnIHRACAT"
- [15] "BF1*BF4=ISSIZLIK_ORANI-7.3*FAIZ_ORANLARI-8.5392"
- [16] "BF1*BF5=ISSIZLIK_ORANI-7.3*TOPLAM SANAYI_URETIM_ENDEKSI-82.9933"
- [17] "BF1*BF7=ISSIZLIK_ORANI-7.3*DOVIZ_KURU-1.2"
- [18] "BF1*BF8=ISSIZLIK_ORANI-7.3*DOVIZ_KURU-2.26"
- [19] "BF1*BF10=ISSIZLIK_ORANI-7.3*lnIHRACAT-17.3359"
- [20] "BF1*BF12=ISSIZLIK_ORANI-7.3*lnITHALAT-17.14"
- [21] "BF1*BF17=ISSIZLIK_ORANI-7.3*82.9933-TOPLAM SANAYI_URETIM_ENDEKSI"
- [22] "BF1*BF18=ISSIZLIK_ORANI-7.3*53.3-TOPLAM SANAYI_URETIM_ENDEKSI"
- [23] "BF1*BF20=ISSIZLIK_ORANI-7.3*2.26-DOVIZ_KURU"
- [24] "BF1*BF21=ISSIZLIK_ORANI-7.3*16.71-lnIHRACAT"
- [25] "BF1*BF22=ISSIZLIK_ORANI-7.3*17.3359-lnIHRACAT"

GCV kriteri kullanılarak toplamda 300 BF arasından 25 temel fonksiyon (BF) içeren bir MARS modeli elde edilmiştir. Bunların 11 tanesi ikili etkileşim fonksiyonu

olarak modele dahil olmuştur. GCV kriteri ile oluşturulan modelde işsizlik, enflasyon, faiz, toplam sanayi üretim endeksi, ihracat, ithalat ve döviz kuru değişkenleri ana etki olarak modele katkı sunmaktadırlar. İşsizlik Oranını gösteren BF1, BF2, BF13 temel fonksiyonlarının sırasıyla 7,3 değerindeki kırılma noktası için modele katkısı pozitif (0.0007471); 10,1 değerindeki kırılma noktası için modele katkısı negatif (-0.0015716) ve ayna görüntüsü olarak modele giren 7,3 değerindeki düğüm değeri için modele katkısı pozitif (0.0007151) olmuştur şeklinde yorumlanır.

GCV kriteri kullanılarak ANOVA tablosunun oluşturulması için yazdırılan kodlar:

GCV ile elde edilen BF ler incelendiğinde ilk değişkenin (BF1) İşsizlik Oranı olduğu görülecektir. Aynı zamanda bu değişken BF2 ve BF13 olarak da modele dahil olmuştur. İşsizlik Oranı için doğrusal regresyon modeli oluşturmak için bu üç BF değeri kullanılmıştır [regGCVX1=lm(y~BF1+BF2+BF13)]. Benzer şekilde modele giren diğer değişken Enflasyon'dur. Enflasyon BF3 ve BF15 olmak üzere iki temel fonksiyonla modele katılmıştır. Regresyon denklemi ise [regGCVX2=lm(y~BF3+BF15)] kodu ile oluşturulmuştur. Bu işlem modeldeki tüm değişkenler için aşağıdaki gibi ayrı ayrı hesaplatılmıştır. Elde edilen regresyon denklemleri kullanılarak standart sapma değeri hesaplanacaktır.

$$\text{regGCVX1}=\text{lm}(y\sim\text{BF1}+\text{BF2}+\text{BF13})$$

$$\text{regGCVX2}=\text{lm}(y\sim\text{BF3}+\text{BF15})$$

$$\text{regGCVX3}=\text{lm}(y\sim\text{BF4})$$

$$\text{regGCVX4}=\text{lm}(y\sim\text{BF5}+\text{BF6}+\text{BF18})$$

$$\text{regGCVX5}=\text{lm}(y\sim\text{BF8}+\text{BF20})$$

$$\text{regGCVX6}=\text{lm}(y\sim\text{BF9}+\text{BF21})$$

$$\text{regGCVX7}=\text{lm}(y\sim\text{BF11})$$

$$\text{regGCVX8}=\text{lm}(y\sim\text{BF1}*\text{BF4})$$

$$\text{regGCVX9}=\text{lm}(y\sim\text{BF1}*\text{BF5}+\text{BF1}*\text{BF17}+\text{BF1}*\text{BF18})$$

$$\text{regGCVX10}=\text{lm}(y\sim\text{BF1}*\text{BF7}+\text{BF1}*\text{BF8}+\text{BF1}*\text{BF20})$$

$$\text{regGCVX11}=\text{lm}(y\sim\text{BF1}*\text{BF10}+\text{BF1}*\text{BF21}+\text{BF1}*\text{BF22})$$

$$\text{regGCVX12}=\text{lm}(y\sim\text{BF1}*\text{BF12})$$

Oluşturulan regresyon denklemleri kullanılarak standart sapma $\sqrt{\frac{\sum(Y-\bar{Y})^2}{n-1}}$

değeri $GCV_{anova1}=\sqrt{\sum(\text{regGCVX1}\$residuals^2)/(n-1)}$ şeklinde kodlanmıştır. Bu işlem 12 BF için tekrar edilmiştir.

$$GCV_{anova1}=\sqrt{\sum(\text{regGCVX1}\$residuals^2)/(n-1)}$$

$$GCV_{anova2}=\sqrt{\sum(\text{regGCVX2}\$residuals^2)/(n-1)}$$

$$GCV_{anova3}=\sqrt{\sum(\text{regGCVX3}\$residuals^2)/(n-1)}$$

$$GCV_{anova4}=\sqrt{\sum(\text{regGCVX4}\$residuals^2)/(n-1)}$$

$$GCV_{anova5}=\sqrt{\sum(\text{regGCVX5}\$residuals^2)/(n-1)}$$

$$GCV_{anova6}=\sqrt{\sum(\text{regGCVX6}\$residuals^2)/(n-1)}$$

$$GCV_{anova7}=\sqrt{\sum(\text{regGCVX7}\$residuals^2)/(n-1)}$$

$$GCV_{anova8}=\sqrt{\sum(\text{regGCVX8}\$residuals^2)/(n-1)}$$

$$GCV_{anova9}=\sqrt{\sum(\text{regGCVX9}\$residuals^2)/(n-1)}$$

$$GCV_{anova10}=\sqrt{\sum(\text{regGCVX10}\$residuals^2)/(n-1)}$$

$$GCV_{anova11}=\sqrt{\sum(\text{regGCVX11}\$residuals^2)/(n-1)}$$

$$GCV_{anova12}=\sqrt{\sum(\text{regGCVX12}\$residuals^2)/(n-1)}$$

Son olarak da her farklı fonksiyona karşılık gelen GCV değeri hesaplatılmıştır.

$$GCVX1=\text{cbind}(BF1,BF2,BF13)$$

$$PM1=(\sum(\text{diag}((GCVX1\%*\%ginv(t(GCVX1)\%*\%GCVX1)\%*\%t(GCVX1))))+1+(3*(3)))$$

$$GCV_crit1=(1/n)*\sum(\text{regGCVX1}\$residuals^2)/((1-PM1/n)^2)$$

cbind ifadesi BF1,BF2,BF13 değişkenlerini sütun olarak birleştir komutu olarak yazılmıştır.

GCV modelinde İşsizlik Oranı için BF1, BF2 ve BF13 temel fonksiyonları elde edilmişti. İşsizlik Oranı için GCV değeri bu üç BF değeri kullanılarak hesaplatılmıştır.

$$GCVX2=\text{cbind}(BF3,BF15)$$

$$PM2=(\text{sum}(\text{diag}((\text{GCVX2}\% \%\% \text{ginv}(\text{t}(\text{GCVX2})\% \%\% \text{GCVX2})\% \%\% \text{t}(\text{GCVX2}))))+1+(3*(2)))$$

$$\text{GCV_crit2}=(1/n)*\text{sum}(\text{regGCVX2}\$residuals^2)/((1-PM2/n)^2)$$

$$PM3=(\text{sum}(\text{diag}((\text{BF4}\% \%\% \text{ginv}(\text{t}(\text{BF4})\% \%\% \text{BF4})\% \%\% \text{t}(\text{BF4}))))+1+(3*(1)))$$

$$\text{GCV_crit3}=(1/n)*\text{sum}(\text{regGCVX3}\$residuals^2)/((1-PM3/n)^2)$$

$$\text{GCVX4}=\text{cbind}(\text{BF5},\text{BF6},\text{BF18})$$

$$PM4=(\text{sum}(\text{diag}((\text{GCVX4}\% \%\% \text{ginv}(\text{t}(\text{GCVX4})\% \%\% \text{GCVX4})\% \%\% \text{t}(\text{GCVX4}))))+1+(3*(3)))$$

$$\text{GCV_crit4}=(1/n)*\text{sum}(\text{regGCVX4}\$residuals^2)/((1-PM4/n)^2)$$

$$\text{GCVX5}=\text{cbind}(\text{BF8},\text{BF20})$$

$$PM5=(\text{sum}(\text{diag}((\text{GCVX5}\% \%\% \text{ginv}(\text{t}(\text{GCVX5})\% \%\% \text{GCVX5})\% \%\% \text{t}(\text{GCVX5}))))+1+(3*(2)))$$

$$\text{GCV_crit5}=(1/n)*\text{sum}(\text{regGCVX5}\$residuals^2)/((1-PM5/n)^2)$$

$$\text{GCVX6}=\text{cbind}(\text{BF9},\text{BF21})$$

$$PM6=(\text{sum}(\text{diag}((\text{GCVX6}\% \%\% \text{ginv}(\text{t}(\text{GCVX6})\% \%\% \text{GCVX6})\% \%\% \text{t}(\text{GCVX6}))))+1+(3*(2)))$$

$$\text{GCV_crit6}=(1/n)*\text{sum}(\text{regGCVX6}\$residuals^2)/((1-PM6/n)^2)$$

$$PM7=(\text{sum}(\text{diag}((\text{BF11}\% \%\% \text{ginv}(\text{t}(\text{BF11})\% \%\% \text{BF11})\% \%\% \text{t}(\text{BF11}))))+1+(3*(1)))$$

$$\text{GCV_crit7}=(1/n)*\text{sum}(\text{regGCVX7}\$residuals^2)/((1-PM7/n)^2)$$

$$\text{GCVX8}=\text{cbind}(\text{BF1}*\text{BF4})$$

$$PM8=(\text{sum}(\text{diag}((\text{GCVX8}\% \%\% \text{ginv}(\text{t}(\text{GCVX8})\% \%\% \text{GCVX8})\% \%\% \text{t}(\text{GCVX8}))))+1+(3*(2)))$$

$$\text{GCV_crit8}=(1/n)*\text{sum}(\text{regGCVX8}\$residuals^2)/((1-PM8/n)^2)$$

$$\text{GCVX9}=\text{cbind}(\text{BF1}*\text{BF5},\text{BF1}*\text{BF17},\text{BF1}*\text{BF18})$$

$$PM9=(\text{sum}(\text{diag}((\text{GCVX9}\% \%\% \text{ginv}(\text{t}(\text{GCVX9})\% \%\% \text{GCVX9})\% \%\% \text{t}(\text{GCVX9}))))+1+(3*(3)))$$

$$\text{GCV_crit9}=(1/n)*\text{sum}(\text{regGCVX9}\$residuals^2)/((1-PM9/n)^2)$$

$$\text{GCVX10}=\text{cbind}(\text{BF1}*\text{BF7},\text{BF1}*\text{BF8},\text{BF1}*\text{BF20})$$

$$PM10=(\text{sum}(\text{diag}((\text{GCVX10}\% \%\% \text{ginv}(\text{t}(\text{GCVX10})\% \%\% \text{GCVX10})\% \%\% \text{t}(\text{GCVX10}))))+1+(3*(3)))$$

$$\text{GCV_crit10}=(1/n)*\text{sum}(\text{regGCVX10}\$residuals^2)/((1-PM10/n)^2)$$

$$\text{GCVX11}=\text{cbind}(\text{BF1}*\text{BF10},\text{BF1}*\text{BF21},\text{BF1}*\text{BF22})$$

$$PM11=(\text{sum}(\text{diag}((\text{GCVX11}\% \%\% \text{ginv}(\text{t}(\text{GCVX11})\% \%\% \text{GCVX11})\% \%\% \text{t}(\text{GCVX11}))))+1+(3*(3)))$$

$$\text{GCV_crit11}=(1/n)*\text{sum}(\text{regGCVX11}\$residuals^2)/((1-PM11/n)^2)$$

$$\text{GCVX12}=\text{cbind}(\text{BF1}*\text{BF12})$$

$$PM12=(\text{sum}(\text{diag}((\text{GCVX12}\% \%\% \text{ginv}(\text{t}(\text{GCVX12})\% \%\% \text{GCVX12})\% \%\% \text{t}(\text{GCVX12}))))+1+(3*(1)))$$

$$\text{GCV_crit12}=(1/n)*\text{sum}(\text{regGCVX12}\$residuals^2)/((1-PM6/n)^2)$$

Elde edilen deęerleri ile ANOVA tablosu hazırlanmıřtır.

Tablo 7.6: GCV kriteri kullanılarak elde edilen ANOVA ayrıřımı

Fonksiyon	St. Sapma	GCV	#Temel Fonk.	#Serbestlik Derecesi	Deęiřkenler
1	0.010177	0.00016599	3	7.5	İřsizlik
2	0.011589	0.00018278	2	5.0	Enflasyon
3	0.011331	0.00015024	1	2.5	Faiz
4	0.009101	0.00013274	3	7.5	Top. San. En.
5	0.011466	0.00017893	2	5.0	Döviz Kuru
6	0.010393	0.00014700	2	5.0	İhracat
7	0.011366	0.00015118	1	2.5	İthalat
8	0.009908	0.00012853	1	2.5	İřsizlik*Faiz
9	0.007044	0.00007952	3	7.5	İřsizlik*Top. San. En
10	0.007392	0.00008757	3	7.5	İřsizlik*Döviz
11	0.006905	0.00007642	3	7.5	İřsizlik*İhracat
12	0.010707	0.00015602	1	2.5	İřsizlik*İthalat

Yukarıda MARS regresyonun GCV kriterine baęlı olarak elde edilmiř anlamlı deęiřkenlerinin ANOVA ayrıřımı sonuçları verilmiřtir. Bu tabloda final modele giren her bir deęiřkenin baęımlı deęiřken üzerine olan genel etkileri görölmektedir. ANOVA tablosundaki standart sapmaların büyüklüęü o deęiřkenin baęımlı deęiřken üzerinde olan genel etkisinin fazla olduęunu gösterir. Üçüncü sütunda yer alan GCV deęerleri, söz konusu deęiřkene ait temel fonksiyonları modelden çıkardıęımızda tahminlerde meydana gelecek kaybı göstermektedir. Bu durumda 2 temel fonksiyon ile modele giren Enflasyon deęiřkeninin standart sapması ve GCV deęeri (st.sapma: 0.011589; GCV: 0.00018278) incelendięinde Takibe Dönüř Oranı üzerinde en etkili deęiřken olduęu söylenebilir. Modele katkı saęlaması bakımından Enflasyon deęiřkenini Döviz Kuru (st.sapma: 0.011466; GCV: 0.00017893) izlemektedir. Beřinci sütunda yer alan serbestlik derecesi deęerleri için ¹³Koç ve Bozdoęanın (2013) makalesi referans alınmıřtır. Buradan yola çıkılarak modele dahil olunan her bir temel fonksiyon için serbestlik derecesi 2,5 olarak belirlenmiřtir.

¹³ The fourth column gives the number of BFs comprising the ANOVA function while fifth column provides and estimate of the additional number of linear degrees-of-freedom; syf:14

7.2 MARS_AIC Model Seçimi

AIC kriterinin RStudio programına yazdırılması işlemi:

$$*AIC = -2\log L(\hat{Q}) + 2k$$

```
fast_forward(databig, crit=aic,maxf = limit)%>%
```

```
multi_backward(crit=aic)->mmAIC
```

Rstudio programında hazır olan AIC formülü sayesinde komut direkt çağrılmış ve işlemde kolaylık sağlanmıştır. Nedeni GCV kriteri model seçiminde bahsedildiği üzere ileri doğru seçim adımında model oluşturma işlemi AIC kriteri ile gerçekleştirilmiştir. Final model oluşum kodu geriye doğru eleme adımında AIC kriteri kullanılarak oluşturulmuştur.

AIC kriteri ile elde edilen MARS modeli

$$Y = 0.0002469 - 0.00059*BF2 + 0.0007217*BF3 - 0.0002371*BF5 - 0.0004872*BF6 + 0.0014673*BF8 + 0.0008853*BF11 - 0.000493*BF18 + 0.0015257*BF20 - 0.0032715*BF21 - 0.000517*BF1*BF4 - 0.0000766*BF1*BF5 + 0.0007414*BF1*BF7 - 0.0004561*BF1*BF10 + 0.0007447*BF1*BF12 + 0.0022892*BF1*BF17 - 0.0020518*BF1*BF18 - 0.0019613*BF1*BF20 + 0.001582*BF1*BF21 + 0.0008637*BF1*BF22$$

AIC için BF gösterimleri

[1] "BF2=ISSIZLIK_ORANI-10.1"

[2] "BF3=ENFLASYON-3.99"

[3] "BF5=TOPLAM SANAYI_URETIM_ENDEKSI-82.9933"

[4] "BF6=TOPLAM SANAYI_URETIM_ENDEKSI-53.3"

[5] "BF8=DOVIZ_KURU-2.26"

[6] "BF11=lnITHALAT-17.8627"

[7] "BF18=53.3-TOPLAM SANAYI_URETIM_ENDEKSI"

[8] "BF20=2.26-DOVIZ_KURU"

[9] "BF21=16.71-lnIHRACAT"

[10] "BF1*BF4=ISSIZLIK_ORANI-7.3*FAIZ_ORANLARI-8.5392"

[11] "BF1*BF5=ISSIZLIK_ORANI-7.3*TOPLAM SANAYI_URETIM_ENDEKSI-82.9933"

[12] "BF1*BF7=ISSIZLIK_ORANI-7.3*DOVIZ_KURU-1.2"

- [13] "BF1*BF10=ISSIZLIK_ORANI-7.3*lnIHRACAT-17.3359"
- [14] "BF1*BF12=ISSIZLIK_ORANI-7.3*lnITHALAT-17.14"
- [15] "BF1*BF17=ISSIZLIK_ORANI-7.3*82.9933-TOPLAM SANAYI_URETIM_ENDEKSI"
- [16] "BF1*BF18=ISSIZLIK_ORANI-7.3*53.3-TOPLAM SANAYI_URETIM_ENDEKSI"
- [17] "BF1*BF20=ISSIZLIK_ORANI-7.3*2.26-DOVIZ_KURU"
- [18] "BF1*BF21=ISSIZLIK_ORANI-7.3*16.71-lnIHRACAT"
- [19] "BF1*BF22=ISSIZLIK_ORANI-7.3*17.3359-lnIHRACAT"

AIC kriteri kullanılarak elde edilen MARS modeli 19 temel fonksiyon içermektedir. Bunların 10 tanesi ikili etkileşime girerek, 9 tanesi ise ana etki olarak modele dahil olmuştur. Modele ana etki olarak katkı sunan değişkenler İşsizlik, Enflasyon, Toplam Sanayi Üretim Endeksi, Döviz Kuru, İthalat ve İhracat'tır. GCV kriterine kıyasla daha az parametrelili bir model oluştuğu görülmektedir.

AIC kriteri kullanılarak ANOVA tablosunun oluşturulması için yazdırılan kodlar:

AIC kriteri için oluşturulan regresyon modeli ve standart sapma değeri hesaplama işlemi GCV kriteri ile benzer mantıktadır. AIC değeri ise

boy1=1

loglik1=as.numeric(logLik(regAICX1)[1])

AIC_crit1=(-2*loglik1)+2*boy1

kodu ile hesaplatılmıştır. Boy1 değeri modeldeki BF sayısına karşılık gelmektedir. Örneğin İşsizlik değişkeni BF2 fonksiyonu ile modele dahil olmuştur. Boy1 kodunun değeri 1'e eşittir şeklinde kodlanmıştır. Toplamdaki 11 adet değişken için aynı işlem tekrarlanıp ANOVA tablosu oluşturulmuştur.

Tablo 7.7: AIC kriteri kullanılarak elde edilen ANOVA ayrışımı

Fonksiyon	St. Sapma	AIC	#Temel Fonk.	#Serbestlik Derecesi	Değişkenler
1	0.011647	-363.1	1	2.5	İşsizlik
2	0.011607	-363.5	1	2.5	Enflasyon
3	0.009101	-388.7	3	7.5	Top. San. En.
4	0.011466	-362.9	2	5.0	Döviz Kuru
5	0.011366	-366.0	1	2.5	İthalat
6	0.011275	-367.0	1	2.5	İhracat
7	0.009908	-382.5	1	2.5	İşsizlik*Faiz
8	0.007044	-419.4	3	7.5	İşsizlik*Top. San. En.

9	0.008928	-393.0	2	5.0	İşsizlik*Döviz
10	0.006905	-421.8	3	7.5	İşsizlik*İhracat
11	0.010707	-373.2	1	2.5	İşsizlik*İthalat

MARS regresyonun AIC kriterine bağlı olarak elde edilmiş anlamlı değişkenlerinin ANOVA ayrışımı sonuçları elde edilmiştir. Standart sapma değerlerine bakıldığında İşsizlik ve Enflasyon değişkenlerinin bağımlı değişken üzerinde daha fazla etkilerinin olduğunu söyleyebiliriz. İşsizlik ve Döviz Kuru değişkenlerinin AIC değerlerinin diğer değişken değerlerine kıyasla daha büyük olması (Döviz Kuru AIC: -362.9; İşsizlik AIC: -363.1) bu değişkenlerin modelden çıkarıldıklarında modeldeki kaybın bu derece büyük olduğunu ifade ediyor.

7.3 MARS_SBC Model Seçimi

$$*SBC = -2\log L(\hat{Q}) + k\log(n)$$

```
fast_forward(databig, crit=aic,maxf = limit)%>%
```

```
multi_backward(crit=bic)->mmSBC
```

Rstudio programında hazır olan SBC formülü sayesinde formülün açık bir şekilde yazılmasına ihtiyaç duyulmamış komut direkt çağrılmış ve işlemde kolaylık sağlanmıştır. MARS model oluşumu geriye doğru eleme adımında SBC kriteri kullanılarak kod oluşturulmuştur.

SBC kriteri ile elde edilen MARS modeli

$$Y = -0.0002254 + 0.0005362*BF3 - 0.000312*BF5 - 0.000254*BF6 + 0.0007473*BF8 - 0.0005031*BF11 + 0.0009622*BF18 - 0.0000504*BF20 - 0.0002505*BF21 + 0.0001434*BF1*BF4 - 0.0005787*BF1*BF5 - 0.0002347*BF1*BF7 + 0.0003357*BF1*BF10 + 0.0002081*BF1*BF17 + 0.0000193*BF1*BF20 + 0.0001388*BF1*BF21 - 0.0000663*BF1*BF22$$

SBC için BF gösterimleri

[1] "BF3=ENFLASYON-3.99"

[2] "BF5=TOPLAM SANAYI_URETIM_ENDEKSI-82.9933"

[3] "BF6=TOPLAM SANAYI_URETIM_ENDEKSI-53.3"

[4] "BF8=DOVIZ_KURU-2.26"

- [5] "BF11=lnITHALAT-17.8627"
- [6] "BF18=53.3-TOPLAM SANAYI_URETIM_ENDEKSI"
- [7] "BF20=2.26-DOVIZ_KURU"
- [8] "BF21=16.71-lnIHRACAT"
- [9] "BF1*BF4=ISSIZLIK_ORANI-7.3*FAIZ_ORANLARI-8.5392"
- [10] "BF1*BF5=ISSIZLIK_ORANI-7.3*TOPLAM SANAYI_URETIM_ENDEKSI-82.9933"
- [11] "BF1*BF7=ISSIZLIK_ORANI-7.3*DOVIZ_KURU-1.2"
- [12] "BF1*BF10=ISSIZLIK_ORANI-7.3*lnIHRACAT-17.3359"
- [13] "BF1*BF17=ISSIZLIK_ORANI-7.3*82.9933-TOPLAM SANAYI_URETIM_ENDEKSI"
- [14] "BF1*BF20=ISSIZLIK_ORANI-7.3*2.26-DOVIZ_KURU"
- [15] "BF1*BF21=ISSIZLIK_ORANI-7.3*16.71-lnIHRACAT"
- [16] "BF1*BF22=ISSIZLIK_ORANI-7.3*17.3359-lnIHRACAT"

Bayesci bilgi kriteri olarak da bilinen SBC, AIC'in bir türevi olarak önerilmiştir. Bayes kuramını temel alan SBC'nin ceza terimi AIC'e göre daha büyüktür. Dolayısıyla SBC ile seçilecek olan modelin, AIC ile seçilecek modelden daha küçük veya eşit boyutta olması beklenir yorumu bilinmektedir. Analiz sonucu da bunu destekler niteliktedir. AIC ile 19 BF içeren model elde edilirken SBC ile bu değer 16 olarak elde edilmiştir. SBC ile elde edilen modelde ise Enflasyon, Toplam Sanayi Üretim Endeksi, Döviz Kuru, İthalat ve İhracat ana etki olarak katkı sunan değişkenlerdir.

SBC kriteri kullanılarak ANOVA tablosunun oluşturulması için yazdırılan kodlar:

SBC kriteri için oluşturulan regresyon modeli ve standart sapma değeri hesaplama işlemi AIC kriteri ile benzerdir. SBC değeri hesaplamada farklılık log(n) eklenmesinden oluşmaktadır.

boy1=1

loglik1=as.numeric(logLik(regSBCX1)[1])

SBC_crit1=(-2*loglik1)+boy1*log(n)

Buradaki boy1 değeri ise modele 1. değişken olarak giren enflasyonun BF sayısıdır. Sadece BF3 olarak modele dahil olduğu için boy1=1 olarak kodlanmıştır.

Toplamda 9 adet deęişken için aynı işlem tekrarlanıp ANOVA tablosu oluşturulmuştur.

Tablo 7.8: SBC kriteri kullanılarak elde edilen ANOVA ayrışımı

Fonksiyon	St. Sapma	SBC	#Temel Fonk.	#Serbestlik der	Deęişkenler
1	0.011607	-361.4	1	2.5	Enflasyon
2	0.009101	-382.4	3	7.5	Top. San. En
3	0.011466	-358.8	2	5.0	Döviz Kuru
4	0.011366	-363.9	1	2.5	İthalat
5	0.011275	-364.9	1	2.5	İhracat
6	0.009908	-380.4	1	2.5	İşsizlik*Faiz
7	0.007779	-405.3	2	5.0	İşsizlik*Top. San. En.
8	0.008928	-388.8	2	5.0	İşsizlik*Döviz
9	0.006905	-415.5	3	7.5	İşsizlik*İhracat

MARS regresyonun SBC kriterine baęlı olarak elde edilmiş anlamlı deęişkenlerinin ANOVA ayrışımı sonuçları elde edilmiştir. Standart sapma deęerlerine bakıldığında sırasıyla 1 ve 2 temel fonksiyonla modele dahil olmuş Enflasyon ve Döviz kuru deęişkenlerinin baęımlı deęişken üzerinde daha fazla etkilerinin olduęu görülmektedir. Enflasyon ve Döviz Kuru deęişkenlerinin SBC deęerlerinin dięer deęişken deęerlerine kıyasla daha büyük olması (Döviz Kuru SBC: -358.8; Enflasyon SBC: -361.4) bu deęişkenlerin modelden çıkarıldıklarında modeldeki kaybın da büyük olacaęı şeklinde yorumlanır.

7.4 MARS_ICOMP(IFIM)_PEU Model Seçimi

ICOMP(IFIM)_{PEU} kriterinin RStudio programına yazdırılması işlemi:

Programda kriterin hazır bulunmaması sebebiyle ilk olarak formül kodlanmıştır.

```
critICOMP <- function(loglik,k,n,Xm)
{
  rss=(exp(loglik*(-2/n))*n)
  sigma2hat=rss/n
  covmat11=sigma2hat*ginv(t(Xm)%*%Xm)
```

```

boy=ncol(Xm)

covmat22=2*(sigma2hat^2)/n

covmat12=matrix(0,boy,1)

covmat21=matrix(0,1,boy)

IFIM1=cbind(covmat11,covmat12)

IFIM2=cbind(covmat21,covmat22)

IFIM=rbind(IFIM1,IFIM2)

C1=(boy+1)*log((sum(diag(covmat11))+2*sigma2hat^2)/n)/(boy+1))-log(abs(det(covmat11)))-
log((2*sigma2hat^2)/n)

ICOMP_crit=(-2*loglik)+boy*(1+log(n))+2*C1

return(ICOMP_crit)

}

```

Hatırlanacağı üzere $ICOMP(IFIM)_{PEU} = -2\log L(\hat{Q}) + k(1 + \log(n)) + 2C_1(\hat{F}^{-1})$ $C_1(\hat{F}^{-1}) = (M + 1)\log \left[\frac{\text{tr} \hat{\sigma}^2 (\mathbf{B}' \mathbf{B})^{-1} + \frac{2\hat{\sigma}^4}{n}}{M+1} \right] - \log \hat{\sigma}^2 (\mathbf{B}' \mathbf{B})^{-1} - \log \frac{2\hat{\sigma}^4}{n}$,

$Cov(\hat{\beta}, \hat{\sigma}^2) = \hat{F}^{-1} = \begin{bmatrix} \hat{\sigma}^2 (\mathbf{B}' \mathbf{B})^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix}$ eşitliğinden oluşuyordu. Bakıldığında C_1

eşitliğindeki $\text{tr} \hat{\sigma}^2 (\mathbf{B}' \mathbf{B})^{-1} + \frac{2\hat{\sigma}^4}{n}$ ifadesinin covaryans matrisinin izine eşit olduğu görülecektir (covmat11). ‘boy’ ifadesi ise modeldeki BF’lerin sayısı olarak tanımlanmıştır. Önceki kriterlerde ifade edilen M değerine eşittir. C1 hesaplamasında yazdırılan ‘abs’ ise mutlak değer olup determinant içerisindeki ifadenin negatif çıkması durumunda logaritmanın hesaplanamaması sorununa çözüm için getirilmiştir. Son olarak da ileriye doğru seçim ve geriye doğru eleme işlemleri yaptırılmıştır.

```
fast_forward(datatabig, crit=aic,maxf = limit)%>%
```

```
multi_backward(crit=critICOMP)->mmICOMP
```

ICOMP(IFIM)_PEU kriteri ile elde edilen MARS modeli

$$Y = 0.0003294 - 0.0001451*BF4 - 0.0001128*BF5 + 0.0003227*BF6 - 0.0000537*BF18 + 0.0002261*BF21 - 0.0002504*BF1*BF4 +$$

$$0.0001778*BF1*BF5 + 0.0002267*BF1*BF7 - 0.000201*BF1*BF17 - 0.0000099*BF1*BF21$$

[1] "BF4=FAIZ_ORANLARI-8.5392"

[2] "BF5=TOPLAM SANAYI_URETIM_ENDEKSI-82.9933"

[3] "BF6=TOPLAM SANAYI_URETIM_ENDEKSI-53.3"

[4] "BF18=53.3-TOPLAM SANAYI_URETIM_ENDEKSI"

[5] "BF21=16.71-lnIHRACAT"

[6] "BF1*BF4=ISSIZLIK_ORANI-7.3*FAIZ_ORANLARI-8.5392"

[7] "BF1*BF5=ISSIZLIK_ORANI-7.3*TOPLAM SANAYI_URETIM_ENDEKSI-82.9933"

[8] "BF1*BF7=ISSIZLIK_ORANI-7.3*DOVIZ_KURU-1.2"

[9] "BF1*BF17=ISSIZLIK_ORANI-7.3*82.9933-TOPLAM SANAYI_URETIM_ENDEKSI"

[10] "BF1*BF21=ISSIZLIK_ORANI-7.3*16.71-lnIHRACAT"

Amacı bir modelin karmaşıklığı ile uyum iyiliği arasındaki en uygun dengeyi sağlayarak en doğru modeli en yalın şekilde ifade etmek olan ICOMP(IFIM)_PEU kriteri bu amacı doğrultusunda daha sıkı bir cezalandırma metodu uygulamaktadır. Diğer kriterlerdeki serbest parametre sayısını cezalandırmak yerine kovaryans karmaşıklığını cezalandırmayı tercih etmektedir. Dolayısıyla diğer bilgi kriterlerine kıyasla daha az BF içeren bir model oluşturması beklenmektedir. ICOMP(IFIM)_{PEU} ile elde edilen Faiz Oranı, Toplam Sanayi Üretim Endeksi ve İhracat değişkenleri ana etki olarak katkı sunmaktadırlar. Ayrıca modele dahil ettiği etkileşimli değişkenler incelendiğinde faiz oranlarıyla işsizlik arasında bir ilişki tespit edip bu ilişkiyi modele yeni bir değişken olarak dahil etmiştir. Diğer etkileşimli değişkenler arasında da benzer ilişkilerden söz edilebilir. ICOMP(IFIM)_PEU kriteri 10 BF içermesiyle ve içerdiği değişkenlerin modele katkısı olan değişkenlerden oluşması sebebiyle, ICOMP(IFIM)_PEU'nun model seçiminde en başarılı kriter olduğunu söyleyebiliriz.

ICOMP(IFIM)_PEU kriteri kullanılarak ANOVA tablosunun oluşturulması için yazdırılan kodlar:

regICOMPX1=lm(y~BF4) ile modele sadece BF4 değişkeniyle dahil olan Faiz Oranı için regresyon modeli oluşturulmuştur. Diğer değişkenler için de benzer şekilde regresyon modelleri elde edilmiştir.

regICOMPX2=lm(y~BF5+BF6+BF18)

```
regICOMPX3=lm(y~BF21)
```

```
regICOMPX4=lm(y~BF1*BF4)
```

```
regICOMPX5=lm(y~BF1*BF4+BF1*BF17)
```

```
regICOMPX6=lm(y~BF1*BF7)
```

```
regICOMPX7=lm(y~BF1*BF21)
```

ICOMPanova1=sqrt(sum(regICOMPX1\$residuals^2)/(n-1)) koduyla da standart sapma değeri hesaplatılmıştır. Bu işlem modeldeki tüm değişkenler için tekrarlanmıştır.

```
ICOMPanova2=sqrt(sum(regICOMPX2$residuals^2)/(n-1))
```

```
ICOMPanova3=sqrt(sum(regICOMPX3$residuals^2)/(n-1))
```

```
ICOMPanova4=sqrt(sum(regICOMPX4$residuals^2)/(n-1))
```

```
ICOMPanova5=sqrt(sum(regICOMPX5$residuals^2)/(n-1))
```

```
ICOMPanova6=sqrt(sum(regICOMPX6$residuals^2)/(n-1))
```

```
ICOMPanova7=sqrt(sum(regICOMPX7$residuals^2)/(n-1))
```

Modele BF4 olarak katılan Faiz Oranı değişkeni için aşağıdaki kod ile ICOMP(IFIM)_PEU değeri hesaplatılmıştır.

```
sigma2hat1=sum(regICOMPX1$residuals^2)/n
```

```
covmat11=sigma2hat1*ginv(t(BF4)%*BF4)
```

```
boy=ncol(as.matrix(BF4))
```

```
covmat22=2*(sigma2hat1^2)/n
```

```
loglik1=as.numeric(logLik(regICOMPX1)[1])
```

```
C1=(boy+1)*log((sum(diag(covmat11)))+(2*sigma2hat1^2)/n)/(boy+1))-
```

```
log(abs(det(covmat11)))-log((2*sigma2hat1^2)/n)
```

```
ICOMP_crit1=(-2*loglik1)+boy*(1+log(n))+2*C1
```

Üç farklı temel fonksiyon (BF5,BF6,BF18) ile modele dahil olan Toplam Sanayi Üretim Endeksi için ise ICOMP(IFIM)_PEU değeri hesabı aşağıdaki gibi yaptırılmıştır.

```
XM2=cbind(BF5,BF6,BF18)
```

```

sigma2hat2=sum(regICOMPX2$residuals^2)/n
covmat11=sigma2hat2*ginv(t(XM2)%*(XM2))
boy=ncol(as.matrix(XM2))
covmat22=2*(sigma2hat2^2)/n
loglik2=as.numeric(logLik(regICOMPX2)[1])
C1=(boy+1)*log((sum(diag(covmat11))+2*sigma2hat2^2/n)/(boy+1))-
log(abs(det(covmat11)))-log((2*sigma2hat2^2)/n)
ICOMP_crit2=(-2*loglik2)+boy*(1+log(n))+2*C1

```

Modele ikili etkileşim şeklinde dahil olan İşsizlik Oranı ve Faiz Oranı değişkenleri için ise ICOMP(IFIM)_PEU değeri aşağıdaki gibi hesaplatılmıştır.

```

sigma2hat4=sum(regICOMPX4$residuals^2)/n
covmat11=sigma2hat4*ginv(t(BF1*BF4)%*(BF1*BF4))
boy=ncol(as.matrix(BF1*BF4))
covmat22=2*(sigma2hat4^2)/n
loglik4=as.numeric(logLik(regICOMPX4)[1])
C1=(boy+1)*log((sum(diag(covmat11))+2*sigma2hat4^2/n)/(boy+1))-
log(abs(det(covmat11)))-log((2*sigma2hat4^2)/n)
ICOMP_crit4=(-2*loglik4)+boy*(1+log(n))+2*C1

```

Yukarıda modele dahil olma şekli farklılık göstermiş (tek değişken, etkileşimli değişken, vb.) değişkenlerin ICOMP(IFIM)_PEU değerlerinin hesaplanması bilgisi ışığında tüm fonksiyonlar için benzer işlemler gerçekleştirilip, ICOMP(IFIM)_PEU için ANOVA tablosu oluşturulmuştur.

Tablo 7.9: ICOMP(IFIM)_PEU kriteri kullanılarak elde edilen ANOVA ayrışımı

Fonksiyon	St. Sapma	ICOMP	#Temel Fonk.	#Serbestlik der	Değişkenler
1	0.011331	-358.1	1	2.5	Faiz Oranı
2	0.009101	-371.4	3	7.5	Top. San. En
3	0.011275	-361.0	1	2.5	İhracat
4	0.009908	-378.6	1	2.5	İşsizlik*Faiz
5	0.007626	-404.3	2	5.0	İşsizlik*Top. San. En.

6	0.009290	-381.3	1	2.5	İşsizlik*Döviz
7	0.010856	-368.4	1	2.5	İşsizlik*İhracat

MARS regresyonun ICOMP(IFIM)_PEU kriterine bağlı olarak elde edilmiş anlamlı değişkenlerinin ANOVA ayrışımı sonuçları elde edilmiştir. Standart sapma değerlerine bakıldığında birer temel fonksiyon ile modele dahil olan Faiz Oranı ve İhracat değişkenlerinin bağımlı değişken üzerinde daha fazla etkilerinin olduğunu görüyoruz. Faiz Oranı ve İhracat değişkenlerinin ICOMP değerlerinin diğer değişken değerlerine kıyasla daha büyük olması (Faiz Oranı ICOMP: -358.1; İhracat ICOMP: -361.0) bu değişkenlerin modelden çıkarıldıklarında meydana gelecek kaybın büyük olacağı anlamını taşımaktadır.

8. SONUÇ

Çok Değişkenli Uyarlanabilir Regresyon Eğrileri (MARS) tekniği bağımsız değişkenlerin ve bu değişkenlerin etkileşimlerinin bağımlı değişken ile olan ilişkisini ortaya koymaktadır. Olası ilişkiyi belirlemede ise düzleştirme uzanımlarını kullanmaktadır. MARS tekniği bu işlemleri iki adımda gerçekleştirmektedir. İleriye doğru seçim ve geriye doğru eleme adımlarından oluşan bu sürecin ilk aşamasında eldeki değişkenlerle ve/veya değişken etkileşimleriyle mümkün olan en karmaşık model oluşturulur. Eleme adımında ise oluşan bu modele gerekli budama işlemi yapılarak modelin karmaşıklığı azaltılır. Buradaki amaç bağımlı değişkeni açıklayan en doğru ve en sade modeli oluşturabilmektir. İleriye doğru seçim adımında model oluşturmada kullanılan farklı kriterler mevcuttur. Modelin uyum eksikliğinin değerlendirilmesi ve parametre sayısının cezalandırılması esasına dayalı bir kriter olan AIC, sahip olduğu uyum hassasiyeti ile özellikle fazla sayıda bağımsız değişken içeren modellerde geniş bir kullanıma sahiptir. Biz de çalışmamızın bu adımında AIC kriterini tercih ettik. Eleme adımında ise MARS yöntemi GCV kriterini kullanmaktadır. GCV kriterinin formülünde kullanılan düzleştirme parametresi için keyfi değer seçilmesi ayrıca yüksek boyutlu modeller elde edilmesi eleştirilen yönleriydi. Eleştirilerden yola çıkarak geriye doğru eleme aşamasında kullanılan GCV kriteri yerine AIC, SBC ve ICOMP(IFIM)_PEU gibi alternatif bilgi kriterleri kullanarak, hem MARS tekniği için keyfi seçim eleştirisini ortadan kaldırmayı en çok da doğru ve daha sade bir model elde etmeyi amaçladık.

Analizlerimiz sonucunda mevcut olan 300 BF arasından;

*GCV kriteri ile 25 temel fonksiyon (BF) içeren bir MARS modeli elde edildi.

*AIC kriteri kullanılarak elde edilen MARS modeli 19 BF,

*SBC kriteri ile elde edilen MARS modeli ise 16 BF içermektedir.

* ICOMP (IFIM)_PEU kriteri ile 10 BF içeren MARS modeli oluşturulmuştur.

Modellere seçilen değişkenler bakımından kriterlere bakıldığında,

-GCV kriteri ile İşsizlik Oranını, Enflasyon, Faiz Oranını, Toplam Sanayi Üretim Endeksi, Döviz Kuru, İthalat ve İhracat değişkenlerini ana etki olarak içeren bir final modeli elde edilmiştir.

-AIC kriteri ile elde edilen modele İşsizlik, Enflasyon, Toplam Sanayi Üretim Endeksi, Döviz Kuru, İthalat ve İhracat ana etki olarak dahil olmuştur.

-SBC kriteri ile elde edilen modele Enflasyon, Toplam Sanayi Üretim Endeksi, Döviz Kuru, İthalat ve İhracat ana etki olarak katkı sunan değişkenlerdir.

-ICOMP(IFIM)_PEU kriteri ile elde edilen MARS modeli ana etki olarak Faiz Oranı, Toplam Sanayi Üretim Endeksi ve İhracat değişkenlerini seçmiştir. Ayrıca İşsizlik ile Faiz Oranı; İşsizlik ile Toplam Sanayi Üretim Endeksi; İşsizlik ile Döviz Kuru ve İşsizlik ile İhracat arasındaki etkileşimlerin de ICOMP(IFIM)_PEU kriteri ile elde edilen MARS modeliyle Takibe Dönüş Oranını etkilediği görülmüştür.

Sonuç olarak, hem yapılan simülasyon çalışması hem de gerçek veri seti üzerinden elde edilen çıktılar göstermektedir ki: Değişkenler arasındaki olası ilişkiyi belirlemede kullandığı, düzleştirme uzanımları tekniği sayesinde, hem yenilik hem de büyük bir kolaylık sağlayan MARS yöntemi algoritmasında; ICOMP(IFIM)_PEU, daha az sayıdaki BF ile doğru modelin seçiminde güçlü bir kriter olarak kullanılabilir.

Mevcut çalışma kod yazılarak gerçekleştirilmiştir. Analizin daha pratik gerçekleştirilebilmesi adına ileriki çalışmalarda kodlar bir program haline getirilerek gerekli regresyon denklemi sadece veri girişi sağlanarak elde edilebilir. Bu şekilde alternatif kriterler ile MARS yönteminin kullanımı daha yaygın ve kolay bir nitelik kazanacaktır.

9. KAYNAKLAR

- Al-Musaylh, S. M., Deo, R. C., Adamowski, J. F., Li, Y. (2018). Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advanced Engineering Informatics*, 35, 1-16.
- Andres, J. D., Lorca, P., Juez, F. J. C., Lasheras, F. S. (2011). Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications*, 38:3, 1866-1875.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory, 267-281, Budapest: Akademiai Kiado.
- Altınöz, U. (2018). Sorunlu krediler bağlamında Türk bankacılığında kredi kayıp karşılığının makroekonomik değişkenlere etkisi: Panel Data ve Zaman Serileri Analizi, Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 63-82, doi:http://dx.doi.org/10.17218/hititsosbil.359834
- Avcı, T. (2020). The effects of macroeconomic factors and bank loans on inflation: the case of Turkey. *Research of Financial Economic and Social Studies (RFES)*, ISSN:2602-2486. Doi: 10.29106/fesa.690721
- Barron, A.R, Xiao, X. (1991). Discussion.
- Bhattacharai, S. (2016). Determinants of Non-Performing Loans: Perception of Nepali Bankers. *Economic Journal of Development Issues*, 17(1-2), 128-148.
- Bayrakdaroğlu, A., Aydın, G. C. (2017). A research on the effects of interest rates volatility on credit volume. *Journal of Economics and Management Research* Vol:6/No:2
- Beak, S. H., Park, D., Bozdoğan, H. (2016). Hybrid kernel density estimation for discriminant analysis with information complexity and genetic algorithm. *Knowledge Based Systems*, 99, 79-91.
- Bearse, P., Bozdoğan, H. (2002). Multivariate regressions, genetic algorithms, and information complexity: A three way hybrid. *Measurement and Multivariate Analysis, Springer Japan*, 269, New York.

- Bozağaç, D., Batmaz, İ., Oğuztüzün, H. (2016). Dynamic simulation metamodeling using MARS: A case of radar simulation. *Mathematics and Computers in Simulation*, 124, 69-86.
- Bozdoğan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, 52:3, 345-370.
- Bozdoğan, H. (1988). ICOMP: A new model selection criterion. *Classification and Related Methods of Data Analysis*: 599-608.
- Bozdoğan, H. (1990). On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics, Theory and Methods*, 19, 221-278.
- Bozdoğan, H., Haughton, D. M. A. (1998). Informational complexity criteria for regression models. *Computational Statistics & Data Analysis*, 28, 51-76.
- Bozdoğan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, 28, 51-76.
- Bozdoğan, H., Bearse, P. (2003). Information complexity criteria for detecting influential observations in dynamic multivariate linear models using the genetic algorithm. *Journal of Statistical Planning and Inference*, 114, 31-44.
- Bozdoğan, H. (2004). *Intelligent statistical data mining with information complexity and genetic algorithm. In Statistical Data Mining and Knowledge Discovery*. H. Bozdogan (ed).15-56, Chapman and Hall/CRC. Florida.
- Castro, V. (2013). Macroeconomic determinants of the credit risk in the banking system: The Case of the GIPSI. *Economic Modelling*, 31, 672-683.
- Capron, X., Walczak, B., Noord, O.E., Massart, D.L. (2005). A modification of the ICOMP criterion for estimation of optimum complexity of PCR models. *Journal of Chemometrics*, 19, 308-316.
- Chaloner, K., Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science*, 10:3, 273-304.

- Charron, N., Annoni, P. (2021). What is the Influence of news media on people's perception of corruption? Parametric and Non Parametric approaches. Springer Link, *Social Indicators Research* 153, 1139-1165.
- Chaibi, H., and Ftiti, Z. (2015). Credit risk determinants: Evidence from a cross-country study. *research in international business and finance*, 33, 1-16.
- Chen, M., Tompson, A. F. B., Mellors, R. J., Ramirez, A. L., ..., Wagoner, J. L. (2014). An efficient Bayesian inversion of a geothermal prospect using a multivariate adaptive regression spline method, *Applied Energy*, 136, 619-627.
- Chang, L., Chu, H., Lin, D., Lui, P. (2012). Analysis of freeway accident frequency using multivariate adaptive regression splines. *Procedia Engineering*, 45, 824-829.
- Chou, S. M., Lee, T. S., Shao, Y. E., Chen, I.F. (2004). Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 27, 133-142.
- Clark, A.E., Troskie, C.G. (2006). Regression and ICOMP, A simulation study. *Communications in Statistics – Simulation and Computation*, 35, 591-603.
- Craven, P., Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik, Verlag*, 31, 377-403.
- Curak, M., Pepur, S., and Poposki, K. (2013). Determinants of non-performing loans—evidence from southeastern banking systems. *Banks & Bank System*, 8(1), 45-54.
- Demirci, N. S. (2016). Relationship between production and bank credits in manufacturing industry sector: cointegration and causality analysis for Turkey. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 19(1). Doi: 10.16953/deusbed.38804
- Demirezen, Ö. (2015). Türkiye’de kredilerin özel tüketim harcamalarına etkisi. Uzmanlık tezi. T.C. Kalkınma Bakanlığı. Yayın No:2941. ISBN 978-605-9041-58-04.

- Deniz, A. E. (2007). Yapısal eşitlik modellerinde bilgi kriterleri. Doktora tezi. Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, 134, İstanbul.
- Deniz, E., Bozdoğan, H., Katragadda, S. (2011). Structural equation modeling (SEM) of categorical and mixed-data using the novel Gifi transformations and information complexity (ICOMP) criterion, *Istanbul University Journal of the School of Business Administration*, 40:1, 86-123.
- Dursun, G. D., ve Bozkır, B. (2018). Türkiye’de faaliyet gösteren ticari bankaların aktif kalitesinin topsis yöntemi ile ölçümü. *Journal of Research in Economics, Politics and Finance*, 3(3):243-25. Doi:10.30784/epfad.422500
- Fernandez, J. R. A., Nieto, P. J. G., Muniz, C. D., Anton, J. C. A. (2014). Modeling eutrophication and risk prevention in a reservoir in the Northwest of Spain by using multivariate adaptive regression splines analysis. *Ecological Engineering*, 68, 80-89.
- Finansal Göz (2018). <https://www.finansalgoz.com/2018/12/takibe-donusum-oran.html> (Erişim tarihi: 17.10.2020)
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19:1, 1-141.
- Ghasemzadeh, A., Ahmed, M. M. (2018). Utilizing naturalistic driving data for in-depth analysis of driver lanekeeping behavior in rain: Non-parametric MARS and parametric logistic regression modeling approaches. *Transportation Research Part C: Emerging Technologies*, 90, 379-392.
- Güney, Y. (2018). Simetrik ve çarpık dağılımlar için bilgi karmaşıklık (ICOMP) kriteri kullanılarak robust model seçimi. Doktora tezi. Ankara Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, 149, Ankara.
- Harris, C., J. (1978). *An information theoretic approach to estimation*. M., J. Gregson (ed.) In *Recent Theoretical Developments in Control*. Academic Press. 563-590, London.
- Heddam, S., Kisi, O. (2018). Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology*, 559, 499-509.
- Investaz, (2018). Sanayi üretim endeksi nedir, nasıl hesaplanır? <https://www.investaz.com.tr/blog/sanayi-uretim-endeksi-nedir-nasil-hesaplanir/> (Erişim tarihi: 17.10.2020)

- Karaman, E. (2019). Optimal ölçkleme teknikleri ve bir uygulama. Doktora tezi. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, Sayısal Yöntemler Bilim Dalı, 108, İstanbul.
- Kavcıoğlu, Ş. (2012). Kriz dönemleri sonrasında Türk bankacılık sektöründe kredi analizi. *Finansal Araştırmalar ve Çalışmalar Dergisi*. 3:6. 23-39.
- Kayhan, B. (2010). Parameter estimation in generalized partial linear models with tikhonov regularization. Master Thesis, Middle East Technical University, Department of Scientific Computing, 101, Ankara.
- Keshtegar, B., Mert, C., Kisi, O. (2018). Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree. *Renewable and Sustainable Energy Reviews*, 81:1, 330-341.
- Kuter, S., Akyurek, Z., Weber, G. W. (2018). Retrieval of fractional snow covered area from MODIS data by multivariate adaptive regression splines. *Remote Sensing of Environment*, 205, 236-252.
- Koç, E. K., Bozdoğan, H. (2014). Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function. *Mach Learn*, 101:1-3, 35-58.
- Kolyshkina, I., Wong, S., Lim, S. (2004). Enhancing generalized linear models with data mining, www.casact.org/pubs/dpp/dpp04/04dpp279.pdf.
- Konstantakis, K. N., Michaelides, P. G., & Vouldis, A. T. (2016). Non performing loans (NPLs) in a crisis economy: Long-run equilibrium analysis with a real time VEC model for Greece (2001– 2015). *Physica A: Statistical Mechanics and Its Applications*, 451, 149-161.
- Kullback, S., Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79-86.
- Kullback, S. (1968). *Information theory and statistics*. Dover Publications, 432, New York.
- Linhart, H., Zucchini, W. (1986). Finite sample selection criteria for multinomial models. *Statistische Hefte*. 27, 173-178.
- Lokuge, W., Wilson, A., Gunasekara, C., Law, D. W., Setunge, S. (2018). Design of fly ash geopolymer concrete mix proportions using Multivariate Adaptive Regression Spline model. *Construction and Building Materials*, 166, 472-481.

- Macit, F. ve Keçeli, B. (2012). Takipteki kredi oranını etkileyen faktörler: Türkiye'de katılım bankaları örneği, *Avrasya İncelemeleri Dergisi*, 1(1), 193-207.
- Magnus, J. R., Neudecker, H. (1980). The elimination matrix: some lemmas and applications. *Society for Industrial and Applied Mathematics*, 1:4, 422-449.
- Magnus, J. R., Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. (Third edition). John Wiley and Sons, 56-60, New York.
- Mehdizadeh, S., Behmanesh, J., Khalili, K. (2017). Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and Electronics in Agriculture*, 139, 103-114.
- Migueis, V. L., Camanho, A., Cunha, J. F. (2013). Customer attrition in retailing: An application of Multivariate Adaptive Regression Splines. *Expert Systems with Applications*, 40:16, 6225-6232.
- Mileris, R. (2012). Macroeconomic determinants of loan portfolio credit risk in banks. *Engineering Economics*, 23(5), 496-504.
- Nacar, S. (2020). İklim değişikliğinin doğu karadeniz havzası sıcaklık ve yağış parametreleri üzerindeki olası etkilerinin incelenmesi. Doktora tezi. Karadeniz Teknik Üniversitesi Fen Bilimler Enstitüsü, İnşaat Mühendisliği Anabilim Dalı, 277, Trabzon.
- Neter, J., Wasserman, W., Kutner, M. H. (1989). *Applied linear regression models*. Irwin Inc. 214-227, Boston.
- Oğuz, A. (2014). Çok değişkenli uyarlanabilir regresyon zincirlerinin irdelenmesi ve bir uygulama. Yüksek Lisans Tezi. Erzincan Üniversitesi, Fen Bilimleri, Enstitüsü Matematik Anabilim Dalı, 74, Erzincan.
- Özmen, A., Yılmaz, Y., Weber, W. W. (2018). Natural gas consumption forecast with MARS and CMARS models for residential users. *Energy Economics*, 70, 357-381.
- Pamukçu, E., Bozdoğan, H., Çalık, S. (2015). A novel hybrid dimension reduction technique for undersized high dimensional gene expression data sets using information complexity criterion for cancer classification. *Computational and Mathematical Methods in Medicine*, 2015.
- Pamukçu, E. (2017). Aşırı derecede küçük örneklem problemi için hibrit regresyon modeli. *Celal Bayar University Journal of Science*, 13:3, 803-813.

- Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57, 120-125.
- Rissanen, J. (1978). Model selection by shortest data description. *Automatica*, 14, 465-471.
- Roy, S. S., Roy, R., Balas, V. E. (2018). Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. *Renewable and Sustainable Energy Reviews*, 82 :3, 4256-4268.
- Schwartz, G. (1978). Estimating the dimension of model. *Annals of Statistics*, 6, 461-464.
- Shahraiyini, H. T., Shahsavani, D., Sargazi, S., Nokhandan, M. H. (2015). Evaluation of MARS for the spatial distribution modeling of carbon monoxide in an urban area. *Atmospheric Pollution Research*, 6:4, 581-588.
- Shannon, C. E. (1948). A mathematical theory of communication, *The Bell System Technical Journal*, 27:3, 623-656.
- Stein, C. (1975). Estimation of covariance matrix, *Rietz Lecture, 39th Annual Meeting IMS*. Atlanta, Georgia.
- Stripling, H. F., McClarren, R. G., Kuranz, C. C., Grosskopf, M. J., ..., Torralva, B. R. (2013). A calibration and data assimilation method using the Bayesian MARS emulator, *Annals of Nuclear Energy*, 52, 103-112.
- Shahsavani, D., Tarantola, S., Ratto, M. (2010). Evaluation of MARS modeling technique for sensitivity analysis of model output. *Procedia – Social and Behavioral Sciences*, 2:6, 7737-7738.
- Sakamoto, W. 2007. MARS: selecting basis functions and knots with an empirical Bayes method. *Computational Statistics*, 22, 583-597.
- Sevimli, Y. (2009). Çok deęişkenli uyarlanabilir regresyon uzanımlarının bir split-mouth çalışmasında uygulaması. Yüksek Lisans Tezi. Marmara Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, 87, İstanbul.
- Şahbaz, N. ve İnkaya, A. (2010). Türk bankacılık sektöründe sorunlu krediler ve makro ekonomik etkileri. *Optimum Ekonomi ve Yönetim Bilimleri Dergisi*, 1, 69-82.

- Tanınmış Yücememiş, B. ve Sözer, İ.A. (2011). Bankalarda takipteki krediler: Türk bankacılık sektöründe takipteki kredilerin tahminine yönelik bir model uygulaması. *Finansal Araştırmalar ve Çalışmalar Dergisi*, 3(5), 43-56.
- TCMB (2020). Türkiye Cumhuriyeti Merkez Bankası Terimler Sözlüğü. <https://www.tcmb.gov.tr/wps/wcm/connect/TR/TCMB+TR/Main+Menu/Banka+Hakkinda/Egitim-Akademik/Terimler+Sozlugu/> (Erişim tarihi: 17.10.2020)
- Theil, H., Fiebig, D. G. (1984). On the maximum entropy approach to undersized samples. *Applied Mathematics and Computation*. 14, 301-312.
- Thomaz, C. E. (2004). Maximum entropy covariance estimate for statistical pattern recognition. A Thesis submitted for the degree of Doctor of Philosophy of the University of London and for the Diploma of the Imperial College. 143, London.
- Tiku, M. L., Suresh, R. P. (1992). A new method of estimation for location and scale parameters. *Journal of Statistical Planning and Inference*. 30:2, 281-292.
- Ünal, B. (2009). Çok değişkenli uyarlamalı regresyon uzanımları. Yüksek Lisans Tezi. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, 131, Ankara.
- Aydoğan Ünal, B. (2020). Türkiye'deki seçmen tercihlerinin coğrafi kümeleme analizi: 24 Haziran 2018 seçimleri. Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 39, 21-32. Doi.org/10.30794/pausbed.688598.
- Yağcılar, G. G. ve Demir, S. (2015). Türk bankacılık sektöründe takipteki kredi oranları üzerinde etkili olan faktörlerin belirlenmesi. *Journal of Alanya Faculty of Business/Alanya İşletme Fakültesi Dergisi*, 7(1), 221-229.
- Yurdakul, E., M. (2014). Türkiye'de ithalatın gelişimi ve ithalatın yapay sınır ağırları yöntemi ile tahmin edilebilirliğine yönelik bir analiz. Doktora Tezi. Adnan Menderes Üniversitesi, Sosyal Bilimler Enstitüsü, İktisat Anabilim Dalı, 244, Aydın.
- Yüksel, S. (2016). Bankaların takipteki krediler oranını belirleyen faktörler: Türkiye için bir model önerisi. *Bankacılar Dergisi*, 98, 41-56.
- Van Emden, M. H. (1971). *An analysis of complexity*. Amsterdam: Mathematisch Centrum, 91, Amsterdam.

- Van Der Horn, E., Mahadevan, S. (2018). Bayesian model updating with summarized statistical and reliability data. *Reliability Engineering & System Safety*, 172, 12-24.
- Williams, J. L., Bozdoğan, H. (1995). Inference problems with equivalent models. *Lawrence Erlboun Associates*. New Jersey.
- Xu, Q. S., Daszykowski, M., Walczak, B., Daeyaert, F., DeJonge MR., Heeres, J., Koymans, LMH., Lewi, PJ., Vinkers, HM., Janssen, PA., Massart, DL. (2004). Multivariate adaptive regression splines, Studies of HIV Reverse Transcriptase Inhibitors. *Chemometrics and Intelligent Laboratory Systems*, 72, 27-34.

EK 1. Kredi Veri Seti

TAKİBE DÖNÜS ORANI	İSSİZLİK ORANI	ENFLASYON	FAİZ ORANLARI	TOPLAM SANAYİ ÜRETİM ENDEKSİ	DÖVİZ KURU	GSYİH	İHRACAT	İTHALAT
5,82%	13,70	11,84	14,54	124,19	5,80	1.189.855.413,16	17.977.779,21	27.822.542,39
5,38%	13,80	9,26	20,64	113,02	5,68	1.145.099.354,34	17.927.740,87	28.427.445,19
4,71%	13,00	15,72	26,40	110,50	5,88	1.023.855.087,17	18.214.793,00	29.165.359,42
4,34%	14,10	19,71	24,18	105,39	5,37	922.029.155,99	19.019.136,60	30.200.269,07
4,15%	13,50	20,30	30,39	117,75	5,53	1.017.190.008,72	19.609.666,43	32.270.116,14
3,43%	11,40	24,52	28,66	112,17	5,59	1.026.648.922,65	21.013.530,91	35.640.203,97
3,20%	10,16	15,39	19,50	114,39	4,37	890.435.944,67	21.681.658,33	35.360.565,85
3,05%	10,12	10,23	17,49	111,72	3,82	790.113.059,53	23.189.347,54	36.353.567,28
3,10%	10,40	11,92	16,68	126,84	3,80	892.228.264,53	24.113.841,31	38.097.509,82
3,20%	10,60	11,20	16,58	112,12	3,52	833.706.740,87	25.975.232,65	40.647.773,12
3,25%	10,20	10,90	15,85	109,48	3,59	735.280.547,65	27.285.761,92	43.088.794,21
3,39%	11,70	11,29	14,71	102,05	3,70	649.434.601,86	30.038.030,95	47.418.207,72
3,41%	12,70	8,53	14,35	114,77	3,28	747.226.024,63	33.639.635,66	52.247.149,42
3,48%	11,30	7,28	14,78	95,29	2,97	666.176.429,47	34.994.753,71	55.387.122,89
3,46%	10,20	7,64	15,62	105,72	2,90	631.232.693,23	37.037.950,51	55.137.537,17
3,45%	10,10	7,46	16,00	97,85	2,95	563.890.602,00	26.285.719,14	39.078.009,06
3,25%	10,80	8,81	15,67	111,04	2,91	646.500.325,08	25.121.683,98	30.833.904,62
3,07%	10,30	7,95	14,59	97,90	2,85	631.512.354,71	22.984.527,13	33.015.698,07
3,01%	9,60	7,20	13,35	100,87	2,67	562.947.770,73	26.132.662,58	37.259.887,91
2,96%	10,60	7,61	12,77	90,19	2,46	497.687.043,16	27.512.471,84	38.918.668,13
2,97%	10,90	8,17	12,79	100,32	2,27	557.419.788,61	26.691.063,68	40.899.122,49
3,05%	10,50	8,86	12,19	94,09	2,17	548.625.834,71	28.444.820,35	43.726.839,94
2,86%	9,10	9,16	14,23	94,29	2,12	487.151.068,27	28.023.658,61	46.785.777,85
2,94%	9,70	8,39	14,79	88,26	2,22	451.269.184,24	30.770.060,02	54.261.741,39
2,87%	9,60	7,40	11,02	96,35	2,03	491.085.106,51	32.280.376,79	58.607.863,90
2,87%	9,20	7,88	10,97	89,20	1,97	491.263.793,92	33.694.262,21	61.259.818,69
2,89%	8,10	8,30	9,78	90,06	1,84	441.539.542,87	35.079.775,10	61.926.342,92
3,14%	9,40	7,29	11,58	80,39	1,79	385.824.643,40	33.797.330,25	57.785.923,55
3,04%	9,30	6,16	12,23	89,30	1,79	429.732.717,44	36.059.975,06	57.976.902,96
3,14%	8,30	9,19	14,79	82,99	1,80	424.705.390,32	38.537.663,95	59.772.547,36
2,83%	7,30	8,87	14,19	84,42	1,81	382.070.001,70	39.782.515,20	60.029.446,36
2,94%	9,10	10,43	14,74	77,38	1,80	333.164.005,46	37.950.624,72	57.542.787,84
2,92%	9,00	10,45	13,39	87,36	1,84	385.734.139,69	38.114.830,20	62.033.201,03
2,95%	8,20	6,15	11,48	80,56	1,73	381.898.595,37	37.678.061,58	65.085.408,69
3,20%	8,70	6,24	9,64	79,52	1,57	336.234.139,95	38.180.766,47	61.773.059,36
3,59%	10,10	3,99	8,54	72,76	1,58	290.610.290,52	38.252.883,95	63.568.884,77
4,08%	10,60	6,40	8,61	77,48	1,46	322.360.447,43	41.340.967,32	60.618.424,37
4,83%	10,60	9,24	8,91	70,14	1,52	318.732.806,05	39.251.904,34	61.082.515,71
5,10%	9,90	8,37	9,15	69,11	1,54	278.647.853,11	39.182.873,71	60.008.683,51
5,81%	12,80	9,56	8,96	60,81	1,51	240.272.871,66	37.493.874,08	60.807.202,91
6,40%	12,60	6,53	9,98	67,73	1,49	274.874.781,66	38.217.761,37	55.117.476,90
6,56%	12,50	5,27	12,19	63,04	1,50	271.840.902,50	35.685.145,70	53.049.625,45
6,04%	12,20	5,73	14,32	61,33	1,57	241.220.604,55	35.577.523,29	50.451.581,42
5,67%	14,70	7,89	19,87	53,29	1,66	211.255.559,34	34.150.724,66	48.381.281,87
4,59%	12,70	10,06	21,67	62,55	1,54	258.624.782,99	35.200.766,54	48.450.877,07
3,74%	9,80	11,13	18,50	68,79	1,21	268.726.652,43	36.128.443,05	50.211.892,34
3,69%	8,70	10,61	17,65	72,38	1,26	249.483.189,18	35.048.448,89	49.867.810,66
3,69%	10,10	9,15	17,04	68,59	1,20	217.948.233,82	36.032.107,84	50.024.107,07
4,16%	9,90	8,39	17,64	71,22	1,19	240.942.451,97	38.481.147,60	51.649.351,64
4,27%	8,90	7,12	18,45	69,50	1,29	237.359.324,52	39.621.932,17	57.234.284,58
4,36%	8,30	8,60	19,19	69,46	1,34	212.565.694,63	39.078.155,26	61.458.594,06
4,41%	9,50	10,86	19,43	63,66	1,41	189.593.408,03	39.692.777,55	61.841.625,58
4,39%	9,50	9,65	20,22	67,49	1,46	217.970.763,48	41.534.955,88	63.833.440,74
4,44%	8,40	10,55	19,85	65,52	1,50	217.039.426,46	41.048.283,36	60.055.096,54
4,49%	8,10	10,12	16,84	65,51	1,46	193.081.372,66	43.007.491,14	52.506.712,88
5,33%	10,00	8,16	18,22	57,41	1,33	161.135.992,52	42.151.159,29	46.705.969,01
5,61%	10,40	7,72	18,54	64,89	1,36	185.766.662,29	43.030.367,14	50.924.472,61
6,21%	9,20	7,99	19,09	61,37	1,34	185.021.847,57	41.718.909,48	49.569.257,75
6,39%	8,70	8,95	20,74	58,90	1,36	161.871.869,54	43.424.691,56	50.534.605,13
6,85%	10,00	7,94	23,93	53,30	1,33	141.042.563,35	42.938.342,66	52.105.678,02

ÖZ GEÇMİŞ

Meryem BEKAR ADIGÜZEL, 17.08.1988 tarihinde Almanya/Duisburg'da doğdu. Kütahya Ali Gural Lisesi'ni bitirdikten sonra Selçuk Üniversitesi Fen Fakültesi İstatistik Bölümü'nden 2011 yılında mezun oldu. Aynı dönemde Selçuk Üniversitesi İİBF Fakültesi İktisat Bölümü'nden de yandal eğitimi aldı. 2015 yılında Anadolu Üniversitesi AÖF İktisat Bölümü lisans eğitimini tamamladı. 2014 yılında Gazi Üniversitesi Fen Bilimleri Enstitüsü İstatistik ABD Yüksek Lisans programını bitirdi. 2016 yılından bu yana Aksaray Üniversitesi Ortaköy Meslek Yüksekokulu'nda Öğretim Görevlisi olarak görev yapmaktadır. (2021)

İletişim Bilgileri

E mail: bekarmeryem@gmail.com

Telefon: 507 133 06 98