

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÇOKLU BAĞINTI PROBLEMİNDE LİU VE RİDGE REGRESYON
KESTİRİCİLERİNİN KARŞILAŞTIRMASI

Semih GÖKÇE

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Doç. Dr. Atıf EVREN

Haziran, 2022

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**ÇOKLU BAĞINTI PROBLEMİNDE LİU VE RİDGE REGRESYON
KESTİRİCİLERİNİN KARŞILAŞTIRMASI**

Semih GÖKÇE tarafından hazırlanan tez çalışması 09.06.2022 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Atıf EVREN
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Atıf EVREN, Danışman
Yıldız Teknik Üniversitesi

Dr.Öğr. Üyesi Erhan USTAOĞLU, Üye
Marmara Üniversitesi

Dr.Öğr. Üyesi Doğan YILDIZ, Üye
Yıldız Teknik Üniversitesi

Danışmanım Doç. Dr. Atıf EVREN sorumluluğunda tarafımca hazırlanan Çoklu Bağıntı Probleminde Liu ve Ridge Kestiricilerinin Karşılaştırılması başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Semih GÖKÇE

İmza

TEŞEKKÜR

Bana maddi ve manevi imkan sağlayan anneme teşekkür ederim. Bana maddi imkan sağlayan ablama teşekkür ederim. Doç.Dr.Atıf EVREN'e değerli vaktini bana ayırdığı, bana yol gösterdiği ve bu çalışmayı bana tamamlama imkanı verdiği için teşekkür ederim.

Semih GÖKÇE



İÇİNDEKİLER

SİMGE LİSTESİ	vi
KISALTMA LİSTESİ	vii
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ	ix
ÖZET	x
ABSTRACT	xii
1 GİRİŞ	1
1.1 Literatür Özeti	1
1.2 Tezin Amacı	5
2 REGRESYON VE MODEL KURMA	6
2.1 Regresyon ve Model Kurma	6
2.1.1 Çoklu Doğrusal Regresyon	6
2.1.2 Zaman Serisi Verileri İçin Regresyon Modellerine Giriş	9
2.2 Çoklu Bağlantı Problemi	11
2.3 Ridge Regresyon	16
2.4 Liu Regresyon	23
3 UYGULAMA	27
3.1 Model ve Veri Seti Liu Regresyon	27
3.2 Ampirik Bulgular	27
4 SONUÇ VE ÖNERİLER	33
KAYNAKÇA	34
TEZDEN ÜRETİLMİŞ YAYINLAR	35

SİMGELİSTESİ

E	Beklenen Değer
I	Birim Matris
Cov	Kovaryans
d	Liu Regresyon Yanlılık Parametresi
$\hat{\beta}_d$	Liu Tahmini
$\hat{\beta}$	Regresyon Katsayısı
$\hat{\beta}_r$	Ridge Regresyon Katsayı Tahmini
k	Ridge Regresyon Yanlılık Parametresi
Var	Varyans
Yan	Yanlılık



KISALTMA LİSTESİ

DESTE	Doğrusal En iyi Sapmasız Tahmin Edici
EKK	En Küçük Kareler
HKO	Hata Kareler Ortalaması



ŞEKİL LİSTESİ

Şekil 2.1 k Sabiti.....	17
Şekil 2.2 Ridge İzi.....	19
Şekil 2.3 Ortak k ve d	26



TABLO LİSTESİ

Tablo 3.1	Normal Dağılıma Uyuma.....	27
Tablo 3.2	Model Anlamlılığı ve Vif.....	28
Tablo 3.3	R Ridge k Çıktısı.....	28
Tablo 3.4	R Liu d Çıktısı.....	29
Tablo 3.5	R Ridge k Değerlerine göre Çıktı.....	29
Tablo 3.6	R Liu d Değerlerine göre Çıktı.....	29
Tablo 3.7	R Ridge k Değerlerine göre R^2 Çıktısı.....	30
Tablo 3.8	R Liu d Değerlerine göre R^2 Çıktısı.....	30
Tablo 3.9	R Ridge k Değerlerine göre Vif Çıktı.....	30
Tablo 3.10	R EKK AIC, BIC Çıktı.....	31
Tablo 3.11	EViews AIC, Durbin-Watson Çıktı.....	31
Tablo 3.12	R Ridge AIC, BIC Çıktı.....	32
Tablo 3.13	R Liu AIC, BIC Çıktı.....	32
Tablo 4.1	R Liu Katsayı Çıktı.....	33

Çoklu Bağntı Probleminde Liu ve Ridge Regresyon Kestiricilerinin Karşılaştırması

Semih GÖKÇE

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Doç.Dr.Atıf Evren

Regresyon analizi değişkenler arasındaki ilişkiyi modellemek ve incelemek için kullanılan istatistiksel bir yöntemdir. Birçok uygulama alanları vardır ve bu uygulamalar mühendislik, fizik, sosyal bilimler gibi hemen hemen tüm alanlarda kullanılmaktadır. En küçük kareler metodu, günümüzde hemen hemen bütün bilim dallarında çokça kullanılır. En küçük kareler metodu, birbirine bağımlı olarak değişen iki büyüklük arasındaki matematiksel bağıntıyı olabildiğince gerçeğe uygun bir denklem olarak yazmak için kullanılan regresyon metodudur. Çoklu regresyon analizinde ortaya çıkan sorunlardan birisi de çoklu bağıntı sorunudur. Çoklu doğrusallık, açıklayıcı değişkenler arasındaki güçlü bir ilişkinin varlığı olarak tanımlanabilir. Çoklu doğrusallık parametre tahminlerinin varyansını büyütür. Ayrıca kestirilen regresyon katsayıları işaretlerinde buna ilave olarak büyüklükleri dahi yanlış olabilir. Dolayısı ile açıklanan, açıklayıcı değişkenler arasındaki ilişkiyi yanlış tarif edilebilir. Bu çalışmada öncelikli çoklu doğrusallık teşhis yöntemleri anlatılıp, Türkiye'nin ihracat modelini Ridge, Liu kestiricisiyle ve zaman serisi modelleri bir uygulaması yapılmıştır. Açıklanan değişken ihracattır. Açıklayıcı değişkenler TÜFE, Yİ-ÜFE, TCMB pasif yükümlülükleridir. Veriler 2003-2020 yılları arası çeyreklik verilerini

kapsamaktadır. Veriler Zaman serisine uygun olduđu için otokorelasyon probleminden şüphelenmiştir. Durbin-Watson testine göre otokorelasyon sorunu gözlemlenmemiştir. AIC değerlerine göre kıyaslama yapıldığında Liu kestiricisi ile elde edilen modelin diğerlerine göre en iyi sonuçları verdiği gözlenmiştir.

Anahtar Kelimeler: Çoklu Doğrusallık, Ridge, Liu Kestiricisi, Zaman Serisi, İhracat



Comparison of Liu and Ridge Regression in Multiple Connection Problem

Semih GÖKÇE

Department of Statistic

Master of Science Thesis

Supervisor: Assoc.Prof.Dr.Atif Evren

Regression analysis is a statistical method used to model and examine the relationship between variables. There are many application areas and these applications are used in almost all fields such as engineering, physics, social sciences. The least squares method is widely used in almost all branches of science today. The least squares method is a standard regression method used to write down the mathematical relationship between two dependently varying quantities as an equation that is as realistic as possible. One of the problems encountered in the multi-regression analysis is multicollinearity case. Multicollinearity in regression models is a result of strong correlations among independent variables. The existence of multicollinearity inflates the variances of the parameter estimates. Multicollinearity may also result in wrong signs and magnitudes of regression coefficient estimates, and consequently in incorrect conclusions about relationships between independent and dependent variables. In this research, firstly, theoretical structure of collinearity diagnostics and Liu estimator is introduced. At the end an application of Ridge, Liu, time series estimator for the Turkey's export model is done. The variable explained is exports. Explanatory variables are TUFE, Yİ-UFE and TCMB's liabilities. The data

includes quarterly data for the years 2003-2020. Suspicious of autocorrelation problem as the data is suitable for Time series According to the Durbin-Watson test, no autocorrelation problem was observed. When compared according to the AIC values, it was observed that the model obtained with the Liu estimator gave the best results compared to the others.

Keywords: Multicollinearity, Ridge Estimator, Liu Estimator, Time Series, Export



1.1 Literatür Özeti

İhracat ya da dışsatım olarak da bilinen kavram bir malın veya hizmetin yabancı ülkelere yabancı para karşılığında satılması işlemine verilen kavramdır. Ülkeler veya şirketler tarafından gerçekleştirilebilir. İthalat ve ihracat bir ülkenin dış ticaret dengesini hesaplamımıza yarar. Ülkede döviz kurunun düşmesi, diğer taraftan yerli paranın yabancı para karşısında değer kazanması, ticarete konu edilen malın uluslararası pazarda daha pahalı konuma getireceğinden malın rekabet gücünü azaltır. Aksi durumda ise ülkede döviz kurunun yükselmesi, yerli paranın değer kaybetmesine ve malın uluslararası piyasada ucuzlaması neden olup ülkenin rekabet gücünü artıracaktır.

Üretici fiyat endeksi ya da kısaltması ÜFE olarak da bilinen kavram belirli bir referans yıl alınarak ülke ekonomisinde üretimi yapılan ve içeriye satışı konu olan malların, üretici fiyatlarını zaman içinde karşılaştırılmasıyla fiyat değişikliklerini ölçmeye yarayan fiyat endeksi olarak ifade edilir. Aylık ,üç aylık,yıllık enflasyon rakamının belirlenmesinde rol oynar.

Tüketici fiyat endeksi ya da kısaca TÜFE olarak da bilinen kavram tipik bir tüketicinin satın aldığı seçili mal ve hizmet grubunun fiyatlarındaki ortalama değişimleri göstermeye yarayan fiyat endeksi olarak ifade edilir. Yıllık enflasyon değişimlerini ölçmeye yarayan bir endekstir. Belirli bir yıl seçilmeli ve bu yıl baz yıl olarak ele alınmalıdır. Endeks değeri olarak 100 olarak kabul edilir. Bundan sonraki yıllardaysa cari yıllar enflasyon değerleri karşılaştırarak değişkenlik gösterecek ve baz alınan endeks için oynaklıklar oluşacaktır.

Merkez bankası bilançosu olarak da bilinen kavram para politikası uygulamalarındaki ortaya çıkan gelişmenin izlenmesi açısından son derece önem arz etmektedir. Para politikası uygulamalarıyla merkez bankasının bilançosu etkilenmektedir; uygulamadaki değişiklikler merkez bankası bilançosuyla

gözetlenir. Pasif kalemler, merkez bankasının bankacılık ve banka dışı kesime olan yükümlülüklerini ifade eder. Pasif kalemlerde yapılan değişikliklerle ekonomideki likidite ayarlanabilirken, likiditenin hangi kalemlerden dolayı kaynaklandığının sağlanması aktif kalemler tarafından izlenebilir.

En küçük kareler metodu ilk kez 1795 yılında Gauss tarafından ortaya konmuştur. Gauss 1801 yılındaki çalışmasında metodu ele alarak, bulunmasından az süre sonra kaybolan Ceres asteroidinin tekrardan gözleyeceği yeri hesaplamış, bunu başararak büyük bir şana ulaşmıştır. Gauss bu metodu ilk 1809 yılında yayınlamıştır. 1806 yılındaki çalışmasında matematikçi Adrien-Marie Legendre'nin ile 1808'deki matematikçi Robert Adrian'ın çalışması, F.Gauss'dan ayrı olarak bu metodu geliştirmişler ve kullanmışlardır. En küçük kareler metodu, günümüzde neredeyse bütün bilim dallarında sık olarak tercih edilmektedir.

Ridge Regresyon Metodu ilk kez 1970'de Teknometri dergisinde yayınlamış oldukları iki makaleyle Hoerl ve Kennard ortaya atılmıştır. Hoerl ve Kennard makalelerinde "Ridge Regresyon: Ortogonal Olmayan Sorunlar İçin Sapmalı Tahmin" başlığıyla, tam ranklı genel lineer hipotez modelini içeren çoklu regresyonda yansız kestirici sorununu detaylıca açıklamıştır. Makaleyle, yanlış kestirici biçiminin seçenек bir karakterizasyonu çıkartmış ve buna dayalı yanlış kestiriciyi oldukça iyi duruma getirmiş k değerini ortaya koyan Banerjee ve Carr tarafından 1971'de kanıtlanmıştır.

Hoerl ve Kennard bu dergideki ikinci makalesinde "Ridge Regresyon: Ortogonal Olmayan Sorunlara Uygulamalar", başlığıyla önceki makalesinde Ridge Regresyon yöntemi uygulamalarını anlatmışlardır. Bu makaleyle özellikle çok etmenli verilerde ilişkiyi saptamak için iki boyutlu uzayda bir grafik uygulaması olarak bilinen Ridge izine önem verilmiştir. Burada ayrıca EKK yöntemiyle bulunan denklemlerden oldukça iyi regresyon denklemi aramak için yapılabilecek işlemlerden söz edilmiştir.

Hoerl ve Kennard'ın Teknometri dergisindeki yayımlanan makalesiyle ilgili 200'den fazla makaleler kaleme alınmıştır. Makalelerin çoğu yine Hoerl ve Kennard'ın Amerikan Matematik ve Yönetim Bilimleri Dergisinde yayımlanan

makalesiyle birlikte gözden geçirilmiştir. 1980 yılına kadar Ridge Regresyon konusunu içeren çalışmalar gözden geçirilmiştir.

Ridge regresyon yönteminde yanlılık parametresi kestiricinin doğrusal bir fonksiyonu değildir, yanlılık parametresi seçimi çok zordur. Bu nedenle yanlılık parametresinin seçimiyle fazla çalışma yapılmıştır. Yapılan çalışmalar şunlardır: Theobald (1974), Lawless ve Wang (1976), Kibria (2003), Khalaf ve Shukur (2005), Alkhamisi vd. (2006), Alkhamisi ve Shukur (2007), Muniz ve Kibria (2009), Muniz vd. (2012), Dorugade ve Kashid (2012), Karaibrahimoğlu vd. (2016), Asar ve Genç (2017).

Theobald(1974), yaptığı çalışmasıyla ridge regresyon kestiricisinin EKK kestiricisinden daha az HKO'su olmasına yarayan gerekli ve geçerli koşulu elde etmiştir.

Kejian(1993), farklı tahmincilerin birleştirilmesi, avantajlı olur düşüncesiyle bir araya getirerek Stein'in 1956 yılında önerdiği tahminciyle ridge regresyon tahmincisini birleştirerek yeni bir yanlı tahminci tanımlamıştır. Kejian(1993), EKK tahminciyle öne sürdüğü yeni tahmincinin hata kareler ortalaması değerlerini karşılaştırdı. Elde edilen tahmincinin hata kareler ortalaması değerinin küçük olduğunu görmüştür.

Liu(1993) yılında önerilen kestirici, Akdeniz ve Kaçıranlar tarafından 1995 yılında Linear Unified kısaltması şeklinde ifade edilen Liu tahminci (Liu Estimator-LE) ifade edilerek tanımlanmıştır.. Liu kestiricisiyle ilgili yapılmış çalışmalardan bazılarıysa şunlardır: Akdeniz ve Kaçıranlar(2001), Akdeniz (2001) ve Sakallıoğlu vd. (2001) tarafından yapılmış çalışmalardır. Kaçıranlar ve Sakallıoğlu(2001), ridge kestiricisi, liu kestiricisinin avantajlarını biraraya getirerek r-d sınıf kestiricisini ifade etmişlerdir. Liu(2003), çalışmasında ridge kestiricisi, liu kestiricisine seçenек olarak bu kestiricilerden oldukça uygun sonuçları veren Liu tip kestiricileri önermiştir.

Mahajan ve ark.(1977), çalışmalarında veri seti içerisinde çoklu bağıntı problemi olması durumunda, EKK yöntemi ile tahmin edilen pazarlama modellerine ait regresyon katsayılarının yüksek varyansa sahip ve yanlış işaretli olduklarını

görmüşlerdir. Çalışmalarında ridge regresyon analizini kullanarak pazarlama verilerinde çoklu bağıntı sorunu ile mücadele edilebileceğini göstermişlerdir.

Kidwell ve Brown(1982), çalışmasında çoklu bağıntı üzerine odaklanmıştır. Sorunu göstermek için Ridge regresyon tekniğini yapay verilere uygulamıştır. Çalışmada kestiricilerin ortogonal olmadığı durumda Ridge regresyon yönteminin en küçük kareler yönteminden farklı sonuçlar verdiğini göstermiştir.

Aktaş ve Yılmaz(2003), çoklu regresyon analizinde açıklayıcı değişkenler arasında çoklu doğrusal bağlantı olması hakkında EKK yöntemi yerine modele konulan bağımsız değişkenleri atmak yerine yanlış kestirimler yapmasına karşılık bağımsız değişkenlerin regresyon katsayılarının standart hatalarını küçülten Ridge regresyon metodunu kullanmışlardır. Alternatif olarak Liu metodunun da kullanılabileceğini söylemişlerdir.

Topçubaşı ve Billor(2003), çalışmasında bağımsız değişkenler lineer bağımlı olması halinde, en küçük kareler tahmincisine çeşitli seçenek içeren yanlış tahmin modellerin olduğunu, bu yanlış tahmin modellerinin çok büyük bir bölümünü içeren bir sınıfı oluşturduğunu anlatmışlardır. Araştırmacılar yanlış tahmincilerin bir bölümünü bu sınıfa, Liu ve genelleştirilmiş Liu tahmincilerini dahi eklemişlerdir.

Yolacan ve ark.(2005), yapmış oldukları çalışmalarında çoklu bağıntı sorununu içeren ekonomik verilere ridge regresyon ve yapay sinir ağları algoritmalarını uygulayarak sonuçları karşılaştırarak yorumlamıştır.

Albayrak(2005), çalışmasında En Küçük Kareler, Ridge ve Temel Bileşenler Regresyon analizleriyle ulaşılan bulguları yorumlamıştır. Kullanılmış veri setindeki çoklu bağıntı sorunu olması durumunda standart hataların yüksek çıkabileceğini ve yanlış kestirim metodlarının en küçük kareler metoduna göre daha tutarlı, geçerli kestirimler olduğunu göstermiştir.

Aktaş(2007), çalışmasında çoklu regresyon analizinde karşılaşılmış sorunların bir tanesi olan çoklu bağıntıyı saptama yöntemlerini anlatmış, enflasyon modeli için Liu tahmincisiyle sınama yapmıştır. Çalışmasında Liu tahmincisi ile çoklu bağıntıdan arındırılmış model için, ilgili dönemdeki enflasyona para arzı ve USD değişkenlerinin etki ettiğini söylemiştir.

Karakaş(2008), yapılan yüksek lisans tez çalışmasındaki çoklu lineer bağıntının kaynakları, saptama metotlarını ve istatistiksel etkilerini gözlemlemiştir. Çalışmasında, Ridge ve Liu regresyon, sapmalı regresyon kestiricilerini çoklu lineer bağıntı onarım metodu şeklinde tanımlanmıştır. Uygulamasında Ridge ve Liu tipi regresyon yaparak istihdam modelini kestirmeyi amaçlanmıştır. Sapmalı regresyon kestiricilerine ait standart errorların formülü bulunmadığını söylemiştir.

Tüylüoğlu ve Albayrak(2010), çalışmasında, illerin genel fiyat düzeyleriyle diğer alt endeksler arasındaki ilişkileri ridge yöntemiyle inceleyerek illerin hayat pahalılığını ortaya çıkaran etkenlerden en etkili ilk üç etkenin sırasıyla giyim, mobilya ve ev eşyası, diğer ürün ve hizmetleriyle konut harcamalarının olduğunu göstermişlerdir.

Regresyon analizinde hatalar birbiriyle bağımlı olmamalıdır. Bu sorunun testi için ($n \geq 15$) olması durumunda Durbin-Watson test istatistiği ile karar verilir. Karagöz(2016), bu değer (1,5-2,5) arasında bulunursa otokorelasyon sorunun olmadığını söylemiştir.

1.2 Tezin Amacı

Tezin amacı, Türkiye'nin ihracat modelini bir uygulama ile açıklamaya çalışmaktır. Bu çalışmada en küçük kareler, ridge regresyon, liu regresyon, zaman serisi yöntemlerini hakkında bilgi verilecektir. Çoklu doğrusallığı belirleme metotları anlatılacaktır. Çoklu doğrusal bağıntı durumunda en küçük kareler, ridge, liu, zaman serisi tahminleri incelemek, yöntemleri karşılaştırarak en iyi yöntemin hangisi olduğuna karar vermektir. Ortaya çıkan modelin tahmin sonuçlarını yorumlamaya çalışmaktır.

2.1 Regresyon ve Model Kurma

Regresyon analizi değişkenler arasındaki ilişkiyi modellemek ve göz önünde bulundurmak için kullanılan istatistiksel bir yöntemdir. Regresyonun değişik sına alanları mevcuttur. Bu sına alanlar mühendislik, fizik ve kimya gibi alanlarda çokça kullanılmaktadır.

2.1.1 Çoklu Doğrusal Regresyon

Kimyasal bir süreçteki dönüşümün pound türünden verimi, sıcaklık ve katalizör konsantrasyonu bağı olsun. Bu bağıntı çoklu regresyon modelinde aşağıdaki gibi tanımlanabilir:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (2.1)$$

Burada y verimi, x_1 sıcaklığı ve x_2 de katalizör konsantrasyonunu göstermektedir. Bu iki bağımsız değişkenli çoklu doğrusal modeldir.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.2)$$

Modeli k tane açıklayıcı değişkenli çoklu lineer regresyon modeli diye adlandırılır. β_j $j=0,1,\dots,k$ parametreleri regresyon katsayıları diye bilinir. Modelde x_j bağımsız değişkenlerinin k boyutlu uzayında bir hiper düzlemi ifade eder. β_j parametresi, x_i ($i \neq j$) açıklayıcı değişkenleri sabit tutulduğu zaman x_j açıklayıcı değişkenindeki bir birimlik değişme karşılık gelen y değişkenindeki beklenmekte olan değişimi göstermektedir. Bu nedenle, β_j , $j=1,2,\dots,k$ parametreleri, genellikle kısmi regresyon katsayıları diye adlandırılır.

Çoklu regresyon modelleriyle ilgilenirken onları matris biçiminde ifade etmek daha uygundur.

$$y = x\beta + \varepsilon \quad (2.3)$$

$$X = \begin{bmatrix} 1, x_{11}, x_{12}, \dots, x_{1n} \\ 1, x_{21}, x_{22}, \dots, x_{2n} \\ 1, \dots, \dots \\ 1, x_{n1}, x_{n2}, \dots, x_{nk} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix} \quad (2.4)$$

Burada matrisler denklem 2.4'deki gibidir.

y , $n \times 1$ boyutlu gözlemler vektörü; X , $n \times p$ boyutlu açıklayıcı değişkenler matrisi; β , $p \times 1$ boyutlu regresyon katsayıları vektörü ve ε , $n \times 1$ boyutlu rastgele hatalar vektörü olarak tanımlanır.

$S(\beta)$ fonksiyonunu minimize ederek $\hat{\beta}$ en küçük kareler kestiriciler vektörünü elde ederiz.

$$S(\beta) = \varepsilon' \varepsilon = (y - X\beta)'(y - X\beta) \quad (2.5)$$

$$= y'y - 2\beta'X'y + \beta'X'X\beta \quad (2.6)$$

$\beta'X'y$, 1×1 'lik bir matris ya da skaler olduğundan tranpozu da aynı şekilde skalerdir. En küçük kareler kestiricileri,

$$\frac{\partial S}{\partial \beta} \xrightarrow{\beta=\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0 \quad (2.7)$$

eşitliğini sağlar. Basitleştirilse,

$$X'X\hat{\beta} = X'y \quad (2.8)$$

Denklem 2.8 en küçük kareler normal denklemleridir. Normal denklemleri çözmek için denklem 2.8'nin denklemin iki tarafını $X'X$ 'in tersiyle çarpılırsa, böylece β 'nin en küçük kareler kestiricisi,

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2.9)$$

olur. Burada $(X'X)^{-1}$ ters matrisinin de mevcut olma koşulu sağlanmalıdır. Eğer bağımsız değişkenler doğrusal bağımsız ise X matrisinin hiçbir sütununun diğer sütunlarla doğrusal bir birleşimi yoktur.

$\hat{\beta}$ en küçük kareler kestiricilerinin istatistiksel özellikleri kolayca gösterilebilir. İlk önce yanlılığı ele alalım.

$$E(\hat{\beta}) = E[(X'X)^{-1}X'y] = E[(X'X)^{-1}X(X\beta + \varepsilon)] \quad (2.10)$$

$$= E[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon] = \beta \quad (2.11)$$

Burada $E(\varepsilon) = 0$ ve $(X'X)^{-1}X'X = I$ 'dir. Bu nedenle $\hat{\beta}$, eğer model doğru ise β 'nin yansız bir kestiricisidir.

$$\text{Cov}(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\} \quad (2.12)$$

Kovaryans matrisi $p \times p$ boyutlu simetrik bir matristir ve j . köşegen elemanı ve $\hat{\beta}_i$ 'nin varyansının, ij . köşegen dışı elemanı ise $\hat{\beta}_i$ ve $\hat{\beta}_j$ arasındaki kovaryansı verir. $\hat{\beta}$ 'nin kovaryans matrisi,

$$\text{Cov}(\hat{\beta}) = \text{Var}(\hat{\beta}) = \text{Var}[(X'X)^{-1}X'y] \quad (2.13)$$

ile bulunur. Burada, $(X'X)^{-1}X'$ sabitler matrisidir ve y 'nin varyans $\sigma^2 I$ 'dir. Böylece

$$\text{Var}(\hat{\beta}) = \text{Var}[(X'X)^{-1}X'y] = (X'X)^{-1}X'\text{Var}(y)[(X'X)^{-1}X']' \quad (2.14)$$

$$= \sigma^2 (X'X)^{-1}X'X(X'X)^{-1} = \sigma^2 (X'X)^{-1} \quad (2.15)$$

elde edilir. Bundan dolayı, eğer $C = (X'X)^{-1}$ alırsak, $\hat{\beta}_j$ 'nin varyansı, $\sigma^2 C_{jj}$ 'dir ve ise $\hat{\beta}_i$ ve $\hat{\beta}_j$ arasındaki kovaryans ise $\sigma^2 C_{ij}$ 'dir.

En küçük kareler kestiricilerinin kalitesine ilişkin bir diğer sonuç $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2$ varsayımına ve ilişkisiz hatalara sahip regresyon modeli için en küçük kareler kestiricisinin, y_i 'nin doğrusal bileşimleri diğer bütün yansız kestiricilerle karşılaştırıldığında yansız oldukları ve minimum varyansa sahip olduklarını belirten Gauss-Markov teoremidir. Genellikle en küçük kareler kestiricileri en iyi

doğrusal yansız kestiriciler olduğunu söyleriz. Burada en iyi demek en küçük varyansa sahip anlamına gelmektedir.

2.2 Zaman Serisi Verileri İçin Regresyon Modellerine Girişi

Regresyon uygulamalarının birçoğu hem yanıt değişkenin hem de bağımsız değişkenlerin zaman odaklı olduğu zaman serilerini içerir. Zaman serisi kullanan regresyon modelleri iktisat, işletme ve birçok mühendislik alanında çokça görülmektedir. Zaman serisi verilerinde hatalar genellikle bir çeşit otokorelasyonlu yapıda olmaktadır. Otokorelasyon terimi ile hataların farklı zaman aralıklarında birbirleri ile ilişkili olduğunu söylenir.

Hatalarda otokorelasyonun varlığı en küçük kareler yöntemi üzerinde çeşitli etkilere sahiptir.

Bunlar aşağıdaki gibidir:

- 1) En küçük kareler regresyon katsayıları hala yansızdır ancak en küçük varyanslı kestiriciler değildirler.
- 2) Hatalar pozitif olarak otokorelasyonlu olduğunda artık kareler ortalaması, σ^2 hata varyansını önemli ölçüde daha küçük yapabilir. Bir ya da daha fazla bağımsız değişkenin gerçekte öyle olmadığını halde anlamlı olduğunu göstererek yanlış yönlendirebilir.
- 3) Güven aralıkları,önkestirim aralıkları t ve F dağılımları dayalı hipotez testleri, kesin olarak söylenebilir ki,artık, artık uygun değildirler.Durbin Watson istatistiği, regresyon modelini kestirdikten sonra artıkların ilişkili olması halinde, ilişkili olup olmadığını ortaya çıkarmayı hedefine koyan istatistiktir.Bu istatistiğin 2 yakınlarında bulunması, “otokorelasyon sorunu vardır” null hipotezini kabul edeceğimizi belirtir. Bu açıklamaya göre e= hata terimini, t= zamanı ifade etmek üzere Durbin Watson test istatistiği:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} \quad (2.16)$$

d değerinin aralığı 0 ile 4 arasında değer alabilir. Artıkların örneklem otokorelasyon değeri r olmak üzere d değeri yaklaşık $2(1-r)$ 'ye eşit olacak şekilde $d=2$ olması otokorelasyon probleminin olmadığı sonucunu verir. Genel kabul gören kanı, eğer Durbin-Watson değeri 1'den azsa bir alarm durumu ortaya çıkacaktır. Küçük olan d değeri, ortalamaya göre ardışık hata terimlerinin pozitif bir ilişkili içinde olduğunu gösterir.

Durbin Watson testi, seri ilişkisinin varlığını ortaya çıkarmak için geliştirilmiş ilk testlerden biri olması sebebiyle önemlidir ama test istatistiğinin hesaplanmış olan D dağılımı, kararsız kalınan bazı bölgelere sahip olduğundan testin etkin bir test olmadığı kanısına varılmıştır. Ayıryeten regresyon modeli, açıklanan değişkenin gecikmeli değerlerini içeren açıklayıcı değişken olarak yer veriliyorsa test sapmalı sonuç göstermektedir; bu durum için seçenek bir test olarak Durbin h test istatistiği önerilmiştir. Durbin h testi aşağıdaki gibidir:

$$h = (1 - \frac{1}{2}d) \sqrt{\frac{T}{1 - T \cdot \text{Var}^{\wedge}(\hat{\beta}_1)}} \quad (2.17)$$

Burada $\text{Var}^{\wedge}(\hat{\beta})$ zaman gecikmeli açıklanan değişkenin eğim katsayısının standart hatasının karesi, T ise gözlem sayısını ifade etmektedir. Fakat bu test istatistiği şu koşulda geçerli olmaktadır.

$T \cdot \text{Var}^{\wedge}(\hat{\beta}) < 1$ matematiksel olarak paydasının eksi işaretli olması durumunda karekök işlemi yapılamaz.

VAR yöntemi modellerinde değişkenlerin bir bütün olarak ele alarak birbirleriyle etkileşimlerini ortaya çıkarması açısından çok mühim bir tekniktir. VAR analizinde elde edilen katsayıların yorumlanabilmesi güç olmasına karşılık, F testi sonuçları sağ tarafta yer alan değişkenlerin beklenen etkileri konusunda bilgi vermesi dolayısıyla çok faydalıdır. VAR analizinde bulunan değişkenlerin durağan olması veya olmaması büyük bir tenkit söz konusudur. Fark alımındaki tenkit, verilerle hareket bilgilerdeki kayba sebep olmaktadır. Bu konuyla ilgili çokça ifade edilen görüşün analizde yer alan değişkenlerin gerçek veriden yararlanılarak yapılması şeklinde ifade edilmektedir. Günümüzdeki deneysel çalışmalarda model kestiriminde veri setilerindeki ilk farklarının alınmasının

daha uygun olup olmayacağını test etmeyi amaçlamışlardır. Durağan olmayan serilerde yapılmış analizler, durağan seriler kullanarak yapılmış analizlerden apayrı değerlendirilmiştir. Durağan olmayan seriler kullanılarak yapılmış analizler, istatistiksel açıdan yanlış sonuçları ifade etmektedir. Birim köke sahip olsalar dahi fark alımına karşıt görüş vererek VAR analizinin hedefinin parametre kestirimi olmadığı, ama değişkenler arası ilişkilerin belirlenmesinde yardımcı olduğu sonuca varılmıştır.

2.2 Çoklu Bağntı Problemi

Çoklu bağntı denklemin bağımsız değişkenleri arasındaki ilişkinin ciddiyetiyle alakalı problemidir. İki türlü olarak karşımıza çıkar. Bağımsız değişkenler arasında tam bir ilişki varsa “tam çoklu bağntı” problemi vardır. Bağımsız değişkenler tam olmasa dahi ilişki içindeyse “tam olmayan çoklu bağntı” problemi ortaya çıkar.

Denklem 2.4'deki X matrisinde X_2 ile X_3 arasında tam bir lineer ilişki varsa bu değişkenler arasındaki korelasyonu 1 olarak bulunacaktır. Bu halde $\text{rank}(X)=k$ varsayımı sağlanamayacak $\text{rank}(X)<k$ sonucuna ulaşılacaktır. Yine bu problem $|X|=0$ ve $|X'X|=0$ sonucunu ortaya çıkardığından X^{-1} ve $(X'X)^{-1}$ bulunamayacaktır ve denklem tahmin edilemeyecektir.

Tam çoklu bağntı problemi şu şekilde tespit edilebilir. Verilerin türetilişinde ve/veya kullanılışında farkında olmayıp yapılan bir yanlışlıktan kaynaklanabilir. Denklemde bir sabit terim vardır ve bağımsız değişken sabit kalabilir. Bir bağımsız değişken diğer bağımsız değişkenlerin lineer bileşimi olarak türetilmiştir. Kukla değişken hilesine yakalanılmış olabilir.

Klasik modelin varsayımları sağlandığında regresyon katsayılarının en küçük kareler tahmin edicileri DESTE(Doğrusal En iyi Sapmasız Tahmin Edici)'dir. Çoklu bağntı tama yakın çoklu bağntıdaki gibi çok yüksek olsa bile en küçük kareler tahmin edicileri DESTE özelliklerini korumayı sürdürürler. Öyle ise neden çoklu bağntı önem kazanıyor. Aslında çoklu bağntı hiçbir regresyon koşulunun dışına çıkmaz. Ayıryeten sapmasız, tutarlı kestirimler bulunur, bu kestirimlerin standart hataları dahi doğru hesaplanır. Çoklu bağntının tek etkisi

küçük standart sapmalı katsayı kestirimleri elde etmeyi zorlaştırır. Gözlem sayısı tahmin edilecek katsayı sayısının üstündeyse çoklu bağıntı sorunu ortaya çıkar.

Çoklu bağıntı uygulamada doğurduğu sonuçlar şunlardır. EKK tahmin edicilerin DESTE olmalarına karşın varyansları ve ortak varyansları büyüktür, bu da kesin tahmini güçleştirir. Güven aralıkları çok geniş olma eğilimindedir, bu da “sıfır” (yani ana kütledeki gerçek katsayısının sıfır olduğu) yokluk hipotezi kolayca red edilememesine yol açar. Bir ya da daha çok katsayının t oranları istatistik bakımından anlamsız olmasına karşın, bütünün uyum iyiliğinin ölçüsü çok yüksek olabilir. EKK tahmin edicileriyle onların standart hataları, verilerdeki değişimlere karşı duyarlı olabilirler.

Bir ya da daha çok katsayı yani t oranları

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}} \quad (2.18)$$

istatistik bakımından anlamsız olur.

Çoklu bağıntı sorununu düzeltici önlemler şunlardır; Önsel bilgi, kesit verileriyle zaman serisi verilerini bir araya toplama yani dışsal ya da önsel bilgi tekniğinin bir biçimi de karma veri yaratmak diye bilinen, kesit ile zaman serisi verilerini bir araya toplamaktır. Değişken(ler)i atmak ve model kurma sapması yani çoklu bağıntı sorununda en basit yol ortak doğrusal değişkenlerden birini modelden çıkarmaktır. Bir değişkeni modelden çıkarmak da model kurma hatası (sapması) nın büyümesine yol açar. Değişkenlerin dönüştürülmesi, ek ya da yeni veri alınabilir. Çoklu regresyonlarda ortak doğrusallığı azaltmak gerekir.

Çoklu bağıntıyı düzeltmek için önlemler şunlardır; Faktör analizi (etmen çözümlemesi) yapmak, temel bileşenler analizi yapmak, ridge (sırt) regresyon yapmak gerekebilir.

Son olarak bir çatı altında derlemek istersek çoklu bağıntı hakkında özet ve sonuçlar şunlardır.

Klasik regresyon analizinde çoklu doğrusallık olmadığı varsayılsa da X değişkenleri arasında ya tam ya da tama yakın doğrusal ilişkiler bulunabilir.

Çoklu bağıntının olması sonucu itibariyle regresyon katsayıları belirlenemez ve standart hataları tanımlanamaz. Ortak bağıntı tam değil ama yüksekse regresyon katsayıları tahmin edilebilir ama bunların standart hataları yüksektir. Amaç regresyon katsayılarının doğrusal bileşiminin tahmin edilebilir fonksiyonlarını tahmin etmekse, tam çoklu bağıntı varken bile bu yapılabilir.

Çoklu bağıntının göstergesi R^2 'nin yüksek olması, regresyon katsayılarının t testine göre istatistik bakımından anlamlı olmamasıdır. İki değişken arasında sıfıncı dereceden korelasyon katsayısının değeri yüksek ise nedeni çoklu bağıntıdır. İki'den fazla değişken için bulunan sıfıncı dereceden bulunan korelasyon katsayısı yanılgıya sebep olabilir. Bu durumda kısmi korelasyon katsayılarına başvurulabilir. R^2 yüksekken kısmi korelasyonlar düşükse çoklu bağıntı olasıdır. Eğer R^2 yüksekken kısmi korelasyonlarda yüksek ise çoklu doğrusallığın varlığı hemen anlaşılamayabilir. X_i değişkeninin modeldeki diğer X değişkenleri üzerine bulunan regresyon modelinden bulunan belirlilik katsayıları R_i^2 bulunabilir. R_i^2 'nin büyük olması X_i 'nin diğer X 'lerle yüksek bir korelasyonu olduğunu gösterir. Bu durumda X_i model kurma hatasında ciddi bir büyümeye sebep olmayacaksa modelden çıkarılır.

Çoklu bağıntının saptanması maçı kazanmanın yarısı, diğer yarısı ise nasıl giderileceğidir. Dışsal ya da önsel bilgidен yararlanılabilir. Kesit ya da zaman serisi verilerini bir araya getirilebilir. Yüksek ortak doğrusallık gösteren değişkeni dışlanılabilir. Veriler dönüştürülebilir. Ek ya da yeni veri derlenilebilir. Bu kurallardan hangisinin uygulamada işe yarayacağı, kuşkusuz verinin niteliğine ve ortak doğrusallık sorununun ciddiyetine bağlıdır.

Çoklu bağıntının kestirimdeki rolüne dikkat çektik ve ortak doğrusallık yapısı gelecekteki örneklemelerde de varlığını sürdürmeyecekse, çoklu bağıntıya bulaşmış regresyon tahminlerini geleceğin kestiriminde kullanmanın sakıncalı olabileceğine söylemek gerekir.

Her ne kadar çoklu bağıntı sorunsal da, küçük örnekleme en az onun kadar sorundur.

2.2.1 Çoklu Bağıntının Saptanması

Çoklu bağıntıyı saptamak ve ciddiyetini ölçmek için oldukça fazla metot ileri sürülmüştür. Bazı metotların bir bölümü çoklu bağıntının varlığı için bir gösterge olarak önerilmişken bazı metotlar katı kurallar içermektedir. Tüm metotların ortak özelliği çoklu bağıntının ciddiyeti hakkında bilgi vermesidir.

Açıklayıcı değişkenlerin ilişki katsayıları yüksek olması durumunda çoklu bağıntının kanıtı görülür. Bu metotta ilişki matrisinin XX' köşegenleri haricindeki birimleri, değişkenler arasındaki ilişki katsayısını gösterir. X_i , X_j değişkenleri arasında tama yakın bağıntı var ise, $|r_{ij}|$ bire yakınlaşır.

Kuralı ifade etmek gerekirse sadece iki açıklayıcı değişken arasındaki lineer ya da lineere yakın bağıntıyı ortalığa çıkarır. Değişkenler arası ilişki katsayıları zayıfken de ciddi çoklu bağıntı ile karşılaşmak olası olduğu için, ikiden çok açıklayıcı değişken arasındaki ilişkiyi ilişki katsayıları ile açıklamak zayıf kalır. Açıklanan değişkenle açıklayıcı değişkenlerin tüm bileşimleri arasındaki kısmi ilişki katsayıları hesaplanarak çoklu bağıntı sorunu olduğu ortaya konabilir.

Çoklu bağıntı sorununun teşhisinde çokça kullanılan metotlardan biri Varyans Büyütme Faktörü (VBF)'dür. $(XX')^{-1}$ matrisini açıklamak gerekirse asal köşegeni VBF değerini ifade etmektedir

Hoerl ve Kennard(1970), çoklu bağıntının teşhisinde VBF değerleri gösterge olduğu söylemiştir. Fakat VBF değeriyle çoklu bağıntının ciddiyetine karar vermek için kaynaklarda ortak bir sayı söylenememiştir. Bazı kişiler VBF değerinin 5'ten büyük olması durumunda çoklu bağıntının ciddi olacağını belirtirken, bazı kişiler çoklu bağıntının ciddiyetinin VBF değerinin 10'dan büyük olması durumunda araştırılması gerektiğini söylemiştir. Bazı kişiler, 4'ün üzerindeki bir VBF daha fazla araştırmanın gerekli olduğunu söylemişlerdir.

Özdeğerlerin sıfıra yakın olması çoklu bağıntı belirti gösterse de ciddiyetini bize sunmaz. Bundan ötürü özdeğerlere tek tek bakılması yerine birbirleriyle kıyaslanmasını içeren bir metot önerilmiştir. En büyük özdeğerin en küçük özdeğere oranı şeklinde ifade edilmiştir. Koşul sayısı 100 ile 1000 arasında ise çoklu bağıntı orta derece, 1000'den büyükse yüksek derecede sorun vardır.

$X'X$ ilişki matrisinin determinanı çoklu bağıntının ciddiyetini ortaya çıkarmak için bir göstergedir. Matrisin determinanı 0 ile 1 arasında değer alacaktır. Sıfır olması durumu, tam çoklu bağıntıyı ifade ederken, 1'e eşit olması çoklu bağıntının olmadığı söyleyecektir. Determinanı sıfıra yaklaşan değer olması tama yakın çoklu bağıntının varlığını gösterir ve bu durum en çok karşımıza çıkan bir durumdur. Ancak bu gösterge çoklu bağıntının ciddiyeti hakkında görüş vermekte ve hangisinden kaynaklı bu sorunun ortaya çıktığı hakkında bilgi vermez.

Yardımcı regresyon, modeldeki açıklayıcı değişkenlerin sırası ile açıklanan değişken olarak ifade etmekte, kalanlarıysa açıklayıcı değişkenler ile oluşturulan bir regresyon modelinin kestirimine dayanan metottur. Bu modellere ait R^2 'ler ile hesaplanacak F oranının anlamlılığı çoklu bağıntının varlığının saptanmasında yardımcı olacaktır..

Açıklayıcı değişkenler arası ilişki katsayıları, modelin R^2 'sinden yüksekse, çoklu bağıntı sorunu ciddidir.

Theil-M ölçütünde, öteki metotlardan farklılığı, açıklayıcı değişkenlerin kendi aralarındaki ilişkiyi değil açıklanan değişken ile olan ilişkilerini öne sürerek çoklu bağıntı olup olmadığının araştırmasına dayanır. Bu metotta, modelden sırası ile açıklayan değişkenler çıkartılıp kestirilen modellerin çoklu ilişki katsayılarıylada ilk kestirilen modelin R^2 'si arasındaki ilişkiyi ortaya koyan bir ölçüdür.

Bu yöntemin diğer yöntemlere göre zayıflığı, çoklu bağıntının miktarı hakkında bilgi vermez. Varyans büyütme faktörü çoklu bağıntı hakkında daha fazla bilgi veren çok kullanılan yöntemdir.

Örneklemdaki çoklu bağıntının sebebi, diklikten sapma olarak ortaya çıkabilir. Diklikten sapma arttıkça $X'X$ ilişki matrisinin determinantının sıfıra yaklaşması ve çoklu bağıntıya sebebi bulgusunu izleyerek bu metot, açıklayıcı değişkenlerin dikliğini ortaya çıkarmayı konu eder. Farrar ve Glauber(1967), normal dağılıma uyduğunu gösterdikleri p tane açıklayıcı değişken arasında çoklu bağıntıyı ortaya çıkarmak için, $X'X$ ilişki matrisinin determinantına dikkat çeken χ^2 test istatistiği geliştirmiştir.

Curto ve Pinto(2007), çoklu bağıntının EKK kestiricileri üzerindeki etkilerinden bahsederken ayrıca temel bileşenler ve etmen analizi gibi çok değişkenli istatistiksel yöntemlerdeki kötü etkilerini ortaya çıkaran bir yeni metot ortaya koymuştur.

2.3 Ridge Regresyon

Dik olmayan verilere en küçük kareler yöntemi uygulandığında regresyon katsayılarının çok zayıf kestirimleri elde edilir. En küçük kareler yönteminde sorun, $\hat{\beta}$ 'nin β yansız bir kestirici olma zorunludur. Gauss-Markov özelliği en küçük kareler kestiricisinin yansız lineer kestiriciler sınıfındaki en küçük varyansa sahip olduğundan emin olmamızı sağlamaktadır ancak bu varyansın küçük olacağını garantisi yoktur. Sorunu hafifletmenin yolu, β 'nın kestiricisinin yansız olması zorunluluğundan vazgeçmektir. Regresyon katsayılarının yanlış kestiricilerini elde etmek için farklı yöntemler geliştirilmiştir. Bunlardan biri ridge regresyondur. İlk olarak Hoerl ve Kennard tarafından önerilmiştir.

$$\hat{\beta}_r = (X'X + kI)^{-1} X' y \quad (2.19)$$

Denklem 2.19'da ridge regresyon kestiricisi gösterilmektedir. Denklem 2.9'da EKK tahmin edicisine göz önünde bulundurulduğunda yanlış olmasına ayrıyeten varyansı küçültmesi sebebiyle tercih edilebilecek bir metot olan ridge regresyon metodunun başlıca kullanım amaçlarını şu şekilde ifade etmek gerekir:

- Çoklu regresyon modelinde açıklayıcı değişkenler birbirleriyle ilintili olduklarında, en küçük kareler tahmin edicisinden daha küçük varyanslı tahmin edici elde edilmesi için,
- Güçlü çoklu lineer bağlantı etkisiyle regresyon katsayılarında oluşmuş kararsızlıkların grafik ortamında gösterilmesi için,
- Modeldeki gereksiz değişkenlerin çıkarılması için,
- Ridge regresyon metodunun en küçük karelerden bir diğer farklılığıysa, k Ridge yanlışlık parametresinin olmasıdır. 0 ile 1 arasında değer alan her k için hesaplanan parametre tahminleri arasından uygun kriterlere sahip olanları seçmek için kullanılır.

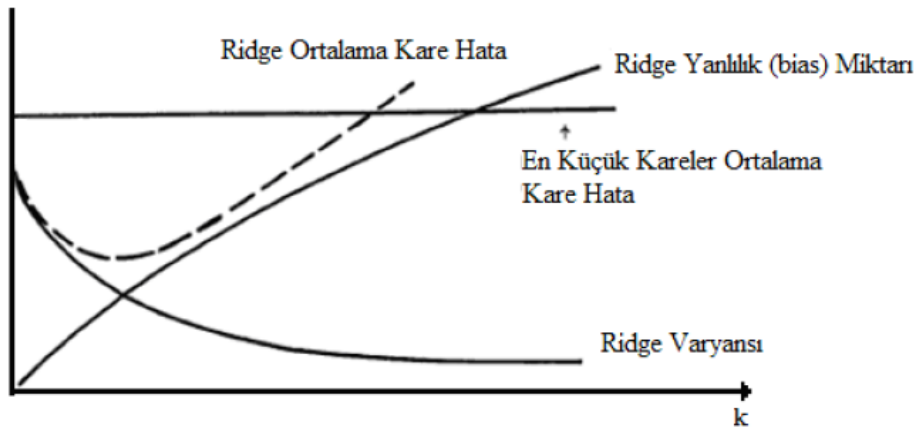
$$E(\hat{\beta}_r) = \beta - k(X'X + kI)^{-1}\beta \quad (2.20)$$

$\hat{\beta}_r$, denklem 2.19'deki Ridge regresyon kestiricisini ifade etmektedir. Denklem 2.20'deki Ridge regresyon beklenen değeri gösterilmektedir.

$$yan(\hat{\beta}_r) = E(\hat{\beta}_r) - \beta \quad (2.21)$$

Denklem 2.21'de yanı bulabilmek için denklem 2.20'deki ifadeyi yerine koyduğumuzdaki denklem 2.22'deki ifadeye ulaşırız.

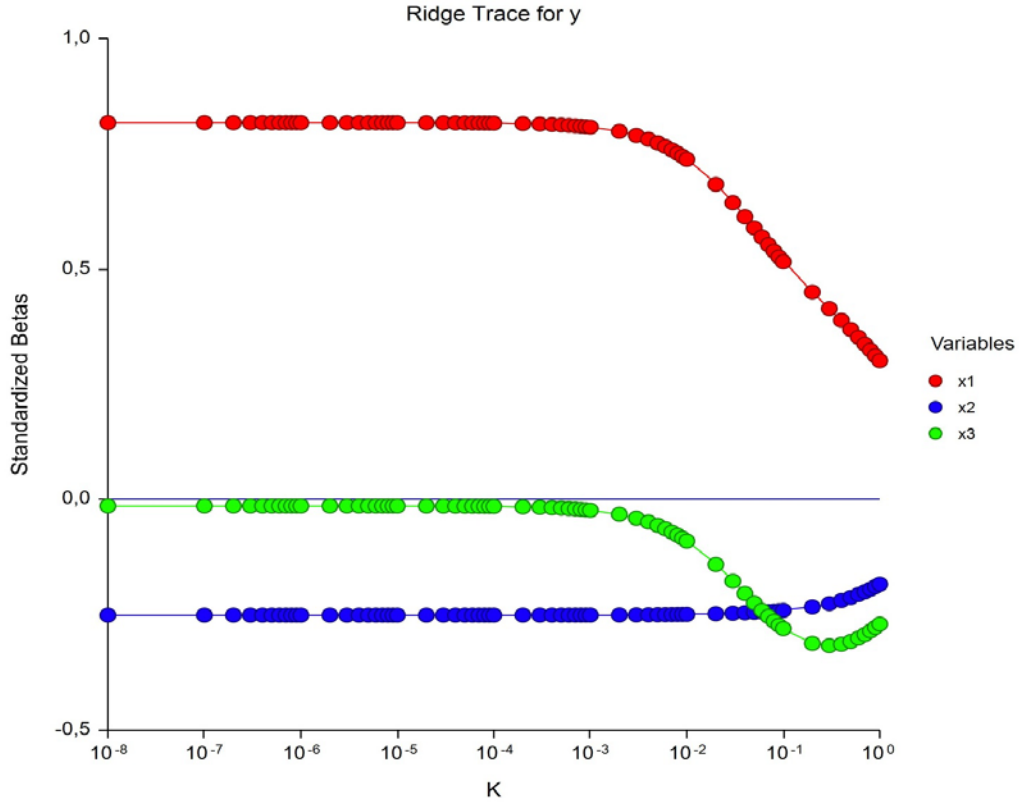
$$yan(\hat{\beta}_r) = -k(X'X + kI)^{-1}\beta \quad (2.22)$$



Şekil 2.1 K sabiti

Varyans, sapma kare ve k arasındaki ilişkiyi şekilde göstermeye çalışılmıştır. Şekil 2.1'deki en küçük kareler varyansı hata kareler ortalamasına eşittir ve k'dan etkilenmediği için yatay doğruyla ifade edilmiştir. Şekil 2.1'de görüleceği üzere, k artarken ridge'in varyansı azalmakta, sapma kare artmaktadır. Şekil 2.1'de üzerinde kesikli çizgiyle gösterilen ridge kestirici metodunun hata kareler ortalaması, uygun bir k'da minimum değeri almakta ve daha sonra tekrar bekleneneği gibi en küçük karelerin hata kareler ortalamasını geçecektir. Böylelikle, küçük bir sapmaya göze alarak, kestirimin artık kareler ortalamasında meydana gelecek az bir artışa karşılık varyansta kayda değer bir azalma sağlayacağından dolayı k değerinin seçilmesine yardımcı olacaktır.

Çoklu lineer regresyon probleminde çoklu bağlantı olduğu durumda katsayı kestirimleri hassas olur. Yani veri kümesine birkaç gözlemin ilave edilmesi ile bu kestiricilerde değişikliklerin olabileceği gözlenmektedir. Bu sebeple duyarlılık analizi Ridge regresyonun temel amacı olarak karşımıza çıkar. Ridge regresyonun en büyük avantajlarından birisi çalışılan veri için duyarlı olan katsayıları görmemize yardımcı olabilir. Ridge izi (Ridge Trace) olarak ifade edilen bu grafik gösterim duyarlı katsayıları görmemizi sağlar. Ridge izi araştırmacıya hangi katsayıların verilere hassas olduğu konusunda bilgi verir. Ridge izi grafiklerinde yanlı regresyon katsayıları, k 'nın bir fonksiyonu olarak ele alınmaktadır. Ridge izi, her bir katsayı değerinin grafiksel gösterilmesidir. Her bir katsayı için iz oluşturur. Ridge izinde amaç EKK'dan daha küçük HKO'suna sahip k değerini elde etmektir. Ridge izinde bir diğer amaç kararlı katsayılar kümesini oluşturmaktır. Kararlı katsayılar kümesinin anlamıysa verideki küçük değişikliklere karşı katsayıların hassas olmamasıdır. Eğer açıklayıcı değişkenler arasında yüksek bir ilişkili iseler k 'nın küçük değerleri için katsayılar çok hızlı değişir ve k 'nın daha büyük değerlerinde derece derece kararlı olur. Katsayıların kararlı olduğu k değeri katsayıların istenen veri kümesinin elde edilmesinde yardımcı olur.



Şekil 2.2 Ridge izi

Burada x_1 'e dikkat edelim. Ridge izinde en çok hassasiyet x_1 'dir. Bu aralıkta bir k değeri seçilmelidir. Çoklu doğrusal bağıntı durumunda Ridge regresyon metoduyla yapılacak tahminlerin kararlılığı, ridge yanalılık parametresi k için uygun değerin bulunmasına bağlıdır. Optimum k değerinin bulunması halinde ridge tahmincilerinin HKO'sı EKK tahmincisinin HKO'dan daha az olur. Fakat en optimal k değeri, k 'nın bilinmeyen β parametre vektörüne ve varyansa σ^2 bağlı olmasına karşın kesin bir biçimde belirlenmiş olmaz.

Yapılan çalışmalarda ridge parametresi k 'nın belirlenmesi için kesin bir kural konulmamasına karşın k 'nın tahmini için bir oldukça fazla öneri ortaya konulmuştur. Hoerl ve Kennard, 1970 yılında k 'nın seçimi için Ridge İzini (Ridge Trace) önermiştir. Marquardt ve Snee, Varyanas Büyütme Faktörü'nün 1 ile 10 değerleri arasındaki k 'nın seçilebilmesini önermişlerdir. Anderson, 1998 yılında açıklayıcı değişkenler için birlikte 1'e yaklaşan VBF değerlerine karşılık gelen k sabitinin alınmasını ifade etmişlerdir.

Hoerl, Kennard ve Baldwin 1975 yılında,

$$k^* = \frac{\hat{\sigma}^2}{\hat{\beta}' \hat{\beta}} \quad (2.23)$$

önermiştir. Buradaki k açıklayıcı değişken sayısını göstermesi sebebiyle ridge yanlılık parametresi k^* ifade edilmiştir.

McDonald ve Galarneau 1975 yılında,

$$\hat{\beta}^* \hat{\beta}^* = \hat{\beta}' \hat{\beta} - \hat{\sigma}^2 \sum_{j=1}^k \lambda_j^{-1} \quad (2.24)$$

Denklem 2.24'i sağlayacak k değerinin varlığının uygun olabileceği belirtmişlerdir. Bunların dışında çoklu bağıntı çok ciddi olması durumunda EKK kestirimleri doğru sonuçlar vermeyeceği için denklem 2.24'deki formülasyona yazılırsa,

$$k^* = \frac{\hat{\sigma}^2}{\sum_{i=1}^k \hat{\alpha}_i^2} \quad (2.25)$$

önerilmiştir. Burada $\hat{\alpha}$ temel bileşenler kestiricisidir.

Lawless ve Wang'ın 1976 yılında önerdiği bir formüldür.

$$k_{lw} = \frac{p \hat{\sigma}^2}{\sum_{j=1}^p \lambda_j \hat{\alpha}_j^2} \quad (2.26)$$

Anderson(1998), 1'e yaklaşan VBF değerlerine karşılık gelen k sabitinin alınabileceğini belirtmiştir. Albayrak(2005), bu yaklaşımın kullanabileceğini belirtmiştir

Kibria(2003), yılında k yanlılık parametresi için önerdiği formüller aşağıdaki gibidir.

$$\frac{1}{p} * \sum \frac{\hat{\sigma}^2}{\hat{\beta}_j^2} \quad (2.27)$$

$$\frac{\sigma^{\wedge 2}}{(\prod (\alpha_j))^{\wedge 2} (1/p)} \quad (2.28)$$

$$\frac{median(\frac{\sigma^{\wedge 2}}{\alpha_j})}{\alpha_j} \quad (2.29)$$

Muniz ve Kibria(2009), k yanlılık parametresi için önerdiği formüller aşağıdaki gibidir.

$$\max \left(\frac{1}{\sqrt{\frac{\sigma^{\wedge 2}}{\alpha_j}}} \right) \quad (2.30)$$

$$\max \left(\sqrt{\frac{\sigma^{\wedge 2}}{\alpha_j}} \right) \quad (2.31)$$

$$\left(\prod \frac{1}{\sqrt{\frac{\sigma^{\wedge 2}}{\alpha_j}}} \right)^{1/p} \quad (2.32)$$

$$\left(\prod \sqrt{\frac{\sigma^{\wedge 2}}{\alpha_j}} \right)^{1/p} \quad (2.33)$$

$$median \left(\frac{1}{\sqrt{\frac{\sigma^{\wedge 2}}{\alpha_j}}} \right) \quad (2.34)$$

Muniz vd.(2012), k yanlılık parametresi için önerdiği formüller aşağıdaki gibidir.

$$\max\left(\frac{1}{\sqrt{\frac{\lambda_{\max}^2 \sigma^2}{(n-p)\sigma^2 + \lambda_{\max} \alpha_j}}}\right) \quad (2.35)$$

$$\max\left(\sqrt{\frac{\lambda_{\max}^2 \sigma^2}{(n-p)\sigma^2 + \lambda_{\max} \alpha_j}}\right) \quad (2.36)$$

$$\max\left(\sqrt{\frac{\lambda_{\max}^2 \sigma^2}{(n-p)\sigma^2} + \lambda_{\max} \alpha_j}\right) \quad (2.37)$$

$$\left[\Pi \left(\frac{1}{\sqrt{\frac{\lambda_{\max}^2 \sigma^2}{(n-p)\sigma^2 + \lambda_{\max} \alpha_j}}}\right)\right]^{1/p} \quad (2.38)$$

$$\left[\Pi \left(\sqrt{\frac{\lambda_{\max}^2 \sigma^2}{(n-p)\sigma^2 + \lambda_{\max} \alpha_j}}\right)\right]^{1/p} \quad (2.39)$$

$$\text{median}\left(\frac{1}{\sqrt{\frac{\lambda_{\max}^2 \sigma^2}{(n-p)\sigma^2 + \lambda_{\max} \alpha_j}}}\right) \quad (2.40)$$

Dorugade ve Kashid(2012), k yanlılık parametresi için önerdiği formüller aşağıdaki gibidir.

$$\max\left(0, \frac{p \sigma^2}{\alpha' \alpha} - \frac{1}{n(VBF_j)_{\max}}\right) \quad (2.41)$$

$$\text{HarmonicMean} \left[\frac{2p}{\lambda_{\max}} \sum \left(\frac{\sigma^2}{\alpha_j^2} \right) \right] \quad (2.42)$$

Karaibrahimoğlu, Asar ve Genç(2014), Hacettepe Matematik ve İstatistik dergisinde önerdikleri tahmin ediciler aşağıdaki gibidir.

$$\frac{p^2}{\lambda_{\max}^2} \frac{\sigma^2}{\sum_{i=1}^p \alpha_i^2} \quad (2.43)$$

$$\frac{p^3}{\lambda_{\max}^3} \frac{\sigma^2}{\sum_{i=1}^p \alpha_i^2} \quad (2.44)$$

$$\frac{p}{\lambda_{\max}^{1/3}} \frac{\sigma^2}{\sum_{i=1}^p \alpha_i^2} \quad (2.45)$$

$$\frac{2p}{\lambda_{\max}^{1/2}} \frac{\sigma^2}{\sum_{i=1}^p \alpha_i^2} \quad (2.46)$$

$$\frac{p}{(\sum_{i=1}^p \sqrt{\lambda_i})^{1/3}} \frac{\sigma^2}{\sum_{i=1}^p \alpha_i^2} \quad (2.47)$$

Hacettepe Matematik ve İstatistik Dergisinde denklem 2.43 ve denklem 2.44' deki kestiriciler en iyi performansı gösterdiğinden bahsedilmiştir.

Göktaş ve Sevinç(2016), türlü türlü simülasyon çalışması türlü türlü p ve n değerleri için yeni iki k yanlılık parametresi ifade etmişlerdir.

2.4 Liu Regresyon

Çoklu bağıntı durumunda alternatif olarak önerilen yanlı kestiricilerden birisi Liu kestiricisidir. Liu(1993), Ridge regresyonun farklı bir kombinasyonu ile Stein tipi kestiricisinin bir kombinasyon oluşturmuştur, bu kestiricilerin avantajlarını da biraraya toplayarak yeni bir yanlı kestirici tanımlamıştır. Liu tanımladığı bu kestiriciyle EKK'yı karşılaştırmıştır, önerdiği kestircinin EKK kestircisinden az hata kareler ortalaması olduğunu ortaya çıkarmıştır.

$$\hat{\beta}_d = (X'X + I)^{-1}(X'y + d\hat{\beta}) \quad (2.48)$$

Denklem 2.48'deki eşitlikte Liu kestiricisi gösterilmektedir. $\hat{\beta}_d$ ile gösterilmiştir. d , yanlılık parametresinin değerleri $0 < d < 1$ arasında değer almaktadır. Liu kestiricisi d yanlılık parametresinin lineer bir fonksiyonunu ifade etmesi itibariyle, d 'nin seçilmesi k 'nın seçilmesine nazaran kolaydır. Bunu göstermesi itibariyle bahsedilen özellik Liu kestiricisini daha önemli yapan bir husustur.

$$X'y = (X'X)^{-1}X'y \quad (2.49)$$

$$A = (X'X + I)^{-1}(X'X + dI) \quad (2.50)$$

$$E(\hat{\beta}_d) = E(A\hat{\beta}) = AE(\hat{\beta}) = A\beta \quad (2.51)$$

Denklem 2.51'deki eşitlikte Liu tahmin edicisinin beklenen değeri ifade edilmeye çalışılmıştır. Denklem 2.48'deki ifadeye 2.49'deki eşitlikte yazılmıştır. $\hat{\beta}$ parantezine alınmıştır. Ortaya çıkan denklem 2.50'deki ifadeye A denmiştir. Liu tahmin edicisinin beklenen değeri hesaplanmıştır.

$$yan(\hat{\beta}_d) = E(\hat{\beta}_d) - \beta \quad (2.52)$$

$$yan(\hat{\beta}_d) = (X'X + I)^{-1}(X'X + dI)\beta - \beta \quad (2.53)$$

$$yan(\hat{\beta}_d) = [(X'X + I)^{-1}(X'X + dI) - I]\beta \quad (2.54)$$

Denklem 2.52'deki yan denklemine Liu kestiricisi yazılmıştır. Denklem 2.53'deki eşitlikte açık hali gösterilmeye çalışılmıştır. Denklem 2.54'de β parantezine almıştır ve yan hesaplanmıştır.

Liu kestiricisinde hedef diğer yanlı kestiricilerdeki gibi olması beklenen az yan karşılığında parametre kestiricilerinin varyansını azaltmaktır. Liu kestiricisini varyansını;

$$Var((X'X + I)^{-1}(X'X + dI)\beta) \quad (2.55)$$

Denklem 2.55'deki eşitlikteki gibi varyansı bulmak için Liu kestiricisi yazılmıştır.

$$((X'X+I)^{-1}(X'X+dI))Var(\beta)((X'X+I)^{-1}(X'X+dI))' \quad (2.56)$$

Denklem 2.56'daki eşitlikte varyans kuralı uygulamıştır.

$$Var(\hat{\beta}_d) = \sigma^2 (X'X + I)^{-1} (X'X + dI) (X'X)^{-1} (X'X + dI) (X'X + I)^{-1} \quad (2.57)$$

Liu tahmin edicisinin varyansı elde edilmiştir.

Liu(1993), EKK kestiricisinden daha küçük HKO sahip $0 < d < 1$ arasında bir d sabiti elde edilebileceğini göstermiştir.

Akdeniz vd.(2006), optimal d parametresi ifade etmişlerdir. $X'X = I$ olması haliyle denklem 2.48'den elde edilmiş Liu kestiricisi,

$$\hat{\beta}_d = \frac{1}{2}(1+d)X'y \quad (2.58)$$

biçiminde elde edilmiştir ve burada d *yanlılık* parametresi,

$$d_{opt} = 1 - \frac{2p\sigma^2}{p\sigma^2 + \beta'\beta} \quad (2.59)$$

ifade edilmiş olur.

Bilinmeyen parametre kestirimleri yerine yazılırsa d *yanlılık* parametresi,

$$\hat{d}_{opt} = 1 - 2p \frac{\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \quad (2.60)$$

şeklinde yazılmıştır. d_{opt} parametresi için önerilen bir üst sınır ifade edilmek istenirse,

$$d^* = 1 - \left(\frac{\lambda_{\min}}{\lambda_{\max}}\right)^2 p \frac{\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \quad (2.61)$$

olmak üzere,

$$\hat{d}_{opt} < 1 - \left(\frac{\lambda_{\min}}{\lambda_{\max}}\right)^2 p \frac{\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} = d^* \quad (2.62)$$

biçimde elde edilmiştir.

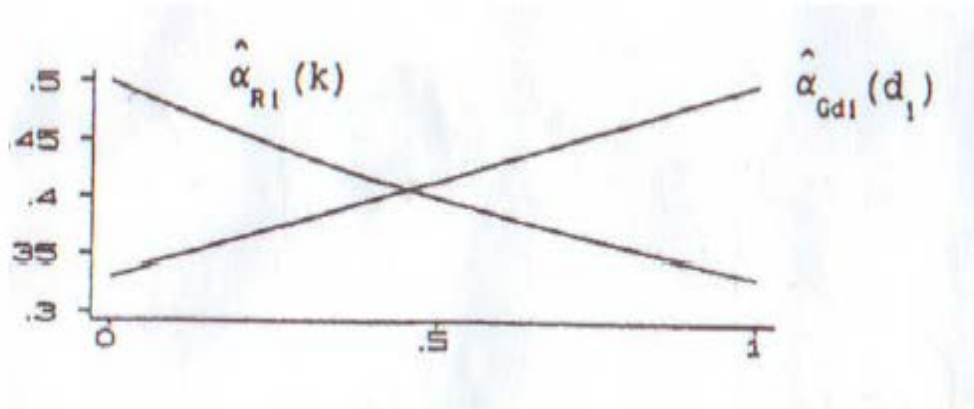
Denklem 2.62'deki eşitsizlik, $p > 2$ olması durumunda,

$$E(\hat{d}_{opt}) < E(d^*) = 1 - \left(\frac{\lambda_{\min}}{\lambda_{\max}}\right)^2 \frac{p}{p-2} \quad (2.63)$$

biçiminde bulunmuştur.

$$d_{opt} = \frac{\sum_{j=1}^p \left[\frac{\alpha_j^2 - \sigma^2}{(\lambda_j + 1)^2} \right]}{\sum_{j=1}^p \left[\frac{\sigma^2 + \lambda_j \alpha_j^2}{\lambda_j (\lambda_j + 1)^2} \right]} \quad (2.64)$$

Denklem 2.64'deki eşitliği Liu tarafından 1993 yılında önerdiği bir eşitliktir.



Şekil 2.3 Ortak k ve d

Optimal k değerinin sıfıra yakın seçilmesi ridge regresyonu konu alan kaynaklarda çokça olarak uygun görülmüştür. Fakat şekil 2.3'de görüleceği gibi ridge kestiricisi ile liu kestiricisinin eşitliğini ortaya koyacak ortak bir k ve d yanlılık parametresi değeri saptandığı görülmüştür.

3

UYGULAMA

3.1 Model ve Veri Seti

Veriler TCMB sitesinden alınmıştır. 2003-2020 yıllarına ait çeyreklik verileri kullanılmıştır. Modelde açıklanan değişkenler ise ihracat, açıklayıcı TÜFE, pasif toplamı, birikimli Yİ-ÜFE endeksi kullanılmıştır

logihr: Logaritma ihracat

yiüfe: Yurtiçi üretim endeksi

tüfe: Tüketici fiyat endeksi

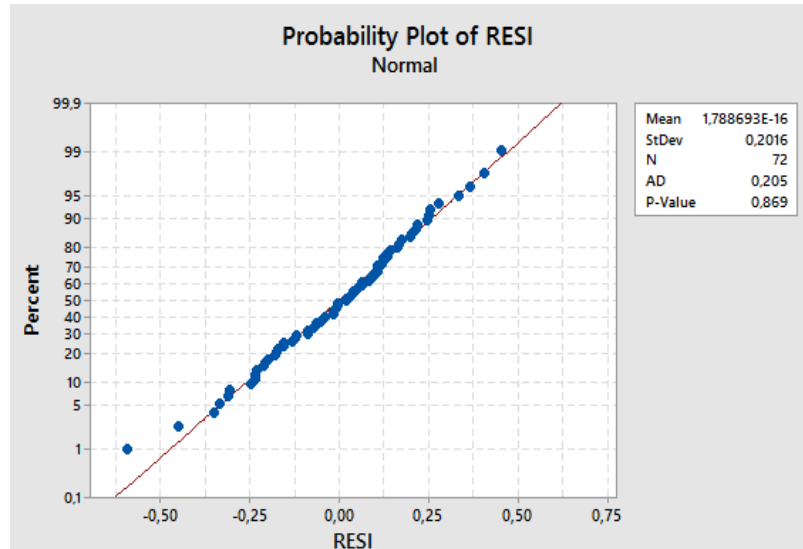
pasif: TCMB pasif yükümlülükleri

Standartlaştırılarak elde edilen model:

$$\text{logihr} = \beta_0 + \beta_1 \text{tüfe} + \beta_2 \text{yiüfe} + \beta_3 \text{pasif}$$

3.2 Ampirik Bulgular

Tablo 3.1 Normal dağılıma uyma tablosu



Regresyon modeli Minitab 18 programında kurulmuştur. Residuals'lerin Anderson-Darling testine göre tablodaki p değeri göz önünde bulundurulduğunda normal dağılımına uyduğu görülmüştür.

Tablo 3.2 Model anlamlılığı ve vif

ANOVA ^a					
Model		Sum of Squares	df	Mean Square	F
1	Regression	68,115	3	22,705	535,139
	Residual	2,885	68	,042	
	Total	71,000	71		
					Sig.
					,000 ^b

a. Dependent Variable: logihr

b. Predictors: (Constant), yiüfe, tüfe, pasif

Coefficients ^a								
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-2,656E-007	,024		,000	1,000		
	tüfe	-,050	,051	-,050	-,982	,330	,234	4,279
	pasif	-,837	,143	-,837	-5,852	,000	,029	34,232
	yiüfe	1,749	,137	1,749	12,809	,000	,032	31,214

a. Dependent Variable: logihr

SPSS 21 programının çıktısında modelin anlamlı olduğu görülmüştür. Bazı katsayıların anlamsız olduğu bazı katsayıların anlamlı olduğu görülmüştür. Çoklu doğrusal bağlantı probleminden şüphelenmiştir. VİF bu sonucu doğrular niteliktedir.

$$k_{lw} = \frac{p \sigma^2}{\sum_{j=1}^p \lambda_j \alpha_j^2} \quad (3.1)$$

Denklem 3.1'deki eşitlikte Lawless ve Wang'ın 1976 yılında k yanlılık parametresi için önerdiği bir eşitliktir. R 4.1.1 sürümünde tablo 3.3'deki değer hesaplanmıştır.

Tablo 3.3 R Ridge k çıktısı

LW (lm.ridge) 0.04420

$$d_{opt} = \frac{\sum_{j=1}^p \left[\frac{\alpha_j^2 - \sigma^2}{(\lambda_j + 1)^2} \right]}{\sum_{j=1}^p \left[\frac{\sigma^2 + \lambda_j \alpha_j^2}{\lambda_j (\lambda_j + 1)^2} \right]} \quad (3.2)$$

Denklem 3.2'deki eşitlikte Liu 1993 yılında d yanlılık parametresi için önerdiği bir eşitliktir. R 4.1.1 sürümünde tablo 3.4'deki değer hesaplanmıştır.

Tablo 3.4 R Liu d çıktısı

```
dopt 0.97582
```

Tablo 3.5 R Ridge k değerlerine göre çıktı

```
      K=0      K=0.05      K=0.1      K=0.15      K=0.2      K=0.25
tüfe -0.04964905 -0.0477060 -0.04602021 -0.04456005 -0.04329869 -0.04221324
pasif -0.83696447 -0.7814974 -0.73063243 -0.68382494 -0.64061306 -0.60060237
yiüfe 1.74940920 1.6953197 1.64560601 1.59975164 1.55731832 1.51793183
```

Ridge regresyon yanlılık parametresi k değerlerine göre regresyon katsayıları R 4.1.1 sürümünde tablo 3.5'deki değerler hesaplanmıştır. Lawless ve Wang'ın 1976 yılında önerdiği eşitlikle hesaplanan k yanlılık parametresi k=0.04420 olarak hesaplanmıştır. Yaklaşık olarak k yanlılık parametresi 0.05 alınmıştır.

Tablo 3.6 R Liu d değerline göre çıktı

```
      d=0.9      d=0.925      d=0.95      d=0.975      d=1
tüfe -0.0484963 -0.04878449 -0.04907267 -0.04936086 -0.04964905
pasif -0.7762979 -0.79146457 -0.80663120 -0.82179784 -0.83696447
yiüfe 1.6891020 1.70417877 1.71925558 1.73433239 1.74940920
```

Liu regresyon yanlılık parametresi d değerlerine göre regresyon katsayıları R 4.1.1 sürümünde tablo 3.6'daki değerler hesaplanmıştır. Liu'nun 1993 yılında önerdiği eşitlikle hesaplanan d yanlılık parametresi dopt=0.97582 olarak hesaplanmıştır. Yaklaşık olarak d yanlılık parametresi 0.975 alınmıştır. Ancak bu tablolar incelendiğinde ridge kestiricisi ile liu kestiricisinin eşitliğini ortaya koyacak sağlayacak ortak bir k ve d değeri saptanabileceği görülmüştür.

Tablo 3.7 R Ridge k değerlerine göre R^2 çıktısı

	Variance	Bias ²	MSE	rsigma2	F	R2	adj-R2	CN
K=0	0.0411	0.0000	0.0411	0.0418	543.0078	0.9594	0.9582	178.6255
K=0.0225	0.0396	0.0013	0.0409	0.0418	542.7420	0.9570	0.9558	175.1246
K=0.045	0.0383	0.0049	0.0432	0.0419	541.9830	0.9548	0.9535	171.7591

Ridge regresyon yanlılık parametresi k değerlerine göre düzeltilmiş R^2 değerleri 4.1.1 sürümünde tablo 3.7'deki değerler hesaplanmıştır. Lawless ve Wang'ın 1976 yılında önerdiği eşitlikle hesaplanan $k=0.0442$ olarak hesaplanmıştır. Yaklaşık olarak k yanlılık parametresi 0.045 alınmıştır. Ridge regresyon $k=0.045$ yanlılık parametresi için düzeltilmiş $R^2=0.9535$ değeri bulunmuştur.

Tablo 3.8 R Liu d değerlerine göre R^2 çıktısı

	EDF	Sigma2	CL	VAR	Bias ²	MSE	F	R2	adj-R2
d=0.9	69.0023	0.0419	4.8915	0.0377	0.0073	0.0450	541.4658	0.9592	0.9574
d=0.925	69.0013	0.0419	4.9182	0.0385	0.0041	0.0426	542.1396	0.9593	0.9575
d=0.95	69.0006	0.0418	4.9452	0.0393	0.0018	0.0412	542.6219	0.9593	0.9575
d=0.975	69.0001	0.0418	4.9724	0.0402	0.0005	0.0406	542.9117	0.9594	0.9576

Liu regresyon yanlılık parametresi d değerlerine göre düzeltilmiş R^2 değerleri 4.1.1 sürümünde tablo 3.8'deki değerler hesaplanmıştır. Liu'nun 1993 yılında önerdiği eşitlikte hesaplanan d yanlılık parametresi $dopt=0.97582$ olarak hesaplanmıştır. Yaklaşık olarak d yanlılık parametresi 0.975 alınmıştır. Liu regresyon $d=0.975$ yanlılık parametresi için düzeltilmiş $R^2=0.9576$ değeri bulunmuştur. Bu sonuç itibarıyla Liu tahmincisini üstün kılan bir sonuçtur. Çünkü Liu tahmincisinin düzeltilmiş R^2 değeri Ridge tahmincisinin düzeltilmiş R^2 değerinden büyüktür.

Tablo 3.9 R Ridge k değerlerine göre vif çıktı

	tüfe	pasif	yiüfe
k=0	4.27852	34.23230	31.21403
k=0.0225	3.28486	6.37817	6.06647
k=0.045	2.64213	2.84232	2.82215

Marquardt ve Snee, 1975 yılında ridge yanlılık parametresinin k'nın saptanması için VBF değerlerinin kullanılmasını söylemişlerdir. VBF değeri k değerine 1 ile

10 arasında karşılık gelecek şekilde, uygun k değeri olarak saptanabileceğini belirtmiştir. Genel olarak, 4'ün üzerindeki bir VBF çoklu doğrusallığın var olabileceğini ve daha fazla araştırmanın daha iyi olduğunu göstermiştir. Yukarıdaki k yanlılık parametresine göre hesaplanmış VBF değerleri göz önünde bulunduğu buradan da elde edeceğimiz sonuç $k=0.045$ bulunmuştur.

Tablo 3.10 R EKK AIC, BIC çıktı

```
> AIC(mod)
[1] -17.30403
> BIC(mod)
[1] -5.920696
```

Tablo 3.10'daki çıktı EKK yönteminin AIC, BIC değerine aittir.

Tablo 3.11 EViews AIC, Durbin-Watson çıktı

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.056952	0.037523	1.517776	0.1344
SERIES01(-1)	0.898634	0.119768	7.503108	0.0000
SERIES01(-2)	-0.671113	0.153011	-4.386043	0.0000
SERIES01(-3)	0.715039	0.138144	5.176037	0.0000
SERIES02(-2)	-0.048474	0.043509	-1.114094	0.2698
SERIES02(-1)	-0.017982	0.046159	-0.389558	0.6983
SERIES03(-1)	0.326850	0.236226	1.383634	0.1717
SERIES03(-2)	-0.457115	0.220285	-2.075106	0.0423
SERIES04(-1)	0.221716	0.539723	0.410797	0.6827
SERIES04(-2)	-0.118869	0.595175	-0.199721	0.8424
R-squared	0.984843	Mean dependent var		0.068761
Adjusted R-squared	0.982531	S.D. dependent var		0.963601
S.E. of regression	0.127359	Akaike info criterion		-1.150331
Sum squared resid	0.956996	Schwarz criterion		-0.826547
Log likelihood	49.68642	Hannan-Quinn criter.		-1.021875
F-statistic	425.9616	Durbin-Watson stat		2.072187
Prob(F-statistic)	0.000000			

Tablo 3.11'deki çıktı zaman serisi modeline aittir. Var analizi göz önünde bulundurularak fark alma işlemleriyle otokorelasyon problem ortadan kaldırılmıştır. AIC değerinin küçük olması istenir. Yukarıdaki AIC değerleri göz önünde bulunduğu buradan da elde edeceğimiz sonuç zaman serisi modeli EKK yöntemi daha kötü olduğu görülmüştür.

Tablo 3.12 R Ridge AIC, BIC çıktı

	AIC	BIC
K=0	-225.6311	89.11891
K=0.0225	-200.2411	112.88419
K=0.045	-187.9036	124.64517

Tablo 3.12'deki çıktı AIC değeri Ridge regresyon yöntemine aittir. AIC değerinin küçük olması istenir. Yukarıdaki k yanlılık parametresine göre hesaplanmış AIC değerleri göz önünde bulunduğunda buradan da elde edeceğimiz sonuç Ridge yönteminden EKK yöntemi daha iyi olduğu görülmüştür.

Tablo 3.13 R Liu AIC, BIC çıktı

	AIC	BIC
d=0.9	-225.5348	89.08910
d=0.925	-225.5976	89.05774
d=0.95	-225.6347	89.05220
d=0.975	-225.6459	89.07253
d=1	-225.6312	89.11878

Tablo 3.13'deki çıktı AIC değeri Liu regresyon yöntemine aittir. AIC değerinin küçük olması istenir. Yukarıdaki d yanlılık parametresine göre hesaplanmış AIC değerleri göz önünde bulunduğunda buradan da elde edeceğimiz sonuç Ridge yönteminden Liu regresyon yöntemi daha iyi olduğu görülmüştür.

Anlatıdan da görüleceği gibi d yanlılık parametresinin seçimi k yanlılık parametresine göre daha kolaydır..Liu kestircisini öne çıkaran özelliklerden biridir.

Teorik olarak yerine yazıldığında k ve d yanlılık parametresi $k=0$, $d=1$ değerleri EKK kestircisine eşittirler.

Uygulama sonuçları da göz önünde bulundurulduğunda ortak bir d ve k yanlılık parametresi kestirici değeri elde edilebilir.

Ridge kestircisinin EKK metoduna göre üstünlüğü daha az standart hatalı kestirimler ortaya çıkarmasıdır. Ridge kestircisinin standart hatalarının hesaplanması için kaynaklarda yer alan formülasyon yoktur.

Seçilecek k yanlılık parametresi değeri yanlılığın artan bir fonksiyonu olması sebebiyle, k değeri büyüdükçe parametreler daha fazla yanlı kestirilecektir.

Ridge kestircisinin EKK kestircisine, liu kestircisiyse hem EKK hem de ridge kestircisine göre etkileyici sonuçlar vereceği ortaya konmuştur.

Tablo 4.1 R Liu katsayı çıktı

```

                                d=0.975
tüfe    -0.04936086
pasif   -0.82179784
yiüfe    1.73433239

```

Liu regresyon kestirimleri tablo 4.1'deki gibidir. Açıklanan değişken ihracattır. TÜFE artarsa, iç talebin artmış olduğu dolayısıyla ihracattın azalacağı sonucuna varılmıştır. ÜFE artarsa, devalüasyon etkisi ile dış talep artmış olduğu dolayısıyla ihracattın artacağı sonucuna varılmıştır. TCMB pasif yükümlülükleri azalırsa, döviz yükümlülükleri azalmış olduğu dolayısıyla ihracattın artacağı sonucuna varılmıştır.

- Aktaş, C. (2007), Çoklu bağıntı ve liu kestiriciyle enflasyon modeli için bir uygulama, Uluslararası Yönetim İktisat ve İşletme Dergisi, 3(6),67-80.
- Aktaş, C ve Yılmaz, V. Mohan, Textbook of pathology. Jaypee Brothers, Medical Publishers Pvt. Limited, 2018.
- Albayrak, A. S. (2005), Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama, Uluslararası Yönetim İktisat ve İşletme Dergisi,1(1),105-126.
- Girginer, N. (2015), Ridge regresyon yöntemiyle Türk tekstil ve konfeksiyon sanayi dışsatımının analizi: (1997-1994)(Doctoral dissertation, Anadolu Üniversitesi.
- Göktaş, A. ve Sevinç, V. (2016), Two new ridge parameters and a guide for selecting an appropriate ridge parameter in linear regression, Gazi University Journal of Science, 29(1), 201-211.
- Hines, W. W. , Montgomery, D. C. and Borror (2008), Probability and statistics in engineering, John Wiley and Sons.
- Hoerl, A. E. And Kennard, R. W. (1970),Ridge regression: applications to nonorthogonal problems, Technometrics, 12(1), 69-82.
- JF, L. (1976) , Sırt ve diğer regresyon tahmin edicilerinin simülasyonu, İstatistik Teori ve Yöntemlerde İletişim, 5(4), 307-323.
- Karakaş, S. (2008), Çoklu doğrusal bağlantı problemi ve yanlı regresyon tahmincileri, Yüksek Lisans Tezi, İstanbul Üniversitesi.
- Marquardt, D. W. And Snee, R. D. (1975), Ridge regression in pratice, American Statistician, 29(1), 3-20
- McDonald, G. C. and Galarneu, D. I. (1975), A Monte Carlo evaluation of some ridge type estimators, Journal of the American Statistical Association, 70(350), 407-416.
- Özgül, V. U. P. A. ve Alma Ö. G. , Regresyon analizinde kullanılan en küçük kareler ve en küçük medyan kareler yöntemlerinin karşılaştırılması, Süleyman Demirel Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi, 3(2), 219-229.

TEZDEN ÜRETİLMİŞ YAYINLAR

Kongre Bildirileri

1. Gökçe S., Evren A., Çoklu Bağıntı ve Ridge, Liu, Zaman Serisi Kestiricisiyle Türkiye'nin İhracat Modeli İçin Bir Uygulama, 13.Uluslararası Bilimsel Araştırmalar Kongresi, 73-81.

