

**UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

Berrin GÜLTAY

YÜKSEK LİSANS TEZİ

OKLU İ İLİŐKİ VE EKOLOJİK REGRESYON

İSTATİSTİK ANABİLİM DALI

ADANA, 2009

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÇOKLU İÇ İLİŞKİ VE EKOLOJİK REGRESYON

Berrin GÜLTAY

YÜKSEK LİSANS TEZİ

İSTATİSTİK ANABİLİM DALI

Bu tez 29/06/2009 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği/
Oyçokluğu İle Kabul Edilmiştir.

İmza.....
Prof.Dr. Selahattin KAÇIRANLAR
DANIŞMAN

İmza.....
Prof.Dr. Hamza EROL
ÜYE

İmza.....
Yard.Doç.Dr.Ebru GÜLER
ÜYE

Bu tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.
Kod No:

Prof. Dr. Aziz ERTUNÇ
Enstitü Müdürü
İmza ve Mühür

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

ÖZ
YÜKSEK LİSANS TEZİ

ÇOKLU İÇ İLİŞKİ VE EKOLOJİK REGRESYON

Berrin GÜLTAY

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

Danışman : Prof. Dr. Selahattin KAÇIRANLAR

Yıl : 2009, Sayfa: 89

Jüri : Prof. Dr. Selahattin KAÇIRANLAR

Prof. Dr. Hamza EROL

Yrd. Doç. Ebru GÜLER

Doğrusal modellerden yararlanarak, toplanmış seçim verilerinin analizi politik bilimlerde geniş bir ilgi görmüştür. Toplanmış verilerden bireysel davranışlar üzerine sonuç çıkarmak ekolojik regresyon yönteminin amacıdır. Bu çalışmada seçmenlerin geçiş olasılıklarını tahmin etmek için Goodman' ın klasik ekolojik regresyon yöntemi geliştirilmiştir. Bunu yaparken, Ridge regresyon yöntemi birkaç adım daha ilerletilerek kısıtlı modified genelleştirilmiş Ridge tahmin edici elde edilmiştir. Bu tahmin yöntemini gösterebilmek için İsveç, Çek Cumhuriyeti ve Türkiye' deki seçim bölgelerinde birbiri ardına gelen iki seçim sonuçları kullanılmıştır.

Anahtar Kelimeler: Geçiş Olasılıkları, Toplanmış Veri, Ridge Regresyon, EKK Yöntemi, SURE Model.

ABSTRACT

MSc. THESIS

MULTICOLLINEARITY AND ECOLOGICAL REGRESSION

Berrin GÜLTAY

DEPARTMENT OF STATISTICS
INSTITUTE OF NATURAL AND APPLIED SCIENCES
UNIVERSITY OF CUKUROVA

Supervisor : Prof. Dr. Selahattin KAÇIRANLAR
Year : 2009, Pages:89
Jury : Prof. Dr. Selahattin KAÇIRANLAR
Prof. Dr. Hamza EROL
Asist. Prof. Ebru GÜLER

The use of the linear models in aggregate voting has now received widespread attention in political science. The aim of ecological regression technique, the use of aggregate data to describe behaviour at the individual – level. In this thesis Goodman’s classical model of ecological regression technique is developed to estimate voters transitions probabilities. In doing this, ridge regression technique further developed steps and have restricted modified generalized ridge estimator. To illustrate proposed estimation technique, results of two consecutive elections in various electrol districts used for Swedish, Czech Republic and Turkey.

KeyWords: Transition Probabilities, Aggregated Data, Ridge Regression, OLS Method, SURE Model.

TEŞEKKÜR

Bugüne kadar tüm eğitim hayatım boyunca başarılı olmanın en büyük kaynağının sabır ve azim olduğunu öğrendim. Fakat bunların kendi başına yeterli olmadığını da gördüm. Eksiklikleri tamamlayan ise her daim size inanan ve destekleyen insanları varlığı olduğunu biliyorum.

Tüm lisans ve yüksek lisans öğrenimim boyunca attığım her adımda güvenini ve desteğini hiçbir zaman esirgemeyen başta sayın danışmanım Prof. Dr. Selahattin KAÇIRANLAR olmak üzere tüm bölüm öğretim üyelerine sonsuz saygı, sevgi ve teşekkürlerimi sunarım.

Sadece seslerini duymanın bile verdiği o güç için beni bu günlere taşıyan sevgili aileme tüm destekleri için sonsuz teşekkürler.

Tüm tez çalışmam boyunca evdeki huzuru ve mutluluğu sağlayarak bana destek olan sevgili eşim Mehmet GÜLTAY' a teşekkür ederim.

Ayrıca benim için çok değerli olan tüm BOZKURT ailesine sonsuz saygı ve teşekkürlerimi bir borç bilirim.

İÇİNDEKİLER

SAYFA

ÖZ	I
ABSTRACT	II
TEŞEKKÜR	III
İÇİNDEKİLER.....	IV
TABLolar DİZİNİ	VI
ŞEKİLLER DİZİNİ	VII
SİMGELER VE KISALTMALAR	VIII
1. GİRİŞ	1
2. EKOLOJİK OLARAK SONUÇ ÇIKARMA VE KULLANILAN BAZI YÖNTEMLER.....	3
2.1. Ekolojik Olarak Sonuç Çıkarmak Nedir?	3
2.2. Toplanmış Verinin Kullanım Amacı	4
2.3. Ekolojik Olarak Sonuç Çıkarmanın Gerekliliği.....	5
2.4. Goodman Ekolojik Regresyon Yöntemi.....	8
2.4.1. Tablonun 2 x 2 Olması Durumu	11
2.4.2. Tablonun n x 2 Olması Durumu	12
2.5. Achen ve Shively' nin Karesel Modeli	19
2.6. Duncan ve Davis Sınırlar Metodu.....	22
2.7. King EI Yöntemi	25
3. EKOLOJİK REGRESYON VE ALTERNATİF ÇÖZÜM YÖNTEMLERİ	29
3.1. EKK Yöntemi ve Ekolojik Regresyon.....	29
3.1.1. EKK Yöntemi Varsayımları	31
3.1.2. EKK Tahmin Edicisinin Özellikleri.....	33
3.1.3. EKK Yönteminin Ekolojik Regresyondaki Sonuçları	34
3.2. Ridge Regresyon Tahmin Edicisi.....	35
3.2.1. k Yanlılık Parametresinin Seçimi	39
3.2.2. k Yanlılık Parametresinin Seçimi İle İlgili Bir Uygulama	40

3.3. Görünüşte İlişkisiz Regresyon Modeli (SURE Model).....	45
3.3.1. SURE Model ve Varsayımları	46
3.3.2. SURE Modelde Ridge Tahmin Edici	49
3.4. Genelleştirilmiş Ridge Tahmin Edici	49
3.5. Modified Ridge Tahmin Edici	52
3.5.1. SURE Modelde Modified Ridge Tahmin Edici (MR)	53
3.5.2. SURE Modelde Modified Genelleştirilmiş Ridge Tahmin Edici (MGR)	54
3.5.3. MGR Tahmin Edicinin Uygulama Şekli	56
3.6. Kısıtlı Ridge Tahmin Edici (RRR)	58
3.6.1. SURE Modelde Kısıtlı Modified Ridge Tahmin Edici	60
4. EKOLOJİK REGRESYON UYGULAMALARI	62
4.1. İsveç Seçimlerinin Geçiş Olasılıkları Tahmini.....	63
4.2. Çek Cumhuriyeti 1992-1996 Seçimlerinin Geçiş Olasılıkları.....	67
4.3. Türkiye’deki 2002-2007 Genel Seçimleri Geçiş Olasılıkları	70
5. SONUÇ VE ÖNERİLER	76
KAYNAKLAR.....	77
ÖZGEÇMİŞ.....	82
EKLER	83

TABLÖLER DİZİNİ**SAYFA**

Tablo 2.1. Parametrelere ilişkin çapraz tablolar.....	8
Tablo 2.2. Hücrelerin beklenen değerleri.....	9
Tablo 2.3. Hücrelerin toplam beklenen değerleri.....	9
Tablo 2.4. İki seçimde oy verenlerin sayısı.....	11
Tablo 2.5. Birbiri ardına gelen iki seçimde seçmenlerin geçişleri.....	12
Tablo 2.6. Geçiş oylarının katsayıları.....	13
Tablo 2.7. Örnek seçim tablosu.....	13
Tablo 2.8. Örnek veri tablosu.....	14
Tablo 2.9. Birbiri ardına gelen iki seçimin sonuçları.....	16
Tablo 2.10. EKK ile tahmin edilmiş geçiş olasılıkları.....	19
Tablo 2.11. Sınırlar Metodu için örnek tablo.....	22
Tablo 2.12. Örnek veri kümesi.....	24
Tablo 3.1. Hald çimento verisi.....	41
Tablo 3.2. k değerlerine karşılık gelen $\hat{\beta}_R$ katsayılarının değerleri.....	43
Tablo 3.3. SPSS programı ile elde edilen varyans analizi tablosu.....	44
Tablo 3.4. SPSS programı ile elde edilen katsayılar tablosu.....	44
Tablo 4.1. İsveç seçimlerindeki partiler.....	63
Tablo 4.2. EKK ile tahmin edilmiş geçiş olasılıkları (yüzde olarak).....	64
Tablo 4.3. Kısıtlı MGR ($\hat{\beta}_{MGR}^{(r)}$) ile tahmin edilen geçiş olasılıkları (yüzde olarak).....	66
Tablo 4.4. Çek Cumhuriyetinde 1992' de oylanan partiler.....	67
Tablo 4.5. Çek Cumhuriyetinde 1996' da oylanan partiler.....	68
Tablo 4.6. EKK yöntemi ile tahmin edilen geçiş olasılıkları.....	68
Tablo 4.7. RMGR ile tahmin edilen geçiş olasılıkları.....	69
Tablo 4.8. 3 Kasım 2002 genel seçimine katılan partiler.....	70
Tablo 4.9. 22 Temmuz 2007 genel seçimine katılan partiler.....	71
Tablo 4.10. Analiz edilecek partiler.....	72
Tablo 4.11. Geçiş olasılıklarının EKK tahminleri.....	73
Tablo 4.12. $\hat{\beta}_{MGR}^{(r)}$ ile tahmin edilmiş geçiş olasılıkları.....	75

ŞEKİLLER DİZİNİ

SAYFA

Şekil 2.1. Aynı veriye ait iki görünüm.....	27
Şekil 3.1. Negatif ilişkili regresyon katsayılarının izi.....	39
Şekil 3.2. Ridge izi 1. Hald çimento verisine ait birRidge izi.....	42
Şekil 3.3. Ridge izi 2. Aralıklar azaltılarak oluşturulmuş bir Ridge izi	42

SİMGELER VE KISALTMALAR

EKK	: En Küçük Kareler
GR	: Genelleştirilmiş Ridge Tahmin Edici (Generalized Ridge Estimator)
MGR	: Modified Genelleştirilmiş Ridge Tahmin Edici (Modified Generalized Ridge Estimator)
MR	: Modified Ridge Tahmin Edici (Modified Ridge Estimator)
MSE	: Hata Kareler Ortalaması (Mean Square Error)
OR	: Ridge Tahmin Edici (Ordinary Ridge Estimator)
RMGR	: Kısıtlı Modified Genelleştirilmiş Ridge Tahmin Edici (Restricted Modified Generalized Ridge Estimator)
RRR	: Kısıtlı Ridge Tahmin Edici (Restricted Ridge)

1. GİRİŞ

Ekolojik olarak analiz etme yöntemi, düzenlenmiş belli bir grup için toplanmış veriyi kullanmayı kapsamaktadır. Buradaki amaç, bireysel seviyeli veriler arasında bağlantı kurmaktır. Belirlenen gruplar ya da veriler coğrafi bilgiler olarak tanımlanır. Örneğin; iller ve seçim bölgelerinden alınan toplanmış veriler coğrafi bilgiler olarak analiz etmek için sosyal bilimlerde sıklıkla kullanılır. Analizler, bireysel seviyeli verinin dışında, seçmenlerin başka bir partiye geçiş eğilimleri ve sosyal sınıfların parti seçimleri gibi konuları da kapsamaktadır. Birbiri ardına gelen seçimler arasındaki seçmen dönüşümlerinin analizi ve tahmini gibi, seçimle ilgili davranışlar hakkındaki çalışmalara, istatistikçilerin yanı sıra siyasi bilimlerde de değerli ve ilginç bir çalışma olarak önem verilmektedir. Böyle bir analiz için, bireysel seviyeli veri ya da toplanmış veriye ihtiyaç duyulmaktadır. Seçimle ilgili toplanmış veri pahalı değildir ve yayınlanmış istatistiklerden kolaylıkla elde edilebilir. Tersine, araştırmalarla elde edilen bireysel seviyeli veri pahalıdır, çoğu zaman hatalı hesaplanır ve birçoğumuzun politik geçmişi için de geçerli değildir.

İstatistiğin temel konularından biri de, toplanmış verilerin tahmin etmeye çalıştığımız parametreler hakkında ne kadar bilgiyi içerdiğidir. Bireysel seviyeli veriyi simgeleyen parametrelerin tahmini için kullanılan standart metotlardan biri olan bilinen En Küçük Kareler Yöntemi kullanıldığında genellikle istenmeyen sonuçlarla karşılaşmaktadır. Bunun nedeni En Küçük Kareler Yöntemi' nin varsayımlarının çok kısıtlayıcı olması ve büyük ihtimalle bu varsayımların sağlanmamasıdır.

Toplanmış veriden bireylerin davranışları üzerine sonuç çıkarma, bazı durumlarda ekolojik hataya götürmektedir (Robinson, 1950). Ekolojik hatanın sebep olduğu mantıksız sonuçlardan kaçınmak için, toplanmış verinin analizi, bilinen metotların yanı sıra özel koşulların da tamamlanmasına ihtiyaç duymaktadır. Bu problemi gidermek için, ilk olarak Goodman (1959) tarafından göz önüne alınan ekolojik regresyon yöntemi vardır. Ancak bu yöntemle seçmenlerin başka bir partiye geçişlerinin olasılığı tahmin edilirken, negatif olasılık gibi hassas olmayan sonuçlarla karşılaşmıştır.

Negatif tahmin probleminin giderilmesi için ilk öneriler Ridge Regresyon metodunun uygulanması olmuştur (Brown ve Payne, 1975).

Ridge regresyon yanlı bir tahmin edici olmakla birlikte literatürde birçok yanlı tahmin edici geliştirilmiştir. Fule (1995)' de bilinen ridge tahmin edici yeniden düzenlenerek, modified genelleştirilmiş ridge ve kısıtlı modified genelleştirilmiş ridge tahmin edicilerine ulaşılmıştır. Bu yöntemler, $X'X$ matrisinin köşegen elemanlarını sıfır yönünde büzmek yerine bazı ön bilgi değerleri yönünde büzmeye dayanır.

Bahsedilen yanlı tahmin ediciler, Türkiye ile birlikte üç ülkeden elde edilmiş toplanmış seçim verileri üzerinde denenmiştir. Uygulamalar kısmında da görüleceği üzere EKK tahmin edicisi ile elde edilen sonuçlar uygun değildir. Bunun yerine yanlı tahmin ediciler kullanılarak problemler ortadan kalkmıştır.

2. EKOLOJİK OLARAK SONUÇ ÇIKARMA VE KULLANILAN BAZI YÖNTEMLER

2.1. Ekolojik Olarak Sonuç Çıkarmak Nedir?

Toplumsal ilişkiler, insan ilişkileri, davranışları ve etkileşimleri neden-sonuç çerçevesinde yıllardır merak edilen ve çalışılan konular arasındadır. Eğitim, hukuk, ekonomi, siyaset, din gibi toplumsal konular hakkında yapılan çalışmalarda, analiz yapmak için kullanılan değişkenleri ekolojik değişkenler olarak adlandırabiliriz. Ekolojik olarak sonuç çıkarmak ise, toplumdaki gruplar için toplanan verileri kullanarak bireysel seviyeli ilişkiler hakkında sonuç çıkarmayı amaçlamaktadır.

Örneğin, bir ailenin gelir seviyesinin yüksek veya düşük olmasına bağlı olarak çocuklarının zayıf ya da şişman olması arasında bir ilişki var mıdır? Ya da bir ailenin gelirinin yoksulluk sınırı altında ya da üstünde olmasının kırsal veya kentsel bir alanda yaşıyor olması arasında bir ilişki var mıdır? Bu gibi soruların cevapları araştırılırken yapılan analizlerde ekolojik değişkenler ile çalışılarak sonuçlar ekolojik olarak yorumlanmaktadır.

Ekolojik olarak sonuç çıkarma yönteminin en sık uygulandığı alanlardan biri de siyasal bilimlerdir. Çeşitli bölgelerde birbiri ardına gelen, iki seçim sonucunda aynı partiye oy veren ya da iki seçimde farklı partilere oy veren, yani parti seçimini değiştiren seçmenlerin özellikleri hakkında elde ettiğimiz toplanmış seçim verilerinden yorum yapmak istediğimizi düşünelim. Genel seçim sonuçlarından açıkça bu geçişlerin olduğu görülmektedir. Bundan dolayı seçim bölgelerindeki bireysel geçişler merak edilmektedir. Bu sürecin büyüklüğünün değerlendirmesi, seçmenlerin hareketinin herhangi bir açıklaması için önemlidir. Bazı bireysel verilere ulaşılır fakat tek tek tüm bireylerin verileri elde edilemez. Bundan dolayı toplanan veriler bir bütün olarak değerlendirilmelidir. İşte bu noktada, kişisel davranışları tanımlamak için, toplanmış verileri kullanmak ekolojik olarak sonuç çıkarmanın bir problemidir.

Ekolojik olarak sonuç çıkarmada, ekolojik değişkenler ile çalışmanın en büyük amacı bireysel davranışlarla ilgili yorum yapmaktır (Goodman, 1953). Ekolojik değişkenler kullanılarak regresyon uygulama metodunun ilk adımları Goodman tarafından atılmıştır. Goodman çapraz tablolardan yararlanarak, toplanmış veriyi EKK yöntemi ile analiz etmeye çalışmıştır.

2.2. Toplanmış Verinin Kullanım Amacı

Toplum içinde bireysel ilişkileri ya da toplumdaki gruplar arası ilişkileri tanımlayan parametreleri tahmin etme yöntemleri günden güne geliştirilmektedir. Lineer ya da lojistik regresyon ya da olasılık tabloları gibi bireysel seviyeli veriler kullanılarak uygulanan yöntemler toplanmış veriler kullanıldığında genellikle yanlış tahminler üretmektedir. Bundan dolayı, ekolojik olarak analiz yapmanın amacının büyük bir kısmı, toplanmış veri kullanılarak yapılan parametre tahminlerinin yansız ya da daha az yanlış olabilmesi içindir.

Toplanmış verinin, tahmin edilmeye çalışılan parametrelerle ve bireysel ilişkiler ile ilgili ne kadar bilgiyi içerdiği, istatistiğin temel konularından biridir. Toplanmış veri ile hipotez testleri de yapılmaktadır. Parametrelerin sabitliğinin test edildiği hipotez testleri toplanmış veri kullanılarak uygulanmaktadır. Ayrıca, toplanmış veriden alınan bilgiler, herhangi bir hipotez testinin gücünün yanı sıra tahminlerin standart hatalarını da etkilemektedir.

Siyaset bilimciler tarafından sıklıkla kullanılan bireysel davranışların ölçümleri ile ilgili anket verileri her zaman mümkün olmayabilir. Bu gibi nedenlerden dolayı bireysel davranışlar hakkında yorum yapabilmek için toplanmış verilerden oldukça sık faydalanılmaktadır. Ayrıca Sosyoloji ve Ekonomi dallarıyla ilgilenen bilim adamları da toplanmış verileri kullanarak uyguladıkları yöntemleri geliştirir ya da yeni yöntemler ortaya çıkarmaktadırlar. Örneğin; ekonomi alanında çalışan bir araştırmacı belli bir bölgedeki bireylerin vergi bilgilerini ilgili yerlerden elde ederek o bölgede yaşayan ailelerin gelirleri ile ilgili bazı sonuçlara varabilir ya da ekonomistler toplanmış verileri kullanarak ev halkının harcamalarına ait regresyon modeli oluşturarak gelecek hakkında tahminlerde bulunabilirler.

Ekolojik olarak sonuç çıkarma yöntemi, bireysel ilişkiler ile ilgili gruplardan elde edilmiş toplanmış verileri kullanarak yorum yapmayı amaçlar.

Analiz için uygun veriler genellikle değişkenlerin ortalaması ya da toplamaları şeklindedir. Toplanmış veriler, seçim bölgeleri gibi coğrafik bölgelerle ilgili olduğu gibi okul ya da hastane gibi kurumlardan da elde edilmiş veriler olabilir. Oluşturulan veri kümelerinin nerden elde edildiği ya da nitelikleri veya kullanım alanları ne olursa olsun her zaman faydalıdır. Bireylerden tek tek bilgi edinmek yerine toplanmış verileri kullanmak, çok daha kolay, ucuz ve daha güvenilir bir yoldur.

2.3. Ekolojik Olarak Sonuç Çıkarmanın Gerekliliği

Bireyler hakkında bir bilgiye sahip olmamamıza rağmen, bireylerle ilgili elde edilen, toplanmış verileri kullanmak çoğu kez faydalıdır. Toplanmış verilerle çalışmanın temel avantajlarından bazıları; planlanan-üzerinde çalışılan bölgelerin politik sonuçları, makro- ekonomik durumların tahmini ve milletler arası bebek ölüm oranlarının karşılaştırılması, karşılaşılan sadece birkaç durumdan biridir.

Siyasal bilimlerin en fazla uygulama yapılan alanlarının temel soruları ekolojik olarak sonuç çıkarmayı gerektirmektedir. Diğer birçok alanlardaki araştırmacılar, tüm dünyadaki siyaset bilimi ile birlikte sürekli olarak toplanmış verilerden bireysel davranışların özellikleri hakkında yorum yapmak istemektedirler. Eğer böyle bir yorum yapabilmek için geçerli bir yöntem mümkün olursa, bilim adamları ekolojik verilerle ilgili bu sorular için kesin cevaplar sağlayabilir. Ayrıca siyasetçilerin kararları, güvenilir ve bilimsel teknikler üzerine dayanır. Ekolojik olarak sonuç çıkarmanın kullanılacağı alanların çeşitliliğini aşağıdaki şekillerde özetleyebiliriz.

- Amerika'daki kamu politikasında ekolojik olarak sonuç çıkarmak, federal yasanın anahtar özelliklerini yerine getirmek için gereklidir. Örneğin; 1965 Amerika Seçim Yasası (ve onun genişletilmiş olanları 1970, 1975, ve 1982) ırk, renk, dil kaynaklı seçim ayrımlarını yasaklamaktadır.

Eğer böyle bir ayırım bulunursa mahkemeler, politik sınırları yeniden düzenleyecek seçim yasalarını yürürlüğe koymaktadır. Irk bakımından kutuplaşmış seçimlerle ilgili anket verileri pek kullanılabilir ve güvenilir değildir. Bu sebeplerden dolayı, Seçim Yasası'nın uygulanması için, seçim ve nüfus verilerinden geçerli bir ekolojik olarak yorum yapmak gerekmektedir.

- Eylül 1930'da Alman Parlamentosunda yapılan bir seçimde, çok fazla önemli olmayan Milliyetçi Sosyalist Alman İşçilerin partisi Weimar Cumhuriyeti'nin en büyük ikinci partisi haline gelmiştir. Milliyetçi Sosyalistler yerel ve başkanlık seçimlerinde hayret verici başarılarına devam ettiler ve en son seçimlerde oyların % 37,3'ünü aldılar. Bu hayret verici ilerlemenin nasıl olduğuna dair birçok soru soruldu. Naziler için kimler oy vermişti? Nazi seçim bölgesi düşük sınıflar tarafından mı destekleniyordu, yoksa çok daha geniş kitleler tarafından da mı destekleniyordu? Hangi dindar gruplar ve işçi sınıfı Milliyetçi Sosyalistleri destekliyordu? Hangi bölgedeki partiler Nazilere karşı oylarını kaybetmişlerdi? Bu soruların cevaplarını, 1200 tane bölgeden alınan seçim ve nüfus verilerinden oluşan toplanmış verilerden elde etmek mümkündür. Bu önemli sorular yalnızca ekolojik olarak sonuç çıkarma yöntemiyle tam ve kesin olarak cevaplandırılabilir. (Hamilton, 1982).

- Salgın hastalıklar bilimi olan Epidemiyoloji üzerine çalışan bilim adamlarının, radyoaktif radonun hangi seviyesinde, akciğer kanseri riski taşıyıp taşımadığını bilmeleri gerekmektedir. (Stidley ve Samet, 1993; Greenland ve Robins, 1994a). Sigaradan sonra en önemli ikinci akciğer kanseri nedeni olarak kabul edilen radon, renksiz, kokusuz ve radyoaktif bir gazdır. Topraktaki radon, yukarı doğru hareket eder ve evlerin tabanlarındaki ya da duvarlarındaki çatlaklardan ve boşluklardan içeri sızar. Özellikle bodrum katlarda ve madenlerde radon gazı yoğun olarak birikebilir. Evlerin içine sızıntı yapan radon çok ciddi sağlık sorunlarına neden olabilir. Araştırmacılar tarafından birçok ülkeden toplanan geniş kapsamlı verilerin analizini yapabilecek yöntem ekolojik olarak sonuç çıkarmadır.

- Pazarlamanın akademik alanında, araştırmacılar, belli ürünleri kimin satın aldığını öğrenmek ve tüketicilerin daha çok satın almaları için uygulanan reklamların nerelerde etkili olduğunu bulmak için çalışmaktadırlar.

Birçok durumda araştırmacılar belli ürünleri alan bireylerin sosyal-ekonomik durumlarına ait ve araştırmacının sorularına direk olarak cevap verecek verilere sahip değildirler. Bunun yerine o bölgedeki bireylerin karakterlerini ortalama olarak belirleyebilecek geniş kapsamlı veriler kullanırlar. Araştırmacılar, belli bir ürünün, o bölgede ilgili firmada ne kadar satıldığına ait verilere sahiptirler. Buradaki soru, her bir bölge için satılan ürünlerin miktarı neye göre değişmektedir? Örneğin; ailelerdeki çocuk sayısı, gelirin yüksek olması, tek ebeveynli aileler için talep miktarları nasıl değişmektedir? Araştırmacılar aradıkları cevapları, tüm alanlarda geçerli olan ekolojik olarak sonuç çıkarma yöntemiyle bulabilirler. Bu bilgiler altında, bilim adamları ürün talebinin bireysel özelliklere ya da aile yapılarına bağlı olup olmadığını bulabilirler ve üretici firmalar ürünlerini kullanan müşterilere nasıl reklam uygulayacaklarına karar verebilirler.

- Tarihçiler modern inceleme ve araştırma teknikleri icat edilmeden önceden de belli grupların politik tercihlerini merak etmektedirler. Örneğin; bilim adamları sadece ekolojik olarak sonuç çıkarma yöntemiyle, işçi sınıfındaki seçmenlerinin, Sosyalist partiyi desteklediği alanı bulabilirler.

- Önemli sosyal sorunlardan biride işsizlik ve suç oranları arasındaki ilişkidir. Uygun veriler genellikle, şehirler ya da ülkeler bazında oluşturulan toplanmış verilerden elde edilmektedir. Kişisel düşüncelere dayanan anketlerden elde edilen bireysel seviyeli veriler genellikle yetersiz ve doğru oldukları şüphelidir. İşsizlik ve suç oranları arasındaki kritik bağlantıyı bulabilmek için ya çok iyi elde edilmiş veriler ya da ekolojik olarak sonuç çıkarma yöntemi kullanılabilir.

- Ekolojik olarak sonuç çıkarma yönteminin kullanıldığı alanlardan bir diğeri ise ekonomidir. Ekonomistler çok uzun süreden beri makroekonomik ve mikroekonomik davranışlar arasındaki bağlantıyı bulabilmek için birçok çalışma yapmaktadırlar. Ekolojik olarak sonuç çıkarma yöntemi ekonomistlerin teorilerinin geçerli olup olmadıklarını değerlendirmelerini sağlamaktadır.

- Eğitim politikasındaki tartışmalı konulardan biride kolejler ve devlet okullarıdır. Devlet veya belediyeler tarafından özel bir bursla kolejlere öğrenciler gönderilmektedir. Böylece kolejler, varlıklı ailelerin çocukları ve burs ile devam eden öğrenciler ile karma bir hale gelmiştir. Araştırmacılar tüm okuldan öğrencilerin durumlarına ait veriler toplamaktadırlar. Acaba hangi öğrenci tipi üniversiteyi kazanabilecektir? Okullardan elde edilen toplanmış veri sayesinde ekolojik olarak sonuç çıkarma metodunu uygulayabilirler.

Üzerinde durulan bu noktalar, ekolojik olarak sonuç çıkarma metodunun çok çeşitli sorular üzerinde uygulanabilir olduğunu göstermektedir. Sadece toplanmış verilere sahip olunan durumlarda ekolojik olarak sonuç çıkarma yöntemi oldukça kullanışlı bir yoldur.

2.4. Goodman Ekolojik Regresyon Yöntemi

Belli bir bölgedeki bir toplumu, bir kategoriye göre N ve W gibi ayrık iki gruba, başka bir kategoriye göre I ve L gibi ayrık iki gruba böldüğümüz varsayılmaktadır. Bu toplumdaki bir bireyin, I ve N gruplarında aynı anda olmasının ortalama olasılığını gösteren, bilinmeyen bir p parametresi olsun.

Benzer şekilde, bir bireyin aynı anda W ve I gruplarında olmasının ortalama olasılığını gösteren bilinmeyen bir r parametresi olsun. Bu grup şekli kısaca bir çapraz tablo ile gösterilebilir.

Tablo 2.1. Parametrelere ilişkin çapraz tablo

	N	W
I	p	r
L		

Bu şekilde gruplandırılmış bir toplumda N kategorisine ait n kişi var ise, I-N hücresinde np kişi olması beklenir. Yani, I-N hücresinin beklenen değerini $E\{u\}$ ile gösterecek olursak $E\{u\} = np$ olur.

Benzer şekilde W kategorisine ait w kişi var ise, I-W hücresinde wr kişi olması beklenir. Yani, I-W hücresinin beklenen değerini $E\{v\}$ ile gösterecek olursak $E\{v\} = wr$ olur.

Tablo 2.2. Hücrelerin beklenen değerleri

	N	W
I	np	wr
L		
Toplam	n	w

Bu durumda yalnızca I kategorisine ait $(np+wr)$ kişi olacaktır. Yani, I kategorisinin toplam beklenen değeri $E\{u+v\} = np + wr$ olarak elde edilir.

Tablo 2.3. Hücrelerin toplam beklenen değerleri

	N	W	Toplam
I	np	wr	$np+wr$
L			
Toplam	n	w	

$E\{u+v\} = np + wr$, eşitliğinin her iki tarafı örneklem büyüklüğü olan $n + w$ toplamına bölünürse,

$E\{Y\} = Xp + (1-X)r$, elde edilir. Burada X , I-N kategorisinde bulunan bireylerin $n/(n+w)$ 'ye eşit olduğu orandır. Eşitlik yeniden düzenlendiğinde $E\{Y\} = r + (p-r)X$ olarak elde edilir.

Buradan yeni bir dönüşümle $E\{Y\} = a + bX$ denklem formuna ulaşılır. Burada açıkça görülüyor ki, $a = r$ ve $p-r = b$ eşitlikleri vardır.

Eğer X ' in değişik değerlerinde birkaç örneklem elde edilirse (yani her bir örneklemdeki N kategorisine dahil bireylerin oranları elde edilirse), her bir örneklemde Y değerlerini gözlemleyebiliriz. Y ve X 'in bu örneklemdeki değerleri ile bir grafik çizildiğinde saçılım grafiği elde edilir. Böylece EKK metodu kullanılarak en uygun denklem kurulur ve a ve b için en küçük kareler tahminleri bulunur.

Buradan $a = r$ ve $p - r = b$ eşitlikleri kullanılarak istenilen p ve r parametreleri, EKK metodu kullanılarak tahmin edilmiş olur. Ayrıca elde edilen p ve r parametrelerinin tahminleri yansızdır.

Yalnızca ekolojik değişkenler olan X ve Y kullanılarak, bilinmeyen p ve r parametreleri EKK metodu yoluyla tahmin edilmiş oldu.

Peki böyle bir analiz için ihtiyaç olan varsayımlar nelerdir?

- a) Örneklemden örnekleme değişmeyen p ve r gibi parametrelerinin tahmini bulunmaya çalışılmalıdır.
- b) X ' in bir oran olarak ifade edildiği, $E\{Y\}$ 'nin belli bir kategorideki bireylerin beklenen oranlar olarak ifade edilebileceği $E\{Y\} = Xp + (1 - X)r$ ilişkisi olmalıdır.

Eğer (a) ve (b) ' deki varsayımlar doğru ise bilinen EKK yöntemi kullanılarak p ve r parametrelerinin yansız tahmini elde edilecektir. Dahası, bunlara ek olarak;

- c) X ' in her değeri için, Y ' nin değerlerinin varyansları eşit ve yaklaşık olarak normal dağılıyorsa bilinen tüm regresyon yöntemleri (hipotez testleri, güven aralıkları gibi) uygulanabilir.

Ekolojik değişkenler ile regresyon analizi yapmak için (a), (b) ve (c) varsayımları, analiz edilmek istenen veri için kontrol edilmelidir.

2.4.1. Tablonun 2 x 2 Olması Durumu

Başlangıç durumu için birbiri ardına gelen iki seçimde de parti 1 ve parti 2 gibi sadece iki parti olduğu durumunu düşünelim. N_{ij} ($i, j = 1, 2$) birinci seçimde i. partiye ve ikinci seçimde j. partiye oy verenlerin sayısı olarak gösterilmektedir. Aşağıdaki tabloda $N_{.j}$ ve $N_{i.}$ 'lerin marjinal değerler olduğu, birinci ve ikinci seçimlerde olabilecek tüm durumlar gösterilmiştir. Burada;

$N_{.j}$ = İkinci seçimde j. partiye oy verenlerin sayısı (Birinci seçimde hangi partiye oy vermiş oldukları fark etmez)

$N_{i.}$ = Birinci seçimde i. partiye oy verenlerin sayısı (İkinci seçimde hangi partiye oy vermiş oldukları fark etmez)

olarak tanımlanmaktadır. Birbiri ardına gelen iki seçimde iki parti arasındaki dağılım çapraz tablo üzerinde aşağıdaki şekilde ifade edilmektedir.

Tablo 2.4. İki seçimde oy verenlerin sayısı

N_{11}	N_{12}	$N_{1.}$
N_{21}	N_{22}	$N_{2.}$
$N_{.1}$	$N_{.2}$	

İki parti ve birbiri ardına gelen ardışık seçimler için marjinal değerler bilinebilir ve tablodaki hücre değerleri bir denklem sistemi ile elde edilebilir. Tabi ki her iki seçimde de iki parti olması durumu yani 2 x 2 şeklindeki bir tablo kullanımı durumu çok gerçekçi değildir. Yine de, böyle bir modelin pratikte nasıl kullanılacağını göstermek için bu tablodan faydalanılmaktadır.

Eğer 1. seçimde oyların büyük kısmını 1. parti ve ardından gelen 2. seçimde oyların büyük kısmını diğer parti alıyorsa yukarıdaki modellenen tablo ile bireylerin geçiş hareketlerini hesaplayabiliriz. Dahası, birbiri ardına gelen iki seçimde seçmenlerin sayısının değişimi oldukça azdır (en azından genel seçimlerde).

Buradan yola çıkarak $N_{.1} + N_{.2} \cong N_{.1} + N_{.2} \cong N$ eşitliği elde edilecektir, bu da olasılık tablosunun dönüşümünde oldukça fayda sağlayacaktır.

Birinci seçimde i. partinin aldığı oyların oranı x_i olarak ve ikinci seçimde j. partinin aldığı oyların oranı y_j olarak ifade edildiğinde $x_i = \frac{N_{i.}}{N}$ ve $y_j = \frac{N_{.j}}{N}$ eşitlikleri elde edilir. Birinci seçimde ve ikinci seçimde aynı partiye oy verenlerin oranını β_1 , her iki seçimde farklı partilere oy verenlerin oranını da β_2 ile gösterdiğimizde elde ettiğimiz geçiş sonuçları aşağıdaki tablodaki gibidir.

Tablo 2.5. Birbiri ardına gelen iki seçimde seçmelerin geçişleri

			<i>TOPLAM</i>
	$\beta_1 x_1$	$(1 - \beta_1)x_1$	x_1
	$\beta_2 x_2$	$(1 - \beta_2)x_2$	x_2
<i>TOPLAM</i>	y_1	y_2	1

Yukarıdaki notasyonlara göre ikinci seçim olan y_j 'ler birinci seçimlerin bir kombinasyonu olarak yazılabilir. Şöyle ki;

$$\begin{aligned}
 y_1 &= \beta_1 x_1 + \beta_2 x_2 \\
 &= \beta_1 x_1 + \beta_2 (1 - x_1) \\
 &= (\beta_1 - \beta_2)x_1 + \beta_2
 \end{aligned} \tag{2.1}$$

Birinci ve ikinci seçimlerdeki oyların oranları arasındaki ilişki genellikle aşağıdaki regresyon fonksiyonları şeklinde gösterilmektedir.

$$\begin{aligned}
 y_1 &= \beta_{11}x_1 + \beta_{12}x_2 + \varepsilon_1 \\
 y_2 &= \beta_{21}x_1 + \beta_{22}x_2 + \varepsilon_2
 \end{aligned} \tag{2.2}$$

Burada $\varepsilon_1, \varepsilon_2$ hata terimleri olmakla birlikte, 2 x 2 şeklindeki bir tablo için geçiş oylarının katsayıları Tablo 2.6.'daki gibidir.

Tablo 2.6. Geçiş oylarının katsayıları

	β_{11}	β_{12}
	β_{21}	β_{22}
Toplam	1	1

2 x 2 Tablolar için Örnek bir Uygulama

Bireylerle ilgili sosyal alanlarda çalışan araştırmacıların ilgilendiği, ırk, sosyal sınıf, din, seçimler gibi ilgilendikleri değişkenler nominal (sözel) ya da ordinal (sıralı) ölçümlerdir.

Leo Goodman, nominal ve ordinal değişkenler kullanılarak uygulanabilecek olan ekolojik regresyon tekniğini bulmuştur. Siyasal bilimlerde oldukça sık kullanılan yöntem genellikle seçimlerle ilgili analizlerde kullanışlıdır. Bu analizlerde bireyler arasındaki ilişkilere yönelik yorum yapmak istenmektedir.

Birbiri ardına gelen iki seçim düşünelim. Burada; kendi içlerinde iki grup olan, 1. seçimdeki oylar ve 2. seçimdeki oylar olmak üzere 2 değişken vardır. Uygulamanın anlaşılır olması açısından, belli bir bölgede, her iki seçimde de Demokratlar ve Cumhuriyetçiler olmak üzere 2 parti vardır. Birinci ve ikinci seçimde Demokratların ve Cumhuriyetçilerin aldığı oy yüzdelerinin verileri elde edilmiştir. X_1, X_2, Y_1, Y_2 değişkenleri ile birlikte her bölge için uygulanacak matris formu aşağıdaki gibidir:

Tablo 2.7. Örnek Seçim Tablosu

Seçim 1 \ Seçim 2	Demokratlar	Cumhuriyetçiler	
Demokratlar	PX_1	RX_2	Y_1
Cumhuriyetçiler	QX_1	SX_2	Y_2
	X_1	X_2	

$$\begin{aligned} P + Q &= \%100 & X_1 + X_2 &= \%100 \\ R + S &= \%100 & Y_1 + Y_2 &= \%100 \end{aligned}$$

Burada Y_1 değişkeni; birinci ve ikinci seçimde Demokrat partisine oy verenlerle birinci seçimde Cumhuriyet partisine oy verip, ikinci seçimde Demokrat partiye oy veren seçmenlerin yüzde olarak toplamıdır.

Goodman (1953), uyguladığı yöntemle göre kısaca, $Y_1 = PX_1 + RX_2$ olarak ifade edilir. Burada $X_1 + X_2 = 1$ olduğundan;

$$\begin{aligned} Y_1 &= PX_1 + RX_2 \\ Y_1 &= PX_1 + R(1 - X_1) \\ Y_1 &= PX_1 + R - RX_1 \\ Y_1 &= R + (P - R)X_1 \end{aligned} \tag{2.3}$$

eşitlikleri elde edilmektedir. Bu da X_1 üzerinde Y_1 'in lineer regresyon denklemi formundadır. $a_{yx} = R$ ve $b_{yx} = P - R$ olmak üzere $Y_1 = a_{yx} + b_{yx}X_1$ regresyon denklemini çözebiliriz.

Örnek olarak, her birinde 1000 bireyin olduğu 100 adet örneklemden alınan toplanmış veriler aşağıdaki tablo ile gösterilmektedir.

Tablo 2.8. Örnek veri tablosu

	X_1	X_2	Toplam
Y_1	20.000	60.000	80.000
Y_2	80.000	40.000	120.000
Toplam	100.000	100.00	

Tablodan X_1, X_2 ve Y_1, Y_2 toplamlarının yüzdelikleri bilinmektedir. Toplanmış verinin analizinden elde etmek istediğimiz bilgi; P, Q, R ve S değerleridir. Tablodaki değerlere göre;

P değerine karşılık gelen; X_1, Y_1 hücresinin olasılık değeri $20.000 / 100.000 = 0.20$,
Q değerine karşılık gelen X_1, Y_2 hücresinin olasılık değeri $80.000 / 100.000 = 0.80$,

R değerine karşılık gelen X_2, Y_1 hücresinin olasılık değeri $60.000 / 100.000 = 0.60$, ve S değerine karşılık gelen X_2, Y_2 hücresinin olasılık değeri $40.000 / 100.000 = 0.40$ olarak hesaplanmaktadır.

Örneğin; Tablo 2.7.'deki Cumhuriyetçilerin aldığı tüm oylara karşılık gelen X_2 değerinin, 800 olduğu düşünülürse X_2, Y_1 hücresinin, yani birinci seçimde Cumhuriyetçilere oy vermişken, ikinci seçimde oyunu Demokratlara verenlerin beklenen değeri $0.60 \times 800 = 480$ kişi olacaktır.

2.4.2. Tablonun n x 2 Olması Durumu

Ekolojik regresyon yönteminin en sık uygulandığı alanlardan birinin siyaset bilimleri olduğu açıkça görülmektedir. Peki, parti sayıları çoğaldıkça olasılık tablolarının durumu nasıl olacaktır?

Birinci seçimde p tane partinin, ikinci seçimde q tane partinin olduğu birbirini ardına gelen iki seçim sonuçlarının toplanmış verileri olsun.

Birinci seçimde oylanan partinin oylarının oranları x_i , ikinci seçimde oylanan partinin oylarının oranları y_i olmak üzere $\sum x_i = \sum y_i = 1$ 'dir.

n tane seçim bölgesinin olduğu ve herhangi bir seçim bölgesinin h ile gösterildiği ($h=1,2,\dots,n$), seçmenlerin bir büyüklüğü $N^{(h)}$ için seçim verisinin çapraz tablosu aşağıdaki Tablo 2.9. ile gösterilmiştir.

Tablo 2.9. Birbiri Ardına Gelen İki Seçimin Sonuçları

<u>1. seçim</u>	<u>2. seçim</u>						Toplam
	1. parti	2. parti	...	j. parti	...	q. parti	
1. parti	...	...		...		...	$N_{1.}^{(k)}$
$\vdots$							$\vdots$
i. parti	...	...		$N_{ij}^{(k)}$		...	$N_{i.}^{(k)}$
$\vdots$							
p. parti	...	...				$N_{pq}^{(k)}$	$N_{p.}^{(k)}$
Toplam	$N_{1.}^{(k)}$	...	...	$N_{j.}^{(k)}$	...	$N_{.q}^{(k)}$	$N^{(k)}$

Tablonun kullanılmasıyla, yayınlanmış istatistiklerden marjinal frekanslar olan $N_{i.}^{(h)}$ ve $N_{.j}^{(h)}$ 'lar bilinmektedir. Yani; seçim bölgesi h olan iki seçimde her politik parti için oyların sayısı bilinmektedir. Buna rağmen 2. seçimde j. partiye oy veren ve bir önceki seçimde i. partiye oy veren bireylerin sayısı olan $N_{ij}^{(h)}$ hücre frekansları bilinmemektedir. Tek grup seviyesindeki birimde marjinal frekansları bilmek, bu birimdeki uygun hücre frekansını tahmin etmek için açıkça yeterli değildir. Marjinal frekanslar bu birimlerin büyük sayıları için bilinirse ($h=1, \dots, n$), bu bilgi, tüm birimlerin hücre frekanslarının ortalamasını tahmin etmekte kullanılabilir. Bu da, ekolojik regresyon yönteminin arkasındaki düşüncedir.

Çok değişkenli çoklu regresyon teorisinin kullanılmasıyla, açıklayıcı değişkenler olan x_{ih} ; birinci seçimde h bölgesinde i. partinin aldığı oyların oranı, yanıt değişkenler olan y_{jh} ; ikinci seçimde h bölgesinde j. partinin aldığı oyların oranlarıdır.

altında, $\left(E\left(\frac{\beta_{ij}}{x_j}\right)=C_{ij}, i=1,\dots,p, j=1,\dots,q\right)$, her bir formun p açıklayıcı değişkenli q regresyon denklemleri sistemi aşağıdaki şekildedir.:

$$\begin{aligned} y_{1h} &= \beta_{11}x_{1h} + \dots + \beta_{p1}x_{ph} + e_{1h} \\ y_{2h} &= \beta_{12}x_{1h} + \dots + \beta_{p2}x_{ph} + e_{2h} \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ y_{qh} &= \beta_{1q}x_{1h} + \dots + \beta_{pq}x_{ph} + e_{qh} \end{aligned} \tag{2.4}$$

n seçim bölgesinden verilen bilgilerle eşitlik sistemini kısaltarak çok değişkenli çoklu lineer regresyon modelinde matris formunda yazılabilir.

$$Y = XB + \varepsilon \quad (2.5)$$

Burada Y ; $n \times q$ tipinde q yanıtın bir matrisi, X ; $n \times p$ tipinde elemanları sabit olan açıklayıcı değişkenler matrisi, $B = (\beta_1, \beta_2, \dots, \beta_q)$ $p \times q$ tipinde bilinmeyen katsayıların matrisi ve $\varepsilon = (e_1, e_2, \dots, e_q)$ matrisi ve $E(e_j) = 0$, $E(e_j e_l') = \sigma_{jl}^2 I_n$ $j, l = 1, \dots, q$ özelliklerini sağlayan $n \times q$ tipinde bir matristir. Matris formları olarak gösterildiğinde;

$$Y = \begin{bmatrix} y_{11} & y_{21} & y_{31} & \cdots & y_{q1} \\ y_{12} & y_{22} & y_{32} & \cdots & y_{q2} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{1n} & y_{2n} & y_{3n} & \cdots & y_{qn} \end{bmatrix} \quad X = \begin{bmatrix} x_{11} & x_{21} & x_{31} & \cdots & x_{p1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & x_{3n} & \cdots & x_{pn} \end{bmatrix}$$

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \dots & \beta_{1q} \\ \beta_{21} & \beta_{22} & \beta_{23} & \dots & \beta_{2q} \\ \vdots & \vdots & \vdots & & \vdots \\ \beta_{p1} & \beta_{p2} & \beta_{p3} & \dots & \beta_{pq} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} e_{11} & e_{12} & e_{13} & \dots & e_{1n} \\ e_{21} & e_{22} & e_{23} & \dots & e_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ e_{q1} & e_{q2} & e_{q3} & \dots & e_{qn} \end{bmatrix}$$

şeklindedir. Model (2.5)' deki denklemler yapısal olarak ilişkisiz görünseler bile, aslında onlar arasında bir bağlantı oluşturan çapraz denklemlerin hataları ilişkilidir.

Herhangi bir seçimle ilgili kaynakta parametreler, 1. seçimde i. parti için, 2. seçimde j. parti için oy veren seçmenlerin olasılıklarını göstermektedir. Bu nedenle parametre matrisi olan $B = (\beta_1, \beta_2, \dots, \beta_q)$, satır toplamaları bire eşit olan geçiş matrisi olarak göz önünde bulundurulabilir. Yani; $\sum \beta_{ij} = 1$ olur.

$$\left(E \left(\beta_{ij} / x_j \right) = C_{ij}, i = 1, \dots, p, j = 1, \dots, q \right), \quad \text{varsayımının sağlanması}$$

durumunda β_{ij} 'ler En Küçük Kareler yöntemi doğrultusunda, $\hat{\beta}_{EKK} = (X'X)^{-1} X'Y$ ile tahmin edilebilir. Fakat β 'lar yeterince durağan olmadığında, yani pozitif ya da negatif bir korelasyonla karşılaşıldığında tahminler kabul edilebilir sınırın (0,1) dışına çıkabilir. Diğer taraftan yansız olarak tahmin edilen bu parametreler çok yüksek varyanslı olup etkinliklerini kaybedebilirler.

Örneğin, E. Fule' nin İsveç seçimleri ile ilgili, EKK yöntemi kullanılarak yapmış olduğu çalışmada elde ettiği sonuçlar Tablo 2.10.'da görülmektedir.

Tablo 2.10. EKK ile tahmin edilmiş geçiş olasılıkları (yüzde olarak)

1991 1988	m	c	fp	kds	s	v	nyd	oth	boe	Toplam
m	73.9	-0.4	2.9	0.4	-1.1	-1.9	5.0	-2.3	23.5	100
c	76.1	64.1	15.7	-0.6	0.9	-2.2	28.3	17.9	-100.1	100
fp	64.7	-2.5	46.8	32.1	9.3	4.3	10.0	7.2	-71.9	100
mp	8.9	2.5	16.9	6.7	-33.8	0.8	3.7	47.7	46.7	100
s	0.4	-0.2	-1.4	0.2	81.0	-0.4	2.5	0.9	16.9	100
v	-21.9	1.6	13.8	-17.7	-24.6	76.3	1.7	20.3	50.5	100
oth	0.8	4.4	-8.8	34.2	8.6	-1.6	-1.7	13.1	51.1	100
boe	15.8	0.8	3.9	0.1	13.6	1.7	5.6	0.5	58.2	100

Ekolojik regresyonda böyle bir durumdan kaçınmak için, Shrinkage tahmin ediciler kullanılmaktadır. Bu durumda, tahminleri orjin yönünde büzmek yerine, sahip olunan ön bilgiye dayalı bir noktaya doğru büzme yolu kullanılmaktadır (Brown ve Payne, 1975).

2.5. Achen ve Shively' nin Karesel Modeli

Achen ve Shively (1995), karesel bir form kullanarak Goodman' ın uyguladığı regresyon yöntemini genişletmiştir. Bu model kullanılarak toplanmış veriden kaynaklanan yanlılığın azaldığı kanıtlanmıştır. Bu modelde, Goodman (1953), uyguladığı regresyon yönteminde, $Y_i = \beta_1 X_1 + \beta_2 X_2$ ile ifade edilen regresyon denklemi, $X_1 + X_2 = 1$ olmasından dolayı $Y_i = \beta_2 + (\beta_1 - \beta_2)X_i$ olarak yeniden düzenlenmektedir.

Achen ve Shively bu modeli kullanarak, her ikisi de X 'in bir lineer fonksiyonu olan β_{Bi} ve β_{Wi} parametrelerini tanımlamışlardır. Tanımlanan bu parametreler:

$$\begin{aligned}\beta_{Bi} &= b_1 + b_2 X + \varepsilon_{Bi} \\ \beta_{Wi} &= b_3 + b_4 X + \varepsilon_{Wi}\end{aligned}\tag{2.6}$$

şeklindedir. Bu eşitlikler Goodman'ın temel modelinde yerine koyulduğunda EKK tarafından tahmin edilebilen yeni bir karesel model elde edilmektedir. Şöyle ki;

$$\begin{aligned}Y_i &= \beta_2 + (\beta_1 - \beta_2)X_i \\ T &= \beta_{Wi} + (\beta_{Bi} - \beta_{Wi})X \\ T &= (b_3 + b_4 X) + (b_1 + b_2 X - b_3 - b_4 X)X \\ T &= b_3 + (b_1 - b_3 + b_4)X + (b_2 - b_4)X^2 \\ T &= B_0 + B_1 X + B_2 X^2 + \varepsilon\end{aligned}\tag{2.7}$$

elde edilmektedir. Burada açıkça görülüyor ki;

$$\begin{aligned}B_0 &= b_3 \\ B_1 &= b_1 - b_3 + b_4 \\ B_2 &= b_2 - b_4\end{aligned}\tag{2.8}$$

eşitlikleri vardır. Yani tanımlanan bu karesel model, $b_2 = b_4 = 0$ olması durumunda, Goodman'ın tanımladığı regresyon denkleminin özel bir halidir.

Bu yeni parametrelerden dolayı, böyle bir modeli tanımlamak için Goodman'ın varsayımlarına ek olarak yeni varsayımlar gerektirmektedir. b_1, b_2, b_3 ve b_4 parametreleri bilinmiyor olsa bile, B_0, B_1 ve B_2 bulunabilir. Bu modeli tanımlamak için eklenen bu yeni parametrelerle ilgili olarak Achen ve Shively aşağıdaki adımları düzenlemişlerdir.

1. $b_2 \geq 0$ ve $b_4 \geq 0$ ise β_B ve β_W arasındaki ilişki pozitif tanımlıdır. Bu varsayım altında, aşağıdaki tanımlar kullanıldığında, Goodman regresyon yöntemindeki toplanmış veriden kaynaklanan yanlışlığın azaldığı kanıtlanmıştır.

$$\begin{aligned} B_2 > 0 \text{ ise } b_4 &= 0; \\ B_2 < 0 \text{ ise } b_2 &= 0; \\ B_2 = 0 \text{ ise } b_2 &= b_4 = 0 \end{aligned}$$

Bu kurallar, tahmin edilen $\hat{b}_1, \dots, \hat{b}_4$ parametrelerinin bulunmasında da kullanılmaktadır.

2. $\hat{b}_1, \dots, \hat{b}_4$ ve (2.6) eşitlikleri kullanılarak, seçim bölgelerinden oluşan örneklem parametreleri $\tilde{\beta}_{Bi}$ ve $\tilde{\beta}_{Wi}$ olarak tahmin edilir.

3. Kitle parametreleri olan $\tilde{\beta}_B$ ve $\tilde{\beta}_W$

$$N_B = \sum N_i X_i \text{ ve } N_W = \sum N_i (1 - X_i)$$

olmak üzere, örneklemelerin ağırlıklandırılmış ortalamaları olan,

$$\begin{aligned} \tilde{\beta}_B &= \sum \tilde{\beta}_{Bi} N_i X_i / N_B \\ \tilde{\beta}_W &= \sum \tilde{\beta}_{Wi} N_i X_i / N_W, \end{aligned} \tag{2.9}$$

kullanılarak kitle parametreleri hesaplanmaktadır.

Ekolojik olarak sonuç çıkarma yöntemiyle geçerli olan birçok uygulamada $b_2 \geq 0$ ve $b_4 \geq 0$ varsayımları mümkündür.

Örneğin; oylamalarla ilgili olarak aynı sınıfı, ırkı veya etnik grubu kapsayan bir bölgedeki gruba ait partinin oyları yükselecektir, yani $b_2 \geq 0$ olacaktır. $b_2 \geq 0$ ve $b_4 \geq 0$ varsayımları, karesel model kullanılmasıyla elde edilen yanlışlık, her zaman Goodman'ın kullandığı yöntemdeki yanlışlıktan daha az veya eşit olacaktır. Eğer karesel regresyon koşulları sağlanıyorsa, karesel modeli kullanmak doğrusal modeli kullanmaktan daha uygun olacaktır.

2.6. Duncan ve Davis Sınırlar Metodu

Duncan ve Davis uyguladıkları bu yöntemde, toplanmış verilerle oluşturulan marjinal tablolarda, olabilecek bireysel seviyeli ilişkinin maksimum ve minimum sınırlarını oluşturmuşlardır. Toplanmış verinin içerisindeki bireyler hakkında hiçbir varsayım yapmadan, bireysel seviyeli ilişkilerin, hatasız değerlerinin olduğu limitler oluşturulmaktadır. Böylece içerisinde ilişkilerin değerlerinin doğru olduğu sınırlar elde edilmektedir.

Örneğin; belli bir bölgede, bireyler gelir düzeylerine göre orta ve yüksek olarak iki gruba ayrılabilir. Bu iki ana grubun desteklediği A ve B gibi iki ana parti olsun. Buradan bireylerin gelir düzeylerine göre desteklenen partiler arasındaki ilişkiyi açıklamak isteyelim. Bu gruplandırma şeklinde belli bir bölgeden elde edilmiş sonuçlar şu şekilde olsun: Gelir düzeyi yüksek olan bireylerin sayısı 10.000, gelir düzeyi orta olan bireylerin sayısı 1000 ve A partisini destekleyen bireylerin sayısı 8000, B partisini destekleyen bireylerin sayısı 3000 olsun. Burada 1000 kişi gelir düzeyi orta olduğundan, gelir düzeyi yüksek olup A partisini destekleyen minimum 7000 ve maksimum 8000 kişi vardır. Benzer şekilde gelir düzeyi yüksek olup B partisini destekleyen minimum 2000 ve maksimum 3000 kişi olacaktır.

Böyle bir çalışma tahmin olamamakla birlikte kesin sonuçlardır. Yani, elde edilen bu veriler doğrultusunda bireysel seviyeli ilişkiler bu limitlerin dışında olamaz. Duncan ve Davis yönteminin en büyük avantajı toplanmış verideki bireyler hakkında hiçbir varsayıma gerek olmamasıdır.

Yöntemin uygulama aşaması, yukarıda bahsedilen örnek ile oluşturulan tablo üzerinde gösterilebilir.

Tablo 2.11. Sınırlar Metodu için örnek tablo

	A Partisi	B Partisi	Toplam
Gelir Yüksek	β_i^b	$1 - \beta_i^b$	X_i
Gelir Orta	β_i^w	$1 - \beta_i^w$	$1 - X_i$
Toplam	T_i	$1 - T_i$	

i. seçim bölgesindeki nüfusun toplamının N_i olduğu yukarıdaki tabloya göre;

X_i : Geliri yüksek olan bireylerin oranı,

$1 - X_i$: Geliri orta olan bireylerin oranı,

T_i : A partisini destekleyen bireylerin oranı,

$1 - T_i$: B partisini destekleyen bireylerin oranı ve bilinmeyen parametreler

β_i^b ve β_i^w olarak gösterilmiştir.

Burada amacımız; yöntemi kullanarak i. seçim bölgesinde bilinmeyen β_i^b ve β_i^w parametrelerini tahmin etmektir.

β_i^b : Geliri yüksek olup, A partisini destekleyenlerin oranı,

β_i^w : Geliri orta olup, B partisini destekleyenlerin oranıdır.

Duncan ve Davis tarafından elde edilen sınırlar aşağıdaki şekilde tanımlanmaktadır.

$$\beta_i^b \in \left[\max\left(0, \frac{T_i - (1 - X_i)}{X_i}\right), \min\left(1, \frac{T_i}{X_i}\right) \right] \quad (2.10)$$

$$\beta_i^w \in \left[\max\left(0, \frac{T_i - X_i}{1 - X_i}\right), \min\left(1, \frac{T_i}{1 - X_i}\right) \right] \quad (2.11)$$

Bu yaklaşım sınırlar metodu olarak bilinmektedir ve belli bir grubun sayısal miktarının minimum ve maksimum değerleri olduğunu ifade etmektedir. Elde edilen değerler minimum noktasından daha düşük ya da maksimum noktasından daha yüksek olamaz.

Örnek bir uygulama:

Washington eyaletinde yapılan nüfus verilerinden elde edilen bilgilere göre; 25 yaş ve üstü bireylerin yerli ve yabancı olmaları göz önünde bulundurularak, düşük ve yüksek gelir seviyeleri arasındaki ilişki için sınırlar oluşturulmak istenmektedir. Burada bireylerin bir yılda kazandıkları para 50.000 \$ veya daha fazla ise geliri yüksek olarak nitelendirilmektedir.

Tablo 2.12. Örnek veri kümesi

	Yerli	Yabancı	Toplam
Gelir Yüksek	?	?	2.182.000
Gelir Düşük	?	?	1.146.000
Toplam	3.066.000	263.000	3.329.000

Sınırlar metodu için sayısal değerleri örnek bir 2 x 2 tipinde bir tablo yukarıda görülmektedir. Tabloya göre değerleri aşağıdaki şekilde oluşturabiliriz.

$$X_i : 2.182.000$$

$$1 - X_i : 1.146.000$$

$$T_i : 3.066.000$$

$$1 - T_i : 263.000$$

Bahsedilen yöntem doğrultusunda, yüksek gelirli ve yabancı bireyleri temsil eden hücrenin değeri 263,000 değerini geçemez (yani sütun toplamı). Tablodan yararlanarak;

$$\beta_i^b \in \left[\max\left(0, \frac{T_i - (1 - X_i)}{X_i}\right), \min\left(1, \frac{T_i}{X_i}\right) \right] \text{ sınırında değerler yerine koyulduğunda;}$$

$$\beta_i^b \in \left[\max\left(0, \frac{3.066.000 - 1.146.000}{2.182.000}\right), \min\left(1, \frac{3.066.000}{2.182.000}\right) \right]$$

$\beta_i^b \in [\max(0, 0.88), \min(1, 1.405)] \rightarrow \beta_i^b \in [0.88, 1]$ olarak elde edilmektedir. Diğer sınır ise;

$$\beta_i^w \in \left[\max\left(0, \frac{T_i - X_i}{1 - X_i}\right), \min\left(1, \frac{T_i}{1 - X_i}\right) \right]$$

$$\beta_i^w \in \left[\max\left(0, \frac{3.066.000 - 2.182.000}{1.146.000}\right), \min\left(1, \frac{3.066.000}{1.146.000}\right) \right] \text{ değerleri}$$

hesaplandığında;

$$\beta_i^w \in [\max(0, 0.77), \min(1, 2.67)] \rightarrow \beta_i^w \in [0.77, 1] \text{ olarak hesaplanmaktadır.}$$

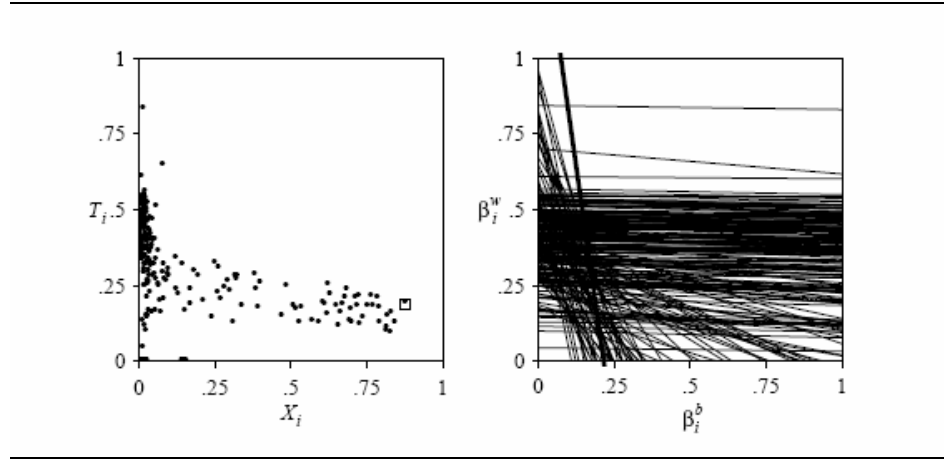
2.7. King' in EI Yöntemi

Ekolojik tahmin yöntemlerinin istatistiksel olarak sağlam dayanakları olması için bugüne kadar Goodman (1953, 1959), Duncan ve Davis (1953) gibi araştırmacıların birçok çabaları ve çalışmaları olmuştur. Ekolojik olarak sonuç çıkarma yöntemleri üzerine çalışan kişilerden biri de Gary King (1997)' dir. Siyasal bilimler araştırmasında, ekolojik birimlerden elde edilen toplanmış verileri kullanarak yorum yapabilme alanında yeni ufuklar açmıştır. Birçok nedenden dolayı, King (1997)' nin ekolojik olarak sonuç çıkarma konusundaki yaklaşımının Goodman' ın klasik ekolojik regresyon yönteminden çok daha iyi olduğu tartışılmaktadır. Kullandığı yöntemlerin çoğu, veri ve sınırlar hakkındaki tüm bilgiyi parametreler üzerinde kullanılabilir hale getirmektedir ve tahmin edilen parametreler hakkında bulduğu sonuçlar genellikle tutarlıdır.

Sınırlar metodu ve genelleştirilmiş halinin amacı, problem ile ilgili bilgiyi özet halinde hesaplamaktır. Örneğin; herhangi bir seçim bölgesinde geliri yüksek olarak gruplandırılan 150 kişi olsun. Bu gruptan oy verenlerin sayısı 87 ise, geliri yüksek olan kişilerin belli bir partiye oyları 0 ile 87 arasında olacaktır. Bu yöntemdeki problem olası sonuçların, yani bulmak istediğimiz cevaplar için sadece uygun bir aralık oluşturmaktır. Ayrıca varsayımlar problem için yanlış ise oluşturulan bu aralık, doğru cevapları içine almayabilir.

King yönteminin anahtar noktası, sınırlar metoduyla fikir ayrımı olmamasıdır. Buradaki amaç ayrı olan bilgi kaynakları birleştirilip, ekolojik olarak sonuç çıkarmayı geliştirmektir. Elde edilen bilgilerin miktarı veri kümelerine bağlıdır ve birçok veri kümesinden alınan bilgiler kayda değer büyüklüktedir. Örneğin; herhangi bir seçim bölgesine ait T_i 'nin X_i 'ye bağlı saçılım grafiği düzgün olarak dağılıyorsa, β_i^b ve β_i^w için oluşturulan aralığı $[0,1]$ aralığından daha dar bir alana indirgenebilir ve böylece problemin bir çok bölümü azaltılabilir. Sınırlar yönteminden alınan bu ek bilgi sayesinde ise kurulan model çok daha güçlü olabilir.

β_i^b ve β_i^w noktalarını göstermek için, uygun olan verileri, hiçbir varsayım yapmadan kullanarak sınırlar metodu olabildiğince genişletilmektedir. Başlangıç noktası olarak, örnek bir veri kümesi alınarak, aşağıdaki grafiklerde sol taraftaki, T_i 'nin X_i 'ye bağlı saçılım grafiği elde edilmiştir.



Şekil 2.1. Aynı veriye ait iki görünüm

Not: Sol taraftaki grafik, T_i 'nin X_i 'ye bağlı saçılım grafiğidir. Sağ taraftaki ise, aynı verilerle elde edilen β_i^b ve β_i^w katsayılarına ait tomografi grafiğidir. Sol tarafta her bir seçim bölgesine denk gelen noktalar, sağ taraftaki grafikte birer çizgi olarak gösterilmiştir. Örneğin; sol tarafta 52. seçim bölgesi olarak gösterilen kare içine alınan nokta, sağ taraftaki grafikte koyu çizgi olarak görülmektedir. Veriler; King (1997, Şekil 5.1, 5.5).

Saçılım grafiğindeki her bir nokta bir seçim bölgesine denk gelmektedir. Sağ taraftaki grafikte ise bilinmeyen iki parametre β_i^b ve β_i^w 'nin değerleri çizgisel olarak belirtilmiştir. β_i^b yatay olarak, β_i^w ise dikey olarak grafiklendirilmiştir. Her bir bölge için, β_i^b ve β_i^w 'nin gerçek değerleri bu birim karenin içinde olmalıdır. Böylece birim kare dışındaki değerler olanaksızdır.

Tüm bu bilgiler altında King' in oluşturduğu EI modeli aşağıdaki şekilde tanımlanmaktadır:

$$T_i = X_i \beta_i^b + (1 - X_i) \beta_i^w \quad (2.12)$$

Yukarıdaki denklem hata terimi olmadığından dolayı regresyon denklemi formunda değildir. Bu eşitlikten, bilinmeyen parametrelerden birini diğerine göre çözersek;

$$\beta_i^w = \left(\frac{T_i}{1 - X_i} \right) - \left(\frac{X_i}{1 - X_i} \right) \beta_i^b \quad (2.13)$$

olarak elde edilir. Burada, (2.13) ile gösterilen eşitlik β_i^w 'nin β_i^b 'ye bağlı olan, sabit terimi ve eğimi bilinen, lineer bir fonksiyonudur. T_i yanıt değişkeni, X_i açıklayıcı değişken olmakla birlikte, β_i^b ve β_i^w bireysel ilişkilerin miktarını tanımlayan katsayılardır.

King' in tanımladığı bu model Duncan ve Davis' in sınırlar metodu ile Goodman' in regresyon modelinin bir kombinasyonu şeklindedir. Bu model için gerekli üç varsayım vardır. Bunlardan birincisi; β_i^b ve β_i^w , iki değişkenli kesikli normal dağılıma sahip olmalıdırlar. King, özellikle iki değişkenli kesikli normal dağılımın kullanılmasıyla, seçim bölgeleri veya gruplar üzerinde kurulan modelin temelini oluşturan varsayımların değiştiğini belirtmektedir. Ayrıca bu dağılım, katsayıların doğru olmadığını, Monte Carlo denemeleriyle de gösterilen tek bir moda sahiptir. Dağılımın bu özelliklerinden dolayı King, sınırları tahmin etmede ve katsayıların etkinliğini bulmada tomografi grafiklerini kullanmaktadır.

İkinci varsayımda, farklı birimlerden elde edilen toplanmış verilerde, X_i üzerine bir koşul konduktan sonra, T_i değerlerinin birbirlerinden bağımsız olmasıdır. Yani çalışılan modelde otokorelasyon olmadığı varsayılmaktadır.

Son varsayım ise, X_i 'nin katsayılarından yani β_i^b ve β_i^w 'den bağımsız olmasıdır.

King, parametre tahminlerinde çok büyük etkilere neden olmayacak şekilde, varsayımlar üzerinde ayarlamalar yapılabileceğini belirtmektedir.

Bu yöntem ile ilgili araştırmacılar, toplanmış veriden kaynaklanan yanlışlığa ve otokorelasyona dayanıklı olan bu metodun geniş çaplı ve varsayımların sağlandığı verilerde kullanılması gerektiğini vurgulamışlardır. Özellikle coğrafi birimlerden alınan verilerde, coğrafi grupların oransal dağılımını araştırmada, daha çok homojen yapılı ve şans faktörünün az olduğu verilerde kullanışlı bir yöntemdir.

3. EKOLOJİK REGRESYON VE ALTERNATİF ÇÖZÜM YÖNTEMLERİ

3.1. En Küçük Kareler Yöntemi ve Ekolojik Regresyon

Birinci seçimde p tane partinin, ikinci seçimde q tane partinin olduğu birbiri ardına gelen iki seçim sonuçlarının toplanmış verileri olsun. Birinci seçimde oylanan partinin oylarının oranları x_i , ikinci seçimde oylanan partinin oylarının oranları y_i olmak üzere $\sum x_i = \sum y_i = 1$ ' dir. N tane seçim bölgesinden alınan oy oranları verisi elimizde ise, x 'in açıklayıcı değişken olduğu, y 'nin yanıt değişken olduğu çok değişkenli çoklu regresyon modeli oluşturabiliriz. Her bir h bölgesi için ($h=1, \dots, n$) oluşturulan model aşağıdaki gibidir:

$$\begin{aligned} y_{1h} &= \beta_{11}x_{1h} + \dots + \beta_{p1}x_{ph} + e_{1h} \\ y_{2h} &= \beta_{12}x_{1h} + \dots + \beta_{p2}x_{ph} + e_{2h} \\ &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ y_{qh} &= \beta_{1q}x_{1h} + \dots + \beta_{pq}x_{ph} + e_{qh} \end{aligned} \quad (3.1)$$

Burada bilmek istediğimiz, geçiş oranları olarak ifade edilen β_{ij} katsayılarıdır. Ekolojik değişkenler ile oluşturulan bu modelin en büyük özelliği $\sum \beta_{ij} = 1$ olmasıdır. Ayrıca bu katsayılar birer olasılık anlamına geldiğinden negatif değerler alamazlar.

Böyle bir model için, $\left(E\left(\frac{\beta_{ij}}{x_j} \right) = C_{ij}, i = 1, \dots, p, j = 1, \dots, q \right)$ varsayımı

sağlanıyorsa geçiş oranları EKK yöntemi ile tahmin edilebilir.

Tahmin edilen geçiş matrisi:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} & \hat{\beta}_{13} & \dots & \hat{\beta}_{1q} \\ \hat{\beta}_{21} & \hat{\beta}_{22} & \hat{\beta}_{23} & \dots & \hat{\beta}_{2q} \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{\beta}_{p1} & \hat{\beta}_{p2} & \hat{\beta}_{p3} & \dots & \hat{\beta}_{pq} \end{bmatrix} \quad (3.2)$$

şeklindedir. (3.1) ile gösterilen model matris formunda aşağıda gibi düzenlenebilir.

$$Y = X\beta + \varepsilon \quad (3.3)$$

Burada Y ; $n \times q$ tipinde q yanıtın bir matrisi, X ; $n \times p$ tipinde elemanları sabit olan açıklayıcı değişkenler matrisi, $\beta = (\beta_1, \beta_2, \dots, \beta_q)$ $p \times q$ tipinde bilinmeyen katsayıların matrisi ve $\varepsilon = (e_1, e_2, \dots, e_q)$ matrisi ve $E(e_j) = 0$, $E(e_j e_l') = \sigma_{jl}^2 I_n$ $j, l = 1, \dots, q$ özelliklerini sağlayan $n \times q$ tipinde bir matristir.

Hatalar minimize edilerek, β 'nın EKK tahmini aşağıdaki şekilde bulunmaktadır:

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = (Y - X\beta)'(Y - X\beta) \quad (3.4)$$

$$\begin{aligned} S(\beta) &= Y'Y - \beta' X'Y - Y'X\beta + \beta' X'X\beta \\ &= Y'Y - 2\beta' X'Y + \beta' X'X\beta \end{aligned} \quad (3.5)$$

S fonksiyonu β 'ya göre minimize edilmelidir. Böylece $p \times q$ tane normal denklem oluşur. Bu denklemlerin çözümü $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_q$ şeklindeki EKK tahmin edicileri olacaktır.

$$\left. \frac{\partial S}{\partial \beta} \right|_{\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0 \quad (3.6)$$

Buradan;

$$X'X\hat{\beta} = X'Y \quad (3.7)$$

(3.6) ile gösterilen eşitlik EKK normal denklemleri olarak adlandırılır. Bu denklemlerin çözümü için her iki taraf $X'X$ matrisinin tersi ile çarpılır. Böylece β 'nın EKK tahmini;

$$\hat{\beta}_{EKK} = (X'X)^{-1}X'Y \quad (3.8)$$

olarak elde edilmektedir.

3.1.1. En Küçük Kareler Yöntemi Varsayımları

Klasik çoklu doğrusal regresyonda parametre tahminlerinin yapılabilmesi ve elde edilen sonuçların geçerliliği için model üzerinde bazı varsayımlar yapılmıştır. Bu varsayımlar;

$$\textbf{Varsayım 1 : } E(\varepsilon_i) = E(\varepsilon_i / X_i) = 0$$

Hata terimlerinin beklenen değerleri (ortalaması) sıfıra eşittir. Varsayım sağlanmadığı durumlarda, ($E(\varepsilon_i \neq 0)$), parametre tahminleri gerçek değerlerinden daha büyük veya daha küçük değerler alır. Bu durumda yanlı parametre tahminleri elde edilir (Graybill, 1961).

$$\begin{aligned} \textbf{Varsayım 2 : } Var(\varepsilon_i) &= E[\varepsilon_i - E(\varepsilon_i)]^2 \\ &= E(\varepsilon_i^2) = \sigma^2 \end{aligned}$$

Hata terimleri σ^2 ile sabit varyanslıdır. Bu varsayıma göre hata terimi varyansı, bağımsız değişkendeki değişmelere bağlı olarak değişmeyip aynı kalır (sabit varyanslılık) (Myers,1990). Sabit varyanslılık varsayımının geçerli olması halinde hata terimi varyans-kovaryans matrisi; $E(\varepsilon\varepsilon') = \sigma^2 I_n$ şeklinde ifade edilir. Bu varsayımın sağlanmaması durumunda modelde değişen varyanslılık (heteroscedasticity) sorunu ortaya çıkar.

$$\begin{aligned} \textbf{Varsayım 3: } Cov(\varepsilon_i, \varepsilon_j) &= E\{[\varepsilon_i - E(\varepsilon_i)][\varepsilon_j - E(\varepsilon_j)]\} \\ &= E(\varepsilon_i \varepsilon_j) = 0 \\ i, j &= 1, 2, \dots, n \\ i &\neq j \end{aligned}$$

Hata terimleri arasında ilişki yoktur. Başka bir ifade ile hata teriminin herhangi bir değeri başka bir değerinden bağımsızdır. Bu varsayım sağlanmıyorsa hata terimleri arasında otokorelasyon vardır (Graybill,1961).

Modele alınmayan bağımsız değişkenler, değişkenler arasındaki ilişkiyi ortaya koyan matematiksel modelin yanlış seçilmesi, bağımlı değişkende ölçme hatası olması otokorelasyona sebep olmaktadır. Hata terimleri arasında otokorelasyon bulunduğu, parametrelerin tahmin değerleri ve standart hataları bundan etkilenmektedir. Bunun sonucunda parametre tahminleri yansız olmakla birlikte etkin olmayacak, hata teriminin varyansı olduğundan küçük tahmin edilecek, parametre tahminleri için en küçük kareler yöntemi uygun olmayacaktır (Tarı, 1999).

$$\textbf{Varsayım 4 : } \varepsilon_i \sim N(0, \sigma^2)$$

Hata terimleri sıfır ortalama ve σ^2 varyansı ile normal dağılıma sahiptir. Çoklu doğrusal regresyon modellerinde hata terimlerinin normal dağılıma sahip olması kendi sıfır ortalamaları etrafında simetrik bir dağılım göstermeleridir. Bu varsayımın sağlanması parametre tahminleri için güven aralıkları oluşturmaya ve gerekli hipotezleri test etmeye imkân sağlar (Myers,1990).

Normallik varsayımının sağlanmamasının nedeni, aykırı değerler olabileceği gibi etkili gözlemlerin varlığı da olabilir.

Varsayım 5 : $(X'X)$ matrisi singüler olmayan bir matristir.

X tasarım matrisi olup, $(X'X)$ matrisinin tersinin alınabilir olması ve $\text{rank}(X'X) = p$ olduğu varsayılır.

Varsayım 6 : X matrisinin sütun vektörleri lineer bağımsızdır. ($\text{Rank}(X) = p$)

Bu varsayıma göre çoklu doğrusal regresyon modellerinde bağımsız değişkenler arasında doğrusal veya doğrusala yakın bir ilişki yoktur. Bu varsayımın sağlanmaması durumunda $(X'X)$ matrisinin tersi alınamayacağından, parametre tahminleri yapılamayacak, tersinin alınabildiği durumlarda ise parametre tahminlerinin varyansları büyük olacak ve parametre tahminleri tutarsız olacaktır (Rawlings, 1998).

Varsayım 7 : $n \geq k$ 'dır. (n ; gözlem sayısı, k ; parametre sayısı)

Çoklu doğrusal regresyon modellerinde X matrisinin tam sütun ranklı olabilmesi için gözlem sayısı, modeldeki bağımsız değişken sayısından fazla olmalıdır.

3.1.2. En Küçük Kareler Tahmin Edicisinin Özellikleri

1) $Y = X\beta + \varepsilon$ modelini ele alıp, EKK tahmin edicisinin yanlılığı aşağıdaki şekilde incelenmektedir.

$$E(\varepsilon) = 0 \quad \text{ve} \quad (X'X)^{-1} X'X = I \quad \text{olduğundan};$$

$$\begin{aligned} E(\hat{\beta}) &= E[(X'X)^{-1} X'Y] \\ &= E[(X'X)^{-1} X'(X\beta + \varepsilon)] \\ &= E[(X'X)^{-1} XX'\beta + (X'X)^{-1} X'\varepsilon] \\ &= \beta \end{aligned} \tag{3.9}$$

olmaktadır. Bundan dolayı, $\hat{\beta}$, β 'nin yansız bir tahmin edicisidir.

2) $\hat{\beta}$ 'nin varyans – kovaryans matrisi:

$Cov(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\}$ şeklindedir. $Cov(\hat{\beta})$, $p \times p$ tipinde simetrik bir matristir ve j-nci köşegen elemanı $\hat{\beta}_j$ 'nin varyansını göstermektedir. Ayrıca ij- nci eleman ise, i-nci eleman ile j-nci eleman arasındaki kovaryansı göstermektedir. Varyans - kovaryans matrisinin son hali:

$Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ olur. Eğer $C = (X'X)^{-1}$ olursa $\hat{\beta}_j$ 'nin varyansı $\sigma^2 c_{jj}$ 'dir. Ayrıca $\hat{\beta}_i$ ile $\hat{\beta}_j$ arasındaki kovaryans $\sigma^2 c_{ij}$ olur.

3) β 'nin EKK tahmin edicisi, β 'nin “ en iyi lineer yansız tahmin edicisi ” olarak kabul edilir. Bu da EKK' nin “ BLUE ” özelliği olarak bilinmektedir.

3.1.3. En Küçük Kareler Yönteminin Ekolojik Regresyondaki Sonuçları

Ekolojik değişkenler kullanılarak kurulan regresyon modelindeki parametrelerin çözümünde, EKK yöntemini kullanmak ilk kez Goodman (1953) tarafından önerilmiştir. Ancak ekolojik regresyon tekniğinin en çok kullanıldığı seçimlerde, seçmenler coğrafik bir kritere göre gruplandırılmaktadır ve bu yüzden geçiş oranları modeldeki hiçbir açıklayıcı değişken tarafından açıklanamayan, sistematik bir örnekleme değişmektedir. Ayrıca grupların homojen olmamasından dolayı, değişkenler arasında yüksek derecede korelasyon ile karşılaşmaktadır. Bu da tahminlerin, kabul edilebilir olan (0,1) aralığının dışına çıkmasına ya da parametrelerin tahminlerinin varyanslarının çok yüksek olmasına neden olmaktadır. Dahası EKK tahminleri, değişkenlerdeki çok küçük değişikliklerde bile çok hassastır ve durağan değildirler. Bu gibi birçok nedenden dolayı, çoklu iç ilişki problemi, geliştirilen yeni yöntemlerle mutlaka çözülmelidir.

3.2. Ridge Regresyon Tahmin Edicisi

Çoklu lineer regresyon modelinde, genellikle açıklayıcı değişkenlerin bağımsız olduğu varsayılır. Fakat uygulamada, lineer ilişki olabilir. Bu durumda açıklayıcı değişkenler arasında bağımsızlık varsayımı geçerli olmaz ve bu, çoklu iç ilişki problemine neden olur.

Çoklu iç ilişki EKK tahminlerinin mutlak değerce büyük olmasına neden olmaktadır (Montgomery ve Peck, 1992). Bunu görmek için $\hat{\beta}$ ile β arasındaki uzaklığın karesini, örneğin $L^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta)$ yı inceleyelim.

$$\begin{aligned}
 E(L^2) &= E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) \\
 &= \sum_{j=1}^p E(\hat{\beta}_j - \beta_j)^2 \\
 &= \sum_{j=1}^p \text{var}(\hat{\beta}_j) \\
 &= \sigma^2 \text{tr}(X'X)^{-1} = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j}
 \end{aligned} \tag{3.10}$$

olduğundan eğer $X'X$ çoklu iç ilişkiden dolayı kötü koşullu ise en az bir λ_j küçük olacaktır ve (3.10), $\hat{\beta}$ ' dan β ' ya olan uzaklığın büyük olacağı anlamına gelir. Denk olarak

$$\begin{aligned}
 E(L^2) &= E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) \\
 &= E(\hat{\beta}'\hat{\beta} - 2\hat{\beta}'\beta + \beta'\beta)
 \end{aligned}$$

veya

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \text{tr}(X'X)^{-1} \tag{3.11}$$

olur. Yani, $\hat{\beta}$ vektörü genellikle β vektöründen büyüktür. Bu ise EKK metodunun regresyon katsayılarının mutlak değerce büyük tahminlerini ürettiği anlamına gelir.

Ayrıca çoklu iç ilişkinin olduğu verilerde EKK tahmin edicisi kullanıldığında büyük örneklem varyansları ortaya çıkar. Bunun sonucu olarak parametreler üzerindeki güven aralıkları geniş olur; anlamlı regresyon katsayıları modelden çıkarılır ve nokta tahmini kötü olur. Böyle bir problemin çözümünde kullanılan en yaygın yöntemlerden biri ridge regresyon metodudur.

EKK metodundaki problem, EKK tahmin edicisi $\hat{\beta}_{EKK}$ 'nın, β 'nın yansız tahmin edicisi olmasını gerektirmesidir. EKK tahmin edicisi yansız tahmin ediciler arasında minimum varyansa sahip olanıdır. Ancak bazen bu varyans küçük olmayabilir. Bu problemi çözmek için β 'nın tahmin edicisinin yansız olması gereğinden vazgeçelim. β 'nın yansız olan EKK tahmin edicisi $\hat{\beta}_{EKK}$ 'dan daha küçük varyansa sahip $\hat{\beta}(k)$ gibi bir yanlı tahmin edici bulabileceğimizi varsayalım.

$\hat{\beta}(k)$ 'nın β 'dan olan uzaklığın karesinin beklenen değeri olarak tanımlanan MSE değeri:

$$\begin{aligned} MSE(\hat{\beta}(k)) &= E(\hat{\beta}(k) - \beta)^2 \\ &= tr[Var(\hat{\beta}(k))] + [E(\hat{\beta}(k)) - \beta]^2 \\ &= tr[Var(\hat{\beta}(k))] + [Bias(\hat{\beta}(k))]^2 \end{aligned} \quad (3.12)$$

şeklindedir.

$\hat{\beta}(k)$ 'da küçük bir miktar yanlılığa izin vererek $\hat{\beta}(k)$ 'nın varyansı daha küçük yapılabilir; öyle ki, $\hat{\beta}(k)$ 'nın MSE değeri yansız tahmin edici $\hat{\beta}_{EKK}$ 'nın MSE değerinden daha küçüktür. Yanlı tahmin edici kullanıldığında β üzerindeki güven aralıkları daha dardır. Ayrıca yanlı tahmin edicinin varyansının daha küçük olması $\hat{\beta}(k)$ 'nın $\hat{\beta}_{EKK}$ 'dan daha iyi bir tahmin edici olduğunu gösterir.

Regresyon katsayılarının yanlı tahminlerini elde etmek için pek çok yöntem geliştirilmiştir. Bu yöntemlerden biri Hoerl ve Kennard' ın (1970) önerdiği ridge regresyondur. Bu yöntemde $\hat{\beta}_R$ ile gösterilen ridge tahmin edici:

$$(X'X + kI)\hat{\beta}_R = X'y \quad (3.13)$$

normal denklemlerinin çözülmesiyle;

$$\hat{\beta}_R = (X'X + kI)^{-1} X'y \quad (3.14)$$

olarak bulunmaktadır. Bu prosedür ridge regresyon olarak adlandırılmıştır. Çünkü Hoerl (1959) ikinci dereceden yanıt yüzeylerin yapısını açıklamak için kullandığı ridge analiz metoduna benzer matematiksel yapıya sahiptir.

$k = 0$ iken ridge tahmin edici EKK' ya indirgenmektedir. Ridge tahmin edici EKK tahmin edicinin lineer dönüşümü olarak yazılabilir. Şöyle ki;

$$\begin{aligned} \hat{\beta}_R &= (X'X + kI)^{-1} X'y \\ &= (X'X + kI)^{-1} X'X\hat{\beta} \\ &= Z_k\hat{\beta} \end{aligned} \quad (3.15)$$

$E(\hat{\beta}_R) = E(Z_k\hat{\beta}) = Z_k\beta$ olduğundan $\hat{\beta}_R$, β ' nın yanlı bir tahmin edicisidir. k sabiti yanlılık parametresi olarak adlandırılır.

$\hat{\beta}_R$ ' nin kovaryans matrisi;

$$Var(\hat{\beta}_R) = \sigma^2 (X'X + kI)^{-1} X'X(X'X + kI)^{-1} \quad (3.16)$$

$\hat{\beta}_R$ ' nin yanlılığı;

$$\begin{aligned} Bias(\hat{\beta}_R) &= E(\hat{\beta}_R) - \beta \\ &= (X'X + kI)^{-1} X'X\beta - \beta \\ &= (X'X + kI)^{-1} [X'X\beta - (X'X + kI)\beta] \\ &= -k(X'X + kI)^{-1} \beta \end{aligned} \quad (3.17)$$

$\hat{\beta}_R$ ' nin hata kareler ortalaması;

$$\begin{aligned} MSE(\hat{\beta}_R) &= tr[Var(\hat{\beta}_R)] + Bias[(\hat{\beta}_R)]^2 \\ &= \sigma^2 tr[(X'X + kI)^{-1} X'X(X'X + kI)^{-1}] + k^2 \beta'(X'X + kI)^{-2} \beta \end{aligned} \quad (3.18)$$

$k > 0$ olduğundan k arttıkça $\hat{\beta}_R$ 'nin yanlılığı da artar. Fakat k arttıkça varyans azalacaktır. Ridge regresyon kullanılırken varyanstaki azalış, yanlılığın karesindeki artıştan daha büyük olacak şekilde k değeri seçilmelidir. Bu durumda $\hat{\beta}_R$ 'nin MSE değeri $\hat{\beta}_{EKK}$ 'nin varyansından daha küçük olacaktır. Hoerl ve Kennard bu şekilde bir k değeri olduğunu göstermişlerdir. $\hat{\beta}_R$ 'nin rezidü kareler toplamı:

$$\begin{aligned} SS_{Res} &= (Y - X\hat{\beta}_R)'(Y - X\hat{\beta}_R) \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\hat{\beta}_R - \hat{\beta})X'X(\hat{\beta}_R - \hat{\beta}) \end{aligned} \quad (3.19)$$

Bu eşitlikte ilk ifade $\hat{\beta}_{EKK}$ 'nin Rezidü kareler toplamı olduğundan k arttıkça SS_{Res} artar. Çünkü $k = 0$ iken $\hat{\beta}_R = \hat{\beta}_{EKK}$ 'dır ve k arttıkça $\hat{\beta}_R - \hat{\beta}_{EKK}$ farkı da artar. Sonuç olarak k arttıkça R^2 azalır. Çünkü kareler toplamı (SS_T) sabittir

($R^2 = 1 - \frac{SS_{Res}}{SS_T}$). Bu nedenle ridge tahmin edici veriye genellikle en iyi uyumu

sağlamaz. Fakat parametre tahminlerinin durağan bir kümesini elde etmekle ilgilendiğimizden dolayı bu durumla genel olarak ilgilenmeyiz. Ridge tahmin edici sonraki gözlemlerin ön tahminlerinde (prediction) iyi bir sonucunu vermektedir.

Hoerl ve Kennard k 'nın uygun bir seçimi için Ridge izini (ridge trace) önermişlerdir. Ridge izi, k 'nın genellikle $[0,1]$ kapalı aralığındaki değerleri için k 'ya karşı $\hat{\beta}_R$ 'nin elemanlarının grafiğidir.

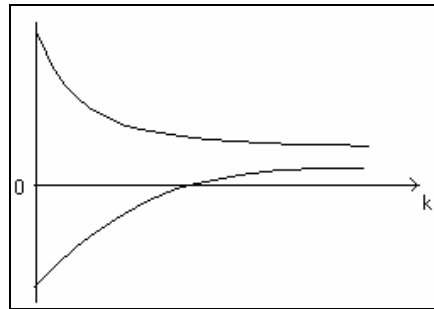
Marquard ve Snee (1975), $[0,1]$ kapalı aralığındaki k 'nın yaklaşık 25 değerini kullanmayı önermişlerdir. Çoklu iç ilişki şiddetli ise regresyon katsayılarındaki durağanlık ridge izinden görülecektir. k arttıkça bazı ridge tahminleri birbirinden oldukça farklı olacaktır. Bazı k değerlerinde ise $\hat{\beta}_R$ değerleri durağan olacaktır. Genel düşünce $\hat{\beta}_R$ tahminlerinin durağan olduğu en küçük k değerini belirlemektir. Bu k değerinin EKK' dan daha küçük MSE' ye sahip tahminleri vermesi beklenir.

3.2.1. k Yanlılık Parametresinin Seçimi

Ridge regresyonda k nın seçimi için pek çok prosedür regresyon katsayılarının tahmininde önemli bir gelişme olacak şekilde önerilir. k stokastik iken bunlardan bazıları şu şekildedir:

1. Ridge İzi (Ridge Trace): Hoerl ve Kennard (1970), k ' nın seçimi için ridge izi metodunu önermiştir. Ridge izi k ' ya karşı $\hat{\beta}(k)$ nın elemanlarının grafiğidir. k arttıkça tahminler küçülür ve belli bir k değerinden sonra durağan (stable) olurlar. $\hat{\beta}(k)$ ' nın durağan olduğu en küçük k değeri uygun bir değer olarak alınır. Hoerl ve Kennard (1970) $0 \leq k \leq 1$ aralığını önermişlerdir. Fakat her veri için bu aralık geçerli olmayabilir (Vinod ve Ullah, 1981).

Ridge izin değişkenliği (instability) çoklu iç ilişkiden kaynaklanan açıklayıcı değişkenler arası iç ilişkiyi gösterir. Örneğin, Şekil 3.1.' de zıt yönde hareket eden iki açıklayıcı değişken görülmektedir. Bunun sebebi bu iki açıklayıcı değişken arasındaki negatif ilişkidir. Buradan da görüldüğü gibi ridge izinin değişken olduğu bölge çoklu iç ilişkiden etkilenmektedir.



Şekil 3.1. Negatif ilişkili regresyon katsayılarının ridge izi (Vinod ve Ullah, 1981)

Hoerl ve Kennard (1970) farklı birimlerdeki regresyon katsayılarını karşılaştırmak amacıyla veriyi standartlaştırdıktan sonra ridge izi grafiklendirmeyi önermiştir.

2. Hoerl ve Kennard (1970) ridge tahmin edicinin EKK'dan daha küçük hata kareler

ortalamasına (MSE) sahip yanalılık parametresinin tahminini $\hat{k}_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{maks}^2}$ olarak

vermiştir. Burada $\hat{\alpha}_{maks}$, kanonik formda regresyon katsayısının tahmininin maksimum değeridir.

3. Theobald (1974) her W için $E[(\hat{\beta}(k) - \beta)'W(\hat{\beta}(k) - \beta)]$, $E[(\hat{\beta} - \beta)'W(\hat{\beta} - \beta)]$

dan daha küçük olacak şekilde yanalılık parametresinin tahminini $\hat{k}_T = \frac{2\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$ olarak

vermiştir.

4. Hoerl, Kennard ve Baldwin (1975) ridge tahmin edicinin MSE değerini minimum

yapan $k_i = \frac{\sigma^2}{\alpha_i^2}$ değerlerinin harmonik ortalaması $\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$ yı önermiştir.

Simülasyon çalışması yaparak $\hat{k}_{HKB}$ için ridge tahmin edicinin EKK dan daha iyi MSE değerine sahip olduğunu göstermişlerdir.

5. Lawless ve Wang (1976) Bayesian yaklaşımla $\hat{k}_{LW} = \frac{p\hat{\sigma}^2}{\hat{\beta}'X'X\hat{\beta}}$ tahminini

vermişlerdir. Monte Carlo simülasyon çalışması yaparak $\hat{k}_{LW}$ için ridge tahmin edicinin $\hat{k}_{HKB}$ den ve EKK' dan daha iyi bir performans sergilediğini göstermişlerdir.

3.2.2. k Yanalılık Parametresinin Seçimi İle İlgili Bir Uygulama

Aşağıdaki veri, çimentonun bileşiminin, sertleştiği sırada yaydığı ısı üzerine etkileri ile ilgili mühendislik dalında bir uygulamadır. Bu veri; Hald çimento verisi (Hald Cement Data) olarak bilinen ve oldukça sık kullanılan bir veridir. Veri 4 açıklayıcı ve 1 yanıt değişken olmak üzere aşağıdaki şekilde tanımlanmaktadır:

Y : bir gram çimentonun yaydığı ısı (kalori başına)

X_1 : trikalsiyum aliminyum (tricalcium aluminate) miktarı

X_2 : trikalsiyum silikat (tricalcium silicate) miktarı

X_3 : tetrakalsiyum aliminyum demir (tetracalcium alumino ferrite) miktarı

X_4 : dikalsiyum silikat (dicalcium silicate) miktarı

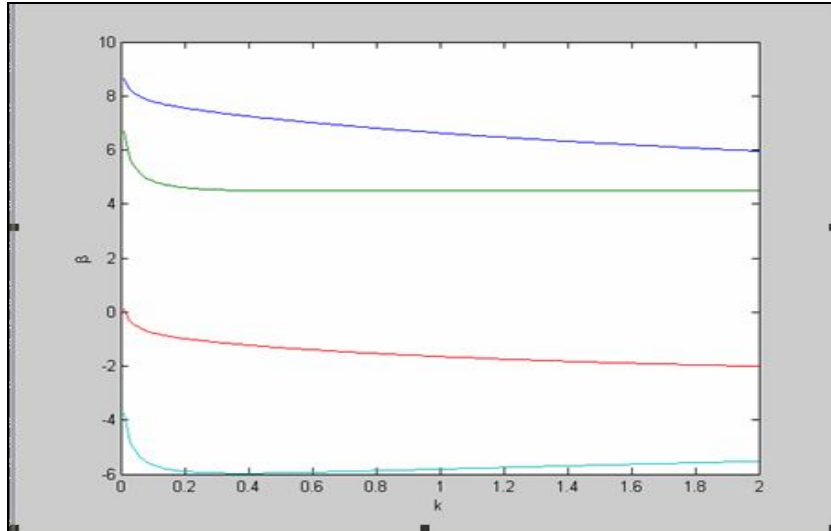
Tablo 3.1. Hald çimento verisi

(C., Montgomery, Introduction to Linear Regression Analysis, syf; 303)

Y	X_1	X_2	X_3	X_4
78,5	7	26	6	60
74,3	1	29	15	52
104,3	11	56	8	20
87,6	11	31	8	47
95,9	7	52	6	33
109,2	11	55	9	22
102,7	3	71	17	6
72,5	1	31	22	44
93,1	2	54	18	22
115,9	21	47	4	26
83,8	1	40	23	34
113,3	11	66	9	12
109,4	10	68	8	12

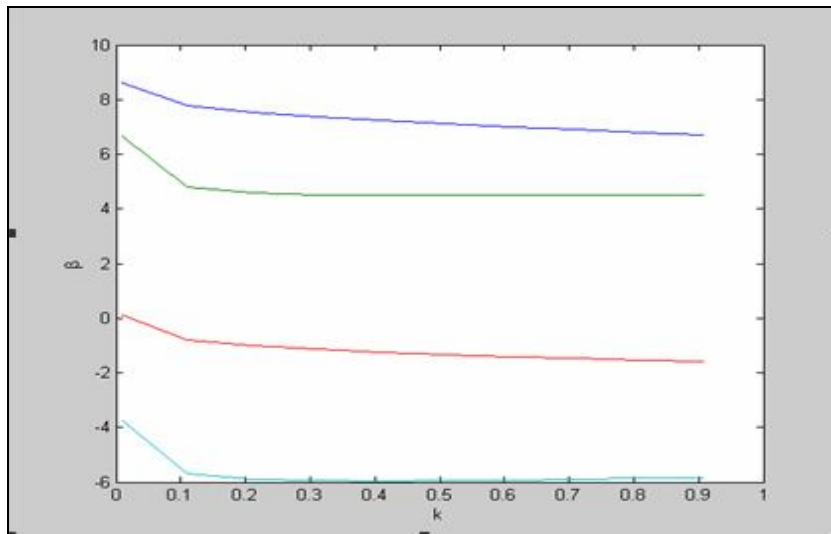
Yukarıdaki Hald çimento verisini kullanarak uygun bir k değeri seçip, bu k değerine karşılık gelen $\hat{\beta}(k)$ katsayılarını da elde edip, veriler için bir denklem oluşturalım.

Ek1’ de verilen program kodu Matlap programında çalıştırıldıktan sonra elde edilen ridge izi (ridge trace) aşağıdaki şekildeki gibidir.



Şekil 3.2. Ridge izi 1. Hald çimento verisine ait bir ridge izi (ridge trace)

Burada k değeri 0,01den başlatılarak 0,01 artırılarak ilerlenmiştir. Ridge İzine baktığımızda katsayıların durağanlaştığı nokta 0,4 civarındadır. k değeri genellikle 0 ile 1 arasında bulunması uygun olur. Bundan dolayı aralıkları azaltarak (program kodundaki değişme ile) Ridge izini yeniden oluşturursak aşağıdaki şekli elde ederiz.



Şekil 3.3. Ridge izi 2. Aralıklar azaltılarak yeniden oluşturulmuş ridge izi

Ridge izi 2.' de k değeri 0,01 den başlatılarak 0.1 arttırılmıştır. Yani 2. grafikte k değerinin aralıkları azaltılmıştır. Böylece k değeri daha net bir şekilde görülmektedir. Ridge izi 2' ye baktığımızda da k için uygun değer 0,4 olarak görülmektedir. Aşağıdaki tabloda her bir k değerine karşılık gelen $\hat{\beta}_R$ katsayıları verilmiştir. Buna göre 0,41 değerine karşılık gelen katsayılar tablonun 5. sütunundadır.

Tablo 3.2. k değerlerine karşılık gelen $\hat{\beta}_R$ katsayılarının değerleri

b =									
8.6316	7.7659	7.5369	7.3734	7.2353	7.1120	6.9993	6.8950	6.7976	6.7061
6.6724	4.8001	4.5894	4.5236	4.4990	4.4906	4.4895	4.4917	4.4954	4.4996
0.1097	-0.7991	-1.0010	-1.1317	-1.2358	-1.3249	-1.4035	-1.4741	-1.5381	-1.5966
-3.7451	-5.7032	-5.9073	-5.9571	-5.9621	-5.9488	-5.9268	-5.9003	-5.8715	-5.8415

Matlap programı ile elde edilen bu katsayılar 0,01' den başlatılarak 0,1 arttırıldığından dolayı 0,41 değerine karşılık gelen $\hat{\beta}_R$ katsayıları tablonun 5. sütunundadır. Böylece yazacağımız model aşağıdaki gibidir:

$$\hat{Y} = 7,2353X_1 + 4,4990X_2 - 1,2358X_3 - 5,9621X_4$$

Diğer taraftan k ' nın seçimi için Hoerl, Kennard ve Baldwin (1975) önerdiği

$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$ prosedürünü uygulayacak olursak elde ettiğimiz sonuçlar aşağıdaki

gibi olacaktır. $\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$ eşitliğinde $\hat{\beta}$ ve $\hat{\sigma}^2$ verinin EKK çözümüdür.

Standartlaştırdığımız verinin EKK çözümlerinden $\hat{\sigma}^2$ ve $\hat{\beta}$ 'ları elde edelim. SPSS analizi sonucunda elde ettiğimiz değerler Tablo 3.3. ve Tablo 3.4.' deki gibidir.

Tablo 3.3. SPSS programı ile elde edilen varyans analizi tablosu

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11,789	4	2,947	111,479	,000(a)
	Residual	,211	8	,026		
	Total	12,000	12			

a Predictors: (Constant), S_4, S_3, S_1, S_2

b Dependent Variable: Y_S

σ^2 'nin tahminini *MSE* (Hata kareler ortalaması) ile yapabiliriz. Veriler için elde edilen *MSE* değeri tabloda işaretlendiği gibidir. ($MSE = 0,026$)

Tablo 3.4. SPSS programı ile elde edilen katsayılar tablosu

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	,000	,045		-,006	,995			
	S_1	,606	,291	,607	2,083	,071	,731	,593	,098
	S_2	,528	,749	,528	,705	,501	,816	,242	,033
	S_3	,043	,321	,043	,135	,896	-,535	,048	,006
	S_4	-,160	,789	-,160	-,203	,844	-,821	-,072	-,010

a Dependent Variable: Y_S

(Not: Tablo 3.3. ve Tablo 3.4.' deki analizler için veriler standartlaştırılmıştır.)

Yukarıdaki tabloda standartlaştırılmış verilerin EKK çözümü olan $\hat{\beta}_{EKK}$ değerlerini görmektediriz.

Son olarak $\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$ tahminini elde etmek için, kolay bir şekilde Matlap programını kullanarak $\hat{\beta}'\hat{\beta}$ değerini 0.6735 olarak hesaplarız. Böylece;

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} = \frac{4.0,0624}{0,6735} = 0,3706$$

olarak bulunmaktadır. Bulduğumuz $\hat{k}_{HKB}$ değeri ridge izinden tahmin ettiğimiz k değerine oldukça yakındır. ($k = 0,41$ olarak tahmin edilmişti.)

3.3. Görünüşte İlişkisiz Regresyon Modeli (SURE Model)

Görünüşte ilişkisiz regresyon modeli, bazı bağımlı değişkenlerin, bağımsız değişkenlerin lineer bir fonksiyonu olarak ifade edilebildiği durumlarda ortaya çıkar ki bu da birbirlerinden bağımsız olmayan hata terimleri ile çoklu regresyon denklemlerinin oluşturduğu bir sisteme sebep olur. Bu sistemi meydana getiren çoklu regresyon denklemleri yapısal olarak ilişkisiz gibi görünseler de (eşanlı denklem sistemi özelliğinde olmasalar da) aşağıdaki sebeplerden dolayı istatistiksel olarak ilişkili olabilirler.

- Bazı katsayılar denklemler arasında ortak kullanılmaktadır.
- Aynı zaman noktasında, denklemlerdeki karşılıklı hata terimleri ilişkilidir.
- Denklemlerdeki bağımsız değişkenlerin alt kümesi aynıdır, bununla beraber tüm gözlemler aynı değildir. Başka bir ifade ile denklemlerdeki bağımsız değişkenler aynı olmakla beraber bu bağımsız değişken değerleri aynı değildir.

Sistemde yer alan denklemlerin birbirleri ile ilişkileri, denklemlere ait hata terimlerinin ilişkili olmasından kaynaklanmaktadır. Karşılıklı hataları ilişkili denklemler sisteminin parametre tahminleri yapılırken sistem bir bütün olarak ele alınır. Hata terimleri ilişkili denklemlerin aynı zamanda birlikte ele alınarak parametre tahminlerinin yapıldığı çok değişkenli regresyon yöntemi Zellner (1962) tarafından öne sürülen “Görünüşte İlişkisiz Regresyon” yöntemidir. Bu yönteminin temeli Genelleştirilmiş En Küçük Kareler yöntemine dayanır.

İlişkili olan denklemler sistemine ayrı ayrı EKK yöntemi uygulanırsa yansız ve tutarlı ancak etkin olmayan tahminler elde edilir ki bu da yanlış istatistiksel sonuç çıkarımlarına sebep olur. Bu yüzden hata terimleri ilişkili bu denklemler sisteminin parametre tahminlerini yaparken sistemi bir bütün olarak ele alan ve hata terimi varyans-kovaryans matrisini de regresyona dahil eden SURE yöntemini kullanmak yansız, tutarlı ve etkin tahmin ediciler elde edilmesini sağlar (Dogan,1998).

Hata terimleri ilişkili bu tür regresyon denklemlerine iktisadi modellerde sık sık rastlanmaktadır. Çeşitli malların talep fonksiyonları veya çeşitli endüstri dallarının üretim fonksiyonlarında hata terimleri ilişkili denklemler görülebilmektedir. Örneğin; A malı talep fonksiyonu hata terimi; B, C malları talep fonksiyonları hata terimleri ile ilişkili olabilir. Ayrıca, bağımlı ve bağımsız değişken verilerinin zaman serisi veya anket verisi olduğu durumlarda da SURE modeli ile karşılaşılabilmektedir. Benzer biçimde biyokimya, göç ve nüfus hareketleri, farmakoloji vb. gibi konularda SURE modellerine sık başvurulur.

3.3.1. SURE Model ve Varsayımları

(3.3) ile gösterilen çok değişkenli çoklu lineer regresyon modeli aşağıdaki formda yeniden düzenlenmektedir.

$$y = Z\beta + e \quad (3.20)$$

Burada y , $qn \times 1$ tipinde y vektörü, hata matrisi olan ε , $qn \times 1$ tipinde e vektörü olarak düzenlenmiştir. Bilinmeyen β parametreleri $pq \times 1$ tipinde bir vektör ve Z , $q(n \times p)$ tipinde köşegeni üzerinde orijinal X_j , $j = 1, 2, \dots, q$, matrisinin olduğu bir blok matristir. $\otimes$ matrislerin Kroneker çarpımını göstermek üzere $Z = I_q \otimes X_j$ dir. Hata vektörü 0 (sıfır) ortalamalı ve $\Sigma = \{\sigma_{ij}^2\}$ $q \times q$ tipinde pozitif tanımlı simetrik bir matris olmak üzere dağılım matrisidir (dispersion matrix).

$$E(ee') = \Sigma \otimes I_n = \Gamma$$

Regresyon modeli (3.20) formundaki modeller, görünüşte ilişkisiz model (SURE) olarak adlandırılır. Model (3.20)' deki eşitlikler yapısal olarak ilişkisiz görünseler bile, (herhangi bir eşitlikteki bağımlı değişken başka bir eşitlikte regresör gibi görünmüyor.) gerçekten çapraz eşitliklerle ilişkilendirilen hatalar onların arasında bir bağlantı oluşturur.

SURE modelinin varsayımlarını aşağıdaki şekilde özetleyebiliriz:

$$\bullet \quad Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = \sigma_{ij} \quad \begin{matrix} i \neq j & i = 1, 2, \dots, p \\ & j = 1, 2, \dots, q \end{matrix}$$

Varsayımına göre denklemlerin hata terimleri arasında ilişki vardır.

$$\bullet \quad Cov(\varepsilon_i, \varepsilon_i) = E(\varepsilon_i \varepsilon_i) = \sigma_{ii}$$

Varsayımına göre modeli meydana getiren her bir denklemde sabit varyanslılık varsayımı geçerlidir.

- Her bir denklemin hata terimleri normal dağılıma sahiptir.
- $E(\varepsilon_i) = 0$, her bir denklemin hata terimlerinin beklenen değeri sıfıra eşittir.

Bu varsayımlar altında i . ve j . denklemler arasındaki varyans-kovaryans matrisi aşağıdaki gibidir:

$$\begin{aligned} Cov(\varepsilon_i, \varepsilon_j) &= E(\varepsilon_i \varepsilon_j') = E \left[\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \vdots \\ \varepsilon_{in} \end{pmatrix} (\varepsilon_{j1} \ \varepsilon_{j2} \ \dots \ \varepsilon_{jn}) \right] \\ &= E \begin{pmatrix} \varepsilon_{i1} \varepsilon_{j1} & \varepsilon_{i1} \varepsilon_{j2} & \dots & \varepsilon_{i1} \varepsilon_{jn} \\ \varepsilon_{i2} \varepsilon_{j1} & \varepsilon_{i2} \varepsilon_{j2} & \dots & \varepsilon_{i2} \varepsilon_{jn} \\ \vdots & \vdots & & \vdots \\ \varepsilon_{in} \varepsilon_{j1} & \varepsilon_{in} \varepsilon_{j2} & \dots & \varepsilon_{in} \varepsilon_{jn} \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
 &= \begin{bmatrix} \sigma_{ij} & 0 & \dots & 0 \\ 0 & \sigma_{ij} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \sigma_{ij} \end{bmatrix} \\
 &= \sigma_{ij} I_n \quad i, j = 1, 2, \dots, n \quad (3.21)
 \end{aligned}$$

Sonuç olarak N denklemlili bir SURE modelde varyans-kovaryans matrisi Ω olmak üzere aşağıdaki şekildedir:

$$Cov(\varepsilon) = E(\varepsilon\varepsilon') = \Omega$$

$$= \begin{bmatrix} E(\varepsilon_1\varepsilon_1') & E(\varepsilon_1\varepsilon_2') & \dots & E(\varepsilon_1\varepsilon_N') \\ E(\varepsilon_2\varepsilon_1') & E(\varepsilon_2\varepsilon_2') & \dots & E(\varepsilon_2\varepsilon_N') \\ \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_N\varepsilon_1') & E(\varepsilon_N\varepsilon_2') & \dots & E(\varepsilon_N\varepsilon_N') \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11}I_{n \times n} & \sigma_{12}I_{n \times n} & \dots & \sigma_{1N}I_{n \times n} \\ \sigma_{21}I_{n \times n} & \sigma_{22}I_{n \times n} & \dots & \sigma_{2N}I_{n \times n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1}I_{n \times n} & \sigma_{N2}I_{n \times n} & \dots & \sigma_{NN}I_{n \times n} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 & \sigma_{12} & 0 & \dots & 0 & \dots & \dots \\ 0 & \sigma_{11} & \dots & 0 & 0 & \sigma_{12} & \dots & 0 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_{11} & 0 & 0 & \dots & \sigma_{12} & \dots & \dots \\ \sigma_{12} & 0 & \dots & 0 & \sigma_{22} & 0 & \dots & 0 & \dots & \dots \\ 0 & \sigma_{12} & \dots & 0 & 0 & \sigma_{22} & \dots & 0 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_{12} & 0 & 0 & \dots & \sigma_{22} & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots & & \ddots \\ \sigma_{1N} & 0 & \dots & 0 & & & & & & \\ 0 & \sigma_{1N} & \dots & 0 & & & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & & & \\ 0 & \vdots & \vdots & \sigma_{1N} & \dots & \dots & & & & \end{bmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix} \otimes I_{n \times n} \\
&= \Sigma \otimes I_{n \times n}
\end{aligned} \tag{3.22}$$

3.3.2. SURE Modelin Ridge Tahmin Edicisi (OR)

(3.3) ile gösterilen lineer regresyon sistemi yeniden düzenlenerek;

$$y = Z\beta + e$$

SURE model olarak adlandırılan yeni bir form elde edilmiştir. Burada matris formları Bölüm 3.3.1.' de belirtildiği gibidir. $y = Z\beta + e$ formu, $E(ee') = \Sigma \otimes I_n = \Gamma$ varyans-kovaryans matrisli olup, $\Sigma = \{\sigma_{ij}^2\}$ $q \times q$ tipinde pozitif tanımlı simetrik bir matristir. SURE model (3.20)' deki β 'nın ridge tahmin edicisi (OR);

$$\hat{\beta}_{OR} = (Z'\Gamma^{-1}Z + kI)^{-1}Z'\Gamma^{-1}y \tag{3.23}$$

ile tanımlanmaktadır. (Srivastava ve Giles, 1987).

$\hat{\beta}_{OR}$ ' yi elde etmek için iki farklı yol vardır. Bu yollardan biri $(y - Z\beta)'\Gamma^{-1}(y - Z\beta)$ ile gösterilen ağırlıklandırılmış hata karelerin verilen bir seviyesinde $\beta'\beta$ 'nın minimize edilmesi, diğeri ise kısıtlı regresyon probleminin bir çözümüdür.

3.4. Genelleştirilmiş Ridge Tahmin Edici (GR)

Hoerl ve Kennard (1970) Ridge regresyon prosedürünün, her bir bağımsız değişken için ayrı ayrı yanlılık parametresi kullanımını sağlayan bir genişlemesini kullanmışlardır. Genelleştirilmiş ridge regresyon (GR); verinin ortogonal değişken uzayına dönüşümü ile ilgilenir.

$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, köşegen elemanları $X'X$ ' in özdeğerleri olan bir köşegen matris ve T , $T'X'XT = \Lambda$ ve $TT' = T'T = I_p$ koşullarını sağlayan $X'X$ ' in özvektörlerinden oluşan bir $p \times p$ tipinde matris olsun. Böylece çoklu lineer regresyon modeli:

$$\begin{aligned} Y &= X\beta + \varepsilon, & \varepsilon &\sim N(0, \sigma^2 I) \\ Y &= Z\alpha + \varepsilon, & \varepsilon &\sim N(0, \sigma^2 I) \end{aligned} \quad (3.24)$$

şeklinde yeniden yazılabilir. Burada $Z = XT$ ve $\alpha = T'\beta$ ' dir. Ayrıca $Z'Z = T'X'XT = \Lambda$ olduğu da görülmektedir. Dolayısıyla $\alpha = T'\beta$ bağıntısından faydalanarak istenildiğinde kanonik formda tahmin edilen parametrelerden orijinal parametrelere geçilebilir.

$$Y = Z\alpha + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I)$$

kanonik formunu ele alalım. Burada α ' nın EKK tahmin edicisi;

$$\begin{aligned} \hat{\alpha}_{EKK} &= (Z'Z)^{-1} Z'Y \\ &= \Lambda^{-1} Z'Y \end{aligned} \quad (3.25)$$

olarak tahmin edilir. (3.24) formunun GR tahmin edicisi ise;

$$\begin{aligned} \hat{\alpha}_{GR}(K) &= (\Lambda + K)^{-1} Z'Y \\ &= (\Lambda + K)^{-1} \Lambda \hat{\alpha}_{EKK} \\ &= [I - K(\Lambda + K)^{-1}] \hat{\alpha}_{EKK} \end{aligned} \quad (3.26)$$

$\hat{\alpha}_{EKK}$ ile $\hat{\alpha}_{GR}(K)$, arasındaki ilişki aşağıdaki şekilde tanımlanmaktadır.

$$\hat{\alpha}_{GR}(K) = (I - KA^{-1}) \hat{\alpha}_{EKK} \quad (3.27)$$

Burada; $A = \Lambda + K$ ve $K = \text{diag}(k_1, k_2, \dots, k_p)$, $k_i \geq 0$ $i = 1, 2, \dots, p$ ' dir.

$\alpha = T' \beta$ olduğundan β ' nın GR tahmin edicisi:

$$\begin{aligned}\hat{\beta}_{GR}(K) &= T \hat{\alpha}_{GR}(K) \\ &= T A^{-1} Z' Y \\ &= (X' X + T K T')^{-1} X' Y\end{aligned}\quad (3.28)$$

olarak elde edilir.

$\hat{\alpha}_{GR}(K)$, kanonik formda regresyon katsayılarının tahmini olmak üzere

$$k_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_{GR}^2(i)}, \quad i = 1, \dots, p \quad \text{genelleştirilmiş ridge tahmin edicinin HKO değerini}$$

minimum yapar. Hoerl ve Kennard (1970), k_i değerlerinin ardışık seçim yöntemini vermiştir. k_i nin ilk tahmini için α ' nın EKK tahmini ile başlanır:

$$k_i^0 = \frac{\hat{\sigma}^2}{\hat{\alpha}_{EKK}^2(i)}, \quad i = 1, \dots, p \quad (3.29)$$

k_i^0 kullanılarak $K = \text{diag}(k_1^0, \dots, k_p^0)$ olmak üzere başlangıç genelleştirilmiş ridge tahmin edici ($\hat{\alpha}_{GR}^0$) hesaplanır ve bunun yardımıyla bir sonraki k_i değeri bulunur:

$$k_i^1 = \frac{\hat{\sigma}^2}{(\hat{\alpha}_{GR}^0(i))^2}, \quad i = 1, \dots, p. \quad (3.30)$$

Bu şekilde durağan parametre tahminleri elde edilene kadar iterasyona devam edilir. Durağanlığın bir ölçüsü olarak regresyon katsayılarının tahminlerinin uzunluklarının karesi kullanılır. Eğer $i - 1$ den i adımına geçerken parametre tahminlerinin boylarının karesinde herhangi bir değişiklik olmuyorsa iterasyon sonlandırılır. Aksi halde devam edilir.

Genelleştirilmiş ridge tahmin edicide farklı k_i değerleri olduğundan k_i ' lerin seçimi için ridge izi (ridge trace) kullanılamaz.

3.5. Modified Ridge Tahmin Edici (MR)

Swindel (1976) ridge tahmin edici ile ön (prior) bilgiyi birleştirerek modified ridge tahmin ediciyi önermiştir:

$$\hat{\beta}(k, b) = (X'X + kI)^{-1}(X'y + kb), \quad k \geq 0 \quad (3.31)$$

Burada b , p boyutlu stokastik olmayan bir vektör olup β üzerindeki ön bilgiyi temsil edecek şekilde seçilmelidir. $\hat{\beta}(k, b)$ 'nin bazı özelliklerini ise şu şekilde vermiştir:

i) $\hat{\beta}(k, b)$, aynı rezidü kareler toplamına sahip β 'nin tahmin ediciler sınıfı içerisinde b 'ye en yakın olandır. Yani $(y - X\tilde{\beta})'(y - X\tilde{\beta}) = c$ olacak şekilde $(\tilde{\beta} - b)'(\tilde{\beta} - b)$ yi minimum yapmak isteriz. $1/k$ Lagrange çarpanı olmak üzere

$$\Psi_1 = (\tilde{\beta} - b)'(\tilde{\beta} - b) + \frac{1}{k}[(y - X\tilde{\beta})'(y - X\tilde{\beta}) - c]$$

nın $\tilde{\beta}$ ya göre türevini alıp sıfıra eşitlersek (3.32) ile verilen $\hat{\beta}(k, b)$ yi elde ederiz.

ii) $\hat{\beta}(k, b)$, b ye eşit uzaklıktaki β 'nin tahmin ediciler sınıfı içerisinde en küçük rezidü kareler toplamına sahiptir. Yani $\hat{\beta}(k, b)$, $(\tilde{\beta} - b)'(\tilde{\beta} - b) = c$ olacak şekilde $(y - X\tilde{\beta})'(y - X\tilde{\beta})$ yi minimum yapan $\tilde{\beta}$ tahmin edicisidir.

$$\Psi_2 = (y - X\tilde{\beta})'(y - X\tilde{\beta}) + k[(\tilde{\beta} - b)'(\tilde{\beta} - b) - c]$$

nın $\tilde{\beta}$ 'ya göre türevini alıp sıfıra eşitlersek (3.31) ile verilen $\hat{\beta}(k, b)$ 'yi elde ederiz.

iii) $\hat{\beta}(0, b) = \hat{\beta}$ dır ve $\lim_{k \rightarrow \infty} \hat{\beta}(k, b) = b$ olduğundan k arttıkça $\hat{\beta}(k, b)$, $\hat{\beta}$ dan b ye bir eğri çizer. Bu eğri rezidü kareler toplamında küçük bir artış olduğunda $\hat{\beta}(k, b)$ nin b ye olan uzaklığında büyük bir azalma olacak şekilde tek türlü belirlenir.

Swindel (1976) $\hat{\beta}(k, b)$ nin EKK dan daha küçük MSE değerine sahip olacak şekilde bir k ' nin varlığını göstermiştir. Ayrıca b , $\hat{\beta}$ ' dan bağımsız bir rasgele vektör iken de böyle bir k ' nin var olduğunu ispatlamıştır. b , $\hat{\beta}$ ' dan bağımsız değilse böyle bir k ' nin varlığı ise garanti değildir.

3.5.1. SURE Modelde Modified Ridge Tahmin Edici (MR)

Fule (1994), tahminleri orijin yönünde büzmek yerine, β_0 gibi verilen bir parametre vektörü yönünde büzme yöntemini ortaya koymuştur. Yöntem $(y - Z\beta)' \Gamma^{-1} (Y - Z\beta)$ ile gösterilen ağırlıklandırılmış hata karelerin verilen bir seviyesinde $\beta' \beta$ ' nin minimize edilmesi ile başlamaktadır. Burada, stokastik olmayan vektör β_0 formunda, β üzerinde ön bilgiye sahip olduğu varsayılmaktadır.

Minimize etme yaklaşımı düşünüldüğünde, önceki ön bilgiye dayanarak bazı değişkenlere diğerlerinden daha fazla ağırlık verildiği varsayılır. O zaman minimize etmek, orijinden başka verilen bir noktaya uzaklığın minimize edilmesi şeklinde olacaktır.

Böylece $\beta' \beta$ yı minimize etmek yerine, $(y - Z\beta)' \Gamma^{-1} (Y - Z\beta) = \lambda$, sabit kısıtına bağlı, uzaklığın ağırlıklandırılmış karesi olan $(\beta - \beta_0)' \Theta^{-1} (\beta - \beta_0)$ ' nin minimize edilmesi seçilir. Burada ağırlık matrisi $\Theta^{-1/2}$, Θ^{-1} bilinen pozitif elemanlı köşegen matristir.

1/k çarpanıyla Langrange çarpanının gösterilmesiyle ve türevin alınmasıyla SURE modelde, β ' nin modified ridge tahmin edicisi;

$$\hat{\beta}_{MR} = (Z' \Gamma^{-1} Z + k \Theta^{-1})^{-1} (Z \Gamma^{-1} y + k \Theta^{-1} \beta_0) \quad (3.32)$$

şeklinde elde edilir.

Modified Ridge tahmin ediciyi elde etmenin ikinci yolu incelendiğinde, her iki bilginin kaynaklarının birleştirildiği görülür. Temel SURE model (3.20) ve $\beta_0 = \beta + v$ stokastik kısıtlaması ile, stokastik bileşen v , $E(v) = 0$, $E(vv') = k^{-1}\Theta$ ve $k > 0$ olmak üzere bir tek modelde birleştirdiğinde,

$$y^* = \begin{bmatrix} y \\ \beta_0 \end{bmatrix} = \begin{bmatrix} Z \\ I \end{bmatrix} \beta + \begin{bmatrix} e \\ v \end{bmatrix} = Z^* \beta + e^* \quad (3.33)$$

elde edilmektedir. Burada e ve v ' nin bağımsız oldukları varsayılır ve

$$PP' = \Omega^{-1} = \begin{bmatrix} \Gamma & 0 \\ 0 & k^{-1}\Theta \end{bmatrix}^{-1} \quad (3.34)$$

$P'\Omega P = I$ olacak şekildeki P dönüşüm matrisi kullanılır. P matrisini kullanarak (3.34) eşitliği;

$$P' y^* = P' Z^* \beta + P' e^* \quad (3.35)$$

eşitliğine dönüştürülür ve EKK yöntemi bu dönüştürülmüş modele uygulandığında (3.32) ile gösterilen tahmin edici elde edilmektedir.

3.5.2. SURE Modelde Modified Genelleştirilmiş Ridge Tahmin Edici (MGR)

Ekolojik regresyonda EKK tahmin edicisinden kaynaklanan problemlerden kaçınmak ve mümkün olmayan bazı varsayımlarından dolayı bir çok yöntem geliştirilmiştir. Bunlardan biri de Fule (1994)' nin geliştirdiği (3.32) ile gösterilen tahmin edicinin genelleştirilmiş halidir.

$X_j = X \quad j = 1, 2, \dots, q \quad Z = I_q \otimes X$ ise, (3.1) ile gösterilen çoklu lineer regresyondaki denklem sisteminin tahminin etkinliğinde hiç bir kazanç olmayacaktır. Çünkü denklemlerin her biri ayrı ayrı, eş zamanlı bir şekilde tahminlerin yüksek olması problemine neden olacaktır (over estimating). Dahası, $\Sigma = \text{diag}\{\sigma_{ii}^2\}$ varsayarak, Γ_0^{-1} ($pq \times pq$) tipinde bir köşegen matris için $Z'\Gamma^{-1}Z$ ifadesi $Z'Z\Gamma_0^{-1}$ ifadesine eşit olacaktır ve $Z'\Gamma^{-1}y$ ifadesi de $\Gamma_0^{-1}Z'y$ gibi yazılabilecektir. Yani;

$$\Gamma_0^{-1} = \Sigma = \text{diag}\{\sigma_{ii}^2\} \text{ olduğundan dolayı,}$$

$$Z'\Gamma^{-1}Z = Z'Z\Gamma_0^{-1},$$

$$Z'\Gamma^{-1}y = \Gamma_0^{-1}Z'y$$

eşitlikleri vardır. Bu eşitliğin kullanılmasıyla modified ridge tahmin edici, $k\Theta^{-1}\Gamma_0$ ifadesiyle ($pq \times pq$) tipindeki köşegen K matrisi tarafından yeniden düzenlenerek,

$$\hat{\beta}_{MR} = (\underbrace{Z'\Gamma^{-1}Z}_{Z'Z\Gamma_0^{-1}} + k\Theta^{-1})^{-1}(\underbrace{Z'\Gamma^{-1}y}_{\Gamma_0^{-1}Z'y} + k\Theta^{-1}\beta_0)$$

$$\hat{\beta}_{MR} = (Z'Z\Gamma_0^{-1} + \underbrace{k\Theta^{-1}}_K)^{-1}(\Gamma_0^{-1}Z'y + \underbrace{k\Theta^{-1}\beta_0}_K)$$

$$\hat{\beta}_{MGR} = (Z'Z + K)^{-1}(Z'y + K\beta_0) = (Z'Z + K)^{-1}(Z'Z\hat{\beta}_{EKK} + K\beta_0) \quad (3.36)$$

olarak modified genelleştirilmiş ridge tahmin edici (MGR) olarak adlandırılan yeni bir tahmin edici elde edilir.

Burada, EKK tahmin edicisi $\hat{\beta}_{EKK} = (Z'Z)^{-1}Z'y$ olarak tanımlanmaktadır. MGR tahmin edici, değişik shrinkage parametrelerinin her birini, $\hat{\beta}_{EKK}$ ile bağlantı kurmaya izin veren genelleştirilmiş ridge (GR) tahmin edicinin bir formudur. MGR tahmin edici, β_0 ' ın önceden belirlenmiş değerleri yönünde EKK ile tahmin edilmiş parametrelerin büzüştüğü modified ridge (MR) tahmin edici olarak da göz önünde bulundurulabilir.

3.5.3. MGR Tahmin Edicinin Uygulama Şekli

(3.36) ile gösterilen MGR tahmin edicinin uygulama şekli iki faktörle ortak olarak hesaplanmaktadır. Bunlar;

- (i) K matrisinin elemanlarıyla hesaplanmış shrinkage derecesi,
- (ii) β_0 vektörü yönünde EKK tahmin edicisinin büzüştürdüğü noktadır.

Shrinkage parametre matrisi olan K genellikle bilinmez ve (3.32) ile gösterilen denklemde Θ ve Γ matrisleri ile gösterilen kovaryans yapıları hakkında hiçbir bilginiz yoktur. Pratikte onları verilerden elde etmeye çalışırız. Shrinkage parametre matrisi K 'nin yapısı; $K = I_q \otimes K_j$ şeklindedir. Burada K_j , $j = 1, 2, \dots, q$ her biri j . eşitliğe uyan p satırın köşegen matrisidir.

Böylece Hoerl, Kennard ve Baldwin (1975) tarafından önerilen, $K = \text{diag}\{k_l\}$ $l = 1, 2, \dots, pq$ olan shrinkage parametrelerinin genel bir şekli aşağıdaki gibi ifade edilmiştir.

$$\hat{k}_i = \frac{pq\hat{\sigma}^2}{\left(\hat{\beta}_{EKK(i)} - \beta_{0(i)}\right)} \quad (3.37)$$

Burada $\hat{\beta}_{EKK(i)}$ EKK tahminini, $\beta_{0(i)}$ uygun olan amaç değerler ve $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = \frac{(y - Z\hat{\beta}_{EKK})'(y - Z\hat{\beta}_{EKK})}{q(n - p)} \quad (3.38)$$

ile verilen σ^2 'nin yansız tahmin edicisidir.

Amaç değerlerin vektörü β_0 , shrinkage yönteminin adımlarıyla elde edilmektedir. EKK tahminleri başlangıç değerleri gibi kullanılmaktadır. $\beta_0 = c\hat{\beta}_{EKK}$ $0 < c < 1$ seçilerek shrinkage için, β_0 değerlerine yakın, beklenen değerleri içeren uzay oluşturulur.

(3.36) ile gösterilen tahmin edici; geliştirilmiş bir tahmin edici için sınırların etkilerinin olup olmadığını görmeye bize test uygulayabilmemizi sağlayan, kısıtlı bir tahmin edicidir. (3.36) ile gösterilen tahmin edicinin hata kareler ortalaması (MSE) EKK tahmin edicisinden daha düşüktür. Baksalary (1990), EKK tahmin edicisinin üzerine Ridge tahmin edicisinin baskınlığını test etmek için F istatistiğinin uygulamaları üzerine çalışmıştır.

Benzer olarak MSE ile (3.36) ile gösterilen MGR tahminci karşılaştırılması bizi aşağıdaki eşitliğe götürür.

$$F = \frac{(\hat{\beta}_{EKK} - \beta_0)'[(Z'Z)^{-1} + K]^{-1}(\hat{\beta}_{EKK} - \beta_0)/v_1}{(y - Z\hat{\beta}_{EKK})'(y - Z\hat{\beta}_{EKK})/v_2} \quad (3.39)$$

Yukarıdaki F istatistiği dağılımının, y 'nin çoklu normal dağılımı varsayımı altında, merkezi olmama parametresi:

$$\theta = \frac{(\beta - \beta_0)'[(Z'Z)^{-1} + K]^{-1}(\beta - \beta_0)}{2\sigma^2} \quad (3.40)$$

ile tanımlanmaktadır. (3.39) ile gösterilen istatistik 0,5 merkezi olmama parametresi (ya da $\frac{pq}{2}$ zayıf kriter ile verilir, Wallace (1972)) ile merkezi olmayan F dağılımıdır ve $v_1 = qp$, $v_2 = q(n - p)$ serbestlik derecelidir. Bu istatistik $[y, Z\beta, \sigma^2(I_{n \times q} + ZKZ')]$ modeli altında $H_0 : \beta = \beta_0$ hipotez testini sağlar.

İki MGR tahmin edici karşılaştırılırken farklılıklar iki kaynaktan doğabilir. Bunlar farklı β_0 ve farklı K değeridir. K 'nın seçimini ve (3.37) ile verilen tahminini değiştirdiğimizde farklılık sadece β_0 seçimleri üzerine olacaktır. Bilinmeyen (ya da bir şekilde önce seçilen) β_0 değerlerini hesaplayabilmek için, (3.39) ile verilen F istatistiğini uygulayabiliriz.

Eğer MSE' de bir gelişme sağlanırsa, K ve β_0 'ın tüm kombinasyonlarını test edebiliriz ya da tam tersi MSE' de önemli bir gelişme sağlayacak β_0 ve K 'nın kombinasyonlarını bulabiliriz.

3.6. Kısıtlı Ridge Tahmin Edici (RRR)

Parametreler üzerindeki kısıtlamalar doğru olduğunda, orijinal parametrelerin yeniden düzenlenmiş halleri elde edilir. Yeniden düzenlenmiş modelde hala çoklu iç ilişki problemi varsa kısıtlı EKK tahmin edicisinin Ridge tahmin edicisi doğrultusunda düzenlenmesi gerekmektedir. Başka bir deyişle Ridge regresyon düşüncesinin, kısıtlı EKK tahmin edicisine uygulanmasıyla yeni bir tahmin edici olan kısıtlı ridge tahmin edicisi (RRR) oluşturulur.

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I)$$

lineer regresyon modelini ele alalım. İki veya daha fazla değişkenin; modelde çoklu iç ilişki problemi yaratacak şekilde ilişkili olduğunu kabul edelim.

$\hat{\beta}_r$, β 'nın kısıtlı EKK tahmin edicisi olmak üzere aşağıdaki şekilde tanımlanmaktadır.

Parametreler üzerindeki kısıtlamalar kümesi:

$$R\beta = r \quad (3.41)$$

olsun. Burada r , $q \times 1$ tipinde bilinen elemanlar vektörü; R ise $q \times q$ tipinde ayrı ayrı parametreler veya parametrelerin bazı lineer kombinasyonlarının yapısı ile ilgili bilgiyi veren ön bilgi matrisidir. Böylece kısıtlı EKK tahmin edicisi;

$$\hat{\beta}_r = \hat{\beta}_{EKK} - (X'X)^{-1} R^{-1} [R(X'X)^{-1} R']^{-1} (R\hat{\beta}_{EKK} - r) \quad (3.42)$$

olarak tahmin edilir.

Sarkar (1992) ridge tahmin edici ile kısıtlı EKK tahmin edicilerini birleştirerek

$$\hat{\beta}_{RRR}(k) = W\hat{\beta}_r, \quad W = (I_p + k(X'X)^{-1})^{-1}, \quad k \geq 0 \quad (3.43)$$

ile verilen kısıtlı ridge tahmin ediciyi tanımlamıştır. Burada $\hat{\beta}_r$, (3.42) ile verilen kısıtlı EKK tahmin edicidir ve (3.41) ile verilen kısıtlamalar geçerlidir. $\hat{\beta}_{RRR}(k)$ 'nin bazı özellikleri ise $\hat{\beta}_{RRR}(0) = \hat{\beta}_r$, $\lim_{k \rightarrow \infty} \hat{\beta}_{RRR}(k) = 0$ ve

$$E(\hat{\beta}_{RRR}(k)) = W\beta + W(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\beta) \quad (3.44)$$

şeklinde dir. (3.44) dikkate alınırsa $k=0$ ve $R\beta=r$ olmadığı sürece kısıtlı ridge tahmin edici yanlı bir tahmin edicidir. Sarkar (1992) kısıtlamalar hem tam olarak sağlanıyor iken hem de sağlanmıyorken MSE anlamında kısıtlı ridge tahmin edicinin ridge ve kısıtlı EKK dan iyi olması için gerekli koşulları vermiştir.

Gross (2003), $\hat{\beta}_{RRR}(k)$ 'nin y 'nin her sonucu için $R\hat{\beta}_{RRR}(k)=r$ kısıtlamasını sağlamadığını ve bu nedenle $R\beta=r$ kısıtlamasına göre bir kısıtlı ridge tahmin edici olmadığını belirtmiştir. Dolayısıyla $\hat{\beta}_{RRR}(k)$, $R\beta=r$ eşitliği ile üretilen alt uzayın dışında değerler alabilir.

Bu düşünceler, Gross'un (2003) β için yeni bir ridge-tipi tahmin edici vermesine neden olmuştur. Bu tahmin ediciyi $S_k = X'X + kI_p$, $\beta_0 = R'(RR')^{-1}r$ ve $\hat{\beta}(k, \beta_0) = S_k^{-1}(X'y + kR'(RR')^{-1}r)$ olmak üzere;

$$\hat{\beta}_r(k) = \hat{\beta}(k, \beta_0) - S_k^{-1}R'[RS_k^{-1}R']^{-1}(R\hat{\beta}(k, \beta_0) - r), \quad k \geq 0 \quad (3.45)$$

olarak vermiş ve kısıtlı ridge tahmin edici olarak adlandırmıştır. Burada $\hat{\beta}(k, \beta_0)$, (3.32) ile verilen modified ridge tahmin edicide $b = \beta_0$ alınarak elde edilmiştir. $\lim_{k \rightarrow \infty} \hat{\beta}_r(k) = \beta_0$ ve $\lim_{k \rightarrow 0} \hat{\beta}_r(k) = \hat{\beta}_r$ olduğundan k arttıkça $\hat{\beta}_r(k)$, $\hat{\beta}_r$ den β_0 a bir eğri çizmektedir. Gross (2003) önerdiği kısıtlı ridge tahmin edicinin kısıtlı EKK tahmin ediciden matris MSE anlamında üstün olması için gerek ve yeter koşulları vermiştir. Ayrıca $Q = I_p - R'(RR')^{-1}R$ olmak üzere $QX'XQ = Q$ iken $\hat{\beta}_r(k)$ yi minimum yapan k değerini $k = \frac{(p-J)\sigma^2}{\hat{\beta}_r' Q \hat{\beta}_r}$ olarak bulmuştur. $QX'XQ = Q$ sağlanmasa bile k 'nın bir tahmin edicisini $\hat{\sigma}_r^2 = \frac{(y - X\hat{\beta}_r)'(y - X\hat{\beta}_r)}{n - (p - J)}$ olmak üzere

$$\hat{k} = \frac{(p - J)\hat{\sigma}_r^2}{\hat{\beta}_r' Q \hat{\beta}_r} \quad (3.46)$$

olarak önermiştir. Herhangi bir kısıtlama yoksa (3.46), Hoerl, Kennard ve Baldwin'in (1975) önerdiği $\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$ ye eşit olmaktadır. Gross (2003), (3.46) da $\hat{\beta}_r' Q \hat{\beta}_r$ yerine $\hat{\beta}_r(k)' Q \hat{\beta}_r(k)$ alarak $\hat{k}$ 'nın iteratif yöntemle belirlenmesini de önermiştir.

3.6.1. SURE Modelde Kısıtlı Genelleştirilmiş Modified Ridge Tahmin Edici (RMGR)

Fule (1994)' de (3.36) ile gösterilen β 'nın modified genelleştirilmiş ridge tahmin edicisi elde edilmiştir. Burada β_0 'ın önceden belirlenmiş değerleri yönünde EKK ile tahmin edilmiş parametreler büzülmektedir (shrinkage).

Ancak önceden belirlenmiş β_0 değerleri yönünde parametre değerlerini büzmek, orijinal modelin birçok önemli özelliklerini kaybetmesine neden olmaktadır. Çünkü geçiş matrisinin satır toplamaları bire eşit olmalı ve elemanların her biri geçiş olasılıkları olarak yorumlanabilir olmalıdır. β parametre vektörü üzerine (3.41) ile gösterilen,

$$R\beta = r$$

lineer eşitlikli kısıt tanımlanmaktadır. Burada her ikisi de bilinen sabitler olmak üzere; R , $(p \times pq)$ tipinde bir matris ve r , $(p \times 1)$ tipinde bir vektördür. Regresyon parametrelerinin geçiş olasılıkları olarak yorumlanabilmesi durumunu örneklendirebilmek için, (3.3)' deki lineer regresyon denklemindeki β matrisinin sadece birleri içeren r vektör gibi gösterilerek (satır toplamaları), satır toplamaları üzerine kısıtlamalar getirilmektedir. c' birlerden oluşan satır vektörü olmak üzere, R matrisi $c' \otimes I$ olarak yazılabilir ve R matrisi satır toplamını hesaplar. (3.41) ile gösterilen kısıt (3.20) ile gösterilen SURE model üzerine uygulanmaz fakat (3.33) ile gösterilen birleştirilmiş model üzerine uygulanabilir. Son olarak, (3.41) kısıtına karşı, $(y^* - Z^* \beta)' \Omega^{-1} (y^* - Z^* \beta)$ denklemi minimize edildiğinde aşağıdaki tahmin edici elde edilir.

$$\hat{\beta}_{MGR}^{(r)} = \hat{\beta}_{MGR} + W^{-1} R' (R W^{-1} R')^{-1} (r - R \hat{\beta}_{MGR}), \quad (3.47)$$

Burada $W = (Z'Z + K)$ ve $\hat{\beta}_{MGR}$, MGR tahmin edicidir. $\hat{\beta}_{MGR}^{(r)}$, shrinkage yöntemi yoluyla EKK ile tahmin edilmiş parametre değerlerini sağlamak şartıyla β için kısıtlanmış bir çözüm olarak gösterilmiştir. Böylece (3.47) numaralı tahmin edici ekolojik regresyon modelindeki tahminler için tüm şartları sağlamaktadır. Ridge regresyon yöntemi, EKK ile tahmin edilen parametrelerin negatif olması probleminin çözümü için bir yol gösterirken, kısıtlı model parametre değerlerinin toplamalarının bir olması özelliğini korumaktadır.

4. EKOLOJİK REGRESYON UYGULAMALARI

Önceki bölümlerden bilindiği gibi, ekolojik regresyon yönteminin kullanıldığı en yaygın uygulama alanı seçimlerdir. Birbiri ardına gelen iki seçimin toplanmış verileri kullanılarak ekolojik regresyon yöntemi ile bireylerin partiler arası geçiş eğilimleri incelenmektedir. Bu oranları elde etmek için bir çok yöntem ele alındı. Yaygın olarak kullanılan yöntemlerin başında Goodman (1953) EKK yöntemi gelmektedir. Bundan sonra, bu yöntemin eksiklikleri görülerek tahmin ediciler geliştirilmiştir. Bu bölümde geliştirilen bu tahmin edicileri, seçimlerin toplanmış verileri üzerinde uygulayarak elde edilen sonuçlar incelenecektir.

İlk uygulama Fule (1994) de ele alınan birbiri ardına gelen, (1988 - 1991) İsveç seçimlerindeki bireylerin geçiş olasılıklarının hesaplanması şeklindedir. Elde edilen toplanmış veriye EKK yöntemi uygulandığında parametre değerlerinin olması gereken (0,1) aralığının dışına çıktığı görülmektedir. Bunun üzerine yöntemin eksikliklerini tamamlayan ve birkaç adımdan sonra elde edilen, kısıtlı modified ridge tahmin edici ile istediğimiz sınırdaki parametre değerlerine ulaşılmaktadır.

İkinci uygulamada ise Çek Cumhuriyetindeki (1992-1996) seçimleri ele alınmıştır. Bu seçimlerdeki geçiş oranlarını hesaplamak için kısıtlı genelleştirilmiş ridge tahmin edici kullanılmıştır. Bu tahmin edici kullanılarak elde edilen sonuçlar kabul edilebilir (0,1) aralığındadır. Dahası, değerler anket sonuçlarıyla benzerlik gösterdiğinden dolayı güvenilirdir.

Son olarak bu uygulamalar doğrultusunda ülkemizden elde edilen (2002-2007) seçimlerinin toplanmış verileri kullanılmıştır. EKK uygulaması, veride istemeyen sonuçlara neden olmuştur. Daha sonra geliştirilen tahmin ediciler uygulanarak mümkün olabilecek parametre değerleri elde edilmiştir. Buradaki geçiş oranları yerel seçimlerle aynı paralellikte olduğundan dolayı tahmin ediciler oldukça kullanışlıdır.

4.1. İsveç Seçimlerinin Geçiş Olasılıklarının Tahmini

Fule (1994) de, İsveç’ te bulunan Malmö şehri için birbiri ardına gelen 1988 ve 1991 seçim yıllarındaki meclis seçimlerinin toplanmış verileri İsveç istatistiklerinden elde edilmiştir. Bu veride, Malmö belediyesi ayrı ayrı 152 ve 154 seçim bölgesine ayrılmıştır. Uygulamanın kolay olması açısından, iki seçimde de seçmenlerin ve seçim bölgelerinin sayılarının eşit olduğu, birbirine yakın topluluklar oldukları varsayılmaktadır. Bu nedenle, her iki seçimde seçim bölgelerinin aynı numaralı olanlarının seçim sonuçları yeniden hesaplanmaktadır. Bunu yapabilmek için, her iki seçimde grup sınırları aynı kalsın diye bazı bölgeler birlikte gruplanmaktadır. Bu yolla, Malmö şehrinde 142 bölge elde edilmiştir.

Mecliste temsil edilen ana partileri içeren ve birinci seçimde 8 parti, ikinci seçimde 9 parti vardır. Bir de oy verenlerin; ölüm, göç ve yaşlılık gibi nedenlerden dolayı değişikliklerini kapsayan özel bir kategoride bir parti olmalıdır. Bu partiye ‘yapay parti’ denmektedir ve yapay parti boş oyları ve çekimserleri de kapsamaktadır. Modelde ve tablolarda kullanılan partilerin kısaltmaları ise aşağıdaki şekildedir:

Tablo 4.1. İsveç seçimlerindeki partiler

m	Sağcılar (Muhafazakarlar)
c	Merkez Parti
fp	Liberal Parti
kds	Hristiyan Demokratlar
mp	Yeşil (Çevresel) Parti
s	Sosyal Demokratlar
v	Komünist Parti
oth	Diğer Tüm Partiler
boe	Yapay Parti

Analize geçiş olasılıklarının EKK tahminleriyle başlanmaktadır (Tablo 4.2.). Homojen olmayan bir nüfusu göz önünde bulundurulduğunda, bazı geçiş olasılıklarının EKK tahmininin negatif olması ve -1'den daha düşük bir tahminle karşılaşılması şaşırtıcı değildir.

Tablo 4.2. EKK ile tahmin edilmiş geçiş olasılıkları (yüzde olarak)

1991 1988	m	c	fp	kds	s	v	nyd	oth	boe	Toplam
m	73.9	-0.4	2.9	0.4	-1.1	-1.9	5.0	-2.3	23.5	100
c	76.1	64.1	15.7	-0.6	0.9	-2.2	28.3	17.9	-100.1	100
fp	64.7	-2.5	46.8	32.1	9.3	4.3	10.0	7.2	-71.9	100
mp	8.9	2.5	16.9	6.7	-33.8	0.8	3.7	47.7	46.7	100
s	0.4	-0.2	-1.4	0.2	81.0	-0.4	2.5	0.9	16.9	100
v	-21.9	1.6	13.8	-17.7	-24.6	76.3	1.7	20.3	50.5	100
oth	0.8	4.4	-8.8	34.2	8.6	-1.6	-1.7	13.1	51.1	100
boe	15.8	0.8	3.9	0.1	13.6	1.7	5.6	0.5	58.2	100

EKK tahminlerini geliştirmek için, (3.47) ile gösterilen ve isteğe bağlı iki parametre tarafından belirlenen, $\hat{\beta}_{MGR}^{(r)}$ tahmin edici uygulanmaktadır. Şöyle ki; 1988 seçim sonuçları (1.seçim) olan açıklayıcı değişkenler X_j , blok matris $Z = I_q \otimes X_j$ olarak düzenlenirken; 1991 seçimlerinde paylaşılan oylar (2.seçim) bağımlı değişkenler olan y , $(qn \times 1)$ sütun vektörü olarak yeniden düzenlenmiştir. Böylece (3.20) ile gösterilen SURE modeli elde edilmektedir.

Model oluşturulduktan sonra shrinkage parametrelerini K matrisinde ve amaç değerleri β_0 vektöründe birleştirilmektedir. β_0 değerlerini belirlemek için EKK tahminlerinden yararlanılarak $0 < c < 1$ olmak üzere $\beta_0 = c\hat{\beta}_{EKK}$ hesaplanır. EKK tahminleri çoğu zaman yanlış işaretli olmakla birlikte mümkün olan $(0,1)$ aralığını aşmaktadır.

Bundan dolayı, β_0 'ın bazı alternatif seçimleri, farklı c değerlerinde ($c = 0,33$ veya $c = 0,25$ gibi) göz önünde bulundurulmaktadır. İsveç seçimlerinin verileri için en iyi tahminler ((3.40) ile gösterilen test istatistiğini kullanarak) EKK değerinin altında birine eşit olan β_0 için elde edilmiştir ($\beta_0 = \frac{1}{6} \hat{\beta}_{EKK}$). Ayrıca, oldukça düşük β_0 değerleri ve 1'den düşük shrinkage sabitleri; EKK ile tahmin edilmiş parametrelerin shrinkage yöntemi için yüksek derecede hassas olduğu anlamına gelmektedir.

Anlamlı olmayan EKK tahminleri için $\beta_0=0$ olarak belirlenmektedir. K matrisindeki shrinkage sabitlerinin değerleri (3.37)' de gösterildiği gibi hesaplanmaktadır. Burada dikkat edilmesi gereken nokta ise, EKK tahminleri sıfıra yakın olduğu zaman aşırı yüksek $\hat{k}_i$ değerleri elde edilmektedir. Bu nedenle, isteğe bağlı olarak yüksek $\hat{k}_i$ değerleri 1'e indirgenmektedir. Düşük $\hat{k}_i$ değerleri için ise, (3.40) ile gösterilen F testi sonucunda, önemli bir anlamlılık elde edilmiştir. Burada kullanılan , $\frac{pq}{2}$ merkezi olmama parametrelili F dağılımı için kritik değerler Goodnight ve Wallace (1972) tarafından verilmiştir.

$\hat{\beta}_{MGR}^{(r)}$ tahmin ediciyi hesaplayabilmek için tüm bu bilgiler elde edildikten sonra $\hat{\beta}_{MGR}^{(r)}$ tahmin edici ;

$$\hat{\beta}_{MGR}^{(r)} = \hat{\beta}_{MGR} + W^{-1} R' (R W^{-1} R')^{-1} (r - R \hat{\beta}_{MGR}),$$

şeklinde hesaplanır.

$\hat{\beta}_{MGR}^{(r)}$ tahmin edici ile elde edilen geçiş olasılıklarının sonuçları Tablo 4.3. de gösterildiği gibidir.

Tablo 4.3. Kısıtlı MGR ($\hat{\beta}_{MGR}^{(r)}$) ile tahmin edilen geçiş olasılıkları (yüzde olarak)

1991 1988	m	c	fp	kds	s	v	nyd	oth	boe	Toplam
m	*78.2	0.4	*7.6	*5.0	0.1	0.0	*6.7	1.8	0.2	100
c	*48.0	*32.9	2.5	1.8	1.1	1.3	*9.4	2.4	0.6	100
fp	*46.1	1.6	*31.1	*13.8	0.0	1.0	3.8	2.6	0.0	100
mp	4.6	3.9	6.1	4.0	3.8	4.4	4.8	*37.5	*30.9	100
s	1.7	0.0	0.2	*1.0	*79.2	0.0	*3.1	*2.6	*12.2	100
v	3.5	2.0	3.5	1.6	0.1	*52.9	2.6	4.9	*28.9	100
oth	5.4	5.0	5.1	*19.9	4.8	4.6	5.3	5.4	*44.5	100
boe	*11.4	*1.9	*5.4	0.8	7.5	*3.4	*5.6	*3.8	*60.2	100

Tablo 4.3’ de yıldız ile işaretli değerler, % 0,05 anlam düzeyinde F testi sonucunda anlamlı olan tahminlerdir (burada $F=2.52$ dir). Yani bu parametreler üzerinde, EKK tahminine dayanan anlamlı bir gelişme elde edilmiştir. Tablodan görüldüğü üzere kısıtlı MGR ($\hat{\beta}_{MGR}^{(r)}$) tahmin edici negatif tahminlerin işaretlerini değiştirmektedir. Ayrıca belli değerler üzerinde durağanlaşmasını da sağlamaktadır. Diğer taraftan parametrelerin tahmini $\hat{\beta}_{MGR}^{(r)}$ tahmin edici ile yapıldığında rezidülerin varyansı 0.00428 iken EKK yöntemi kullanıldığında rezidülerin varyansı 0.00372 olmaktadır. Yani kısıtlı MGR tahmin edici ile rezidülerin varyansı hala düşük bir seviyededir. Daha ayrıntılı bir sonuç almak için tahminlerin ilgili standart hatalarına bakılabilir.

Burada asıl ilgilendiğimiz ise, sadık seçmenler olarak tanımlayabileceğimiz seçmenler yani her iki seçimde de oyunu aynı partiye veren seçmenlerin oranları ve parti dışından eklenen seçmenler olan tablonun köşegenindeki değerlerdir.

Genel olarak seçim sonuçlarını ele aldığımızda ise Sosyal Demokratlar (s) ve Sağcılar (m) oyların neredeyse % 80’ ine sahiptirler. Diğer partiler daha düşük oy oranlarına sahiptirler.

4.2. Çek Cumhuriyeti 1992 - 1996 Seçimlerinin Geçiş Olasılıkları

Çek toplumunun seçim sürecinin uzun bir geçmişi yoktur. Kısa bir süre önce yapılan devrim niteliğindeki değişim, özgür ve demokratik bir seçimin ihtiyacı olan şartları sağlamıştır. Bu gelişmelerden sonra politik parti sayısı oldukça yükselmiştir ve hatta 1996 seçimlerinde parti sayısı on altıya kadar ulaşmıştır.

Çek Cumhuriyetindeki ilk demokratik seçimler 1992 yılında yapılmıştır. Bundan dört yıl sonra 1996 yılında ikinci seçimler yapılmıştır. Her iki seçimde de oransal seçim sistemi kullanılmıştır. Bu sisteme göre herhangi bir partinin meclise girebilmesi için, geçerli oyların en az yüzde beşine sahip olması gerekmektedir.

Bu uygulamaya için ilgili yıllara ait gerekli olan toplanmış veriler Çek İstatistik kurumundan elde edilmiştir. Buradan alınan parti bilgileri için ayrıca bir düzenleme yapılmalıdır. Bu düzenlemelerden ilki; topladıkları oy miktarı çok düşük olan partiler bir grup olarak belirtilmektedir. Bu tablolarda ‘oth’ olarak isimlendirilmiştir (others = diğerleri). Yapılan düzenlemelerden bir diğeri ise, çekimser olarak nitelendirilen bir gruptur. Bu grupta tablolarda ‘abs’ olarak belirtilmektedir (abstainers). Böylece birbiri ardına gelen iki seçimde de partiler belli bir düzende gruplandırılmıştır. Aşağıdaki tablolarda her iki seçimde oylanan partiler verilmiştir.

Tablo 4.4. Çek Cumhuriyetinde 1992’ de oylanan partiler

CSSD_92	Çek Sosyal Demokrat Parti
ODS_92	Demokratik Yurttaş Partisi
KDU_92	Demokratik ve Hristiyan Birliği
ODA_92	Demokratik Yurttaş İttifakı
LB_92	Sol Kesim
REP_92	Cumhuriyetçiler
OTH_92	Diğer Tüm Partiler
ABS_92	Çekimserler

1996 seçimlerinde ise LB_92 ile gösterilen bu seçimlerde komünist parti olarak değişmiştir. Aşağıdaki tabloda 1996 seçimlerindeki partiler belirtilmektedir.

Tablo 4.5. Çek Cumhuriyetinde 1996’ da oylanan partiler

CSSD_96	Çek Sosyal Demokrat Parti
ODS_96	Demokratik Yurttaş Partisi
KDU_96	Demokratik ve Hristiyan Birliği
KSCM_96	Komünist Parti
REP_96	Cumhuriyetçiler
OTH_96	Diğer Tüm Partiler
ABS_96	Çekimserler

Birbiri ardına gelen bu iki seçimde belirtilen bu partiler için Çek İstatistiklerinden alınan toplanmış veriler alınarak; analize geçiş oranlarının EKK tahminleriyle başlanmaktadır. Elde edilen sonuçlar Tablo 4.6’ da gösterildiği gibidir.

Tablo 4.6. EKK yöntemi ile tahmin edilen geçiş olasılıkları

1992 1996	CSSD_92	ODS_92	KDU_92	LB_92	REP_92	ODA_92	OTH_92	ABS_92
CSSD_96	0.93	-0.16	0.05	0.05	0.75	0.33	0.39	0.06
ODS_96	-0.22	0.72	0.07	0.08	-0.32	0.51	0.09	0.07
KDU_96	0.10	0.06	0.90	-0.13	0.03	-0.01	0.04	-0.02
KSCM_96	0.02	-0.04	-0.02	0.76	-0.02	-0.04	0.06	-0.02
REP_96	0.40	-0.10	-0.14	0.06	0.57	-0.01	0.12	0.03
ODA_96	-0.35	0.17	0.12	-0.01	0.05	0.09	0.03	0.04
OTH_96	0.47	-0.01	0.14	-0.03	-0.14	0.09	0.15	0.09
ABS_96	-0.35	0.35	-0.12	0.23	0.08	0.05	0.10	0.74
Toplam	1	1	1	1	1	1	1	1

Tablo 4.6’ da görüldüğü üzere EKK yöntemi yine önemli derecede negatif tahmin değerleri vermiştir. Bunlar elbette olasılık değerleri için geçerli olabilecek sonuçlar değildir. Bunun üzerine daha anlamlı sonuçlar elde etmek için (3.47) ile gösterilen, kısıtlı modified genelleştirilmiş ridge tahmin edici (RMGR) uygulanmıştır (Fule, 1997). Elde edilen sonuçlar Tablo 4.7’ deki gibidir.

Tablo 4.7. RMGR ile tahmin edilen geçiş olasılıkları

1992 1996	CSSD_92	ODS_92	KDU_92	LB_92	REP_92	ODA_92	OTH_92	ABS_92
CSSD_96	0.45			0.18	0.32		0.45	0.09
ODS_96	0.05	0.64			0.07	0.59	0.08	0.05
KDU_96			0.79				0.05	
KSCM_96				0.56				
REP_96	0.18			0.05	0.45		0.07	
ODA_96	0.07	0.06	0.05			0.09		
OTH_96	0.15					0.2	0.21	
ABS_96	0.05	0.26		0.08	0.05		0.06	0.78
Toplam	1	1	1	1	1	1	1	1

Not: Tabloda % 5’ ten büyük olan geçiş olasılıkları gösterilmiştir.

Her iki seçimde de aynı partiye oy veren seçmenler sadık seçmen olarak nitelendirilmektedir. Tahmin edilen geçiş olasılıkları gösteriyor ki, ODA partisi dışındaki tüm partiler önemli derecede sadık seçmen oylarına sahiptirler. En büyük sadık seçmen oylarına sahip olanlar, Çek Cumhuriyetinde halkın büyük bir kesiminin desteklediği KDU (0.79) ve ODS (0.64) partileridir. Tabloyu daha ayrıntılı bir şekilde inceleyecek olursak; örneğin 1. seçimde OTH partisine oy vermişken (OTH_92), ikinci seçimde CSSD partisine (CSSD_96) oy verenlerin geçiş olasılığı 0.45 olarak tahmin edilmiştir. Bu da oldukça yüksek bir olasılıktır. Yani birinci seçimde OTH partisinin içindeki partilerden, CSSD partisine oldukça yüksek bir geçiş eğilimi söz konusudur.

4.3. Türkiye’ deki 2002 – 2007 Genel Seçimleri Geçiş Olasılıkları

Genel seçim, bir ülke parlamentosunun yenilenmesi, dolayısıyla gerektiğinde iktidarın değişmesini sağlayan seçimdir. Türkiye’de seçimler, siyasi bakımdan, [1876 Birinci Meşrutiyet](#) Anayasasıyla ilk defa uygulanmıştır. Çok partili siyasal yaşamda, 1950’den 2002 yılına kadar, 13 Milletvekili Genel Seçimi, 8 Milletvekili Ara Seçimi, 11 Yerel Yönetimler Seçimi, 8 Cumhuriyet Senatosu Kısmi Seçimi olmak üzere 41 seçim yaşanmıştır.

3 Kasım 2002’de, 14. kez Milletvekili Genel Seçimi için sandık başına gidilmiştir. Tablo 4.8.’ de 3 Kasım 2002’ de yapılan genel seçime katılan partiler listelenmiştir. Analiz için Yüksek Seçim Kurumundan alınan veriler kullanılmıştır.

Tablo 4.8. 3 Kasım 2002 genel seçimine katılan partiler

Adalet ve Kalkınma Partisi	AKP	Liberal Demokrat Parti	LDP
Anavatan Partisi	ANAP	Millet Partisi	MP
Bağımsız Türkiye Partisi	BTP	Milliyetçi Hareket Partisi	MHP
Büyük Birlik Partisi	BBP	Özgürlük ve Dayanışma Partisi	ÖDP
Cumhuriyet Halk Partisi	CHP	Saadet Partisi	SP
Demokratik Halk Partisi	DEHAP	Türkiye Komünist Partisi	TKP
Demokratik Sol Parti	DSP	Yeni Türkiye Partisi	YTP
Doğru Yol Partisi	DYP	Yurt Partisi	YP
Genç Parti	GP	Bağımsızlar	
İşçi Partisi	İP		

2002 yılında yapılan seçimler Türkiye için büyük bir sürpriz olmuştur. Yıllardır Türkiye’nin siyasal gündeminde ve ülke yönetiminde yer alan pek çok parti, bu seçimlerde çok düşük oylar alarak meclis dışı kalmıştır.

2002 seçimlerinden Adalet ve Kalkınma Partisi birinci olarak çıktı. 34.2 oranında oy alan AKP 363 milletvekili çıkardı. CHP ise 19.3 oy oranıyla meclise giren ikinci parti oldu.

Türkiye yıllardan sonra sadece iki partinin yer aldığı bir meclis ile karşılaşmıştır. Sonraki yapılan genel seçim ise 22 Temmuz 2007 seçimleridir. Bu seçime katılan partiler ise Tablo 4.9. ile gösterilmiştir.

Tablo 4.9. 22 Temmuz 2007 genel seçimine katılan partiler

Adalet ve Kalkınma Partisi	AKP	İşçi Partisi	İP
Aydınlık Türkiye Partisi	ATP	Liberal Demokrat Parti	LDP
Bağımsız Türkiye Partisi	BTP	Milliyetçi Hareket Partisi	MHP
Cumhuriyet Halk Partisi	CHP	Özgürlük ve Dayanışma Partisi	ÖDP
Demokrat Parti	DP	Saadet Partisi	SP
Emek Partisi	EMEP	Türkiye Komünist Partisi	TKP
Genç Parti	GP	Bağımsızlar	
Halkın Yükselişi Partisi	HYP		

22 Temmuz 2007 tarihinde yapılan erken genel seçimlerden de AKP birinci parti olarak çıkmıştır ve % 46.5 oy oranıyla Meclis'e 341 milletvekili sokmayı başarmıştır. CHP bu seçimlerde ikinci parti olmuştur ve % 20.7 oyla 112 milletvekili kazanmıştır. Meclis'e girebilen bir diğer parti olan MHP'de % 14.2 oyla 70 milletvekiline sahip olmuştur. Ayrıca bu seçimlerde bağımsız milletvekili olan kişilerin sayısındaki artış dikkat çekicidir. 26 kişi de meclise bağımsız milletvekili seçilerek girmiştir.

Yüksek Seçim Kurumundan toplam 84 seçim bölgesi için alınan 3 Kasım 2002 ve 22 Temmuz 2007 genel seçim sonuçlarına göre analiz yapılacaktır. Analize başlamadan önce, her iki seçim için parti sayılarını alınan oyları baz alarak düzenleyeceğiz. Bu düzenleme Tablo 4.10. ile gösterilmiştir.

Tablo 4.10. Analiz edilecek partiler

3 Kasım 2002		22 Temmuz 2007	
Adalet ve Kalkınma Partisi	AKP	Adalet ve Kalkınma Partisi	AKP
Cumhuriyet Halk Partisi	CHP	Cumhuriyet Halk Partisi	CHP
Milliyetçi Hareket Partisi	MHP	Milliyetçi Hareket Partisi	MHP
Doğru Yol Partisi	DYP	Demokratik Parti	DP
Genç Parti	GP	Genç Parti	GP
Bağımsızlar	Bağ	Bağımsızlar	Bağ
Diğer		Diğer	

3 Kasım 2002 genel seçimi için Diğer kategorisine giren partiler kısaca; ANAP, BTP, BBP, DEHAP, DSP, İP, LDP, MP, ÖDP, SP, TKP, YTP, YP partileridir. 22 Temmuz 2007 genel seçimlerinde ise bu kategorideki partiler; ATP, BTP, EMEP, HYP, İP, LDP, ÖDP, SP, TKP ile belirtilen partilerdir. Bu partiler aldıkları oyların diğer partilere göre az olması nedeniyle, analiz sonuçlarını etkilemeyeceği düşünülerek Diğer partiler kategorisine alınmıştır.

Burada, her iki seçim için alınan veriler kullanılarak, her bir parti için yüzdelik oranlar elde edilmiştir. Bu da her bir parti için 84 bölgede uygulanmıştır. Diğer kategorisindeki partiler için ise aldıkları oylar her bir bölge için toplanıp, tek bir parti gibi düşünülerek yüzdelik oranları hesaplanmıştır. Elde edilen veriler Ek-2’deki gibidir.

Veriler üzerinde bu düzenlemeler yapıldıktan sonra, 3 Kasım 2002 seçimi (1. seçim) X matrisi, 22 Temmuz 2007 Seçimi (2. seçim) Y matrisi olmak üzere, verilerin EKK tahmini ile analize başlanmaktadır.

Verilerin EKK tahminleri Tablo 4.11’deki gibidir.

Tablo 4.11. Geçiş olasılıklarının EKK tahminleri

2007 2002	AKP	CHP	MHP	DYP	GP	Bağ	Diğer	Toplam
AKP	0.99	0.04	0.03	0.03	-0.02	-0.15	0.08	1
CHP	0.38	0.67	0.01	0.02	0.02	-0.11	0.02	1
MHP	0.02	-0.17	1.25	-0.05	-0.06	0.02	0.00	1
DP	0.21	0.34	0.19	0.34	0.05	-0.29	0.16	1
GP	-0.45	0.69	0.29	0.18	0.51	-0.24	0.02	1
Bağ	0.68	0.03	0.01	0.17	0.02	0.02	0.08	1
Diğer	0.33	-0.07	-0.09	0.00	0.00	0.83	0.00	1

EKK ile tahmin edilmiş geçiş olasılıkları göz önünde bulundurulduğunda, diğer uygulamalarda karşılaşılan problemle burada da karşılaşılmaktadır. Tablo 4.11'i incelediğimizde birden büyük bir değer ve negatif değerler söz konusudur. Elbette ki olasılıktan bahsederken birden büyük veya negatif değerler geçerli değildir. Bu sorunu gidermek için geliştirilen tahmin ediciler uygulanacaktır.

(3.5.3) ile verilen MGR Tahmin Edicinin Uygulama Şekli kısmında anlatıldığı üzere, $\hat{\beta}_{MGR}^{(r)}$ tahmin ediciye ulaşmak için oluşturulması gereken bazı matrisler vardır. Buna göre önemli olan iki faktör vardır:

(i) K matrisinin elemanlarıyla hesaplanmış shrinkage derecesi,

(ii) β_0 vektörü yönünde EKK tahmin edicisinin büzüştürdüğü noktadır.

Bunlardan birincisi, uyguma için ele alındığında $K = \text{diag}\{k_l\}$ $l = 1, 2, \dots, pq$ olmak üzere shrinkage parametrelerinin genel bir şekli aşağıdaki şekilde elde edilmiştir.

$$\hat{k}_i = \frac{pq\hat{\sigma}^2}{\left(\hat{\beta}_{EKK(l)} - \beta_{0(l)}\right)}$$

İkinci faktör ise yine bahsedildiği üzere $\beta_0 = c\hat{\beta}_{EKK}$ uygun bir c seçerek β_0 'a yakın değerler elde edilmiştir. Bunu da , $H_0 : \beta = \beta_0$ hipotezini sağlayan, $v_1 = qp$, $v_2 = q(n - p)$ serbestlik dereceli olan aşağıdaki F testi ile kontrol edebiliriz.

$$F = \frac{(\hat{\beta}_{EKK} - \beta_0)'[(Z'Z)^{-1} + K]^{-1}(\hat{\beta}_{EKK} - \beta_0)/v_1}{(y - Z\hat{\beta}_{EKK})'(y - Z\hat{\beta}_{EKK})/v_2}$$

Gerekli düzenlemeler yapıp, verilerle ilgili bu bilgilerde elde edildikten sonra (3.36) ile verilen;

$$\hat{\beta}_{MGR} = (Z'Z + K)^{-1}(Z'y + K\beta_0) = (Z'Z + K)^{-1}(Z'Z\hat{\beta}_{EKK} + K\beta_0)$$

elde edilir. Burada, $\hat{\beta}_{EKK}$ ile verilen tahmin edici artık SURE modelin EKK tahmin edicisidir ve $\hat{\beta}_{EKK} = (Z'Z)^{-1}Z'y$ şeklinde tahmin edilmektedir. Daha sonra tahmin ediciye bazı kısıtlar koyup elde etmek istediğimiz (3.47) ile verilen ve $\hat{\beta}_{MGR}^{(r)}$ tahmin ediciyi hesaplayabiliriz. Kısıtlı Modified Genelleştirilmiş Ridge Tahmin Edici olarak adlandırılan bu tahmin edici aşağıdaki formda elde edilmektedir.

$$\hat{\beta}_{MGR}^{(r)} = \hat{\beta}_{MGR} + W^{-1}R'(RW^{-1}R')^{-1}(r - R\hat{\beta}_{MGR})$$

Burada bahsedilen kısıt; β parametre vektörü üzerine (3.41) ile gösterilen,

$$R\beta = r$$

lineer eşitlikli kısıttır. Her ikisi de bilinen sabitler olmak üzere; R , $(p \times pq)$ tipinde bir matris ve r , $(p \times 1)$ tipinde bir vektördür. Ayrıca W ile gösterilen matris, $W = (Z'Z + K)$ şeklinde oluşturulmaktadır.

3 Kasım 2002 – 22 Temmuz 2007 genel seçim sonuçlarının verileri için bu adımlar uygulandıktan sonra $\hat{\beta}_{MGR}^{(r)}$ ile tahmin edilen geçiş olasılıkları Tablo 4.12 ile gösterilmiştir.

Tablo 4.12. $\hat{\beta}_{MGR}^{(r)}$ ile tahmin edilmiş geçiş olasılıkları

2007 2002	AKP	CHP	MHP	DYP	GP	Bağ	Diğer	Toplam
AKP	0.82	0.04	0.03	0.03	0.00	0.00	0.08	1
CHP	0.27	0.67	0.01	0.02	0.02	0.00	0.02	1
MHP	0.02	0.00	0.97	0.00	0.00	0.02	0.00	1
DP	0.11	0.25	0.19	0.34	0.05	0.00	0.06	1
GP	0.00	0.33	0.19	0.08	0.40	0.00	0.00	1
Bağ	0.68	0.03	0.01	0.17	0.02	0.02	0.08	1
Diğer	0.33	0.00	0.00	0.00	0.00	0.67	0.00	1

Elde edilen tahmin sonuçlarına baktığımızda EKK tahmin edicisinde karşılaşılan problemlerin çözüldüğünü görmekteyiz. Tabloda negatif ya da birden büyük değer yoktur. Bu da elde edilmek istenilen sonuçtur.

Tablonun köşegeni üzerindeki olasılıklar sadık seçmenler olarak nitelendirilen yani her iki seçimde de aynı partiye oy verenlerin olasılığıdır. Tabloya göre yapabileceğimiz bir kaç yorum şöyledir: 2002 seçimlerinde AKP' ye oy verip 2007 seçimlerinde de AKP' ye oy verenlerin olasılığı 0.82' dir. Diğer taraftan en yüksek olasılık olan 0.97 her iki seçimde de MHP partisini oylayanların olasılığıdır. Bu olasılığın yüksek olmasının sebebi, Türkiye' de en büyük kemikleşmiş oy (sadık seçmen) oranlarının MHP' ye ait olduğu düşünülerek açıklanabilir.

Köşegen üzerinde olmayan 0.68 olasılığı da oldukça yüksek bir değerdir. Bu olasılık 2002 seçimlerinde oyunu Bağımsızlardan yana kullanıp, 2007 seçimlerinde ise AKP' yi destekleyen seçmenlerin geçiş olasılıklarıdır.

5. SONUÇ VE ÖNERİLER

Bu çalışmada, toplanmış veriyi kullanarak ekolojik olarak analiz etme yöntemi ele alınmıştır. Elde edilen bir toplanmış veri kümesinin analizi EKK yöntemi ile başlamaktadır. Fakat bu yöntem, olasılık olarak ifade edilen parametre değerlerinin kabul edilebilir (0,1) aralığının dışında sonuçlar vermektedir. Bundan dolayı parametre değerlerini elde etmek için alternatif tahmin yöntemlerine ihtiyaç vardır.

Ekolojik regresyonda böyle bir durumdan kaçınmak için yanlı tahmin ediciler kullanılmıştır. Burada bu tahmin edicilerden; ridge, genelleştirilmiş ridge, modified ridge ve kısıtlı ridge gibi bazı tahmin ediciler incelenmiştir. Ayrıca ekolojik regresyon yönteminde SURE modeller kullanıldığı için bu tahmin edicilerin SURE modelde uygulama şekilleri de incelenmiştir. EKK yönteminden kaynaklanan problemler yanlı tahmin ediciler sayesinde ortadan kalkmıştır ve anlamlı sonuçlar elde edilmiştir.

Ekolojik regresyonun en geniş uygulama alanı seçimlerdir. Seçimlerin elde edilen toplanmış verileri SURE model olarak ele alınıp analizleri yapılmaktadır. Bu doğrultuda uygulama için Türkiye ile birlikte üç farklı ülkenin seçim sonuçlarının analizleri yapılmıştır. Uygulamalarda, EKK yöntemi kullanıldığında istenmeyen sonuçlarla karşılaşmıştır. Daha sonra elde edilen kısıtlı modified genelleştirilmiş ridge tahmin edici uygulanarak problemler ortadan kalkmıştır.

Elde edilen sonuçlara göre bazı yanlı tahmin edicilerin EKK tahmin edicisinden daha kullanışlı olduğu görülmektedir. Burada parametre tahminleri yapıldıktan sonra, parametrelerin standart hataları, t istatistiği değerleri ya da p değerleri incelenebilir. Diğer taraftan ekolojik regresyon yöntemi sadece politik olaylarda değil, tüketici davranışlarındaki marka değişimi gibi diğer sosyal olaylarda da uygun veriler elde edilerek kullanılabilir.

KAYNAKLAR

- ACHEN, C. D., SHIVELY, W. P., 1995. Cross-Level Inference. University of Chicago Pres.
- ANSOLABEHERE, S., RIVERS D., 1995. Bias In Ecological Regression Departman of Political Science, Standford University, Standford.
- BAKSALARY, J. K., PORDZIK P. R. and TRENKLER G., 1990. A Note On Generalized Ridge Estimators. Communications in Statistics – Theory and Mothods, 19, 2871-2877.
- BROWN, P. and PAYNE, C., 1975. Election Night Forecasting (with discussion). Journal of the Royal Statistical Scienty, A 138, 463-483.
- BROWN, P. J., Firth, D. ; Payne, C.D., 1999. Forecasting on British election night 1997. J. R. Statist. Soc. A (162) Part 2, 211-226.
- BROWN, P. J., Payne, C. D., 1986. Aggregate Data, Ecological Regression, and Voting Transitions. Journal of the American Statistical Association. Vol.81, No.394 452-460.
- CLEAVE, N., BROWN P. J., PAYNE, C. D., 1995. Evaluation of Methods for Ecological Inference. Journal of the Royal Statistical Society. Series A, Vol. 158, No. 1, 55-72.
- DOGAN, N., 1998. Görünüşte İlişkısız Regresyon Modeli ve Türkiye İhracat Fonksiyonu Üzerine Bir Uygulama. Gazi Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı Yüksek Lisans Tezi, Ankara.
- DUNCAN, O. D. and DAVIS, B., 1953. An Alternative to Ecological Correlation. American Sociological Review, Vol. 18, No. 6, 665-666.
- FREEDMAN, D. A., KLEIN, S. P., SACS, J., SMYTH, C. A., EVERETT, C. G., 1991. Ecological Regression and Voting Rights. A Journal of the Applied Social Research, Vol. 15, No. 6, 673-711.
- FREEDMAN, D. A., 1999. Ecological Inference and the Ecological Fallacy. International Encyclopedia of the Social & Behavioral Sciences Technical Report No. 549.

- FREEDMAN, D. A., KLEIN, S. P., OSTLAND, M., ROBERTS, M. R., 1998. A Solution to the Ecological Inference Problem by G. King. *Journal of the American Statistical Association*, Vol. 93, No. 444, 1518-1522.
- FREEDMAN, D. A., OSLAND, M., ROBERTS, M. R., KLEIN, S. P., 1999. Response to King's Comment. *Journal of the American Statistical Association* Vol. 94, No. 445, 355-357.
- FULE, E., 1995. On Ecological Regression And Ridge Estimation. *Commun. Statist.- Simula.*, 24(2), 385 – 389.
- FULE, E., 1997. Changes on the Czech Political Scene. *Department of Statistics, University of Lund*, Vol. 16, No. 3, 341-347.
- GELMAN, A., PARK, D. K., ANSOLABEHERE, S., PRICE, P. N., MINNITE, L.C., 2001. Models, Assumptions and Model Checking in Ecological Regression. *J.R. Statist. Soc. A*(2001) 164,Part 1, 101-118.
- GOODMAN, L. A., 1953. Ecological Regressions and Behavior of Individuals. *American Sociological Review*, Vol. 18, No. 6, 663-664.
- GOODMAN, L. A., 1959. Some Alternatives to Ecological Correlation. *The American Journal of Sociology*, Vol. 64, No. 6, 610-625.
- GOODNIGHT, J. and WALLACE, T. D., 1972. Operational Techniques and Tables for Making Weak MSE Tests for Restrictions in Regressions. *Econometrica*, Vol. 40, 4, 699-709.
- GRAYBILL, F. A., 1961. *An Introduction to Linear Statistical Models*. Mc Graw Hill, New York.
- GROSS, J., 2003. Restricted Ridge Estimation. *Statistics and Probability Letters*, 65, 57-64.
- HALD, A., 1952. *Statistical Theory with Engineering Applications*. New York: Wiley.
- HAMILTON, R., 1982. *Who Voted for Hitler?* Princeton, NJ: Princeton University Press
- HOERL, A. E., 1959. Optimum Solution of Many Variable Equations. *Chem. Eng. Prog.*, 55, 69

- HOERL, A. E., KENNARD, R. W., 1970a. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 1, 55-67.
- HOERL, A. E., KENNARD, R. W., 1970b. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*, 12, 1, 69-82.
- HOERL, A.E., KENNARD, R.W., and BALDWIN, K. F., 1975. Ridge Regression: Some Simulation. *Communication in Statistics*, 4, 105-123.
- HUSTON, J. L., 1991. Weighting, Confidence Intervals and Ecological Regression. *Journal of Interdisciplinary History*, Vol. 21, No. 4, 631-654.
- KING, G., 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.
- KOUSSER, J. M., 1973. Ecological Regression and the Analysis of past Politic. *Journal of Interdisciplinary History*, Vol. 4, No. 2, 237-262.
- LANCASTER, G., GREN, M., 2002. Deprivation, III-Health and the Ecological Fallacy. *Journal of the Royal Statistical Society Series Vol. 165, No. 2*, 263-278
- LAWLESS, J. F. and WANG, P., 1976. A Simulation Study of Ridge and Other Regression Estimators. *Communication in Statistics*, 7, 139-164.
- LICHTMAN, A. J., 1974. Correlation, Regression and the Ecological Fallacy: A Critique. *Journal of the Interdisciplinary History*, Vol. 4, No. 3, 417-433.
- MARQUARDT, D. W., 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics*, 12, 591-612.
- MCCUE, K. F., 1995. *Individual Choise and Ecological Analysis*. California Institute of Technology.
- MONTGOMERY, D. C. and PECK, E. A., 1992. *Introduction to Linear Regression Analysis*. John Wiley and Sons, New York, 526p.
- MYERS, R. H., 1990. *Classical and Modern Regression with Application*. Duxbury Press, California
- NARWA, D., 1996. *Application of Ecological Regression to Czech Parliamentary Elections*. Discussion paper series No.3/96, CERGE-EI, Prague NJ: Princeton University Press.

- ÖZKALE, M. R., 2007. Çoklu İç İlişki ile İlgili Problemler. Çukurova Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Doktora Tezi, Adana.
- PRIEVOZNIC, P., Estimation Of Voters Transition Rates Between The Czech Parliamentary Elections from 1992 to 2006 (Ecological Regression Technique). The Institute of Economic Studies, Faculty of Social Science Charles University in Prague.
- RAWLINGS, J. O., 1998. Applied Regression Analysis. A Research Tool, California.
- ROBINSON, W. S., 1950. Ecological Correlations and the Behavior of Individuals. American Sociological Review, Vol. 15, No.3, 351-357.
- ROSS, F. A., 1933. Ecology and the Statistical Method. The American Journal of Sociology, Vol. 38, No.4, 507-522.
- SARKAR, N., 1992. A New Estimator Combining the Ridge Regression and the Restricted Least Squares Methods for Estimation. Comm. Statist. Theory Methods, 21, 7, 1987-2000.
- SCHUESSLER, A. A., 1999. Ecological Inference. Proceedings of the National Academy of Sciences of the USA, Vol. 96, No. 19, 10578-10581.
- SEZER, D., 2006. Görünüşte İlişkisiz Regresyon Modelleri ve Bir Uygulama. Selçuk Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Yüksek Lisans Tezi, Konya
- SHIVEL, W. P., 1969. Ecological Inference: The Use of Agggregate Data to Study Individuals. The American Political Science Review, Vol. 63, No. 4, 1183-1196.
- SRIVASTAVA, V. K., GILES, D. E. A., 1987. Seemingly Unrelates Regression Equations Models, 80, Marcel Dekker Inc., New York.
- SWINDEL, F. F., 1976. Good Ridge Estimators Based on Prior Information. Comm. Statist. Theory Methods, A5 (11), 1065-1075.
- TARI, R., 1999. Ekonometri. Alfa Basım Yayım Dağıtım, İstanbul.

- THEOBALD, C. M., 1974. Generalizations of Mean Square Error Applied to Ridge Regression. Journal of the Royal Statistical Society. Series B (Methodological), 36, 1, 103-106.
- VINOD, H. D. and ULLAH, A., 1981. Recent Advances in Regression Methods. Marcel Dekker, New York, 361p.
- WAKEFIELD, J., 2003. Sensitivity Analyses for Ecological Regression. Departments of Statistical and Biostatistics, University of Washington, Seattle, Biometrics 59, 9-17.
- WALLACE, T. D., 1972. Weaker Criteria and Tests for Linear Restrictions in Regression. Econometrica, Vol.40, No. 4, 689-698
- WILSON, B. K., 1985. Simultaneity and Its Impact on Ecological Regression Applications. Biometrics, Vol. 41, No.2, 435-445.
- ZELLNER, A., 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. Journal of the American Statistical Association. 57: 348-368.

ÖZGEÇMİŞ

1985 Çanakkale’ de doğdum. İlk ve orta öğrenimimi Bursa Sönmez İlköğretim Okulu’nda ve Bursa Dr. Ayten Bozkaya okullarında tamamladım. Liseyi Bursa Erkek Lisesi / Yabancı Dil Ağırlıklı bölümünden 2003 yılında tamamladım. Aynı yıl Çukurova Üniversitesi İstatistik Bölümünü kazandım. 2007 yılında bu bölümden ikincilikle mezun olduktan sonra yüksek lisans eğitimime başladım. Eğitimime doktora ile devam ettirmek istiyorum.

EKLER

EK-1

Ridge Trace Matlap Kodu

```
load hald
b=zeros(4,100)
kvec=0.01:0.01:2
count=0
for k=0.01:0.01:2
    count=count+1
    b(:, count)=ridge(k)
end
plot(kvec,b), xlabel('k'), ylabel('b', 'Fontname', 'Symbol')
```

EK-2**3 Kasım 2002 Seçimlerinin Oy Yüzdeleri**

AKP	CHP	MHP	DHP	GP	Bağımsızlar	Diğer
26,823	21,301	11,407	11,494	8,811	0,130	20,034
41,418	13,386	4,774	9,761	2,107	0,020	28,535
42,634	12,400	14,023	13,708	5,403	0,036	11,795
17,697	9,605	5,289	7,796	1,469	4,942	53,201
39,616	23,145	10,078	10,755	2,829	0,000	13,577
33,243	32,059	8,729	5,666	6,514	0,033	13,756
43,158	23,864	8,441	5,161	7,079	0,055	12,242
21,161	30,928	10,368	14,664	5,501	0,061	17,318
23,211	25,402	10,243	20,804	1,548	0,000	18,793
18,266	20,733	11,975	19,797	9,730	1,867	17,632
33,581	19,670	7,941	14,895	8,920	0,035	14,958
24,885	18,382	9,979	21,805	9,635	0,000	15,315
31,732	7,875	8,779	6,522	0,997	9,559	34,536
17,673	4,522	6,441	9,298	0,920	16,908	44,239
43,081	11,665	7,671	12,012	6,990	0,000	18,581
27,255	19,626	12,700	20,057	6,162	0,021	14,180
41,305	15,253	6,488	11,479	8,691	0,127	16,657
25,841	22,883	9,888	16,925	8,947	0,109	15,407
49,928	6,364	14,121	15,106	2,828	0,000	11,653
48,426	22,156	8,681	7,859	3,001	0,000	9,877
24,166	20,754	10,208	17,991	10,227	0,048	16,606
15,959	5,920	1,521	7,193	1,737	1,354	66,316
8,730	28,820	9,433	16,483	18,577	0,051	17,906
41,988	8,856	6,623	1,798	1,611	20,534	18,590
40,552	27,742	7,253	10,121	2,289	0,000	12,043
0,255	54,649	10,863	6,287	0,216	0,000	27,729

29,589	22,589	8,856	14,043	11,429	0,530	12,964
40,044	19,204	6,939	6,354	7,191	0,025	20,242
45,457	14,096	6,868	9,964	3,155	4,444	16,015
42,468	8,792	16,592	13,303	1,195	0,000	17,650
6,808	8,147	3,985	1,509	0,685	23,212	55,652
29,883	25,823	10,890	8,540	5,737	3,739	15,387
41,652	13,182	11,512	18,944	4,546	0,031	10,133
18,069	24,736	18,418	8,736	6,891	0,242	22,909
36,337	27,561	4,819	3,222	6,642	0,061	21,358
37,516	22,237	4,741	4,112	9,266	0,056	22,072
37,867	21,860	5,542	3,630	9,014	0,053	22,033
17,660	28,924	7,819	6,713	18,359	0,045	20,480
16,673	29,200	7,773	11,953	16,662	0,032	17,706
17,239	16,739	11,799	10,655	3,078	11,393	29,098
30,900	11,405	13,131	20,269	5,683	0,000	18,612
54,341	11,349	11,429	5,243	4,961	2,647	10,030
12,542	32,808	7,277	14,142	12,832	0,258	20,142
31,163	21,764	21,333	3,513	5,597	0,532	16,097
42,855	18,645	6,046	6,235	6,680	0,043	19,495
54,944	8,597	9,457	7,285	2,943	0,891	15,883
48,969	7,765	6,954	12,723	5,769	0,049	17,771
50,649	18,072	6,024	7,248	2,352	0,300	15,355
30,141	17,616	8,635	19,163	9,470	0,029	14,946
53,885	11,236	8,580	9,005	2,165	0,000	15,128
15,434	11,616	2,772	8,222	1,544	8,430	51,981
13,167	27,709	11,317	22,660	10,336	0,000	14,811
16,899	6,232	3,089	11,594	0,955	8,967	52,264
43,627	14,394	12,406	10,268	3,570	0,000	15,735
42,168	18,863	9,683	11,227	3,029	0,000	15,030

41,499	15,751	7,902	12,623	6,369	0,158	15,698
44,186	7,426	3,165	7,399	0,754	0,044	37,025
43,842	6,873	4,834	6,158	25,211	0,000	13,082
44,844	14,785	6,954	12,968	6,111	0,608	13,730
17,564	8,921	7,130	4,762	1,138	13,655	46,831
32,458	18,440	9,442	11,258	6,240	0,684	21,479
45,062	15,989	9,150	4,216	2,359	0,045	23,178
17,007	27,109	8,686	12,189	13,081	0,000	21,927
43,974	19,708	11,222	8,968	2,894	2,465	10,768
43,929	14,635	8,396	10,721	1,657	0,350	20,310
6,666	24,616	9,846	13,234	0,761	0,553	44,323
22,904	9,903	7,942	16,489	1,708	4,384	36,671
26,949	24,442	10,502	10,825	11,247	0,053	15,982
25,862	5,152	3,212	6,482	2,160	2,285	54,848
30,526	20,895	4,173	12,587	9,325	0,359	22,136
48,091	11,271	11,436	9,771	4,092	0,041	15,297
26,737	2,489	12,214	14,621	1,050	27,666	15,224
38,191	18,150	13,171	7,579	3,434	0,000	19,475
49,116	12,851	14,101	9,868	4,464	0,013	9,587
20,620	6,881	1,979	9,589	0,786	0,000	60,146
14,016	4,709	2,877	5,173	1,256	9,680	62,289
28,590	13,748	13,198	14,203	6,083	0,000	24,177
11,758	21,894	8,807	17,967	2,097	3,125	34,351
6,496	14,845	13,274	9,907	0,757	9,808	44,913
34,570	18,011	15,158	7,908	7,024	0,026	17,303
45,191	10,791	6,459	14,110	4,722	0,000	18,728
36,393	6,540	9,993	15,181	2,103	12,744	17,046
31,179	13,761	29,187	9,549	2,773	0,063	13,489
52,456	8,579	3,936	12,079	7,445	0,000	15,505

22 Temmuz 2007 Seçimlerinin Oy Yüzdeleri

AKP	CHP	MHP	DP	GP	Bağımsızlar	Diğer
36,89	22,83	22,91	5,28	3,05	5,92	3,11
65,31	13,95	5,24	3,49	0,36	8,04	3,61
53,72	12,34	21,00	6,80	1,60	0,32	4,21
63,02	2,57	3,87	3,02	0,23	24,36	2,93
47,10	25,62	15,26	5,52	0,93	0,39	5,18
43,14	31,78	15,42	2,79	1,89	1,75	3,22
52,28	23,81	14,88	2,89	1,84	0,70	3,60
33,17	28,35	19,46	9,59	2,12	4,26	3,05
37,78	28,43	10,97	13,93	1,50	0,05	7,33
29,19	24,40	25,89	9,94	4,58	2,75	3,25
41,47	24,32	15,84	8,86	4,24	0,72	4,55
39,97	23,24	17,89	7,69	4,26	0,08	6,87
71,12	4,15	2,54	4,22	0,23	14,28	3,47
58,82	8,65	5,73	1,74	0,26	21,77	3,03
54,89	14,68	12,18	6,62	2,46	1,44	7,73
41,44	21,31	18,63	9,69	2,57	0,08	6,27
50,75	18,42	14,33	5,63	3,73	1,08	6,07
35,35	25,68	20,67	7,49	7,22	0,14	3,44
60,74	5,36	21,57	5,17	0,89	0,08	6,20
57,40	22,25	10,77	3,82	0,77	0,11	4,89
43,03	21,98	17,73	8,75	3,99	0,97	3,55
40,90	2,01	2,45	5,22	0,31	47,01	2,11
20,72	35,83	16,83	9,57	12,54	0,33	4,18
57,20	7,07	6,45	20,69	0,34	3,06	5,19
54,32	23,92	11,43	1,22	0,60	4,46	4,04
68,29	4,72	13,14	2,62	0,31	5,37	5,56
44,48	25,38	15,61	7,07	3,28	0,18	4,00

59,25	17,57	11,38	2,40	1,17	5,05	3,18
51,28	16,18	15,98	3,90	1,45	4,79	6,42
60,64	7,94	19,26	4,00	0,41	0,18	7,58
33,49	3,54	1,85	1,99	0,37	56,24	2,51
40,84	31,15	16,79	4,12	2,35	2,18	2,57
43,87	13,34	23,67	12,30	1,54	0,13	5,15
27,16	24,67	30,56	5,47	2,74	6,60	2,81
44,07	30,02	9,87	3,08	2,77	5,46	4,72
45,74	25,36	9,93	3,10	4,60	6,16	5,12
45,85	25,09	11,47	3,04	4,14	6,07	4,33
31,60	35,09	13,51	4,76	7,73	3,98	3,34
29,45	35,93	14,30	5,65	7,28	3,96	3,43
41,18	13,37	20,03	4,07	1,11	15,63	4,61
46,70	13,20	16,46	9,28	1,85	0,14	12,36
65,73	9,82	15,44	2,96	0,91	0,44	4,68
20,17	40,12	15,87	9,01	11,22	0,20	3,41
44,26	15,35	25,53	2,14	0,58	6,74	5,40
49,26	19,38	10,45	5,33	3,32	2,05	10,20
65,31	8,15	13,01	5,92	0,62	0,91	6,08
62,35	8,73	15,83	6,71	1,48	0,11	4,80
66,72	18,87	7,56	1,33	0,65	1,83	3,05
40,94	19,31	20,63	9,07	4,30	1,72	4,03
68,00	9,83	12,39	4,43	0,68	1,35	3,31
44,06	6,92	1,33	5,59	1,05	38,77	2,27
26,33	34,49	20,32	10,06	5,17	0,56	3,08
38,60	3,13	2,31	6,12	0,63	45,81	3,41
55,72	15,80	18,32	3,23	0,69	0,16	6,07
48,60	15,85	18,85	7,41	0,79	1,77	6,73
55,79	15,91	14,75	5,20	2,96	0,16	5,23

53,70	7,63	5,87	3,22	0,30	22,71	6,57
53,14	9,23	11,38	6,15	11,60	0,47	8,03
57,90	17,06	12,18	5,84	1,68	0,34	5,00
48,78	3,46	2,77	2,91	0,16	39,51	2,41
44,04	24,36	12,32	9,07	2,20	0,00	8,01
55,47	15,92	9,45	1,99	0,49	11,21	5,46
29,34	34,41	13,18	8,19	10,97	0,73	3,17
51,95	18,95	17,05	3,08	0,68	2,88	5,42
56,76	13,65	14,06	6,24	0,79	0,23	8,28
12,27	16,58	6,41	0,89	0,55	59,96	3,34
59,78	4,75	5,38	6,32	0,35	20,14	3,26
43,28	22,45	19,57	5,16	4,11	0,14	5,29
53,22	4,06	3,24	2,74	0,61	32,60	3,53
41,02	24,40	7,77	13,83	5,12	0,07	7,80
63,43	10,40	16,05	4,25	0,55	0,74	4,59
60,75	2,91	19,33	5,91	0,20	0,00	10,91
51,32	15,18	18,65	6,69	1,14	0,09	6,94
57,99	11,50	22,50	3,16	0,98	0,00	3,87
46,41	3,91	0,94	5,77	0,20	39,42	3,35
26,93	6,72	2,87	8,61	0,68	51,83	2,36
38,57	22,70	21,69	7,71	1,47	0,08	7,78
40,60	32,29	6,08	3,19	1,08	9,27	7,49
28,91	8,36	15,15	4,27	0,19	40,53	2,59
40,77	24,38	12,34	12,52	3,64	1,73	4,62
54,69	13,60	13,41	6,27	1,84	0,28	9,92
56,21	10,79	20,46	5,96	1,34	0,13	5,13
38,71	9,71	44,90	3,32	0,71	0,04	2,61
60,60	10,64	10,78	6,72	3,51	0,09	7,66

