

# RIDGE REGRESYON-RIDGE ANALİZ İLİŞKİSİ VE YAPILAN ÇALIŞMALARIN İÇERİKLERİ

Dr. Ercan BALDEMİR\*

## ÖZET

*Çoklu Doğrusal Regresyon Modellerinde parametre tahminlerinin bulunmasında en yaygın olarak kullanılan yöntem En Küçük Kareler Yöntemidir. Ancak, eğer açıklayıcı değişkenler arasında Çoklu Doğrusal Bağlantı varsa EKK tahminleri çok büyük varyanslara sahip ve dolayısıyla anlamsız sonuçlar vermektedir. Bunların ters işaretli olmaları da mümkündür. Ridge Regresyon Yöntemi EKK yöntemine alternatif bir metod olarak geliştirilmiştir. Modelde Çoklu Doğrusal Bağlantı varsa bu yöntem sapmalı fakat anlamlı ve tutarlı tahmin sonuçları vermektedir. Bu makalede Ridge Regresyon Yönteminin tarihi gelişimi ve yöntem üzerine yapılan bir kısım uygulama örnekleri incelenmektedir.*

## 1. Giriş

İktisadi olaylarda, sosyal bilimlerle ilişkisinden dolayı, kesin bir kural geliştirmek mümkün değildir. Her teori bir takım varsayımlara dayanır. İktisadi olayların kantitatif incelemesi olan ekonometri de kurduğu modellerde, sosyal bilimlerden kaynaklanan bir hata terimini modele ilave eder ve modelin, üzerine bina edildiği bu hata terimiyle ilgili bir takım varsayımlar ileri sürer. Bu varsayımlar altında model parametrelerini tahmin etmeye çalışır. Model parametrelerinin tahmininde en yaygın kullanılan yöntem En Küçük Kareler (EKK) yöntemidir. Gerek hata terimiyle ilgili ve gerekse modelle ilgili varsayımlar gerçekleştiğinde, EKK tahmincisi, tüm olası doğrusal sapmasız tahminler içerisinde en iyisidir. Normallik varsayımı sağlandığında EKK tahmincisi aynı zamanda en yüksek olabilirlik tahmincisi olur.

EKK da karşılaşılan en önemli problemler genel olarak, normallik, sabit varyans ve hataların bağımsızlığı temel varsayımlarıyla ilgili olan ve bu varsayımlardan kaynaklanan problemlerdir. Diğer problemler ise etkili veri noktaları, sapanlar, modelin fonksiyonel biçiminin zayıf spesifikasyonu, bağımsız değişkenler arasındaki yakın doğrusal bağımlılık ve bağımsız değişkenlerdeki hatalardır (Rawlings 1988).

Bu varsayımlar sağlanmadığında EKK tahmincileri yine sapmasız fakat tutarsız, güvenilir olmayan sonuçlar verir. Böyle durumlarda etkili, tutarlı ve güvenilir parametreler elde etmek için bir takım yöntemler geliştirilmiştir. Bu yöntemlerden bir kısmı, özellikle çoklu doğrusallık durumunda kullanılan Sapmalı Tahmin Yöntemleridir.

Çoklu Doğrusal Regresyon Yöntemi, istatistik metodları içerisinde oldukça yaygın olarak kullanılan bir yöntemdir. Veri analistleri bilim ve teknolojiye yakın her

\* D.E.Ü.İ.İ.B.F., Ekonometri Bölümü

alanda bunu kullanmışlardır. Bununla birlikte, açıklayıcıların veri matrisleri ortogonal olmadığında, regresyon katsayıları tahmininin bazı problemler getirdiği görülmüştür. Katsayılar mutlak değer olarak çok büyük olma eğilimindedirler, hatta bazılarının yanlış işaretli olmaları bile mümkündür. Bu gibi zorlukların olması ihtimalleri, açıklayıcı değişkenlerin ortogonaliteden daha fazla sapmalarına yol açar (Hoerl and Kennard 1981). Hoerl ve Kennard (1970) bu zorluklar için teorik temel vermişler ve bu kusurları olmayan yeni bir tahmin metodu, "Ridge Regresyon" u geliştirmişlerdir.

## 2. Tarihsel Gelişim ve Yapılan Çalışmalar

### 2.1. Ridge Analizinin Gelişimi ve Ridge Regresyonla Farkları

Ridge Analiz 1959 yılında ilk olarak Hoerl tarafından ikiden fazla boyutlu quadratik cevap yüzeylerinin incelenmesi için geliştirilmiştir. Ridge Regresyonla karşılaştırıldığında ki, Ridge Regresyon yöntemi çoklu regresyonda EKK tahminlemesinin bir alternatifidir, Ridge Analiz bu yüzeylerin davranışını grafiksel olarak ortaya koyar ve genel ve yerel optimum bölgeleri belirler. Bir tepki yüzey tekniği olan Ridge Analiz, açıklayıcı değişkenlerin sınırlandırılmamış ve karmaşık olmayan bir yapısı içerisinde geliştirilmiştir (Hoerl 1987). Hoerl, Dupont şehrinde istatistik grubunda çalışırken, kendisinden ısrarla iki ya da üç değişkenden daha fazla değişken içeren ve tepki yüzeyi literatüründe sık sık görülen endüstriyel süreçlerin optimum yapılması için gayret sarfetmesi istenmiştir. O zamanlar Kanonik Analiz yöntemi geliştirilmiş olmasına rağmen, bazı nedenlerden dolayı bu yöntemin kullanılması çok boyutlu yüzeyler için uygun olmamıştır (Carter, Chinchilli, Myers and Chambell 1986). Mühendislik eğitimi almış olan Hoerl, tahminlenen fonksiyonun sayısal optimizasyonundan ziyade başka şeylere de ihtiyaç duyuyordu. Faktörlerin etkilerinin detaylı grafiklerini vermesine ve aynı zamanda belirlenen bir merkezden herhangi bir uzaklık için yüzeyi optimize etmesine rağmen, bu teknik hemen kendisini gösterememiştir. O tarihlerde tepki yüzeyler analizinin daha yeni geliştirilmiş olmasının, bu yöntemin anlaşılmasında etkisi olmuş olabilir. Box ve Wilson (1951)'in, klasik çalışmaları da yalnızca sekiz yıl önce yayınlanmıştır (Hoerl 1985).

Ridge Regresyonla Ridge Analizinin karıştırılması Ridge Analizinin gelişmesini olumsuz etkilemiştir. Ancak Ridge Analiz, Ridge Regresyonun gelişmesine imkan sağlamış fakat kendi gelişmesini hak ettiği halde nazarlar Ridge Regresyona çevrilmiş ve bu yöntem daha hızlı gelişmiştir.

Çoklu regresyonda, çoklu doğrusal bağlantılı modellerde EKK 'nın anlamsız tahminlerinin sıklıkla ortaya çıkması, Ridge Analizinin gelişimi esnasında Hoerl'i şaşırtıyordu. Hoerl bir makalesinde (Hoerl 1962), regresyondaki hata kareler toplamının, katsayıların quadratik fonksiyonu olarak yazılabileceğini belirtmektedir. O halde Ridge Analizinde, katsayıların orjine daha yakın sabit sonuçlarına, hata kareler toplamının minimum ridgesi boyunca hareketinin hesaplanması ve işaretlenmesi için EKK yöntemi kullanılabilir. Fakat, katsayıların minimumdan ne kadar uzaklaştırılacağı problemi bu fikrin yayılmasını önlemiştir. Ancak 1970 yılında Kennard'ın yardımıyla, Ridge Regresyonu üzerine ilk makalenin yayınlanması gerçekleşmiştir (Hoerl ve Kennard 1970 a.b).

## 2.2. Ridge Regresyonun Tarihi Gelişimi

Ridge Regresyon Yöntemi ilk defa 1970 yılında Technometrics dergisinde yayınladıkları iki makale ile Hoerl ve Kennard tarafından geliştirilmiştir. Hoerl ve Kennard bu makalelerinde "Ridge Regresyon : Ortogonal Olmayan Problemler İçin Sapmalı Tahmin" başlığı altında, tam ranklı genel doğrusal hipotez modeline uyan çoklu regresyonda sapmasız tahminleme probleminin detaylı bir tartışmasını ortaya koymuşlardır. Bu makalede, sapmalı tahminleyici formunun alternatif bir karakterizasyonu ortaya konulmuş ve bu karakterizasyona dayalı sapmalı tahminciyi daha geçerli hale getirecek bir k sayısının varlığını gösteren "mevcudiyet (varlık) teoremi" (Banerjee and Carr 1971) de ispat edilmiştir.

Hoerl ve Kennard aynı dergideki ikinci makalelerinde de "Ridge Regresyon: Ortogonal Olmayan Problemlere Uygulamalar", başlığı altında, önceki makalede bahsedilen Ridge Regresyon metodu uygulamalarını izah etmeye çalışmışlardır. Bu açıklama için literatürden iki örnek esas alınmıştır. Bu makalede de özellikle çok faktörlü verilerdeki komple ilişkiyi tarif etmek için iki boyutlu bir grafik prosedürü olan Ridge İzi'ne dikkat çekilmiştir. Burada ayrıca klasik EKK kullanılarak elde edilen eşitliklerden daha iyi regresyon eşitliği bulmak için yapılabilecek birtakım işlemlerden bahsedilmiştir.

Hoerl ve Kennard'ın Technometrics dergisinde yayınlanan bu makaleleriyle ilgili ikiyüzden fazla makale yayınlanmıştır. Bu makalelerin büyük bir kısmı yine Hoerl ve Kennard'ın "American Journal of Mathematical and Management Sciences" de 1981, Vol.1, No.1 de yayınlanan makaleleriyle ele alınıp incelenmiştir. 1980 yılına kadar Ridge Regresyon üzerine yapılan çalışmalar incelenmiş ve;

- 1) Teoremin ortaya çıkışı,
- 2) simülasyonla desteklenen sapma parametresi k'nın seçimi algoritması,
- 3) kimya mühendisliği, hava kirliliği, Jeoloji, Pazarlama, Ekonomi ve Sosyoloji gibi alanlara (Hoerl and Kennard 1981) uygulamaları ihtiva eden basılmış makalelerin açıklamalı bibliyografyası verilmiştir.

## 2.3. Ridge Regresyon Üzerine Yapılan Çalışmaların Muhtevası

Hoerl ve Kennard (1981), makalelerinde 1980 yılına kadar Ridge Regresyon üzerine yayınlanmış 120 den fazla makalenin özetini verirler. Burada ridge analiz üzerine ilk çalışmanın 1962 yılında Hoerl'a ait olduğu görülmektedir. Bu makale Chemical Engineering dergisinde yayınlanmış ve kimya üzerine yapılmış bir uygulamayı içermektedir. 1970 yılına kadar yayınlanan dört makaleden en önemlisi yine Hoerl ve Kennard'ın yayınlamış oldukları ve Ridge Regresyonda ilk bilinen makale olan 1970 yılındaki makalenin materyallerini oluşturan 1968 deki raporlar'dır. Bu raporda Ridge Regresyonun özünü teşkil eden dört esas (Hoerl and Kennard 1981);

i)  $L^2 = (\beta^* - \beta)' X' X (\beta^* - \beta)$  kriteri üzerine yorumlar.

Burada  $\beta^*$  ridge tahmincisidir.

ii) Ridge'nin bayesian yorumu ve izahı,

iii) b sınırlı konveks set ile sınırlandırıldığında ridge karşılaştırması ve b tahmini,

iv) k seçim metoduyla k'nın bir başlangıç tahmini değeri olan  $k_a$  ve takip eden tahmin değeri  $k_{at}$  nin bulunması, olarak belirlenmiştir.

İkinci önemli bir kaynak ise Draper ve Smith tarafından 1969 yılında yayınlanan çalışmadır. Burada ise, büyük örneklerden açıklayıcı değişkenlerin en iyi alt kümesinin seçiminde kullanılan ve Ridge Regresyonda önemli bir yeri olan Ridge İzi açıklanmıştır.

Bu iki makale esas olmak üzere 1970 yılında Hoerl ve Kennard'ın bilinen meşhur makaleleri, biri teori diğeri de uygulama olmak üzere, "Technometrics" dergisinde yayınlanmıştır. Bu makalelerden sonra Ridge Regresyon üzerine yapılan çalışmalarda hızlı bir artış görülmektedir. Bu çalışmalar, Hoerl ve Kennard (1970 a.b) deki;

$$\beta^*(k) = [X'X + kI_p]^{-1} X'y \quad (1)$$

Ridge tahmincisi üzerine yoğunlaşırlar. Ridge Regresyonda amaç,

$$E(L_1^2) = E[(\beta^* - \beta)'(\beta^* - \beta)] = E[(\alpha^* - \alpha)'(\alpha^* - \alpha)] \quad (2)$$

ile ifade edilen tahminin ortalama karesel hatası (MSE) yi minimum yapacak k değerinin bulunmasıdır. Hoerl ve Kennard (1970a)'nın özü de, ridge tahmincisinin EKK dan daha küçük MSE si olan bir  $k > 0$  değerinin her zaman mevcut olduğunu göstermektedir (Dwivedi, Srivastava and Hall 1980). Sonraki yıllarda yapılan çalışmaların bir kısmı da, Ridge Regresyon hesaplamaları için bilgisayar programları hazırlamak şeklindedir.

Yine 1970 yılında Marguardt, D.W.'nin "Genelleştirilmiş Tersler, Ridge Regresyon, Sapmalı Doğrusal Tahmin ve Doğrusal Olmayan Tahmin" makalesi de, genelleştirilmiş tersler kullanılarak elde edilen sapmalı doğrusal tahminlerin tartışıldığı önemli bir çalışmadır.

Ridge Regresyon konusunda yapılan çalışmalardan k değeri bulmak için yapılmış ve yayınlanmış olanların sayısı ise bir kaç düzineyi bulmaktadır (Golub, Heath and Wahba 1979). Bu makalelerde özellikle k değerinin sıfıra çok yakın bir değer olarak belirlenmesi için çalışılmış ve k ile ilgili bir takım formüller geliştirilmiştir. Bu formüller simülasyon ve Monte-Carlo incelemeleriyle sınanarak en iyi k değeri tesbit edilmeye çalışılmıştır. Optimum k nın seçilmesi konusunda Lee ve Campbell'in 1985 yılında Commun Statist. Theory Meth. dergisinde yayınlanan makalelerinde istatistik literatüründeki k elemesi için yapılan bir takım mekanik kurallar tarif edilerek, bu kurallar optimal k nın seçimi için denenmişlerdir. Bu makalede amaç ise, optimal k yı seçmektir. Özellikle böyle bir optimal k nın varlığı ispat edilmiş ve genel olarak optimal k nın bir cebirsel kapalı forma sahip olmamasından dolayı bir iterative sayısal metod, optimal k hesaplamaları için kullanılmıştır (Golub, Heath and Wahba 1979). Bu stokastik optimal k nın örnek dağılımının matematiksel olarak belirlenmesinin güç olmasından dolayı, belirlenen optimal ridge kurallarının istatistiksel özelliklerini sınamak için bir Monte-Carlo incelemesi yürütülmüş ve simülasyon sonuçları verilmiştir.

Bu çalışmanın ardından 1986 yılında Singh, Chaubey ve Dwivedi, Genelleştirilmiş Ridge Regresyon tahmincisinin sapmasını azaltmak için "Jack-knife" denilen bir prosedür uygulayarak, Yaklaşık Sapmasız Genelleştirilmiş Ridge Regresyon tahmincisini geliştirmişlerdir. Singh ve Chaubey 1987 de yayınladıkları bir makalede Yaklaşık Sapmasız Genelleştirilmiş Ridge Regresyon tahmincisinin özelliklerini tartışmışlar ve Monte-Carlo deneyleriyle bu tahmincilerinin işlemsel uygulamalarını sınımlardır (Golub, Heath and Wahba 1979). Ancak onların Monte-Carlo deneyleri parametrelerin dar bir sınırı üzerinde yürütülmüştür. Hem de Klasik Ridge Regresyon tahminine karşılık gelen Yaklaşık Sapmasız Klasik Ridge Regresyon tahmincisinin işlemsel uyarlamaları dikkate alınmamıştır.

Daha sonra 1988 yılında yayınlanan bir makalede Masuo Nomura, ridge parametresinin stokastik olmadığı varsayımı altında ortalama karesel hata terimlerinde Yaklaşık Sapmasız Genelleştirilmiş Ridge Regresyon, Genelleştirilmiş Ridge Regresyon ve En Küçük Kareler arasında karşılaştırma yapmış ve Monte-Carlo deneyleriyle işlemsel Yaklaşık Sapmasız Klasik Ridge Regresyon tahmincilerinin küçük örnek özelliklerini denemiştir. Bu deneylerde, makalede anlatılan işlemsel ridge parametresi ve Hoerl, Kennard ve Baldwin (1975) ve Lawless ve Wang (1976) tarafından geliştirilen işlemsel ridge parametrelerini kullanmıştır.

Ridge Regresyon alanında yapılan çalışmalarda mümkün olduğu kadar sıfıra yakın  $k$  değeri arayışlarının yanısıra Ridge izinin bazı özellikleriyle ilgili de bir takım makaleler yayınlanmıştır. Bu makalelerde Ridge izinin karakteristikleri cebirsel olarak belirlenmeye çalışılmış ve (Golub, Heath and Wahba 1979);

- i) ferdî ridge tahminlerinin sırası,
- ii) bu tahminlerin değişim oranı sırası,
- iii) ferdî bir tahminde işaret sayısının değişmesi,

gibi özellikleri incelemiştir.

### **3.Ridge Regresyon Yöntemi Üzerine Yapılan Uygulama Örnekleri**

#### **3.1. Ridge İzine Dayanan Bir Uygulama**

Ridge Regresyonun nasıl kullanılacağına örneğini Hoerl ve Kennard (1970b) de daha önceden bilinen iki probleme uygulayarak göstermişlerdir. Bu örnekler Gorman ve Toman'ın ortogonal olmayan faktörlere sahip 10 faktör regresyon problemi ve Jeffer'in maksimum sıkılaşmayı incelediği 13 fiziksel faktör içeren 13 faktör örneğidir.

Hoerl ve Kennard (1970b), bu verileri tekrar türetmişler ve özdeğerleri bulduktan sonra  $k$ 'nın (0,1) arasında değişen 15 değeri ve  $\beta^*$ 'ları kullanarak Ridge izi noktalarını elde etmişler ve Ridge izini çizmişlerdir. Ridge izi faktör korelasyonlarının etkilerinin iki boyutlu bir tanımını verir ve hatta bütün regresyonların hesaplanıp hesaplanamayacağını gösteren mümkün değerlendirmeleri yapar (Hoerl and Kennard 1970b).

On-faktör örneğine Ridge Regresyon Yöntemi uygulandığında KEKK ile elde edilen katsayıların kesinlikle aşırı tahminlenmiş oldukları tesbit edilmiş ve bir

k sabiti ilavesiyle bir kısım değişkenlerin işaretlerinin değiştiği, bir kısmının ise süratle sifıra yaklaştığı, görülmüştür. Bir kısım faktörlerin ise farklı isimde olmalarına rağmen aslında aynı faktör oldukları ortaya çıkmıştır.

Hoerl ve Kennard 10 faktör örneğinden Cp istatistiği ile 9, 10, 4 ve 1 olmak üzere dört faktörü elimine etmişler ve kalan faktörler için tekrar Ridge İzi uygulaması sonucu faktör 1'in dışındaki bütün faktörlerde hala uygunsuzluklar ve aşırı tahminin mevcut olduğunu görmüşlerdir. Fakat 5 ve 7 oldukça durağandılar ve  $k > 0$  itibariyle hızla sifıra giderler. Bu faktörler açıklama güçlerini karşılayamadıkları için elimine edilecek faktörler olacakları görülmüştür. 5 ve 7 nin elimine edilmesiyle bulunan Ridge İzinde gereksiz tahminler ve uygun olmayan bileşikler gözle görülür bir şekilde bastırılmıştır. Elde edilen Ridge İzlerinden  $k=0.25$  in en uygun çözümü veren sabit olduğu görülmüş ve bu değeri kullanarak  $\beta^*$ 'i parametreleri elde edilmiştir.

Onüç-faktör örneğinde ise, yine onüç faktörle ilgili özdeğerler bulunmuş ve daha sonra Ridge İzi çizilmiştir. Buradan, katsayıların genellikle büyüme temayülünde oldukları, katsayı vektörünün kare uzunluğu KEKK uzunluğunun sadece %31 i olan 0.1 noktasında  $k=0$  daki 1.6'dan  $k=0.1$ 'deki 0.5'e düştüğü görülmüştür. Faktör 4 ve 5'in yanlış işaretlere sahip olduğu tesbit edilmiş ve  $k=0.2$  etrafında sistemin durgun olduğu ve faktörlerin değerlerinin daha iyi bir açıklama eşitliği verecek (0.2,0.4) aralığındaki katsayılara dayandığı görülmüştür.

Uygulamadan elde edilen sonuç ise şudur:

i) Jeffers'in Temel Bileşenler çözümü için  $k=0.2$  noktasındaki varyasyonun yüzde değeri 63.97 iken ridge çözümünde bu değer 68.67 olmuştur.

ii) İkinci bir mukayese ise hata kareler toplamının sabitliğidir.

Ridge İzinin kullanılması, sistemin durağan olduğu yeri ve bu noktada iyi bir katsayılar kümesinin bulunabileceği k değerini gösterirken temel bileşenlerde bu mümkün değildir. Durgunluk  $k=0.2$  ve  $k=0.4$  arasında meydana gelir ve iyi bir durağan nokta  $k=0.3$  dür. Boyut indirgeme yani, faktör elimine etme işlemi yapılması gerekirse, bu durumda Ridge İzinin nasıl uygulanacağına dair bazı prosedürler de verilmiştir. Bunlar;

1. Durağan katsayılar sınınsın ve en küçük açıklama gücüne sahip faktörler elensin.

2. Durağan olmayan katsayılar sınınsın ve açıklama güçleriyle yakalanamayacak bu faktörler elensin.

3. Geri kalan durağan olmayan katsayılardan biri veya daha fazlası silinsin.

Bu işlemler sonucunda 13 faktör örneğinde 2,4,9,10,11,12 ve 13 nolu faktörler silinmiş ve Ridge İzi çizilmiştir. Bu izden görünen ise kaba uygunsuzlukların elendiği, hata kareler toplamındaki anlamsız artışın yok olduğu ve varyasyon kıymetinin bütün faktörler için %73.09'dan %69.55'e düştüğü görülmüştür. Bazı aşırı tahminlemeler hala mevcuttur ve hata kareler toplamı  $k=0$ 'ın yakınında durgundur.

Özellikle, Çoklu Doğrusal Regresyonda tahminlenen vektörlerin ortogonal olmamalarının etkisi, tahminlenmeye çalışılan gerçek katsayılardan uzakta regresyon katsayılarının EKK tahminlerini verecektir. Katsayılar hem mutlak değer içinde çok fazla büyük ve hem de sadece işarete bakmak hatalı olabilir. Ayrıca EKK çözümü kullanışsızdır. Bu noktadan uzakta önemsiz bir hareket, katsayıların tamamen farklı tahminlerini verebilir. Yaygın olarak kullanılan görüntü ve boyut indirgeme süreçleri iki hataya sahiptir. İlki, bunlarda bilgi eksikliği vardır ve ortogonal olmamalarından dolayı kararsızlığa, aşırı tahminlere ve yanlış işarete ne ölçüde sebep olduğunu göstermezler. Ve ikincisi, bunlar gerçekten ortogonal olmayan veriler için KEKK'nın eksikliklerini büyütebilirler (Hoerl and Kennard 1970b). Yukarıdaki iki örnekte Ridge İzinin ortogonal olmama etkisini gerçek olarak ortaya koyduğu ve daha iyi bir nokta tahmini için yol gösterdiği ispat edilmiştir.

### 3.2. Karışık Verilere Ridge Uygulaması

Roger ve Hoerl 1987 yılında karışık verilere Ridge tekniğinin uygulanmasını gerçekleştirmişlerdir. Bu uygulama esas olarak Ridge Analizine dayanmaktadır. Görünüşte Ridge Analizinin karmaşık verilere uygulanması güç olmasına rağmen bu incelemede Ridge Analizi ile Quadratik Karışık Tepki Yüzeyler Analizi için iki metod anlatılmıştır. Thompson ve Myers (1968), Draper ve Lawrance (1965 a,b) tarafından geliştirilen bu metodlardan ilkinin, standart tepki yüzey dizaynlarıyla kullanılabilirliğini göstermek için ve ikincisinin de sınırlandırılmış dizaynlarda ferdi bileşenlerin etkilerini test etmede bir alternatif metod olmak üzere yapıldığı, ifade edilmektedir. Her iki metod da gerçek veriler üzerinde uygulanmıştır.

Çalışmada tepkiye bir quadratik yüzeyle yaklaşıldığı farzedilmekte ve Ridge Analizinin lokal optimumu ve hepsinin üzerinde bu lokalleştirme esnasında quadratik tepki yüzeylerin ridgelerini grafiksel olarak tanımladığı görülmektedir (Hoerl 1987). Ayrıca Ridge Analiz eşanlı olarak bütün verilerin etkilerini gösterir, bütün ridgelerin koordinatlarını işaretler ve ikincil (lokal) alanlarında optimal bölgeler bulur. Bu tekniğin kanoniksel analize karşı avantajlarından birisi, kullanışlı grafikler sağlayan ve dizayn bölge içinde ikincil (lokal) optimal bölgeler bulan muayyen noktadan (tecrübi bölgenin dışına taşabilen) çok, esas noktaya karşı gelen yüzeyleri yorumlamasıdır.

Örnekte  $q$  karışımli kısıtlanmış değişkenden  $q-1$  bağımsız değişkene kullanışlı bir transformasyon yapılmıştır. Ve Ridge Analiz, karışım olmayan durumlarda olduğu gibi yapılır (eğer quadratik model yüzeye yaklaştırmaya elverişli değilse). Ridge Analiz işaretleri orijinal  $q$  maddelerini ters dönüşüme geçirmekle yapılır. Ridge'lerin yorumu esasında aynıdır, yani karışım kısıtına karşılık yüzeyin lokal optimasının  $q$  boyutlu uzayda merkez noktasından uzağa hareket etmesidir.

Dizayn bölgesi  $r, W$  değişkenlerindeki birim alana karşılık gelir. Thompson ve Myers (1968)'in verdiği  $T$  türetme metoduyla açıklanmış tepki varyans fonksiyonunun ve değişkenlerinin değişebilir olduğundan bunların dizayn edilmemiş  $T$  ye göre bağımsız olmadıkları gösterilmiş ve bu dönüşen dizaynı ile tahmin varyansının sadece merkez noktasından tahminin yapıldığı noktaya bağlı

olduğu yani  $v$  den  $W$  ye eksenlerin dönüşümünün yüzeyin yorumu üzerinde etkisinin olmadığı ispat edilmiştir (Hoerl 1987). Öncelikle bu yaklaşım kesinlikle dönüşebilme dizaynına bağlı değildir fakat,  $T$  seçiminin daha az önemli olduğu dönüşümlere kapalıdır. Bir dizayn seçiminden önce bir formül yardımıyla orijinal hallere dönüştürülebilir (Hoerl 1987).

Çalışmanın sonucunda genellikle lokal optimayı ifade eden ikincil Ridge'lerin merkez noktası ile deney sınırları arasında bulunduğu görülmüştür. Optimal Ridge koordinatları ani olarak optimal seviyelerde elde edilmeli ve durağanlık belirlenmelidir. Kararsız davranış, modelde bir anormallik belirtisi, belki de uyum eksikliğidir.

Analiz arzu edildiğinde satır karışım özelliklerinde yapılan Ridge Analiz işaretleriyle  $W$  ve  $Z$  değişkenlerinin terimlerinde yapılabilir.

#### 4. Ridge Regresyon Hesaplamalarında Geliştirilen Bilgisayar Programları

Ridge Regresyon için hesaplamalar nisbeten tutarlı olup, basit modifikasyonlarla bir standart doğrusal regresyon programı yapılabilir. Hesaplamalarla ilgili temel bir takım bilgiler Evans (1973), Bolding ve Houston (1974) ve Gunst (1979) tarafından verilmiştir. Ridge hesaplamaları için yaygın olarak kullanılan programlarda mevcuttur.

Özel Ridge programları Bradley ve McGann (1977), Jain, Mahajan ve Bergier (1977), Hoerl (1979), Hui ve Sappal (1979) ve Bush (1980) tarafından geliştirilmiştir. İstatistikler için yaygın olarak kullanılan iki ticari program Health Sciences Computing Facility, California Üniversitesi Los Angeles'den Bromedial Computer Program (BMDP) ve SAS Enstitüsü, Güney Karolina'dan Statistical Analysis System (SAS) dır. Hill (1975) regresyon katsayılarını hesaplamak için farklı  $k$  değerleriyle bir Ridge izi çizebilen BMDP sisteminde bulunan programın bir kullanım örneğini vermektedir. Carmer ve Hirsch (1979), Sinha ve Handy (1979) ve Rogers ve Hildebrand (1980), Ridge katsayı hesapları için özellikle SAS Macros'u tavsiye ederler. Massachusetts Teknoloji Enstitüsünün elinde bulunan mevcut program istatistiksel araştırmalar için kapasiteli bir yön ve zaman paylaşımlı sistem olan TROLL'dur (Hoerl and Kennard 1981). Yapılan iş, milli bir data haberleşme ağına dayandırılmaktadır. Ridge etkinliği MIT (1975) de belirlenmiştir.

Hızlı gelişmeler yeni yeni bulunan ve geliştirilen  $k$  tahmin prosedürleri, geniş bir uygulama için daha yeni bir programa ihtiyaç göstermektedir. Henüz bütün  $k$  algoritmalarını kullanarak Ridge Tahminlerini bulan bir program mevcut değildir. Çünkü çalışmalar program üzerinde değilde en küçük  $k$ 'nin seçimi üzerinde yoğunlaşmıştır.

#### 5. Sonuç

1970 yılından beri Ridge Regresyon üzerine yapılan çalışmalar üçlü rakamlarla ifade edilebilecek boyutlardadır. Bunların 120'den fazlası 1981 yılında Hoerl ve Kennard tarafından yayınlanan bir makalede özetlenmektedir. Çalışmaların önemli bir kısmı Ridge Regresyon parametresi olan  $k$  nın bulunuşu üzerinedir. Çalışmaların büyük bir kısmı da Ridge izi ve  $k$  seçim yöntemlerinin uygulamalarını içermektedir.



## Abstract

Least Squares Method is used common in Multivariable Linear Regression Model. However, if there is a multicollinearity among the explanatory variables, Least Squares Estimation will give great variances and illogical results. These estimations may occur vice versa. Ridge Regression Method has developed as an alternative to the Least Squared Method. If there is a multicollinearity in the model, this method may give biased but meaningful and stable results.

In this article, the historical development of Ridge Regression and some of applications of methods are analysed.

## Kaynakça

Banerjee K.S. and Carr R.N., "A Comment On Ridge Regression Biased Estimation for Non Orthogonal Problems", **Technometrics**, Vol.13, No.4, November 1971, s.895-898.

Diane I. Gibbons, "Some Characterizations of The Ridge Trace", **Commun. Statist. Theor. Meth.** 13 (2), 173-182, 1984.

Gene H.Golub, Michael Heath, Grace Wahba, "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter", **Technometrics**, Vol.21, No.2, May, 1979.

Hoerl E.A. Kennard R.W, Ridge Regression - 1980, Advances, "Algorithms and Applications", **American Journal of Mathematical and Management Sciences**, 1981, Vol.1, No.1.

Hoerl R.W. , "Ridge Analysis 25 Years Later", **The American Statistician**, August 1985, Vol.39, No.3.

Masua Nomura, "On The Almost Unbiased Ridge Regression Estimator", **Commun. Statist. - Simula**, 17 (3), 729-743, 1988

Rawlings, **Applied Regression Analysis**, 1988, Brooks Golc, California.

Roger W. Hoerl, "The Application of Ridge Techniques to Mixture Data: Ridge Analysis", **Technometrics**, May 1987, Vo.29, No.2, s.161-172.

T.D.Dwivedi, V.K.Srivastava, R.L.Hall, "Finite Sample Properties of Ridge Estimators", **Technometrics**, Vol.22, No.2, May 1980.

Tje-San Lee and Don B. Chambell, "Selecting The Optimum K in Ridge Regression", **Commun Statist-Theory Meth.**, 14 (7) , 1589-1604, 1985

W.H.Carter and V.M.Chinchilli, R.H.Myers, E.D.Chambell, "Confidance Intervals md an Improved Ridge Analysis of Response Surfaces", **Technometrics**, November 1986, Vol.28, No.4.