

758457

T.C
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ BİLİM DALI

**KANTİL REGRESYON VE
ALTERNATİF REGRESYON MODELLERİ İLE
KARŞILAŞTIRILMASI**

Yüksek Lisans Tezi

İrem SAÇAKLI

Danışman: PROF. DR. SELAHATTİN GÜRİŞ

İstanbul, 2005

İÇİNDEKİLER

	Sayfa No
TABLO LİSTESİ	i
ŞEKİL LİSTESİ	ii
GİRİŞ	1
BİRİNCİ BÖLÜM	
ROBUST TAHMİN VE BAŞLICA ROBUST TAHMİNLERİ	
1.1 Robust Tahminler.....	4
1.2 Nonparametrik ve Dağılımdan Bağımsız Robust Yöntemler.....	5
1.3 Yerleşim ve Ölçek Parametrelerinin Robust Tahminleri	7
1.3.1 Yerleşim Parametresinin Basit Tahmincileri.....	8
1.3.2 Yerleşim Parametresinin Tek Değişkenli Robust Tahminleri	9
1.3.2.1 M Tahminleri	9
1.3.2.2 L Tahminleri.....	12
1.3.2.2.1 Trimmed ve Winsorized Ortalamalar.....	13
1.3.2.2.2 Üç Sıra İstatistiğinin Kombinasyonları.....	15
1.3.2.3 R Tahminleri.....	16
1.3.2.4 W Tahminleri.....	17
1.3.2.5 D Tahminleri.....	18
1.3.2.6 Tahminlerin Karşılaştırılması	20
1.3.3 Ölçek Parametresinin Basit Tahmincileri.....	21
1.3.4 Ölçeğin Tek Değişkenli Robust Tahminleri.....	25
1.3.4.1 M Tahminleri.....	25
1.3.4.2 L Tahminleri.....	27

1.3.4.3 R Tahminleri.....	28
---------------------------	----

İKİNCİ BÖLÜM

ALTERNATİF REGRESYON MODELLERİ

2.1 En Küçük Kareler Regresyonu.....	30
2.1.1 Basit Doğrusal Regresyon Modeli.....	30
2.1.2 Çoklu Doğrusal Regresyon Modeli.....	33
2.2 En Küçük Mutlak Sapmalar Regresyonu	37
2.2.1 Basit LAD Regresyonu.....	37
2.2.2 Çoklu LAD Regresyonu.....	41
2.2.3 En Küçük Mutlak Sapmalar Regresyonu' nun Özel Hali Olan Medyan Regresyon.....	45
2.3 En Küçük Medyan Kareler Regresyonu.....	48
2.4 M Regresyon.....	49
2.4.1 Basit M Regresyon.....	49
2.4.2 Çoklu M Regresyon.....	52
2.5 Rank Regresyon.....	55
2.5.1 Basit Rank Regresyon	55
2.5.2 Çoklu Rank Regresyon.....	59
2.6 Alternatif Regresyon Modellerinin Karşılaştırılması.....	62
2.7 Aşırı Değerlerin Belirlenmesi.....	67
2.7.1 Kutu Çizimleri.....	67
2.7.2 Histogram.....	68
2.7.3 Standartlaştırılmış Artıklar.....	69
2.7.4 Studentlaştırılmış Artıklar.....	69
2.7.5 Hat Matrisin Köşegenleri.....	70

2.7.6 Etki Ölçüleri.....	71
2.7.6.1 Cook'un D Değeri	71
2.7.6.2 DFBETAS.....	72
2.7.6.3 DFFITS.....	72

ÜÇÜNCÜ BÖLÜM

KANTİL REGRESYON

3.1 Bölenler.....	74
3.2 Anakütlenin Modellenmesi.....	75
3.2.1 Kümülatif Dağılım Fonksiyonu.....	75
3.2.2 Olasılık Yoğunluk Fonksiyonu.....	76
3.2.3 Kantil Fonksiyonu.....	77
3.2.4 Kantil Yoğunluk Fonksiyonu.....	79
3.3 Kantil Regresyon	80
3.4 Kantil Regresyonun Özellikleri	84
3.5 Kantil Regresyonun Doğrusal Programlama Gösterimi.....	85
3.5.1 Asimtotik Kovaryans Matris Tahmini.....	87
3.5.1.1 Sıra İstatistiği Tahmincisi	87
3.5.1.2 Bootstrap Tahmincileri.....	89
3.5.1.3 Kernel Tahmincisi.....	92
3.6 Kantil Regresyonda Sabit Varyansın İncelenmesi.....	93
3.7 Kantil Regresyonda Simetrinin İncelenmesi.....	95
3.8 Otoregresif Kantil Modeller.....	98
3.9 I_1 -Tahmincileri ile Kantil Regresyon.....	99
3.10 Gastwirth ve Tukey Tahmincileri ile Kantil Regresyon.....	102
3.11 Trimmed En Küçük Kareler Tahmincileri ile Kantil Regresyon.....	103

DÖRDÜNCÜ BÖLÜM

UYGULAMA

4.1 İktisadi Büyüme Kavramı.....	105
4.1.1 Neoklasik Büyüme Teorisi.....	107
4.1.2 İçsel Büyüme Teorisi.....	108
4.2 Uygulamada Kullanılan Veri ve Değişkenler.....	109
4.2.1 Birinci Model Denemesi.....	111
4.2.2 İkinci Model Denemesi.....	121
4.2.3 İki Modelin Karşılaştırılması.....	134
SONUÇ.....	136
Çalışmada Kullanılan Veriler.....	138
KAYNAKÇA.....	139

GİRİŞ

Değişkenler arasındaki ilişkiler incelenmek istendiğinde korelasyon ve regresyon analizi akla gelmektedir. Regresyon analizinde değişkenler arasındaki ilişkiler matematiksel bir model ile ortaya konmaktadır. Tahmin edilen regresyon modelleriyle nedenselliğin tek yönlü olmasına rağmen değişikliklerin etkisinin ne olabileceği açıklanabilmektedir. Gerek değişkenler arasındaki ilişkilerin iyi açıklanması, gerekse gelecek devre değerlerinin iyi tahmini, oluşturulan regresyon modelinin değişkenler arasındaki ilişkiyi iyi açıklamasına bağlıdır.

Günümüzde regresyon dendiğinde parametrik regresyon modelleri kastedilmektedir. Değişkenler arasındaki ilişkiyi açıklamak için parametrik modellerde, modelin fonksiyonel şeklinin doğru seçilmesi, modelde gerekli tüm değişkenlerin yer alması gibi bazı kriterler bulunmaktadır. Sonuç olarak en uygun parametrik regresyon modeli seçilse bile , “değişkenler arasındaki ilişkiyi açıklamak yeterli midir? Değişkenler arasındaki ilişkiyi daha iyi açıklayacak farklı modeller olabilir mi?” soruları her zaman düşünülebilir. Ayrıca bazı durumlarda modellerin tahmin edileceği verinin yapısı, aşırı değerlerin olması, serilerin normal dağılmaması gibi nedenlerle parametrik regresyon modelleri iyi sonuç vermeyebilir. Bu durumda alternatif regresyon modelleri gerekli olabilir. Parametrik regresyon modelleri normal dağılım varsayımını gerektirmektedir. Dağılımın normal olmadığı durumlarda alternatif regresyon modelleri gerekebilir.

Yukarıda sayılan nedenlerle alternatif regresyon modellerinin kullanılması gerekli olabilmektedir. Parametrik modellerde en uygun modelin seçilmesinden sözettiğimiz gibi, parametrik regresyon modellerinin de dahil olduğu alternatif regresyon modellerinden en uygununun seçilip kullanılması da sözkonusu olabilir. Kantil Regresyon, alternatif regresyon modellerinden biridir ve diğer regresyon yöntemleri gibi değişkenler arasındaki ilişkiyi açıklamakta kullanılmaktadır. Bu çalışmanın amacı; alternatif regresyon modeli olarak Kantil Regresyon’u açıklayarak diğer regresyon yöntemleriyle karşılaştırarak aradaki farklılıkları ve benzerlikleri ortaya koymaktır. Ayrıca çalışmalarımızda parametrik ve parametrik olmayan alternatif

regresyon modellerinin parametrelerinin tahmininde kullanılan yöntemlere de yerverilmiştir.

Çalışmanın girişten sonraki birinci bölümü Alternatif Regresyon Modelleri' ne altyapı sağlanması amacı ile oluşturulmuştur. Bu bölümde Robust Tahmin ve Robust Tahmin çeşitleri incelenmiştir.

Alternatif Regresyon Modelleri ve bunlarla ilgili tahmin yöntemleri çalışmamızın girişten sonraki ikinci bölümünü oluşturacaktır. Bu bölümde En Küçük Kareler (EKK) Regresyonu olarak adlandıracağımız parametrik regresyon modellerinden sonra ,En Küçük Mutlak Sapmalar (LAD), Medyan, En küçük Medyan Kareler (LMS), M, Rank Regresyon Yöntemleri ve bunların tahmininde kullanılan yöntemlere kısaca yerverilecektir. Bu bölümde amaç, sözkonusu yöntemlerle ilgili açıklamalar yapmanın yanında, alternatif regresyon modelleri içinde Kantil Regresyon'un yerini ve durumunu ortaya koymaktır.

Üçüncü bölüm "Kantil Regresyon" başlığını taşımaktadır. Bu bölümde kantil regresyonla ilgili açıklamalar ayrıntılı olarak yapılmaya çalışılmıştır. Kantil regresyonun tahmininde kullanılan yöntemler ve ilgili testler bu bölümde yer almaktadır.

Çalışmanın son bölümü uygulamaya ayrılmıştır. Uygulamada OECD ülkelerinin büyüme oranlarının incelendiği farklı model tahminleri yapılmıştır. Çalışmamızda ülkelerin büyümelerinin seçilmesinin nedeni, ülkeden ülkeye farklılıklar gösterecek büyüme oranlarında oluşacak aşırı değerlerin bulunabilmesidir. Bu bölümde büyüme ile ilgili En Küçük Kareler, M ve Kantil Regresyon Modelleri tahmin edilerek sonuçlar karşılaştırılmıştır.

BİRİNCİ BÖLÜM

ROBUST TAHMİN VE BAŞLICA ROBUST TAHMİNLERİ

Ekonometrik modeller ile ilgili tahminlerin yapılabilmesi için istatistiksel tahmin teorisinin bilinmesi ve tahmincilerin geçerli olan veya olmayan bazı özelliklerinden söz edildiğinde bunların ne olduğunun anlaşılması önemlidir.

Klasik tahmin teorisi normal dağılım temeline dayanmaktadır ve dağılımın normal olup olmadığının test edilmesi gerekir. Normal olmayan dağılımla belirtilmek istenen, normal dağılımdan daha kalın kuyruklu hata dağılımlarıdır. Bu kalın kuyruklu dağılımlarda, bağımlı değişkendeki aşırı gözlem değerlerini ifade eden aşırı değerlere¹ normal dağılıma göre daha çok rastlanır.

Aşırı değerler diğer gözlem değerlerinden uzakta bulunan ve onlarla tutarlılık göstermeyen değerlerdir. Bunlar diğer gözlem değerleriyle karşılaştırıldıklarında çok büyük veya çok küçük değerlerdir². Bu değerler daha çok gözlem değerlerinin oluşturduğu frekans dağılımı aşırı basık olduğunda ortaya çıkarlar. Aşırı değerler seri içinde bulundukları yere göre analiz sonuçlarını etkileyebilecekleri gibi, analiz sonuçları üzerinde herhangi bir etkileri de olmayabilir. Bu nedenle bir gözlem değeri aşırı değer ise bunun nedeninin araştırılması gerekir.

Aşırı değer, özel bir duruma sahip gerçek gözlem değeri olabileceği gibi veri girişi sırasında yapılan bir hatadan da kaynaklanabilmektedir. Aşırı gözlem değerlerini dışlayarak, geriye kalan gözlemler ile klasik tahmin ve testleri yaparak aşırı değerlerin zararlı etkisini önleyebilmek mümkündür; ancak çok parametre söz konusu olduğunda

¹ Aşırı değer outlier için kullanılmıştır.

² Kenneth D. Lawrence, **Robust Regression**, New York and Basel: Marcel Dekker, Inc., 1990.

bu iki adımı birbirinden ayırmak pek mümkün değildir. Bu nedenle klasik normallik teorisi aşırı değerlerden arındırılmış örneğe uygulanamamaktadır.

Aşırı değerleri dışlayan yöntemler yerine robust yöntemler de kullanılmaktadır. Yapılan uygulamalarda en iyi dışlama yöntemlerinin robust yöntemlerin performansına ulaşamayacağı görülmektedir. Robust yöntemler etkiel gözlemlerin etkisini azaltmaya çalışırlar³. Etkiel gözlemler (influential observations) analiz sonuçları üzerinde büyük etki yaratan aşırı değerlerdir.

Robust yöntemler gözlem değerlerini tamamen kabul etmek ile tamamen dışlamak arasında geçiş oluşturmaktadırlar. Bu nedenle aşırı değerleri dışlayan yöntemlere göre daha üstündürler⁴. Ayrıca robust yöntemler klasik yöntemlere göre daha asimtotik etkindir ve klasik yöntemlerin aşırı gözlem değerlerine duyarlılığına karşın, robust yöntemler duyarsızdır.

1.1. Robust Tahminler

Sözlüklere bakıldığında “Robust” kelimesi; sağlam, güçlü, kuvvetli olarak ifade edilmektedir. Robustness ise; sağlamlık, güçlülük, kuvvetlilik anlamlarına gelmekte olup tahminlerdeki küçük sapmalara karşı hassasiyet olmamasını ifade etmektedir.

Uygulamalarda genellikle dağılımın normal dağılım olduğu varsayılmaktadır; fakat dağılımın gerçek şekli normal dağılıma uymayabilir. Ayrıca normallik varsayımı yanında, bağımsızlık, eş dağılımlılık, tesadüfilik, Bayesgil analizlerde önbilginin doğruluğu⁵ gibi istatistiksel varsayımların tümü veya bir kısmı da sağlanamayabilir. Bu gibi durumlarda robust tahminler kullanılabilir⁶.

Robust tahmincileri dirençlilik ve etkinliğin robustness’ı olmak üzere iki özellik taşımaktadır. Bu özelliklerden birincisi tahmincinin birkaç aşırı değerden minimum derecede etkilenmesini ifade etmektedir. Dirençlilik özelliğini incelemek için

³ Ian T Jolliffe, *Principal Component Analysis*, New York :Springer-Verlag, 2002, s.232.

⁴ Frank R.Hampel, “*The Influence Curve and its Role in Robust Estimation*”, *Journal of the American Statistical Association*, Vol.69,No 346, June 1974, s.383.

⁵ Peter J. Huber, *Robust Statistics*, New York: John Wiley & Sons, 1981.

⁶ S. J. Finch, *Testing for Normality*, New Yor: Marcel Dekker Incorporated, 2001, s.288.

hassaslık eğrisi (SC) kullanılır. Hassaslık eğrisi, tahmincinin örnek büyüklüğüne göre eklenen bir gözlem değeri için hassaslığını ifade eden eğridir⁷. Buna göre, n-1 gözlemleri örnek için, T_{n-1} tahmincisi için direnç ölçüsü, x in her değeri için hassaslık eğrisi kullanılarak,

$$SC(x/T_n) = n(T_n(x_1, \dots, x_{n-1}, x) - T_{n-1}(x_1, \dots, x_{n-1}))$$

olarak elde edilir. Bu, x gözlem değeri için T_{n-1} tahmincisinin aşırı değerlere hassaslığını gösterir.

Robust tahmincinin ikinci özelliği olan etkinliğin robustness'ı dağılımlar için minimum varyansa sahip olmayı ifade etmektedir. Robust tahmincinin her iki özelliği de asimtotik olarak aynıdır.

1.2 Nonparametrik ve Dağılımdan Bağımsız Robust Yöntemler

Geleneksel olarak robust yöntemler nonparametrik robust yöntemler ve dağılımdan bağımsız robust yöntemler olarak ikiye ayrılır.

Parametrik ve parametrik olmayan yöntem ayrımı parametre kavramına dayanmaktadır. Parametre, anakütleleri birbirinden ayırmaya yarayan ortalama, oran, varyans gibi ölçülerin genel ifadesidir⁸. Genel anlamı ile temeli ölçülere dayanan yöntemlere “parametrik yöntemler” adı verilmektedir. Bu yöntem grubunun dışında kalan ve uygulanması için çeşitli parametrelerin hesaplanmasına gerek olmayan yöntemlere ise “parametrik olmayan (nonparametrik) yöntemler” adı verilmektedir. Diğer bir ifade ile, parametrik olmayan yöntemler katsayısal olmayan veya dağılımda bağımsız olan yöntemlerdir⁹.

Uygulamada parametrik yöntemlerin parametrik olmayan yöntemlere tercih edildiği ve kullanımlarının daha yaygın olduğu görülmektedir. Bunun nedeni parametrik

⁷ Finch, s. 289.

⁸ John E Freund, *Mathematical Statistics*, 5. Edition, New Jersey :Prentice-Hall International Editions, 1999, s:183.

⁹ Paul Newbold, *İşletme ve İktisat için İstatistik*, Çev. Ümit Şenesen, Literatür Yayınları, No.44, İstanbul, 2000, s:431.

yöntemlerin parametrik olmayan yöntemlere göre kullanılmalarının ve yorumlanmalarının daha kolay olmasıdır. Parametrik olmayan yöntemlerin uygulanması için gerekli varsayımların sayısının azlığı nedeni ile parametrik yöntemlere göre daha geniş uygulama alanı sağlanabilmektedir¹⁰.

Bir parametrik testin uygulanabilmesi için; veri türü, dağılım türü, örnek sayısı bilinmelidir. Bu şartları sağlayan bir anakütle ve örnek varsa parametrik bir test uygulanabilir ve anakütlenin bir parametresine ilişkin bazı sonuçlara varılabilir. Ancak her zaman istenen dağılımda bir anakütleyle, istenen türde verilere ve birim sayısı yeterli örneğe sahip olunmayabilir. Böyle durumlarda parametrik olmayan testler kullanılabilir.

Parametrik yöntemler örneğin alındığı anakütle ile ilgili varsayımların geçerliliğine dayandığından anakütle dağılımı ve parametreler ile ilgili varsayımların kesin olarak geçerli olması gerekmektedir. Herhangi bir varsayımın sağlanamaması ilgili yöntemin kullanılmasını engellemektedir. Ayrıca örneğin alındığı anakütle ile ilgili varsayımların geçerliliğini sağlamak araştırmacı veya araştırmacılara bağlı değildir. Varsayımların geçerli olması halinde sözkonusu parametrik yöntemi uygulayabilecek olan araştırmacı, varsayımlardan bir veya birkaçının geçerli olmaması durumunda araştırmasını yapmak için seçmek istediği parametrik yöntemi uygulayamayacaktır¹¹.

Bir testin dağılımdan bağımsız olarak adlandırılabilmesi için sıfır hipotezini reddetme olasılığının tüm mümkün sürekli dağılımlar için aynı olması gerekir. Dağılımlar arasındaki eşitliği test etmek için kullanılan iki örnek için yapılan rank testi* bunun tipik örneğidir. Dağılımdan bağımsız testlerin çoğu kararlı** (stable) güce sahiptirler.

¹⁰ Ebru ÇAĞLAYAN, “Yarıparametrik Regresyon Modelleri ile Yaşam Boyu Sürekli Gelir Hipotezinin Türkiye Uygulaması”, MÜ Sosyal Bilimler Enstitüsü (Yayınlanmamış Doktora Tezi), İstanbul 2002.

¹¹ William Jay Conover, *Practical Nonparametric Statistics*, Newyork: John Wiley & Sons Inc., 1971, s.2.

* Parametrik olan hipotez testlerinde iki bağımsız örneğin aynı anakütleden gelip gelmediğini test etmek için t testi kullanılır. Parametrik hipotez testlerinde gerekli olan varsayımların yerine getirilemediği durumlarda t testi yerine Mann-Whitney U Testi (Wilcoxon Two Sample Test) kullanılır.

** Kararlılık; denemelerin sonucunda farklı sonuçlar olmamasıdır.

Dağılımdan bağımsız testlerle yapılan tahminler çoğu zaman “dağılımdan bağımsız” olarak adlandırılırsa da bu adlandırma yanlıştır. Nokta tahminlerinin stokastik davranışları temel testlerin gücü ile bağlantılıdır ve dağılıma bağlıdır bunun tek istisnası rank testlerinden türetilmiş aralık tahminleridir. Örneğin, belirlenen iki örnek kantili arasındaki aralık sabit olasılıkla gerçek medyan değeridir. Bu durumda da aralığın uzunluğu dağılıma bağlıdır.¹²

Kendilerini dağılıma uyarlayabilen tahminciler veren yöntemler “adaptive (uyarlayıcı) yöntemler” olarak tanımlanırlar¹³. Adaptivity ile robustness arasındaki bağ net değildir. Robustness’ ta güven, adaptivity’de etkinlik ön plandadır. Tahminin ya da test istatistiğinin değeri örnekteki küçük değişikliklere karşı duyarsızsa, istatistiksel yöntem “resistant (direncili)” olarak adlandırılır.

1.3 Yerleşim ve Ölçek Parametrelerinin Robust Tahminleri

Daha önce söz edildiği gibi dağılımın normal olmadığı durumlarda robust tahmin yöntemleri kullanılır. Normallik varsayımının sağlanmaması, dağılımların çarpık olması, simetrik olmaması ve aşırı gözlem değerlerinin olması robust tahminlerin yapılmasına neden olmaktadır. Bu bölümde robust tahmin yöntemleri için önemli olan yerleştirme ve ölçek parametrelerine* yer verilmiştir.

Bir dağılımın merkeze olan uzaklığını belirlemek amacı ile kullanılan parametrelere “yerleşim parametresi” adı verilir. Yerleşim parametresi olarak herhangi bir parametre kullanılabilirse de, yaygın olarak aritmetik ortalama ve medyan kullanılmaktadır.

Simetrik dağılımlarda genellikle ortalama veya medyan değeri olan “merkez (center)” etrafında simetrik dağılım söz konusudur. Simetrik dağılımlar için uygulanan robust yöntemler için merkezden eşit uzaklıktaki gözlem değerlerinin eşit ağırlıklandırılması gerekmektedir. Merkezden uzakta bulunan gözlemler, özellikle

¹² Huber, s. 6.

¹³ Robert V. Hogg, “Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory”, *Journal of the American Statistical Association*, Vol.69, No.348, (Dec. 1974),s.909.

* Location için yerleştirme, Scale için ölçek kelimesi kullanılmıştır.

çarpık dağılımlarda ve aşırı değerlerin olması durumunda yerleşim tahminini daha az etkilerler¹⁴. Buna karşılık simetrik olmayan dağılımlarda yerleşim'in neresi olacağı, diğer bir ifade ile yerleşim parametresinin ne olacağının belirlenmesi güçtür. Bu nedenle yerleşim parametresinin robust tahmini normallikten simetrik sapmalara veya simetrik bir dağılımın aşırı değerlere sahip olarak bozulmasını dikkate almaktadır.

Dağılımın merkeze olan uzaklığı yerleşim parametresi ile belirlenirken, bu uzaklığın ölçümü de önemlidir. Uzaklığı ölçmek için kullanılan parametreye “ölçek parametresi” adı verilmektedir. Ölçek parametresi olarak genellikle standart sapma kullanılmaktadır. Normallik varsayımı altında örnek ortalaması ve standart sapması en iyi yerleşim ve ölçek tahminleridir; ancak normallik varsayımının sağlanmadığı durumlarda bunlar iyi sonuç vermezler ve robust değildir.

1.3.1 Yerleşim Parametresinin Basit Tahmincileri

Bir dağılımın merkeze olan uzaklığını belirlemek amacı ile kullanılan yerleşim parametresinin en yaygın olarak kullanılan yerleşim tahmincileri örnek ortalaması ve örnek medyanıdır. Bunların hesaplanması kolaydır ve uygun varsayımlar altında istenen tüm özellikleri sağlarlar. Diğer basit yerleşim tahmincileri ise örnek modu ve midrange'dır.

Örnek ortalaması klasik yerleşim tahmincisidir. Örnekleme dağılımı normal dağılıma uyduğu zaman en etkin tahmincidir ve En Çok Benzerlik Tahmincisi'dir. Örnek ortalamasında örnekteki tüm gözlem değerlerine eşit ağırlık verildiği için merkezden uzak olan gözlemlerin tahmin üzerine fazla etkisi olmamaktadır. Örnek ortalamasının hassaslık eğrisi;

$$SC(x/\bar{x}) = n(\bar{x}_n - \bar{x}_{n-1}) = \frac{n}{n-1}(x - \bar{x}_n)$$

olarak elde edilir. Burada $\bar{x}_n$ ve $\bar{x}_{n-1}$, n ve n-1 birim sayılı örneklerin örnek ortalamalarıdır. x ortalama dan uzaklaştıkça hassaslık doğrusal olarak artar, bu da örnek

¹⁴ Finch, s. 290.

ortalamasının özellikle aşırı değerlere ve çarpık dağılımlara dirençli olmadığını göstermektedir.

Örnek medyanı yerleşim'in klasik robust tahmincisidir. Medyan hesaplanırken örnek hacmi tek sayı olduğunda merkezi sıra istatistiğine 1 ağırlığı, diğer tüm gözlemlere de 0 ağırlığı verilmektedir. Örnek medyanı sadece merkezdeki bir ya da iki gözlem değerine dayalı olduğundan aşırı değerlere ve çarpık dağılımlara karşı robust'tır. Medyan aşırı değerleri göz ardı ettiği gibi aynı zamanda örnek ile ilgili diğer bilgileri de göz ardı etmektedir ve bazı durumlarda diğer tahmincilere göre daha büyük varyanslıdır.

Örnek midrange'ı ise,

$$MR = (x_{(n)} - x_{(1)})/2$$

olarak tanımlanmaktadır. Maksimum ve minimum gözlem değerlerine 0,5 ağırlığı, diğer gözlem değerlerinin hepsine 0 ağırlığı verilir. Midrange aşırı değerler ve çarpık dağılımlardan en çok etkilenen yerleşim tahmincisidir, bu nedenle robust değildir.

1.3.2 Yerleşim Parametresinin Tek Değişkenli Robust Tahminleri

Literatürde birçok tek değişkenli robust yerleşim parametresi tahminleri bulunmaktadır. Yerleşim parametresinin elde edilmesinde en yaygın olarak kullanılan tahminler M-Tahminleri, W-Tahminleri, L-Tahminleri, R-Tahminleri ve Adaptive Tahminler'dir.

1.3.2.1 M Tahminleri

En çok benzerlik yöntemine dayanan ve bu tahminde kullanılan tahminci M-Tahminci olarak adlandırılır. Bunlar, $\rho(x_i, T_n)$ şeklindeki amaç fonksiyonunun minimizasyonu ile belirlenirler.

T_M ile gösterilen M-Tahmini,

$$\sum_{i=1}^n \rho(x_i; T_n)$$

ifadesini minimize eden T_n değeridir¹⁵. Burada, $x_1, \dots, x_n$ 'in gözlem değerleri, T_n yerleşim tahminidir.

$$\psi(x_i, T_n) = \delta \rho(x_i, T_n) / \delta T_n$$

şeklinde tanımlanarak M-Tahmini,

$$\sum_{i=1}^n \psi(x_i; T_n) = 0,$$

denkleminin çözümü ile elde edilir.

Yerleşim'in sabit olması için ρ ve ψ da $(x_i, T_n); (x_i - T_n)$ şeklinde olmalıdır. Scale'in sabit olması için M-Tahmini gözlemlerin scale'ini de hesaba katmalıdır. Böylelikle $\sum \psi(x_i - T_n)$ genellikle örneğe dayalı S_n yardımcı scale ölçüsünü de içerir. Burada S_n sabittir, c ayarlama sabitidir, gözlemleri yeniden ölçeklendirir ve S_n ' in çarpanıdır. S_n ' in en çok kullanılan şekli örnek medyanından (MAD) ve dördüncü yayılımdan (d_F) belirlenen medyan mutlak sapmadır.

Bu durumda M-Tahminleri,

$$\sum \psi(y_i) = 0$$

denkleminin çözümüdür. Burada $y_i = \frac{x_i - T_n}{c \cdot S_n}$, dir. En uygun c ve S_n seçilerek T_M ,

$$T_M = \left(\sum_{i=1}^n w_i y_i \right) / \left(\sum_{i=1}^n w_i \right)$$

¹⁵ Finch, s.291.

olacaktır. Burada $w_i = |\psi(y_i)/y_i|$ ’ dir. M-Tahminleri aşamalı yöntemler ile elde edilebildiklerinden belirlenmesi en zor robust tahminlerdir.

Her bir gözlem değerinin tahminci üzerindeki etkisi, etki eğrisiyle belirlenebilir¹⁶. M-Tahmincisi T_M için; T parametresiyle etki eğrisi fonksiyonunun etkisi,

$$T_M = T + n^{-1} \sum_{i=1}^n IC(y_i) + R_n$$

olarak elde edilir. Burada yer alan R_n artığı ihmal edilebilir, $IC(y_i)$ payı $\psi(y_i)$ olan bir fonksiyondur. Böylelikle her bir x_i gözlem değerinin tahminci üzerindeki etkisinin $\psi(y_i)$ ile orantılı olduğu görülür.

M-Tahmincilerinin özel bir hali örnek ortalaması (normal dağılım için En Çok Benzerlik Tahmincisi) ve örnek medyanıdır (Çift Üssel ya da Laplace Dağılımı için En Çok Benzerlik Tahmincisi). Yaygın olarak kullanılan M-Tahmincileri Huber’in, Tukey’in, Hampel’in, Andrew’un M-Tahmincileridir. Daha sonra M-Regresyon inceleneceği ve orada gerekli olduğu için burada sadece Huber’in M-Tahmincilerine yer verilecektir.

Amaç fonksiyonunun artıkların karesi olması durumunda;

$$\rho\left(\frac{x_i - T_n}{c.S_n}\right)$$

denklemini sabitler ihmal edilerek $T_M = \bar{x}$ ’yı veren,

$$\sum_{i=1}^n (x_i - T_n) = 0$$

¹⁶ Hampel, s.383.

denklemini ile çözülür. Dikkat edilirse bu tahmin $c.S_n$ ' ye bağlı değildir. $\psi(y) = x - T_n$ olmasından dolayı aşırı gözlemlerin etkisi sınırlandırılmamıştır böylelikle ortalama, robust değildir. Amaç fonksiyonu T_n ' den mutlak sapma olarak alındığında, $c.S_n$ ' e bağlı olmayan M-Tahmini medyan olacaktır.

Huber'in M-Tahmincisi (T_H);

$$\psi(y) = \begin{cases} y & |y| \leq k \\ k \operatorname{sgn}(y) & |y| > k \end{cases}$$

fonksiyonuna dayalı monoton bir tahmindir. Yani $y_i < y_j$ ise $\psi(y_i) \leq \psi(y_j)$ olacaktır.

1.3.2.2 L Tahminleri

Bu tahminler için kullanılan L-Tahmincileri, sabit ağırlıklı örnek sıra istatistiklerinin doğrusal kombinasyonları olarak adlandırılırlar. L-Tahmincileri aşırı değerlere ağırlık vererek dirençliliklerini veya aşırı olmayan değerlere yeteri kadar ağırlık vermeyerek etkinliklerini kaybedebilirlerse de, hesaplanması en kolay robust tahmincilerdir.

$$Y_i = \beta + e_i$$

şeklindeki basit yerleştirme modelinde kantiller kullanıldığında, F sürekli ise ise; $Q(\theta)$ ile gösterilen θ .' cı kantili ($0 < \theta < 1$), y ' nin $P(y < Q(\theta)) = F(Q(\theta)) = \theta$ eşitliğini sağlayan değeridir. θT tamsayı değilse, θ .' cı örnek kantili $y_{(n)}$ ' dir. Burada $n = [\theta T] + 1$ ' dir ve $[\theta T]$ de θT ' yi aşmayan en büyük tamsayı değerini göstermektedir. θT bir tamsayı ise θ .' cı örnek kantili tek bir değer olarak bulunamaz ve $y_{(\theta T)}$ ile $y_{(\theta T+1)}$ arasında herhangi bir değer olarak alınabilir. Bundan dolayı örnek

kantilleri ve bunların herhangi doğrusal fonksiyonları, yerleştirmenin L-tahmincileridir¹⁷.

Bu tahminler ağırlıkların aldıkları değerler yerine gözlemlerin sırasına dayanmaktadır ve bu tahminlerin ağırlıkları örnek büyüklüğüne göre sabittir. L-Tahmini T_L olarak ifade edilirse;

$$T_L = T(F_n) = \sum_{i=1}^n a_i x_{(i)}$$

olarak hesaplanır. T_L , her i için $0 \leq a_i \leq 1$ aralığında yer alır ve $\sum_{i=1}^n a_i = 1$ ' dir. Simetrik dağılımlar için merkezi yerleşim tahmininin en iyi doğrusal kombinasyonları, ağırlıkları simetrik olarak $a_i = a_{n-i} + 1$ şeklinde belirler. L-Tahmininin robustness'lığı belirlenen ağırlığa bağlıdır. Örneğin düşük ya da yüksek sıra istatistikleri için belirlenen küçük ağırlıklar, aşırı değerler ve çarpık dağılımlara karşı robust olacaktır.

En çok bilinen L-Tahminleri, trimmed ortalamalar, Winsorized Ortalamalar, doğrusal ağırlıklandırılmış ortalamalar ile medyan ve iki simetrik sıra istatistiğinin kombinasyonlarıdır.

1.3.2.2.1 Trimmed ve Winsorized Ortalamalar

Trimmed (düzeltilmiş) ortalama, %5' lik en küçük ve en büyük değerlerin belli bir yüzdesinin veri grubundan çıkarılması sonucu elde kalan verinin aritmetik ortalamasıdır. veri grubunda aşırı değerler olması durumunda bu yöntem merkezi konumu daha iyi belirler¹⁸.

Aşırı değerlerin olmadığı normal örneklerde hesaplanan trimmed ortalamanın etkinliği, aritmetik ortalamaya göre daha iyidir. Trimming yüzdesi aşırı değerleri dışladığı için, bunlar L-Tahmincisine çok az etki etmektedir. Örnek ortalaması $\bar{x}_0$ ve

¹⁷ George G. Judge ve Diğerleri, *The Theory and Practice of Econometrics*, 2.Baskı, Canada: John Wiley, 1985, s.838.

¹⁸ R. Koenker, "The Trimmed Mean in the Linear Model", *The Annals of Statistics*, Vol.15, No.1(Mar.1987), s.39.

medyanı, α 'nın yaklaşık olarak %50' ye eşit olduğu trimmed ortalamadır. Trimmed Ortalama Broadened medyan; α -trimmed ortalama ise α -truncated ortalama olarak da adlandırılır.

$K < n/2$ olduğunda α ; k/n oranını göstermesi durumunda 100α %-trimmed ortalama,

$$\bar{x}_\alpha = \frac{\sum_{i=k+1}^{n-k} x_{(i)}}{n-2k}$$

olarak elde edilir. Elde edilen bu değer, en büyük k ve en küçük k gözlemlerinin örnekten atılmasından sonra kalan gözlemlerin $(n-2k)$ ortalamasıdır. α 'nın önceden belirlendiği ve $n\alpha$ 'nın tamsayı olmadığı durumlarda $\bar{x}_\alpha$ trimmed ortalaması en yüksek ve en düşük trimmed gözlemlerin kısmi ağırlıklandırılması ile,

$$\bar{x}_\alpha = \frac{(k - n\alpha)x_{(k)} + (k - n\alpha)x_{(n-k+1)} + \sum_{i=k+1}^{n-k} x_{(i)}}{n(1 - 2\alpha)}$$

olarak hesaplanır. Burada $k = [n\alpha] + 1$ 'dir. Trimmed ortalamanın varyansı daha sonra gösterilecek olan örnek Winsorized varyansı kullanılarak da tahmin edilebilir.

Winsorized ortalamalar trimmed ortalamalara benzerler. Aşırı gözlem değerleri özel olarak ortalamanın hesaplanmasında kullanılmaz, en yüksek ve en düşük "trimmed" değerler, en yüksek ve en düşük değere gizlenir (censored). Böylece ortalama hesabına dahil edilir.

α -Winsorized ortalama,

$$\bar{x}_\alpha^W = \frac{\sum_{i=r+1}^{n-r} x_{(i)} + rx_{(r+1)} + rx_{(n-r)}}{n}$$

olarak tanımlanır. Burada $\alpha = r/n$, $r \leq [n-1/2]$ ’ dir. n tek veya çift olduğunda $r=(n-1)/2$ ’dir. $r=(n-1)/2$ olduğunda Winsorized Ortalama medyana eşit olacak, $r=0$ olduğunda ise ortalama değeri belirlenecektir.

1.3.2.2.2 Üç Sıra İstatistiğinin Kombinasyonları

Üç sıra istatistiğinin doğrusal kombinasyonu, yerleşim tahmincisi olarak kullanılır. Dağılımın simetrik olması durumunda genel tahminci,

$$\bar{x}(r, \gamma) = \gamma(x_{(r+1)} + x_{(n-r)}) + (1-2\gamma)x_M$$

olur. x_M medyandır¹⁹. Tek ve çift örnek hacimleri için uygun tanımlanmıştır. Diğer iki sıra istatistiği medyan etrafında sıralı olarak simetriktir. Trimean’i içeren iki tahminci,

$$\bar{x}_{TRI} = 0,25(h_l + h_u) + 0,5x_M$$

olarak elde edilir. Burada h_l ve h_u sırası ile örneğin en düşük ve en yüksek noktalarıdır. Başka bir tahminci de²⁰,

$$\hat{\beta}_G = 0,3(\hat{Q}(1/3)) + 0,4(\hat{Q}(1/3)) + 0,3(\hat{Q}(1/3))$$

olarak elde edilir²¹. Burada $\hat{Q}(\theta)$, θ ’ cı örnek kantili ve $\hat{Q}(\alpha)$ ve $\hat{Q}(1-\alpha)$ gibi iki kantil arasında yer alan tüm gözlemlerin örnek ortalaması olan Trimmed Mean (kesilmiş ortalama)’ dır. Bir başka yaklaşım ise 0.25, 0.5 ve 0.25 tartılandırma Tukey veya Trimean Metodu’dur²².

¹⁹ Frederick Mosteller, "On Some Useful Inefficient Statistics", *Annals of Mathematical Statistics*, Vol.17, No.4, (Dec.1946), s.377.

²⁰ Joseph L. Gastwirth, "On Robust Procedures", *Journal of the American Statistical Association*, Vol.61, No.316, (Dec., 1966), s.929.

²¹ Edwin L. Crow, And M.M. Siddiqui, "Robust Estimation of Location", *Journal of the American Statistical Association*, Vol.62, No.318, (Jun., 1967), s.353.

²² Ali Hakan Büyüklü, "Robust Tahmin Edicilerin Kullanılma Sebebi ve Sonuçları", IV.Ulusal Ekonometri ve İstatistik Sempozyumu Bildirileri, M.Ü., Antalya- 14-16 Mayıs 1999, s.655.

1.3.2.3 R Tahminleri

Rank Testleri'ne dayalı nonparametrik yerleşim tahmincilerine rank(R) tahmincileri denir. Anakütle simetrikse, yerleşim için herhangi bir rank testi kullanılarak R tahmincisi elde edilebilir.²³.

Gözlemleri $x_1, \dots, x_m$ ve $y_1, \dots, y_n$ dağılımları $F(x)$ ve $G(x) = F(x - \Delta)$ olan iki bağımsız örnek rank testini düşünelim. İki örnek $m+n$ hacminde birleştirilsin, R_i birleştirilmiş örnekte x_i ' in rankı olsun. $a_i = a(i)$ verilmiş olsun, $\Delta > 0$ olduğunda $\Delta = 0$ ' a dayanan test ;

$$S_{m,n} = \frac{1}{m} \sum_{i=1}^m a(R_i)$$

olur. Genellikle, a, j ' nin bir fonksiyonu olarak tanımlanır.

$$a_i = j\left(\frac{i}{m+n+1}\right)$$

ancak a_i ' yi j ' den türetmenin başka yolları da vardır.

$$a_i = (m+n) \int_{(i-1)/(m+n)}^{i/(m+n)} j(t) dt$$

olarak da tanımlanabilir.

$m=n$ olduğunda,

$$S(F, G) = \int_0^1 j\left(\frac{1}{2}S + \frac{1}{2}G(F^{-1}(S))\right) ds$$

olacaktır. Burada $S_{n,n} = S(F_n, G_n)$, F_n ve G_n ayrı ayrı $(x_1, \dots, x_n)$ ve $(y_1, \dots, y_n)$ ' in örnek dağılım fonksiyonlarıdır. a_i daha önce tanımlandığı gibidir.

²³ Finch, s.300.

Δ_n değişiminin ve T_n ' in tahminleri aşağıdaki testlerden elde edilebilir:

(i) Δ_n , $(x_1, \dots, x_n)$ ve $(y_1 - \Delta_n, \dots, y_n - \Delta_n)$ ' den hesaplandığında, $S_{n,n} \approx 0$ olarak düzeltilerek,

(ii) Δ_n , $(x_1, \dots, x_n)$ ve $(2T_n - x_1, \dots, 2T_n - x_n)$ ' den hesaplandığında, $S_{n,n} \approx 0$ olarak düzeltilerek, ($S_{n,n}$ ' in tam olarak sıfır olması olanaksızdır.)

Burada farklı tahminler önerilmiştir. Bunlardan biri,

$$\bar{x}_H = \text{med}_{i < j} M_{ij}$$

olarak, diğeri ise²⁴,

$$\bar{x}_{BH} = \text{med}_{i \leq [n/2]} \frac{x_{(i)} + x_{(n-i+1)}}{2}$$

olarak hesaplanmaktadır²⁵.

1.3.2.4 W Tahminleri

M-Tahminine dayalı robust yöntemlerden biri de W-(ağırlıklı) Tahmindir. Bu tahminler için kullanılan W-Tahminleri merkezdeki gözlem değerine olan uzaklığa dayanan ağırlıklı tahminlerdir.

W-Tahmininin şekli M-Tahmindeki $\sum \psi(x_i - T_n) = 0$ ' a benzer, burada $\psi(y)$ yerine $yw(y)$ yazılırsa, denklemin çözümü T_w için,

$$\sum_{i=1}^n yw(y) = 0$$

olacaktır. $w(y)$, y 'ye bağlı ağırlık fonksiyonudur. Belirli birleşimler yapıldığında,

²⁴ P.J. Bickel, And J.L.Hodges Jr., "The Asymptotic Theory of Galton's Test and Related Simple Estimate of Location", *The Annals of Mathematical Statistics*, Vol.38, No.1, (Feb., 1967),s.73.

²⁵ Peter J. Huber, "The 1972 Wald Lecture Robust Statistics", *The Annals of Mathematical Statistics*, Vol.43, No.4, (Aug. 1972), s. 1041.

$$w(y) = \frac{\psi(y)}{y}$$

olarak elde edilir. Buradan,

$$T_W = \frac{\sum_{i=1}^n x_i w(y_i)}{\sum_{i=1}^n w(y_i)}$$

şeklinde yazılır. Çözüm için M-Tahmincilerindeki gibi karmaşık sayısal yöntemler kullanmak yerine burada stepwise aşamalı yöntemin kullanılması uygun olur. (k+1).’ci adımda W-tahminci,

$$T_W^{(k+1)} = \frac{\sum_{i=1}^n x_i w^{(k)}}{\sum_{i=1}^n w^{(k)}}$$

olur. $w^{(k)} = w(y_i^{(k)})$ ’dır ve bir sonraki adımda,

$$y_i^{(k)} = \frac{x_i - T_W^{(k)}}{cS_n}$$

olacaktır. Burada S_n örneğe dayalı scale ölçüsüdür ve c de sabittir. W-tahmini için önerilen başlangıç noktası medyan değerine eşit olan T_W^0 değeridir. $\sum_{i=1}^n y w(y) = 0$ ’ın çözümü için tek bir değer bulunuyorsa W-tahmincisiyle M-Tahmincisi benzer tahmin değerini vermektedir²⁶.

1.3.2.5 D Tahminleri

D-Tahminleri Adaptive Tahminler olarak da adlandırılır. Bu tahmin için adaptive tahminler örnek özelliğine dayanmaktadır. Basit adaptive tahminci, tek bir aşırı gözlem değerinin standartlaştırılmış değerinin belirlenen c değerinden büyük

²⁶ Finch, s. 295.

olması durumunda atılması olarak tanımlanmıştır. Aşırı değerler atıldıktan sonra Adaptive Trimming tahmin yöntemi kullanılır²⁷.

Bu tahminler için trimming miktarı seçilmektedir. T_α tahmini için,

$$\hat{V}(\alpha) = \frac{\sum_{i=n\alpha+1}^{n-n\alpha} (x_{(i)} - T_\alpha)^2 + \alpha(x_{(n\alpha+1)} - T_\alpha)^2 + \alpha(x_{(n-n\alpha)} - T_\alpha)^2}{(1-2\alpha)^2}$$

ifadesini minimize eden ($\alpha \leq 0,25$) trimming miktarı seçilir.

Hogg örnek basıklığına dayanan D-tahmincisini önermiştir. Bu tahmin basit trimmed ortalama tipi tahmindir ve trimming oranları örneğin kuyrukları belirlendikten sonra,

$$H = \begin{cases} \bar{x}_{0.25}^c & b_2 < 2 \\ \bar{x} & 2 \leq b_2 \leq 4 \\ \bar{x}_{0.25} & 4 < b_2 \leq 5.5 \\ x_M & 5.5 < b_2 \end{cases}$$

olarak belirlenir. Burada b_2 örnek basıklığıdır. $\bar{x}_{0.25}^c$ sıra istatistiklerinin en düşük %25 ve en yüksek %25'inin ortalamasıdır. $\bar{x}$ ortalama, $\bar{x}_{0.25}$ %25 trimmed ortalama ve x_M ise medyandır. Bu en çok benzerlik tahmincisinin yerleştirmesine benzemektedir. Kısa kuyruklu örneklerde aşırı değerlere daha çok ağırlık verilmiş, uzun kuyruklu örneklerde orta değere daha çok ağırlık verilmiştir.

Kuyruk çarpıklığını ölçmek amacıyla Q istatistiği kullanılabilmektedir²⁸. Q istatistiği,

$$Q = \frac{\bar{U}_{0.05} - \bar{L}_{0.05}}{\bar{U}_{0.5} - \bar{L}_{0.5}}$$

²⁷ Louis A. Jaeckel, "Some Flexible Estimates of Location", *The Annals of Mathematical Statistics*, Vol.42, No.5, (Oct., 1971), s. 1540.

²⁸ Robert V. Hogg, "More Light on the Kurtosis and Related Statistics", *Journal of the American Statistical Association*, Vol.67, No.338, (June 1972), s. 422.

olarak hesaplanır. Tahmin,

$$H = \begin{cases} \bar{x}_{0.25} & Q < 2 \\ \bar{x} & 2 \leq Q \leq 2.6 \\ \bar{x}_{0.1875} & 2.6 < Q \leq 3.2 \\ \bar{x}_{0.375} & 3.2 < Q \end{cases}$$

olacaktır. $\bar{x}_\alpha$, α trimmed ortalamadır.

Başka bir adaptive L-Tahmini yerleşim tahmininden önce aşırı değerlerin atıldığı(trimmed) Tahminidir. Sıralı örnekte, f -yayılımı ile noktalar (hinges) belirlenir ve

$$\begin{aligned} c_1 &= h_1 - 2d_f \\ c_2 &= h_u + 2d_f \\ t_1 &= h_1 - 1,5d_f \\ t_2 &= h_u + 1,5d_f \end{aligned}$$

kullanılarak $(c_1, c_2), (t_1, t_2)$ ve (s_1, s_2) aralıkları belirlenir. Burada,

$$\begin{aligned} s_1 &= h_1 - d_f \\ s_2 &= h_u + d_f \end{aligned}$$

olarak belirlenir. h_1 ve h_u , örneğin en düşük ve en yüksek noktalarıdır. Tanımlanan aralıklardan biri kullanılarak, bu aralık dışındaki tüm gözlem değerleri silinir. Aralık sınırları dışında kalan hiçbir değer kalmayana kadar bu süreç aşamalı olarak devam eder.

1.3.2.6 Tahminlerin Karşılaştırılması

Yerleşim'in tek değişkenleri robust tahminlerini karşılaştırmak için tahmincilerinin özelliklerine bakmak gerekecektir. M tahmincileri en çok benzerlik tahminine dayalı, R-Tahmincileri ise rank testlerine dayalı tahmincilerdir. W

tahmincileri merkezdeki gözlem değerine olan uzaklığa dayanan ağırlıklı tahmincilerken, L tahmincileri sabit ağırlıklı sıra istatistiklerinin doğrusal kombinasyonlarıdır. Adaptive Tahminciler ise örneğin özelliğine göre seçilmektedir.

L-Tahminleri ağırlıkların aldıkları değerler yerine sadece gözlemlerin sırasına dayanması sebebiyle W-Tahmincilerinden farklıdır. Yani L-Tahminlerinin ağırlıkları örnek hacmine göre sabittir, W-Tahmincilerinin ağırlıkları tahminci ile aşamalı olarak belirlenir²⁹.

M-Tahminleri ve W-Tahminlerinde olduğu gibi örnek ortalaması ve medyanı L-Tahminlerinin özel halidir. Her i için örnek ortalamasının ağırlığı $\alpha_i = \frac{1}{n}$ ' dir, örnek medyanı da $\alpha_{[n+1/2]} = 1$ ağırlığı hesaplanarak bulunur. N tek olduğunda $\alpha_i = 0$ ' dır n çift olduğunda $\alpha_{n/2} = \alpha_{n/2+1} = 1/2$ ' dir.

L-tahmincileri örnek büyüklüğüne göre sabit ağırlıklı olduğundan, M-tahminleri ise önceden belirlenen sabitler olduğundan adaptive değildir. Örnek özelliğine dayalı bir α değeri seçildiğinde L-tahminleri D-tahminlerine dönüşmektedir³⁰.

1.3.3 Ölçek Parametresinin Basit Tahmincileri

Standart sapma kavramı yaygın olarak kullanılırken daha genel bir kavram olan “ölçek (scale)” ya da “spread (yayılım)” tahmincileri pek kullanılmamaktadır. Örneğin bazı dağılımların anakütle standart sapması sonsuzdur oysa örnek standart sapması her zaman sonludur. Dağılımlar kısıtlandığında farklı scale ölçüleri sabit çarpanlarla birbirleriyle ilişkilidir ancak pratikte hiçbir zaman örneğin geldiği dağılım kesin olarak bilinmemektedir. Scale ölçüsü belirlemenin iki amacı vardır. Bunlardan birincisi ölçek

²⁹ Crow, And Siddique, s. 357.

³⁰ Hogg, s. 423.

tahmininin veri hakkında çok iyi bilgi vermesi, ikincisi ise dirençlilik, etkinlik gibi şartları sağlayan tahminler elde edilmesidir.

Ölçek tahminci S_n , örneğin pozitif değerlerinin bir fonksiyonudur ve $S_n(x_1, \dots, x_n)$ olması durumunda,

$$|S_n(a + bx_1, \dots, a + bx_n)| = |b| S_n(x_1, \dots, x_n)$$

olacaktır. Ölçek tahminci örneğin yerleşim’inde değişikliklere izin vermez, yerleşim sabit, scale ise değişkendir³¹.

Standart sapma en yaygın olarak kullanılan ölçek tahminidir ve birçok özelliği sağlar. Örnek standart sapmasının ölçek tahmini,

$$s = \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \right)^{1/2}$$

olacaktır. Burada $\bar{x}$ örnek ortalamasıdır ve yerleşim tahminidir. Standart sapmanın ölçek tahmini, örnek ortalaması gibi aşırı değerlerin olması durumunda robust değildir.

Verinin özelliğini belirleyebilmek için sıra istatistikleri uzun zamandan beri kullanılmaktadır, iki sıra istatistiği arasındaki fark gözlem değerlerinin yayılımını belirler. Range yayılımın ölçüsüdür, ancak aşırı değerler ve çarpık dağılımlarda güvenilir değildir. Quasi-ranges medyandan simetrik uzaklıktaki iki sıra istatistiği arasındaki farktır³².

Dördüncü yayılım (Tukey tarafından gösterilen H-yayılımı) en düşük ve en yüksek değer arasındaki farktır. Medyan derinliği,

$$d(M) = (n+1)/2$$

³¹ David C. Hoaglin, Frederick Mosteller ve John W. Tukey, *Understanding Robust and Exploratory Data Analysis*, USA: John Wiley & Sons, 1983, s. 283.

³² Finch, s.306.

olarak tanımlanır. $d(M)$ tamsayı ise (n tek sayı ise) medyan $x_{d(M)}$, tamsayı değilse (n çift sayı ise) $(x_{d(M)-0.5} + x_{d(M)+0.5})/2$ olacaktır. En düşük ve en yüksek noktalar ,

$$d(H) = \frac{[d(M)] + 1}{2}$$

derinliği ile sıralı örnekten belirlenen değerler olarak tanımlanmıştır. $d(H)$ tamsayı ise en düşük ve en yüksek noktalar $h_l = x_{d(H)}$ ve $h_u = x_{n-d(H)+1}$ olacaktır. $d(H)$ karmaşık sayı ise en düşük dördüncü değer medyana benzer şekilde şöyle tanımlanır,

$$h_l = (x_{d(H)-1/2} + x_{d(H)+1/2})/2$$

Dördüncü yayılım (f -yayılımı),

$$d_f = h_u - h_l$$

olarak hesaplanır. F-yayılımı, gözlem değerlerinin yarısının ortasının genişliğini verir. Range, aşırı değerler arasındaki farkı vermektedir. Ancak aşırı değerler range'i çok etkiler, böylelikle düşük dirence sahiptir³³. Normal dağılım altındaki dördüncü yayılım yaklaşık olarak $1,349\sigma$ ' ya eşittir. d_f/s oranı verinin normallliği ile ilgili bilgi vermektedir.

Robust ölçek tahminleri arasında en iyi olanı medyandan mutlak sapmadır. Ortalamadan mutlak sapma,

$$\overline{AD} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

olarak elde edilir. Medyandan mutlak sapma ise,

$$MAD_n = 1,483 \text{ med}_j \{|X_i - \text{med}_j X_j|\}$$

olarak elde edilir. Formülü basit, hesaplanması kolaydır.

³³ Hoaglin, Mosteller, And Tukey, s. 38.

Tahminci medyan mutlak sapma, “*MAD*” olarak gösterilir

Örnek medyanı ve *MAD* basit ve hesaplanması kolaydır; ancak kullanışlı değildir. Aşırı değerlere karşı hassas olmamasından dolayı,

$$\frac{|X_i - med X_j|}{MAD_n}$$

formülüyle hesaplanır. Medyan ve *MAD* daha robust tahmincilerin ilk değerleri olarak kullanılırlar. *M* Tahmincilerinin hesaplanması için başlangıç değerleri olarak kullanılırlar. Bundan dolayı *MAD*, *M* tahmincilerine yardımcı ölçek tahminidir³⁴. *M* tahmincilerinin teorisinde simetriklik varsayımı gerekli değildir. Bu durum aşırı çarpık dağılımlarda *MAD* kullanılmasını engellemez; ancak daha az etkin olabilirler³⁵.

Başka bir ölçek tahmini de medyandan ortalama mutlak sapmadır ve,

$$\frac{1}{n} \sum_{i=1}^n |x_i - Me|$$

olarak gösterilir. Bu da normalliğin testinde kullanılan σ ’nın başka bir tahminidir.

Daha önce bahsedilen Trimmed Ortalamanın varyansı örnek Winsorized varyansı kullanılarak tahmin edilebilmektedir. Winsorized varyans,

$$s_{\alpha}^2 = \frac{\sum_{i=k+1}^{n-k} (x_{(i)} - \hat{\mu})^2 + k(x_{(k)} - \hat{\mu})^2 + k(x_{(n-k+1)} - \hat{\mu})^2}{n(1-2\alpha)^2}$$

olarak tanımlanır. $\hat{\mu}$ trimmed ya da Winsorized ortalamadır. Winsorized varyans Trimmed Ortalamanın örneklem değişkenliğinin tahminidir. $\bar{x}_{\alpha}$ ’nın s_{α} ’ya oranı t-testinin ve güven aralıklarının robust halidir.

³⁴ Huber, s.107.

³⁵ Peter J. Rousseeuw, Cristophe Croux, “*Alternatives to the Median Absolute Deviation*”, *Journal of the American Statistical Association*, Vol.88, No.424, (Dec., 1993), s.1273.

1.3.4. Ölçeğin Tek Değişkenli Robust Tahminleri

Ölçek parametresinin elde edilmesinde en yaygın olarak kullanılan tahminler M-Tahminleri, L-Tahminleri ve R-Tahminleridir.

1.3.4.1 M Tahminleri

Medyan ve dördüncü yayılım gibi basit ölçek tahmincileri etkin olmadıkları için çok daha iyi tahmincilerin geliştirilmesine gerek vardır. Yerleşim'in M-tahmincileri basit tahmincilerin çok daha gelişmiş halidir, benzer şekilde çok daha iyi performansa sahip ölçek tahmincileri bulunmalıdır.

Ölçek sadece bilinmeyen bir parametre ise, örneğin yerleşim parametresinin sıfır ya da bilinen başka bir değer olması durumunda, S_n ölçeğinin M-Tahmincisi χ fonksiyonuna dayanır.

$$\sum_{i=1}^n \chi(x_i / S_n) = 0$$

denkleminin eşanlı olarak çözümünü önermiştir. χ genellikle çift bir fonksiyondur; $\chi(-u) = \chi(u)$ 'dur³⁶.

Yerleşim tahmincisi T_n ve ölçek tahmincisi S_n 'in aşağıdaki iki denklemi sağlaması durumunda;

$$\sum_{i=1}^n \psi\left(\frac{x_i - T_n}{cS_n}\right) = 0$$
$$\sum_{i=1}^n \chi\left(\frac{x_i - T_n}{cS_n}\right) = 0$$

yerleşim ve ölçeğin eşanlı M-Tahmincileridir. Simetrik durumlar için, ψ tek fonksiyon, χ de çift fonksiyondur. Burada c sabittir ve ayarlama sabiti olarak adlandırılır.

³⁶ Hoaglin, Mosteller, And Tukey, s. 347.

Yukarıdaki denklemler ölçek ve yerleşim tahmininde kullanılırlar, bunlara alternatif olarak başka bir yaklaşım geliştirilmiştir. Bu yaklaşım yerleşim'in M-tahmincisinin asimtotik varyansı için örneğe uygulanabilen bir formül ortaya koymuştur. Sonucun karekökü de ölçek tahmincisi olarak kullanılır. Veri F dağılımına sahip olduğunda yerleşim ve ölçeğin anakütle değerleri $T(F)$ ve $S(F)$ ile gösterilir. Burada $S(F)$ ikinci denklemdeki S_n 'e benzemez. T_n ' in hesaplanmasında kullanılan medyandan mutlak sapma (MAD) gibi bir yardımcı ölçek tahmincisi için anakütle değerini ifade etmektedir³⁷.

T_n ' in asimtotik varyansı $\sqrt{n}[T_n - T(F)]$ 'in sınırlayıcı dağılımının varyansdır ve

$$A(F, T) = \frac{c^2 S^2(F) \int \psi^2 \left[\frac{x - T(F)}{cS(F)} \right] f(x) dx}{\left\{ \int \psi \left[\frac{x - T(F)}{cS(F)} \right] f(x) dx \right\}^2}$$

olacaktır. Burada $f(x)$ yoğunluk fonksiyonudur.

Ölçek tahmincisi oluşturabilmek için örnekten $A(F, T)$ hesaplanır ve karekökü alınır. $S(F)$ MAD ile, $T(F)$ M ile, $f(x)$ $1/n$ ile değiştirilir.

$$u_i = \frac{x_i - M}{cMAD}$$

olacaktır ve ölçek tahmincisi,

$$s_T = \frac{(cMAD)n^{1/2} \left[\sum_{i=1}^n \psi^2(u_i) \right]^{1/2}}{\left[\sum_{i=1}^n \psi'(u_i) \right]}$$

³⁷ Hoaglin, Mosteller, And Tukey, s. 415.

olarak belirlenir. Bunun gibi ölçek tahmincileri A-Tahmincileri olarak adlandırılır. Burada iki ψ fonksiyonundan söz edilir ve

$$\psi(u) = \begin{cases} |u| < 1 \text{ için} & \sin(\pi u) \\ |u| \geq 1 \text{ için} & 0 \end{cases}$$

olarak belirlenir. Ölçek tahmincisi,

$$s_{wa} = \frac{(cMAD)n^{1/2} \left[\sum_{w_i < 1} \sin^2(\pi u_i) \right]^{1/2}}{\pi \left| \sum_{w_i < 1} \cos(\pi u_i) \right|}$$

şeklinde yerleşim'in tahmincisine dayanır³⁸.

Biweight ψ fonksiyonu,

$$\psi(u) = \begin{cases} |u| < 1 \text{ için} & u(1-u^2)^2 \\ |u| \geq 1 \text{ için} & 0 \end{cases}$$

olacaktır. $|u| < 1$ için $\psi'(u) = (1-u^2)(1-5u^2)$, ölçeğin A-tahmincisi,

$$s_{bi} = \frac{n^{1/2} \left[\sum_{w_i < 1} (x_i - M)^2 (1-u_i^2)^4 \right]^{1/2}}{\left| \sum_{w_i < 1} (1-u_i^2)(1-5u_i^2) \right|}$$

olacaktır. Bu tahminci yerleşim'in biweight tahmincisine dayanmaktadır.

1.3.4.2. L Tahminleri

L-tahminlerinin genel sonuçları pek fazla değişiklik yapılmadan burada da uygulanabilmektedir. Ölçeğin sabit olduğu bilgisi ile sadece şu fonksiyon tiplerinin uygulanması mümkün olmaktadır:

³⁸ Hoaglin, Mosteller, And Tukey, s. 376.

$$S(F) = \begin{cases} \left[\int F^{-1}(t)^q M(dt) \right]^{1/q}, & q \neq 0 \text{ integrali ile,} \\ \left[\int |F^{-1}(t)|^q M(dt) \right]^{1/q}, & q = 0 \\ \exp \left[\int |F^{-1}(t)| M(dt) \right] & , M\{(0,1)\} = 1 \end{cases}$$

Birinci tip tahminin (kantiller arası aralık, trimmed varyans) ve ikinci tip tahminin (medyandan sapma) her ikisiyle karşılaşmak da mümkündür; ancak sadece ilk tip tahmin ele alınacaktır³⁹.

Ölçeğin L tahminleri için etki fonksiyonu,

$$IC(x; F, S) = \frac{1}{S(F)^{q-1}} \left[\int s \frac{F^{-1}(s)^{q-1}}{f(F^{-1}(s))} M(ds) - \int_{F(x)}^1 \frac{F^{-1}(s)^{q-1}}{f(F^{-1}(s))} M(ds) \right]$$

olarak belirlenir. M'in yoğunluğu m ise,

$$\frac{d}{dx} IC(x; F, S) = \frac{x^{q-1}}{S(F)^{q-1}} m(F(x))$$

olacaktır.

1.3.4.3. R Tahminleri

$(x_1, \dots, x_m)$ ve $(y_1, \dots, y_n)$ ' in herhangi iki örnek olması durumunda test istatistiği,

$$\sum_{i=1}^n a(R_i)$$

olacaktır. R_i , x_i ' nin rankıdır, örnek hacmi $N = m + n$ ' dir.

$$a_i = N \int_{(i-1)/N}^{i/N} J(s) ds$$

³⁹ Huber, s. 110.

olarak tanımlanmıştır. $J, \left|t - \frac{1}{2}\right|$, nin bir fonksiyonudur. Bu testler ölçek tahminlerine dönüştürülür. $\lambda; 0 < \lambda < 1$ aralığında sabit olsun (Daha sonra $\lambda = m/N$ seçilebilir) ve $S = S(F, G)$ fonksiyonu,

$$\int J \left[\lambda F(x) + (1 - \lambda) G \left(\frac{x}{S} \right) \right] F(dx) = 0$$

olarak tanımlanır. $F(x) = t$ şeklinde gösterilirse,

$$\int J \left[\lambda t + (1 - \lambda) G \left(\frac{F^{-1}(t)}{S} \right) \right] dt = 0$$

olacaktır. $\int J(t) dt = 0$ varsayılırsa, $S(F, G)$,

$$S(F_{aX}, F_X) = a$$

eşitliğini sağlayan nispi ölçek ölçüsü olacaktır. F_{aX}, aX tesadüfi değişkeninin dağılımını ifade etmektedir.

İKİNCİ BÖLÜM

ALTERNATİF REGRESYON MODELLERİ

En Küçük Kareler Regresyonu’nda artıkların karelerinin toplamı minimize edilir. Artıkların kareleri yerine farklı değerlerin minimizasyonu da sözkonusu olabilir. Bu durumda “Alternatif Regresyon Modelleri” olarak adlandırılan farklı regresyon modelleri kullanılabilir. Alternatif regresyon modelleri ve bunlarla ilgili tahmin yöntemleri bu bölümü oluşturacaktır. En Küçük Kareler (EKK) Regresyonu olarak adlandıracağımız parametrik regresyon modellerinden sonra ,En Küçük Mutlak Sapmalar (LAD) Regresyonu, Medyan Regresyon, En küçük Medyan Kareler (LMS) Regresyonu, M Regresyon, Rank Regresyon ve bunların tahmininde kullanılan yöntemlere kısaca yer verilecektir.

2.1 En Küçük Kareler Regresyonu

En yaygın olarak kullanılan regresyon yöntemidir. En Küçük Kareler Regresyonu’nda artıkların karelerinin toplamı minimize edilerek parametre tahminleri yapılır.

2.1.1 Basit Doğrusal Regresyon Modeli

İki değişken arasındaki ilişkinin doğrusal varsayılarak incelendiği regresyon modeline “Basit Doğrusal Regresyon Modeli” denir.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Regresyon modelinin sol tarafında yer alan Y_i ile ifade edilen değişkene “bağımlı değişken” denir. Denklemin sağında yer alan X_i ile ifade edilen değişkene ise “bağımsız veya açıklayıcı değişken” denir. Diğer bir ifadeyle bağımlı Y_i değişkenindeki değişimler bağımsız X_i değişkeni tarafından açıklanmaktadır. Değişkenler arasında

stokastik ilişki olduğu için değişkenlerden birinin alacağı değere karşılık diğer değişken farklı değer alabilmektedir. Değişkenlerin aldığı değerler modelin matematiksel yapısı ile tam uyum sağlamazlar değerler matematiksel şekilden sapmalar gösterirler. Sapmaları açıklamak için modele hata terimi eklenir⁴⁰.

Bağımlı değişken Y_i hata terimi ε_i 'nin doğrusal bir fonksiyonudur. Temel varsayımlar nedeniyle ε_i normal dağılır, bundan dolayı Y_i de normal dağılacaktır. Dağılımın ortalaması $(\beta_0 + \beta_1 X_i)$ ve varyansı σ^2 'dir ve $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$ olarak gösterilir. Bu nedenle β_0 ve β_1 'in tahmini Y_i 'nin ortalamasının tahminidir.

Regresyon modellerinin değişkenler arasındaki ilişkiyi açıklamak için yeterli olup olmadığı veya ilişkiyi açıklayıp açıklamadığı konusunda karar vermek için hata terimleri ile ilgili temel varsayımların da geçerli olması gerekmektedir. Basit regresyonun temel varsayımları; normallik, sıfır ortalama, sabit varyans, otokorelasyon olmaması, bağımsız değişkenin tesadüfi değişken olmamasıdır. Varsayımların geçerli olmaması durumunda tahminciler taşıması istenilen bazı özellikleri taşımayacaklardır.

Parametrelerin Tahmini

β_0 ve β_1 'i tahmin etmek için birçok yöntem kullanılabilir. Bu yöntemlerden bir tanesi En Küçük Kareler Yöntemi' dir. Burada farkların kareleri toplamı $\sum_{i=1}^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$ minimize edilerek tahminciler elde edilir. Regresyon modeli için bu fark, Y_i ile $E(Y_i)$ arasındaki fark olacaktır. Yapılan işlemler sonucunda,

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i$$

$$\sum X_i Y_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2$$

⁴⁰ Selahattin Güriş, Ebru Çağlayan, Ekonometri Temel Kavramlar, 2. Basım, İstanbul: Der Yayınları, 2005, s. 87.

olarak iki denklem elde edilir. Bu denklemlere “Normal Denklemler” denir. Bunların çözümü ile parametreler tahmin edilir. Denklemlerden türetilen farklı formüller parametre tahmininde kullanılabilir.

EKK Regresyonu aşırı değerlerin etkisinde kalabilir. Hatalar normal dağılıyorsa değişkenler arasındaki ilişkiyi en iyi şekilde açıklar; ancak hataların normal dağılmadığı durumlarda (aşırı değerlerin olması durumunda hatalar normal dağılmayabilir) EKK Tahminleri ve testleri iyi sonuç vermezler.

Belirlilik Katsayısı

Bağımlı değişkendeki değişmelerin bağımsız değişken veya değişkenler tarafından açıklanma oranını belirten katsayıya “belirlilik katsayısı(R^2)” adı verilmektedir. Basit regresyonda tek bağımsız değişken olduğundan, bağımlı değişkendeki değişmeler sadece bu değişken tarafından açıklanmaktadır.

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

olarak elde edilir. $0 \leq R^2 \leq 1$ olur. R^2 ’nin 1’e yaklaşması bağımlı değişkendeki değişmelerin bağımsız değişken tarafından iyi açıklandığını ortaya koyacaktır.

t Testi

Katsayının örnekten elde edilen değeri ile ileri sürülen veya varsayılan değer arasındaki farkın istatistiksel olarak anlamlı olup olmadığını, farkın örnekleme nedeni ile mi yoksa gerçekten mi varolduğunu belirlemek için yapılmaktadır.

Regresyon modeli değişkenler arasındaki ilişkiyi açıklamakta yetersiz kalıyorsa katsayılar veya katsayılardan en az biri sıfır olacaktır. Tahmin edilen değerler sıfırdan farklı da olsa, istatistiksel olarak bu değer anlamsız ise sıfır kabul edilecektir. Böylece model ilişkiyi açıklayabilecek uygun model olmayacaktır. Bu nedenle t-testi katsayıların dolayısı ile modelin anlamlılığının testi için de kullanılabilir. Bu durumda hipotezler,

$$H_0 : \beta_i = \beta_i$$

$$H_1 : \beta_i \neq \beta_i$$

$$H_1 : \beta_i < \beta_i$$

$$H_1 : \beta_i > \beta_i$$

olarak kurulabilir. Test istatistiği,

$$t = \frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$$

olur. Burada β_i katsayının varsayılan değeridir. Katsayıların anlamlılığının testinde $\beta_i = 0$ olacaktır. Bu hesaplanan değer α hata payı normal dağılım veya (n-k) serbestlik dereceli t-tablosu ile karşılaştırılarak karar verilir.

2.1.2 Çoklu Doğrusal Regresyon Modeli

Basit doğrusal regresyon modelinin bağımlı değişkenini etkileyecek, bağımlı değişkendeki değişmelerin nedeni olabilecek değişkenlerin eklenmesi ile oluşturulan modellere “Çoklu Doğrusal Regresyon Modelleri” denir. çoklu regresyon modeli de basit regresyon modeli gibi, doğrusal veya doğrusal olmayan fonksiyonlarla ifade edilebilir. Değişkenler arasında doğrusal ilişki olduğu varsayımı ile, doğrusal

fonksiyonla ifade edilen çoklu regresyon modellerine “Çoklu Doğrusal Regresyon Modelleri” adı verilmektedir.

İncelenecek bağımlı değişken dahil k sayıda değişken olduğu varsayıldığında anakütle çoklu doğrusal regresyon modeli,

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} + \varepsilon_i$$

olarak ifade edilebilir.

Çoklu doğrusal regresyonda da basit regresyonda olduğu gibi elde edilen sonuçların geçerliliği temel varsayımların geçerliliğine bağlıdır. Çoklu regresyonda basit regresyonun varsayımları aynen geçerli olup iki ilave varsayım daha vardır. Çoklu doğrusal regresyonun temel varsayımları; normallik, sıfır ortalama, sabit varyans, otokorelasyon olmaması, bağımsız değişkenlerin tesadüfi değişken olmaması, çoklu doğrusal bağıllık olmaması, $n > k$ olmasıdır.

Bağımsız değişkenlerden ikisi veya daha fazlası arasında kuvvetli doğrusal ilişki olması durumu “Çoklu Doğrusal Bağıllık” olarak adlandırılmaktadır. Tam çoklu doğrusal bağıllık olduğunda sistem çözülememekte, parametreler tahmin edilememekte veya bazı temel varsayımlar geçerli olmamaktadır. Sonuçların geçerliliği için modelde çoklu doğrusal bağıllık olmamalıdır.

Parametrelerin Tahmini

Çoklu doğrusal regresyon modelinin parametreleri basit doğrusal regresyon modelinin parametreleri gibi aynı yöntemlerle tahmin edilebilir. Anakütle çoklu doğrusal regresyon modeli,

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} + \varepsilon_i$$

olarak tanımlanmıştır.

En Küçük Kareler Yöntemi' nde Y_i ile $E(Y_i)$ arasındaki farkların kareleri toplamı minimize edileceğinden, yapılan işlemler sonucunda⁴¹,

$$\begin{aligned}
 \sum_{i=1}^n Y_i &= n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n X_{i2} + \hat{\beta}_3 \sum_{i=1}^n X_{i3} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{ik} \\
 \sum_{i=1}^n X_{i2} Y_i &= \hat{\beta}_1 \sum_{i=1}^n X_{i2} + \hat{\beta}_2 \sum_{i=1}^n X_{i2}^2 + \hat{\beta}_3 \sum_{i=1}^n X_{i2} X_{i3} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{i2} X_{ik} \\
 \sum_{i=1}^n X_{i3} Y_i &= \hat{\beta}_1 \sum_{i=1}^n X_{i3} + \hat{\beta}_2 \sum_{i=1}^n X_{i2} X_{i3} + \hat{\beta}_3 \sum_{i=1}^n X_{i3}^2 + \dots + \hat{\beta}_k \sum_{i=1}^n X_{i3} X_{ik} \\
 &\vdots \\
 \sum_{i=1}^n X_{ik} Y_i &= \hat{\beta}_1 \sum_{i=1}^n X_{ik} + \hat{\beta}_2 \sum_{i=1}^n X_{ik} X_{i2} + \hat{\beta}_3 \sum_{i=1}^n X_{ik} X_{i3} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{ik}^2
 \end{aligned}$$

denklemleri elde edilir. Bu denklemler çoklu doğrusal regresyonun normal denklemleridir. Elde edilen normal denklemler gerçek değerlerle ifade edilmiştir. Bu denklemler yardımı ile,

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 - \dots - \hat{\beta}_k \bar{X}_k$$

olarak tahmin edilecektir.

Çoklu doğrusal regresyonda tahmin edilecek parametre sayısı arttıkça tahmincilerin formüllerle ifadesi güçleşir. Bu nedenle parametrelerin tahmininde matrisler kullanılır⁴².

Belirlilik Katsayısı ve Düzeltilmiş Belirlilik Katsayısı

Çoklu doğrusal regresyonda birden fazla bağımsız değişken olduğundan, belirlilik katsayısı (R^2) bağımlı değişkendeki değişmelerin bağımsız değişkenler tarafından açıklanma oranını verecektir. Basit regresyonda olduğu gibi teorik değerler çoklu regresyonda da hesaplanacak; ancak işlem kolaylığı sağlaması açısından matrisler kullanılacaktır⁴³.

⁴¹ Güriş, Çağlayan, s. 205.

⁴² Güriş, Çağlayan, s. 212.

⁴³ Güriş, Çağlayan, s. 233.

Aynı bağımlı değişkendeki değişmeler farklı regresyon modelleri ile açıklanabilir. Bu modellerin matematiksel yapıları, gözlem sayıları ve değişken sayıları farklı olabilir. Bağımlı değişkeni aynı, bağımsız değişken sayıları farklı regresyon modellerinin karşılaştırılması için düzeltilmiş belirlilik katsayısı $\bar{R}^2$ kullanılır. Düzeltilmiş belirlilik katsayısı,

$$\bar{R}^2 = 1 - \frac{n-1}{n-k} \cdot (1 - R^2)$$

olarak hesaplanır.

t Testi

Basit regresyonda olduğu gibi t-testi katsayıların dolayısı ile modelin anlamlılığının testi için de kullanılabilir. Hipotezler ve test istatistiği aynı şekilde oluşturulmaktadır. Hesaplanan değer α hata payı ile (n-k) serbestlik dereceli t-tablosu ile karşılaştırılarak karar verilir.

F Testi

F Testi'nde t Testi'nden farklı olarak tüm bağımsız değişkenlerin katsayıları birlikte test edilirken, sabit katsayı test edilmez. Basit regresyonda tek bağımsız değişken olduğundan F Testi ile bağımsız değişkenin katsayısının t-Testi arasında fark yoktur.

Hipotezler,

$$H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1 : \beta_2 \neq \beta_3 \neq \dots \neq \beta_k \neq 0$$

olarak kurulur. F Testi'nde tek taraflı test söz konusu değildir. Test istatistiği,

$$F = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 / (k-1)}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / (n-k)}$$

olarak hesaplanır ve α hata payı ile $(k-1)$ ' e $(n-k)$ serbestlik dereceli F tablo değeri ile karşılaştırılarak karar verilir

2.2 En Küçük Mutlak Sapmalar Regresyonu

En Küçük Mutlak Sapmalar Regresyonu (LAD)' nda artıkların karelerinin mutlak değerleri minimize edilerek parametre tahminleri yapılır.

2.2.1 Basit LAD Regresyonu

Öncelikle doğrusal regresyon modeli; $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ' nin grafiği çizilip, doğrusal olduğu belirlendikten sonra, doğrusal regresyon modeli kullanmaya karar verilir.

EKK Regresyonu'nda artıkların kareleri toplamını $(\sum \varepsilon_i^2)$ minimize eden $\hat{\beta}_0$ ve $\hat{\beta}_1$ tahminleri elde ediliyordu, En Küçük Mutlak Sapmalar (LAD) Regresyonu' nda ise artıkların mutlak değerlerinin toplamını $(\sum |\hat{\varepsilon}_i|)$ minimize eden $\hat{\beta}_0$ ve $\hat{\beta}_1$ tahminleri elde edilir⁴⁴.

$$\sum_{i=1}^n |Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)|$$

ifadesi minimize edilir. $Y_i - (\beta_0 + \beta_1 X_i)$, (X_i, Y_i) noktasının $\hat{Y}_i = \beta_0 + \beta_1 X_i$ doğrusundan sapmasını gösterir.

LAD Regresyonda, ε_i^2 yerine $|\hat{\varepsilon}_i|$ ' ler kullanıldığı için hesaplanması daha kolaydır ve artıkların büyüklüğünü ölçmede ε_i^2 ' den daha doğru sonuç verir; ancak

⁴⁴ C Radhakrishna Rao, Helge Toutenburg, **Linear Models: Least Squares and Alternatives**, 2. Basım, USA: Springer-Verlag New York, 1999, s. 273.

tahminlerin hesaplanması çok daha zordur; çünkü LAD Tahminleri için formüller yoktur, bunun yerine bir algoritma yardımıyla tahminler elde edilir LAD Regresyonu'nda amaç EKK Regresyonu'nda olduğu gibi değişkenler arasındaki ilişkiyi en iyi açıklayan doğruyu bulmaktır. Herhangi (X_0, Y_0) noktası verildiğinde, algoritma yardımıyla, bu noktadan geçen birçok doğrudan en iyi olanı bulunur. Algoritma ile regresyon doğrusunun belirlenmesi; bir noktadan geçen çok sayıda doğrudan birinin en iyi doğru olduğu ve en iyi doğrunun aynı zamanda başka bir noktadan geçeceği varsayımlarına dayanır⁴⁵. Belirlenen (X_0, Y_0) başlangıç noktasına göre diğer tüm noktaların eğimleri $(Y_i - Y_0)/(X_i - X_0)$ hesaplanır ve bunlar küçükten büyüğe sıralanır. $X_i = X_0$ olan noktalar ihmal edilir, böylece;

$$(Y_1 - Y_0)/(X_1 - X_0) \leq (Y_2 - Y_0)/(X_2 - X_0) \leq \dots \leq (Y_n - Y_0)/(X_n - X_0)$$

olur.

$$T = \sum (X_i - X_0)$$

olduğunda,

$$|X_1 - X_0| + \dots + |X_{k-1} - X_0| < \frac{1}{2}T$$

$$|X_1 - X_0| + \dots + |X_{k-1} - X_0| + |X_k - X_0| > \frac{1}{2}T$$

olarak k.' cı indeks bulunur.

Ardından yeni bir (X_k, Y_k) başlangıç noktası seçilir. Bu nokta $\frac{1}{2}T$ ' den daha büyük değere sahiptir ve (X_0, Y_0) noktasından geçen en iyi doğrunun bu noktadan da geçtiği kabul edilir. Yeni başlangıç noktasına göre diğer noktaların eğimleri hesaplanarak seçilen ilk nokta için yapılan işlemlerin aynısı bu nokta için de yapılır.

⁴⁵ David Birkes, Yadolah Dodge, *Alternative Methods of Regression*, New York: John Wiley, 1993, s.58.

Sonuçta (X_k, Y_k) noktasından geçen, sapmayı minimize eden en iyi doğru bulunur. Bu doğru aynı zamanda başka bir noktadan daha geçer, bu nokta (X_M, Y_M) ile gösterilirse, bu noktadan geçen, sapmayı minimize eden en iyi doğru bulunur, bu işlemler aynı şekilde devam eder. Bulunan her yeni doğru bir öncekinden daha iyi sonuç verir. En son bulunan doğrunun geçtiği nokta, daha önceki noktalardan biriyle aynı olana kadar işlemlere devam edilir. Aralarda hangi noktalardan geçtiğine bakılmaksızın en son bulunan noktadan geçen doğru en iyi doğrudur. Bu nokta (X_i, Y_i) ise;

$$Y_i = \beta_0^* + \beta_1^* X_i$$

doğrusunda;

$$\beta_1^* = \frac{Y_i - Y_0}{X_i - X_0}$$

$$\beta_0^* = Y_0 - \beta_1^* X_0$$

olacaktır⁴⁶.

Daha önce de belirtildiği gibi algoritmanın uygulanabilmesi için; bir noktadan geçen bir en iyi doğru vardır ve bir noktadan geçen en iyi doğru aynı zamanda sadece bir noktadan daha geçer varsayımları kabul edilir. Algoritma ileriye doğru adım adım oluşturulmaktadır, her adımda verilen noktadan geçen, sapmayı minimize eden en iyi doğru bulunmaktadır. En iyi doğru her zaman başka bir noktadan geçmekte ve bu yeni nokta bir sonraki adımda kullanılmaktadır. Varsayımın aksine bir noktadan geçen birden fazla en iyi doğru olabilir, bir noktadan geçen en iyi doğru aynı zamanda iki ya da daha fazla noktadan da geçebilir. Diğer bir deyişle, bir sonraki adımda kullanmak için birden fazla nokta seçeneği olacaktır. Yanlış seçimler yapıldığında da algoritma ya sürekli tekrarlanıp duracak ya da LAD Regresyon doğrusu olmayan başka bir doğru da son bulabilecektir. Bu sorun;

⁴⁶ Birkes, Dodge, s. 59.

$$|X_1 - X_0| + \dots + |X_{k-1} - X_0| < \frac{1}{2}T$$

$$|X_1 - X_0| + \dots + |X_{k-1} - X_0| + |X_k - X_0| > \frac{1}{2}T$$

koşullarında eşitlik olması durumunda veya $\beta^* = (Y_k - Y_0)/(X_k - X_0)$ eğiminin $(Y_{k-1} - Y_0)/(X_{k-1} - X_0)$ ya da $(Y_{k+1} - Y_0)/(X_{k+1} - X_0)$ ' a eşit olması durumunda ortaya çıkabilir.

Bu durumda başka bir algoritma kullanılabilir. LAD Regresyon doğrusunun en az iki noktadan geçtiği bilinmektedir böylelikle tüm nokta çiftleri arasından doğru bulunabilir. Bu doğruların bazıları çakışır. $\sum_{i=1}^n |Y_i - (\beta_0 + \beta_1 X_i)|$ ' yi her doğru için hesaplayarak en küçük toplamı veren/ler seçilir. Bu algoritmanın uygunluğu örnek hacmi n' e bağlıdır ve bu algoritmada en iyi doğru sadece iki noktadan geçer, bir noktadan geçen birden fazla doğru olması durumunda da bir tanesi seçilir ya da ortalamaları alınabilir. Bu ortalama doğru aynı zamanda LAD Regresyon doğrusudur.

Eğim Parametresinin Anlamlılığının Testi

Öncelikle LAD Regresyon Tahminleri $\hat{\beta}_0$ ve $\hat{\beta}_1$ bulunur. Bulunan $\hat{\beta}_1$ tahmini değerinin gerçek değere eşit olması beklenemez ancak β_1 ' in gerçek değerinin sıfır olması mümkündür bundan dolayı β_1 ' in sıfıra eşit olup olmadığı test edilir.

$e_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$ artıkları hesaplanır, bunlar küçükten büyüğe sıralanır. En ortadan bağımsız değişken sayısından bir fazla (p+1) gözlem değeri atılır, basit regresyonda bir bağımsız değişken olduğundan m=n-2 olur, burada n başlangıç gözlem sayısını, m de sıfıra eşit olmayan artık sayısını yani yeni gözlem sayısını ifade etmektedir. Test istatistiği,

$$|t| = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} \sim t_{\alpha/2; (n-2)}$$

olarak hesaplanır. Burada,

$$S_{\hat{\beta}} = \frac{\hat{t}}{\sqrt{\sum (X_i - \bar{X})^2}}$$

olacaktır. Parametrenin standart hatasının hesaplanmasında kullanılan $\hat{t}$,

$$\hat{t} = \frac{\sqrt{m} [e_{(k_2)} - e_{(k_1)}]}{4}$$

olarak hesaplanır. Burada,

$$k_1 : \left(\frac{m+1}{2} - \sqrt{m} \right) \text{ ' e en yakın tamsayı değeridir.}$$

$$k_2 : \left(\frac{m+1}{2} + \sqrt{m} \right) \text{ ' e en yakın tamsayı değeridir.}$$

$e_{(k_1)}$ küçükten büyüğe sıralanmış artıklar içerisinde k_1 .'ci artığın değerini, $e_{(k_2)}$ de k_2 .'ci artığın değerini göstermektedir.

$\hat{\beta}_1$ tahmini değerinin $\hat{\beta}_1$ 'ya yakın olması beklenir. β_1 ve $\hat{\beta}_1$ arasındaki fark 1 ya da 2 standart sapmadan ($S_{\hat{\beta}_1}$) büyük olmamalıdır. $|t|$ değerinin büyük olması, $\hat{\beta}_1$ ile sıfır arasındaki mesafenin $S_{\hat{\beta}_1}$ 'den büyük olmasını, böylelikle $\beta_1 \neq 0$ hipotezinin reddedilmemesi yönünde karar verilmemesini sağlar.

2.2.2 Çoklu LAD Regresyonu

LAD Regresyon L_1 -regresyon olarak da adlandırılır çünkü $\sum |Y_i - (a + bX_i)|$, sapma vektörlerin L_1 normudur. Bir v vektörünün L_1 normu $\sum |v_i|$ 'dir. Benzer şekilde EKK Regresyonu da L_2 -regresyon olarak adlandırılabilir çünkü sapma vektörlerinin L_2 -normunu minimize eder. v vektörünün L_2 -normu; $\sqrt{\sum v_i^2}$ 'dür.

Basit regresyonda, LAD Regresyon doğrusu iki noktadan geçmekteydi. Benzer şekilde çoklu regresyonda, p açıklayıcı değişkenli LAD Regresyon denklemi, (p+1) veri noktasını sağlar.

Çoklu EKK Regresyonu' nda olduğu gibi ikiden fazla değişken olduğunda kullanılır. Bu regresyon modeli,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$$

olarak ifade edilebilir. Artıkların mutlak değerlerinin toplamını mümkün olduğunca minimum yapan $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ LAD Tahminleri elde edilir. Bu,

$$\sum |Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik})|$$

olarak gösterilebilir. Minimum edecek değerleri bulmak için herhangi bir formül yoktur, bir algoritma kullanılır. Burada da basit regresyonda olduğu gibi algoritmanın uygulanabilmesi için; bir noktadan geçen bir en iyi doğru vardır ve bir noktadan geçen en iyi doğru aynı zamanda sadece bir noktadan daha geçer varsayımları kabul edilir.

Vektör gösterimiyle,

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad X_i = \begin{bmatrix} 1 \\ X_{i1} \\ X_{i2} \\ \vdots \\ X_{ik} \end{bmatrix}$$

olacaktır. Bu durumda mutlak sapma; $\sum |Y_i - \hat{\beta}' X_i|$ olarak yazılır. Bunu minimize edecek $\hat{\beta}'$ vektörünü bulmak amaçtır. Basit LAD Regresyonu'nda olduğu gibi, Çoklu LAD Regresyonu da iteratif olarak çözülür. $\hat{\beta}'$ vektörüyle işlemlere başlanır, sonra $\sum |Y_i - \hat{\beta}' X_i|$ ' nin minimum değerini veren daha iyi bir vektör bulunur. Sonunda

$\hat{\beta}$ 'nın en iyi vektörü bulunur. Her adımda, $\hat{\beta}$ tahminleri vektöründe daha iyi bir $\hat{\beta}^*$ vektörü,

$$\hat{\beta}^* = \hat{\beta} + td$$

olarak bulunur. Bu vektörün bulunması için yön vektörü d ve t değerlerinin elde edilmesi gerekir. Minimumluğu sağlayacak t ' yi bulmak için bir yöntem geliştirilir. Minimize edilecek ifade,

$$\sum |Y_i - (\hat{\beta} + td)' X_i|$$

olacağından, burada,

$$Z_i = Y_i - \hat{\beta}' X_i, \quad W_i = d' X_i$$

dönüşümü yapılarak,

$$\sum |Z_i - tW_i|$$

Olarak elde edilir. Bu gösterim, daha önce $\sum |(Y_i - Y_0) - \hat{\beta}(X_i - X_0)|$ ' u minimize edecek $\hat{\beta}$ ' yi bulmakla aynıdır. Z_i / W_i oranları hesaplanıp, artan sıraya göre dizilir. Z ve W ' yi yeniden indeksleyerek k indeksi bulunur:

$$|W_1| + |W_2| + \dots + |W_{k-1}| < \frac{1}{2}T$$

$$|W_1| + |W_2| + \dots + |W_{k-1}| + |W_k| > \frac{1}{2}T$$

Burada $T = \sum |W_i|$ ' dir. t ' nin minimum yapan değeri Z_k / W_k ' dir.

Algoritmanın her bir adımında p açıklayıcı değişken sayısından bir fazla $(p+1)$ yön vektörü vardır. Her bir d_j vektörü için (+) pozitif yön sözkonusu olduğu gibi (-) negatif yön de sözkonusudur. Bu nedenle açıklayıcı değişken sayısının bir fazlasının iki

katı sayıda yön olacaktır. Bunlar arasından, $\sum |Y_i - (\hat{\beta} + td)'X_i|$ değerini mümkün olduğunca hızlı $t = 0$ değerine yaklaştıran yön seçilir. Bu değerin nasıl hızla azaldığını belirlemek için sağ tarafın $t = 0$ 'daki türevini alırız.

$\sum |Z_i - tW_i|$ ifadesinde $t = 0$ 'da sağ tarafın türevi, $W_- + W_0 - W_+$ 'dır. Burada ,

$W_- : \frac{Z_i}{W_i}$ negatifken $\sum |W_i|$ 'lerin toplamıdır.

$W_0 : \frac{Z_i}{W_i}$ sıfırken $\sum |W_i|$ 'lerin toplamıdır.

$W_+ : \frac{Z_i}{W_i}$ pozitifken $\sum |W_i|$ 'lerin toplamıdır.

Mümkün yönlerin her biri için bu türevler hesaplanır. Türevi en çok negatif olan yön en uygun yöndür. Tüm türevler pozitifse, bu durumda geçerli $\hat{\beta}$ vektörü β ' katsayı tahminlerinin en iyi vektörüdür ve işlemler bu noktada son bulur.

Katsayıların Anlamlılığının Test Edilmesi

LAD Regresyonu' nda katsayıların anlamlılığının test edilebilmesi kısıtlı ve kısıtsız regresyon modellerinin artıklarının mutlak değerlerinin toplamı ile mümkün olmaktadır. Kısıtsız modelin parametre sayısı p, kısıtlı modelin parametre sayısı q' dur. İki model tahmin edilip artıklarının mutlak toplamı bulunarak iki model arasındaki farkı oluşturan (p-q) sayıda parametrenin anlamlılığı birlikte test edilir. Bu durumda temel hipotez,

$$H_0 : \beta_{q+1} = \dots = \beta_p = 0$$

şeklinde oluşturulacaktır. Alternatif hiotez ise temel hipotezin geçerli olmadığını ifade edecektir. Test istatistiği,

$$F_{LAD} = \frac{SAR_{k,l1} - SAR_{k,siz}}{(p-q)(\hat{\tau}/2)}$$

olarak hesaplanır. Burada SAR, artıkların mutlak değerleri toplamıdır yani $SAR = \sum |e_i|$ olacaktır. Ayrıca $\hat{\tau}$,

$$\hat{\tau} = \frac{\sqrt{m} [e_{(k_2)} - e_{(k_1)}]}{4}$$

olarak elde edilir. $\hat{\tau}$ 'nün hesaplanması için değeri sıfır olmayan artıkların sayısı m elde edilir. $m = n - (p+1)$ ' dir. $e_{(k_1)}$ ve $e_{(k_2)}$ basit LAD Regresyon' da açıklandığı gibi uygulanacaktır.

2.2.3 En Küçük Mutlak Sapma Regresyonu' nun Özel Hali Olan Medyan Regresyon

Alternatif regresyon modellerinden biri de LAD Regresyonu' nun özel bir hali olan “Medyan Regresyon” dur. Bu regresyon modelinin açıklanabilmesi için örnek ortalaması ve örnek medyanı ile ilgili bazı özelliklerin hatırlanması gerekmektedir.

Hiç açıklayıcı değişken olmaması durumunda, Y tesadüfi değişkeninin medyanı herhangi γ sayısı ise;

$$\Pr(Y \leq \gamma) \geq 1/2 \text{ ve } \Pr(Y \geq \gamma) \geq 1/2 \text{ olur.}$$

Örnek medyanı da c gibi bir sayı ise;

$$\frac{1}{n} \sum_{i=1}^n 1(y_i \leq c) \geq 1/2 \text{ ve } \frac{1}{n} \sum_{i=1}^n 1(y_i \geq c) \geq 1/2$$

Burada 1(.) karakteristik fonksiyondur ve parantez içindeki ifade doğruysa 1 değerini, diğer durumlarda “0” değerini aldığını ifade etmektedir. Seride gözlemlerin aldığı değerlerin yarısı c' nin altında yarısı üstünde yeralacaktır.

Y' nin dağılımı simetrikse medyanla ortalama aynıdır. Bu da anakütle ortalamasının/meydanının tahmincisi olarak örnek meydanının ya da ortalamasının kullanılabileceğini gösterir. Hangisinin daha iyi sonuç vereceği Y' nin dağılımına bağlıdır. Y'nin dağılımı normal değilse örnek ortalaması etkin değildir ve örnek medyanın asimtotik varyansı küçük olabilir. Aşağıdaki tabloda üç dağılım için asimtotik varyansların karşılaştırılması yapılmıştır⁴⁷.

Tablo 1

Asimtotik Varyansların Karşılaştırılması

Dağılım	Yoğunluk	Örnek Ortalamasının Varyansı	Örnek Meydanının Varyansı
Normal	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right)$	1	1,57
Laplace	$\frac{1}{2} \exp(- y)$	2	1
Cauchy	$\frac{1}{\pi} \frac{1}{1+y^2}$	∞	2,47

Laplace Dağılımı ve Cauchy Dağılımı' nın ikisi de simetrik değildir, yoğunluk fonksiyonlarının sağ ve sol kuyrukları normal yoğunluktan çok daha yavaş sıfıra yaklaşır (heavy-tailed). Bu, ortalama ya da medyan değerinden çok uzakta yeralan aşırı gözlemlerin olabileceği anlamına gelir. Bu durumda örnek medyanı örnek ortalamasından daha iyi sonuç verir. Cauchy Dağılımı' nın belli bir ortalaması olmadığından, örnek ortalamasının momenti yoktur.

⁴⁷ Roger Koenker, Gilbert Bassett, "Regression Quantiles", *Econometrica*, Vol.46, No.1, (Jan., 1978), s.33

EKK Regresyonunda amaç; $\sum_{i=1}^n (y_i - a)^2$ 'yi tahmin etmekte,

LAD Regresyonunda amaç; $\sum_{i=1}^n |y_i - c|$ 'yi minimize etmekte,

Burada amaç fonksiyonunun türevi alınırsa,

$$\frac{\partial}{\partial c} \sum_{i=1}^n |y_i - c| = \sum_{i=1}^n 1(y_i - c < 0) - 1(y_i - c > 0), \quad c \neq y_i$$

olacaktır. İfadeyi sıfıra eşitleyen c değeri örnek medyanıdır. Formüllerden de görüldüğü gibi örnek ortalaması aşırı değerlere örnek meydanından daha duyarlıdır⁴⁸. Gözlem sayısının (n) çok olduğu durumlarda medyan hesabı sıralama gerektirdiğinden zaman alıcı bir işlemdir.

LAD Tahmincisi;

$$\min_{\hat{\beta}} \sum_{i=1}^n |y_i - \hat{\beta} x_i|$$

olarak yazılabilir. LAD Tahmincisi aynı zamanda Y' nin X' e göre şartlı medyanının tahminini veren "Medyan Regresyon" tahmincisidir. Dağılım simetrik değilse, LAD tahmincisi aşırı değerlere EKK Tahmincisinden çok daha röbust olacaktır ve EKK Tahmincisiyle aynı tahmin sonuçlarını vermeyecektir.

⁴⁸ Z.Ying, S.H. Jung, And L.J.Wei, "Survival Analysis with Median Regression Models", Journal of the American Statistical Association, vol 90, no 429,(Mar., 1995), s.178.

2.3 En Küçük Medyan Kareler Regresyonu

EKK Regresyonu'nda artık kareleri toplamı minimize edilerek parametre tahmincileri elde edilmekteydi, En Küçük Medyan Kareler Regresyonu'nda (LMS) ise artık kareleri medyanı kullanılarak daha robust tahminciler elde edilir⁴⁹.

LMS Regresyon doğrusu aşırı değerlere EKK Regresyon Doğrusundan daha az duyarlıdır ve,

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \text{med } SR_i = \text{Median} \{(Y_1 - (\hat{\beta}_0 + \hat{\beta}_1 X_1))^2, (Y_2 - (\hat{\beta}_0 + \hat{\beta}_1 X_2))^2, \dots, (Y_n - (\hat{\beta}_0 + \hat{\beta}_1 X_n))^2\}$$

olarak gösterilir. Medyan sıralamayla elde edildiği için türev almaya ya da sürekli fonksiyonlardaki gibi diğer hesaplamalara olanak vermez. Her bir kesim ve eğim noktası için artık karelerinin hesaplanıp medyan değerinin belirlenebilmesi için sıralanması gerekir. LMS doğrusunu tahmin etmek karışık yöntemleri gerektirir, basit bir formülü yoktur⁵⁰.

Gözlem sayısı büyük olduğunda, gözlemlerden tesadüfi örnekler seçilerek bunlardan en iyi sonucun belirlenmesi önerilmiştir. LMS Regresyonu' nun amacı, aşırı değerlerin olması durumunda regresyon doğrusunu aşırı olmayan değerlerle oluşturmaktır. Aşırı olan gözlemlerin oranı büyük olabilir; ancak % 50 asimtotik limitini aşmamalıdır.⁵¹

p açıklayıcı değişkenli regresyon modeli için parametrelerin en küçük medyan kareler tahmincileri artıklar kararlı oluncaya kadar , p gözlemlerinin altkümelerinin yeniden örnekleme ile yapılır. LMS Regresyonu aşırı değerler olduğunda çok iyi sonuçlar verir; ancak doğrusal modelin parametre tahmincileri istenen özellikleri sağlamamaktadır⁵². Hata dağılımının simetrik olmaması durumunda ortalama regresyon tahminleri etkin değildirler. Hataların dağılımı, Cauchy Dağılımı' na uyuyorsa ortalama

⁴⁹ Peter J. Rousseeuw, "Least Median of Squares Regression", *Journal of the American Statistical Association*, Vol.79, No.388, (Dec., 1984), 871.

⁵⁰ Koenker and Bassett, s. 35.

⁵¹ A.C. Atkinson, "Influential Observations, High Leverage Points and Outliers in Linear Regression", *Statistical Science*, Vol.1, No.3, (Aug., 1986), s.397.

⁵² Rousseeuw, s. 872.

regresyon tahmini sonsuz varyansa, medyan regresyon tahmini sonlu varyansa sahip olur⁵³.

İkinci bir problem de LMS Regresyon katsayılarının güven aralığı için ya da hipotezlerinin testi için üretilmiş herhangi bir teori yoktur. LMS' a dayalı bir mesafe ölçüsünün hesaplanması önerilmiştir. Bunu kullanarak LMS Regresyonu' na göre aşırı değerleri de belirlemiştir. Bu gözlemler atılarak, geri kalan gözlemlere EKK Yöntemi uygulanır. Sonuç yeniden ağırlıklandırılmış EKK Regresyonu' nu verir⁵⁴.

2.4 M-Regresyon

M-Regresyon robust bir istatistiktir. Bir istatistiksel yöntemin “robust” olarak adlandırılabilmesi için, istatistiksel modelin varsayımları geçerli olmasa bile iyi sonuçlar vermesi gerekir. Verilerde aşırı değerler yoksa, EKK Tahminleri en iyi sonuçları verir; ancak tesadüfi hataların dağılımı normal olmadığında robust tahminler elde edilemez. Bu durumda robust tahminler elde edebilmek için M-Regresyon geliştirilmiştir⁵⁵.

2.4.1 Basit M Regresyon

EKK Regresyonu'nda, $\sum e_i^2$ ' yi minimum yapan parametreler tahmin edilmekte LAD Regresyonu' nda ise $\sum |e_i|$ ' yi minimum yapan değer bulunmaktadır. M-Regresyonda ise $\sum \rho(e_i)$ ' yi minimum yapan değerler bulunur. $\rho(e_i)$, artıkların toplam fonksiyonudur ve her hata teriminin amaç fonksiyonuna katkısını gösterir. ρ ' nun şu özellikleri taşıması gerekir⁵⁶:

$$\rho(e) \geq 0$$

⁵³ Sin-Ho Jung, “Quasi-Likelihood for Median Regression Models”, *Journal of the American Statistical Association*, Vol.91, No.433 (Mar., 1996), s. 251.

⁵⁴ Gerard E. Dallal, Ph. D., “Regression Diagnostics”, <http://www.tufts.edu/~gdallal/diagnose.htm> (11 Kasım 2004), s.3

⁵⁵ Peter J. Huber, “Robust Estimation of a Location Parameter”, *The Annals of Mathematical Statistics*, Vol.35, No.1,(Mar., 1964), s.73.

⁵⁶ John Fox, “Robust Regression”, *Appendix to An R and S-PLUS Companion to Applied Regression*, January 2002, <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-robust-regression.pdf> (11 Kasım 2004), s.1

$$\rho(0) = 0$$

$$\rho(e) = \rho(-e)$$

$$|e_i| > |e_{i'}| \text{ için } \rho(e_i) \geq \rho(e_{i'})$$

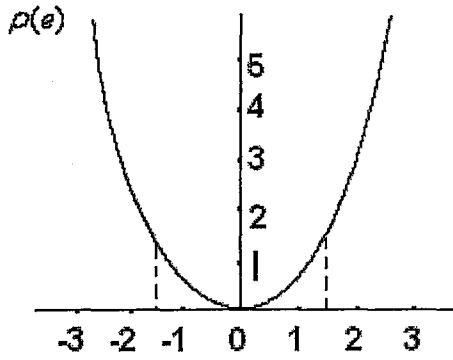
EKK ve LAD Regresyon tahminleri, M Regresyon tahminlerinin $\rho(e) = e^2$ ve $\rho(e) = |e|$ şeklindeki özel durumudur.

LAD Tahminleri aşırı değerlere duyarlı olmamaları açısından EKK tahminlerine göre üstündür; ancak aşırı değerler olmadığında EKK tahminleri daha doğru sonuç verirler. e “0”a yaklaştığında $\rho(e)$ ’yi e^2 ’ye, “0” dan uzaklaştığında da $\rho(e)$ ’yi $|e|$ ’ye eşitleyerek, bu iki yöntemin de avantajları birleştirilmeye çalışılır.

$$\rho(e) = \begin{cases} -k \leq e \leq k & \text{ise } e^2 \\ e < -k \text{ ya da } k < e & \text{ise } 2k|e| - k^2 \end{cases}$$

$k = 1,5\hat{\sigma}$ alınması gerektiğini öne sürmüştür⁵⁷. $\hat{\sigma}$ hataların anakütle standart sapmaları σ ’nın tahminidir. $\rho(e)$ ’yi düzgün (smooth) fonksiyon haline getirmek için $|e|$ yerine $2k|e| - k^2$ kullanılır. Bu fonksiyonun grafiği aşağıdaki şekilde gibidir.

⁵⁷ Peter J. Huber, “Robust Estimation of a Location Parameter”, *The Annals of Mathematical Statistics*, Vol.35, No.1,(Mar., 1964), s.79.



Şekil 1. $\rho(e)$ Fonksiyonunun Grafiği

σ ' yı tahmin edebilmek için $\hat{\sigma} = 1,483 MAD$ alınır. MAD , mutlak sapmaların $|e_i|$ medyanıdır. Tesadüfi hataların normal dağılmış olması durumunda, $\hat{\sigma}$ ' nın σ ' nın iyi bir tahmini olmasını sağlamak amacıyla 1,483 çarpanı seçilir.

M-tahminleri,

$$\sum \rho(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))$$

gösterimini minimize eden değerlerdir. ρ fonksiyonunun içinde $k = 1,5\hat{\sigma}$ değeri yer alır. $\hat{\sigma}$, $Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$ sapmalarından hesaplanır. $\sum \rho(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))$ ' yi minimize etmek için bir algoritma kullanılır.

$\hat{\beta}_0$ ve $\hat{\beta}_1$ ' in ilk tahmin değerleri EKK Yöntemi ile bulunur. Bu değerlerle sapma ve σ ' nın tahmini hesaplanır. Bunlarla da $\hat{\beta}_0$ ve $\hat{\beta}_1$ ' in geliştirilmiş tahminleri yapılır. Bu geliştirilmiş tahminler de yeni sapma ve σ ' nın geliştirilmiş tahminini hesaplamak için kullanılır. Algoritma bu şekilde geliştirilmiş tahminler öncekilerle aynı olana kadar tekrarlanır.

Algoritmanın herhangi bir adımında $\hat{\beta}_0^0$ ve $\hat{\beta}_1^0$ ' in, β_0 ve β_1 'in mevcut tahminleri olması durumunda, sapma $e_i^0 = Y_i - (\hat{\beta}_0^0 + \hat{\beta}_1^0 X_i)$ olarak hesaplanır. Bu sapmalar küçükten büyüğe sıralanarak, sapmaların medyan değeri $\hat{\sigma}^0 = 1,483 MAD$

ifadesinde yerine konulur ve bunun yardımı ile $k = 1,5\hat{\sigma}^0$ hesaplanır. Sapma değerleri hesaplanan bu $k = 1,5\hat{\sigma}^0$ değeri ile karşılaştırılır. Aşağıda belirtilen aralıklarda yeralan e_i^0 sapma değerleri atılarak yerlerine ,bulunduğu aralığa göre e_i^* ile gösterilen yeni sapma değerleri belirlenecektir⁵⁸.

$$-1,5\hat{\sigma} < e_i^0 < 1,5\hat{\sigma}^0 \text{ ise } e_i^* = e_i^0 \text{ olur,}$$

$$e_i^0 < -1,5\hat{\sigma}^0 \text{ ise } e_i^* = -1,5\hat{\sigma} \text{ olur,}$$

$$e_i^0 > 1,5\hat{\sigma}^0 \text{ ise } e_i^* = 1,5\hat{\sigma} \text{ olur.}$$

e_i^* değerleri , e_i^0 , in atılmasıyla belirlenen düzeltilmiş sapma değerleridir, böylelikle mutlak değerce hiçbir sapma $1,5\hat{\sigma}^0$, dan büyük olmayacaktır. Hesaplanan e_i^* değerleri ,

$$Y_i^* = \hat{\beta}_0^0 + \hat{\beta}_1^0 X_i + e_i^*$$

ifadesinde gösterildiği gibi tahmin edilen $\hat{Y}_i = \hat{\beta}_0^0 + \hat{\beta}_1^0 X_i$ değerleri ile toplanarak Y_i^* değerleri elde edilir ve bu değerlere En Küçük Kareler Yöntemi uygulanarak β_0 ve β_1 , in geliştirilmiş tahminleri elde edilmiş olunur. En Küçük Kareler Regresyon Modeli' nde X_i değerleri yerine konularak sapmalar hesaplanır sonra sırası ile yukarıdaki işlemler yapılarak parametrelerin geliştirilmiş tahminleri elde edilir, geliştirilmiş tahminler arasında 10^{-4} gibi çok az bir fark kalınca işlemler son bulur ve M Regresyon' un parametre tahminleri belirlenmiş olur.

2.4.2 Çoklu M Regresyon

Basit regresyonda olduğu gibi Huber-M Tahminleri;

$$\sum \rho(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik}))$$

⁵⁸ Peter J. Huber, "Robust Estimation of a Location Parameter", The Annals of Mathematical Statistics, Vol.35, No.1,(Mar., 1964), s.75.

gösterimini minimize eden $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ değerleridir. Basit regresyonda olduğu gibi,

$$\rho(e) = \begin{cases} -k \leq e \leq k & \text{ise } e^2 \\ e < -k \text{ ya da } k < e & \text{ise } 2k|e| - k^2 \end{cases}$$

olarak tanımlanmıştır. Vektör gösterimi ile:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad X_i = \begin{bmatrix} 1 \\ X_{i1} \\ \vdots \\ X_{ik} \end{bmatrix}$$

olur. Huber Tahminlerinin $\hat{\beta}$ vektörü, $\sum \rho(Y_i - \hat{\beta}' X_i)$ ' yi minimize eden $\hat{\beta}$ vektörü olarak tanımlanır⁵⁹.

$\hat{\beta}$ ile gösterilen regresyon katsayıları vektörü öncelikle EKK tahminleri vektörüyle tahmin edilir. β' 'nin bu tahminleri Basit M Regresyon' da olduğu gibi β' 'nin geliştirilmiş tahminlerini belirlemede kullanılan sapmaları ve σ' 'nin tahminini hesaplamada kullanılır. Algoritma bu şekilde β' 'nin geliştirilmiş tahminleri öncekilerle aynı ya da yaklaşık olarak aynı olana kadar tekrarlanır.

Algoritmanın herhangi bir adımında, $\hat{\beta}^0$ 'ın β' 'nin tahmini olması durumunda, sapmalar $Y_i - (\hat{\beta}^0)' X_i$ olarak hesaplanır. Bu sapmalar küçükten büyüğe sıralanarak, sapmaların medyan değeri $\hat{\sigma}^0 = 1,483 MAD$ ifadesinde yerine konulur ve bunun yardımı ile $k = 1,5\hat{\sigma}^0$ hesaplanır. Sapma değerleri hesaplanan bu $k = 1,5\hat{\sigma}^0$ değeri ile karşılaştırılır. Aşağıda belirtilen aralıklarda yeralan e_i^0 sapma değerleri atılarak yerlerine, bulunduğu aralığa göre e_i^* ile gösterilen yeni sapma değerleri belirlenecektir.

⁵⁹ Sam Efromovich, *Nonparametric Curve Estimation: Methods, Theory & Applications*, Secaucus, USA: Springer Verlag-New York, 1999, s. 150.

$$-1,5\hat{\sigma} < e_i^0 < 1,5\hat{\sigma}^0 \text{ ise } e_i^* = e_i^0 \text{ olur,}$$

$$e_i^0 < -1,5\hat{\sigma}^0 \text{ ise } e_i^* = -1,5\hat{\sigma} \text{ olur,}$$

$$e_i^0 > 1,5\hat{\sigma}^0 \text{ ise } e_i^* = 1,5\hat{\sigma} \text{ olur.}$$

e_i^* değerleri , e_i^0 ' in atılmasıyla belirlenen düzeltilmiş sapma değerleridir, böylelikle mutlak değerce hiçbir sapma $1,5\hat{\sigma}^0$ ' dan büyük olmayacaktır. Hesaplanan e_i^* değerleri ,

$$Y_i^* = (\hat{\beta}^0)' X_i + e_i^*$$

ifadesinde gösterildiği gibi tahmin edilen $\hat{Y}_i = (\hat{\beta}^0)' X_i$ değerleri ile toplanarak Y_i^* değerleri elde edilir ve bu değerlere En Küçük Kareler Yöntemi uygulanarak β_0 ve β_1 ' in geliştirilmiş tahminleri elde edilmiş olunur. En Küçük Kareler Regresyon Modeli' nde X_i değerleri yerine konularak sapmalar hesaplanır sonra sırası ile yukarıdaki işlemler yapılarak parametrelerin geliştirilmiş tahminleri elde edilir, geliştirilmiş tahminler arasında 10^{-4} gibi çok az bir fark kalınca işlemler son bulur ve M Regresyon' un Çoklu parametre tahminleri belirlenmiş olur.

Katsayıların Anlamlılığının Test Edilmesi

Genel doğrusal regresyon modeli,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + e_i$$

'de temel hipotez,

$\beta_{q+1} = \dots = \beta_k = 0$ ' dır. Alternatif hipotez, temel hipotezin doğru olmadığını ifade eder. Test istatistiğinin hesaplanabilmesi için kısıtlı ve kısıtsız modeller tahmin edilir.

$$\text{Kısıtlı model: } Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_q X_{iq} + e_i$$

$$\text{Kısıtsız model: } Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + e_i \quad (q < k)$$

olacaktır. $\hat{\sigma}_{EKK}^2$ 'nın hesabında kısıtsız model kullanılır⁶⁰.

M-Regresyonda test istatistiği,

$$F_M = \frac{STR_{k,l1} - STR_{k,siz}}{(k-q)\hat{\lambda}}$$

'dır. Burada $STR = \sum \rho(\hat{e}_i)$ dönüştürülmüş artıkların toplamıdır. $STR_{k,l1}$ ve $STR_{k,siz}$, kısıtlı ve kısıtsız modele M Regresyon' un uygulanması ile bulunur. Ayrıca,

$$\hat{\lambda} = \frac{(n/m) \sum e_i^{*2}}{(n-k-1)}$$

olarak elde edilir. $\hat{\lambda}$ 'nın hesaplanmasında kısıtsız modelin artıkları kullanılır. $\hat{e}_i^*$ daha önce belirtildiği gibi, e_i^0 'ın atılmasıyla belirlenen düzeltilmiş sapma değerleridir. Burada m düzeltilmemiş (truncation olmamış) sapma değerleridir, yani $|e_i| \leq 1,5\hat{\sigma}$ koşulunu uyan sapma sayısıdır. F test istatistiğinin dağılımı $F, (k-q); (n-k-1)$ serbestlik dereceli F dağılımıdır.

2.5 Rank Regresyon

Alternatif regresyon yöntemi olarak nonparametrik regresyon yöntemleri de kullanılabilir, burada sadece Rank Regresyon' a yer verilmiştir.

2.5.1 Basit Rank Regresyon

Bu regresyonda her bir veri çiftinin eğimi hesaplanarak, bu eğimlerin bir çeşit ortalama ya da meydanının alınması ile doğrunun eğimi bulunur.

β 'nın EKK Yöntemi' nde kullanılan;

⁶⁰ Ronald M. Schrader, And Thomas P. Hettmansperger, "Robust Analyses of Variance Based Upon a Likelihood Criterion", *Biometrika*, Vol 67, No.1 (Apr., 1980), s. 93.

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

formülleri, veri çifti (X_i, Y_i) ve (X_j, Y_j) iken $b_{ij} = (Y_i - Y_j)/(X_i - X_j)$ eğim çiftlerinin ağırlıklı ortalaması olarak,

$$\hat{\beta}_{LS} = \sum w_{ij} b_{ij}$$

$$w_{ij} = (X_i - X_j)^2 / \sum (X_i - X_j)^2$$

olarak ifade edilebilir⁶¹. ($X_i = X_j$ olan noktalar ihmal edilir.)

Diğer bir yol da b_{ij} eğim çiftlerinin ağırlıklı medyanının kullanılmasıdır. Seriyi iki eşit parçaya bölen değere medyan denmektedir. Medyan seriyi iki eşit parçaya böldüğünden birimlerin % 50' sinin değeri medyandan küçük, % 50' sinin değeri medyandan büyük olacaktır. Toplanan verilerin büyüktan küçüğe doğru sıralanmasıyla elde edilen seride en ortadaki değer medyayı verir⁶². X_i seri değerlerinin ağırlıklı medyayı şu şekilde belirlenir: Özellikle X_i değerleri artan sıraya göre dizilir, indisleri değiştirilerek $X_1 < X_2 < \dots < X_n$ şeklinde düzenleme yapılır. Her X_i değerinin w_i ağırlığı bulunmaktadır. Ağırlıklar negatif olmamalıdır ve toplamı "1" olmalıdır.

X_k ağırlıklı medyan değerinin bulunabilmesi için,

$$w_1 + w_2 + \dots + w_{k-1} < 0.5$$

$$w_1 + w_2 + \dots + w_{k-1} + w_k > 0.5$$

⁶¹ J. N. Adichie, "Estimates of Regression Parameters Based on Rank Tests", *The Annals of Mathematical Statistics*, Vol 38, No.3 (Jun., 1967), s. 894.

⁶² Münevver Turanlı, Selahattin Güriş, *Temel İstatistik*, 2. Basım, İstanbul: Der Yayınları, 2005, s. 94.

koşulunu sağlayacak k indeksi bulunur. Bu durumda w_k ağırlığına sahip olan X_k değeri ağırlıklı medyandır. $w_1 + w_2 + \dots + w_{k-1} = 0.5$ eşitliğinin olması durumunda ise $(X_{k-1} + X_k)/2$ ağırlıklı medyan olarak belirlenir. Tüm ağırlıklar birbirine eşit ise, $w_i = 1/n$ olacak ve ağırlıklı medyan, medyana eşit olacaktır.

Basit Rank Regresyon' da tahmin yapabilmek için öncelikle veri çiftlerinin sayısı belirlenir. n gözlem sayısı ise, $n(n-1)/2$ tane veri çifti vardır. bu veri çiftlerinden aynı X değerine sahip olan gözlem değerleri çıkarılır. Kalan veri çiftlerinin her biri için eğimler,

$$b_{ij} = (Y_i - Y_j)/(X_i - X_j)$$

olarak hesaplanır. bunlara karşılık gelen ağırlıklar da,

$$w_{ij} = |X_i - X_j| / \sum |X_i - X_j|$$

olarak hesaplanır. eğimler artan sıraya göre dizilerek, bunların kümülatif ağırlıkları hesaplanır. 0,5 değerini aşan kümülatif ağırlık değerine karşılık gelen eğim değeri regresyon modelinin eğim parametresinin tahmini değeridir. Her gözlem değerinin X ve Y değeri,

$$Y_i - \hat{\beta}_1 X_i = \hat{\beta}_0$$

eşitliğinde yerine konularak elde edilen serinin medyan değeri de regresyon modelinin sabit parametresi olarak belirlenir.

Ranka Dayanarak β ' nin Tahmini

Regresyon modelinin $\hat{\epsilon}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$ artıklarını minimum yapacak β_0 ve β_1 parametrelerinin tahmini değerlerinin seçilmesi amaçlanmaktadır. Artıkların minimumluğu, mutlak değerlerinin ağırlıklı toplamalarının ortalaması $\sum w_i |\hat{\epsilon}_i|$ ile ölçülebilir. w_i ağırlıkları pozitif olmalıdır. En küçük mutlak sapmalar tahmininde β_0 ve

β_1 parametrelerinin tahmini için ağırlıklı toplam $w_i = 1$ ağırlıklarıyla, EKK Tahmininde $w_i = |e_i|$ ağırlıkları ile minimize ediliyordu. Artıkları eşit olarak ağırlıklandırmayla, mutlak değerlerine göre ağırlıklandırma arasında orta yol; mutlak değerlerinin rankına göre ağırlıklandırmaktır. $w_i = \text{rank}(|\hat{\varepsilon}_i|)$ (Sıralama küçükten büyüğe doğru yapılır, $|\hat{\varepsilon}_i|$ 'nin en küçük değeri en küçük sıra sayısı olan 1' dir.) $|\hat{\varepsilon}_i|$ 'nin değeri büyük olabilirken, $\text{rank}|\hat{\varepsilon}_i|$, n' den daha büyük olamaz. Böylelikle büyük artıkların etkisi EKK' de olduğundan daha çok sınırlandırılmış olur⁶³.

$\sum \text{rank}(|\hat{\varepsilon}_i|)|\hat{\varepsilon}_i|$ 'yi minimize edecek tahminler yerine, $\sum \left[\text{rank}(\hat{\varepsilon}_i) - \frac{n+1}{2} \right] \hat{\varepsilon}_i$ 'yi minimize edecek tahminler yapılır. Hatalar simetrik dağılımlı olduğunda ve örnek hacmi n büyük olduğunda, her iki toplam da yaklaşık olarak aynı tahminlerin elde edilmesini sağlar. $\sum \left[\text{rank}(\hat{\varepsilon}_i) - \frac{n+1}{2} \right] \hat{\varepsilon}_i$ toplamının hatalar simetrik dağılımlı olmasa da iyi sonuçlar vermesi beklenir. Nonparametrik $\hat{\beta}$ tahmini, ağırlıklı medyan olarak ifade edildiği gibi, $\hat{\varepsilon} = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$ olarak tanımlandığında, $\sum \left[\text{rank}(\hat{\varepsilon}_i) - \frac{n+1}{2} \right] \hat{\varepsilon}_i$ 'yi minimize eden $\hat{\beta}_1$ değeri olarak da tanımlanabilir.

Eğim Parametresinin Anlamlılığının Test Edilmesi

$\beta = 0$ ' ı test etmek için test istatistiği,

$$|t| = \frac{|U|}{SD|U|}$$

olarak hesaplanır. Burada,

⁶³ David Draper, "Rank-Based Robust Analysis of Linear Models", *Statistical Science*, Vol.3, No.2 (May., 1988), s.239.

$$U = \sum \left[\text{rank}(Y_i) - \frac{n+1}{2} \right] X_i$$

$$SD(U) = \sqrt{\frac{n(n+1)}{12} \sum (X_i - \bar{X})^2}$$

olarak hesaplanır.

Testin p değeri, $\text{Prob} [|Z| \geq |t|]$ ile hesaplanır. Z; standart normal dağılıma sahip tesadüfi değişkendir.

2.5.2 Çoklu Regresyon

Basit regresyonda iki yolla tahmin yapılyordu. Eğim çiftlerinin ağırlıklı medyan toplamalarını minimize ederek, diğer yol da basit bir şekilde çoklu regresyon için genelleştirilebilir. Nonparametrik $\hat{\beta}_1, \dots, \hat{\beta}_k$ tahminleri,

$$\sum \left[\text{rank}(Y_i - (\hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik})) - \frac{n+1}{2} \right] \cdot (Y_i - (\hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik}))$$

toplamını minimize eden değerlerdir.

$\hat{\beta}_0$ 'ın nonparametrik tahmini $Y_i - (\hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik})$ farklarının medyanıyla belirlenir. Yukarıdaki fonksiyonu minimize etmek için bir algoritma hesaplanır. Fonksiyonun vektör gösterimi ile ifade edebilmesi için,

$$g(\hat{\beta}) = \sum \left[\text{rank}(Y_i - \hat{\beta}' X_i) - \frac{n+1}{2} \right] (Y_i - \hat{\beta}' X_i)$$

olarak gösterilsin. Burada,

$$b = (b_1, \dots, b_k)$$

$$X_i = (X_{i1}, \dots, X_{ik})$$

' dir. $g(\beta)$ ' yı minimize edecek $\hat{\beta}$ vektörünü bulmak amaçtır. Bunun için, EKK Tahminleri vektörleriyle iteratif yöntemle başlanır, sonra g fonksiyonunun daha küçük değerlerini veren vektörler bulunur. $\hat{\beta}^0$ başlangıç vektörü iken, $\hat{\beta}^* = \hat{\beta}^0 + t^* d$ oluşturularak daha iyi bir vektör belirlenir. t^* ve d şu şekilde hesaplanır: $z_i = Y_i - (\hat{\beta}^0)' X_i$, $u_i^0 = rank(z_i) - \frac{1}{2}(n+1)$, $u^0 = u_i^0$ ların oluşturduğu $n \times 1$ ' lik vektördür. $X_c = X_{ij} - \bar{X}_j$ lerin oluşturduğu $n \times p$ ' lik matristir. $d = (X_c' X_c)^{-1} X_c' u^0$, $w_i = d' X_i$, t^* da $|w_i - w_j| / \sum |w_i - w_j|$ ağırlıklarıyla $|z_i - z_j| / |w_i - w_j|$ oranlarının ağırlıklı medyanıdır. Bu iterasyon tamamlanır, b^* geçerli vektör halini alır ve iterasyon tekrarlanır.

Katsayıların Anlamlılığının Test Edilmesi

Genel doğrusal regresyon modeli,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + e_i$$

'de temel hipotez,

$\beta_{q+1} = \dots = \beta_k = 0$ ' dır. Alternatif hipotez, temel hipotezin doğru olmadığını ifade eder. Test istatistiğinin hesaplanabilmesi için kısıtlı ve kısıtsız modeller tahmin edilir.

$$\text{Kısıtlı model: } Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_q X_{iq} + e_i$$

$$\text{Kısıtsız model: } Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + e_i \quad (q < k)$$

Rank ağırlıklı artıkların toplamına SRWR denirse,

$$SRWR = \sum \left[rank(\hat{\varepsilon}_i) - \frac{1}{2}(n+1) \right] \hat{\varepsilon}_i. \text{ Nanparametrik test istatistiği,}$$

$$F_{NP} = \frac{SRWR_{k,l1} - SRWR_{k,siz}}{(p-q)c\hat{\tau}}$$

olarak hesaplanır. Burada,

$$c = (n+1) / \sqrt{48}$$

$$\hat{\tau} = f \frac{\sqrt{n} [A_{(k_2)} - A_{(k_1)}]}{2(1,645)}$$

olarak gösterilir. $SRWR_{k,l}$ ve $SRWR_{k,siz}$ artıklar, kısıtlı ve kısıtsız modele nonparametrik regresyon yönteminin uygulanması ile elde edilir.

$$\text{Kısıtlı model: } Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_q X_{iq} + e_i$$

$$\text{Kısıtsız model: } Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + e_i \quad (q < k)$$

SRWR, $g(b) = \sum \left[\text{rank}(Y_i - b'X_i) - \frac{n+1}{2} \right] (Y_i - b'X_i)$ nin minimum değeridir, $SRWR = g(\hat{\beta})$.

$\hat{\tau}$ ' nun hesaplanması için, kısıtsız modelden artıklar alınır ve $1 \leq i \leq j \leq n$ için $A_{ij} = (\hat{\varepsilon}_i + \hat{\varepsilon}_j) / 2$ ortalama çiftleri oluşturulur., Bu $N = n(n+1)/2$ tane değer, artan sıraya göre dizilir; $A_{(1)} \leq A_{(2)} \leq \dots \leq A_{(N)}$, $a = n(n+1)/4$, $b = \sqrt{n(n+1)(2n+1)/24}$ iken, $k_1 = \frac{1}{2} + a - (1,645)b$ ' e en yakın tamsayı değerini, $k_2 = \frac{1}{2} + a + (1,645)b$ ' e en yakın tamsayı değerini, $f = \sqrt{n/[n-(p+1)]}$ değerini ifade etmektedir.

EKK Testi'nde olduğu gibi nonparametrik testin yaklaşık p değeri $\text{Pr ob}[F \geq F_{NP}]$ olacak şekilde hesaplanır. F, (k-q) ve (n-k-1) serbestlik dereceli F dağılımına sahip tesadüfi değişkendir⁶⁴.

⁶⁴ J. N. Adichie, "Rank Tests of Sub-Hypotheses in the General Linear Regression", *Annals of Statistics*, Vol 6, No.5 (Sep., 1978), s. 1012.

2.6 Alternatif Regresyon Modellerinin Karşılaştırılması

EKK Regresyonu diğer regresyonlara göre daha kolay hesaplanır. Burada amaç, artıkların kareleri toplamını minimize etmektir. Hatalar normal dağıldığında sapmasız tahminciler arasından minimum varyanslı olanı EKK tahmincisidir. Hatalar normal dağılmadığında EKK tahmincilerinin varyansları gerçek değerlerinden daha büyük olacaktırlar⁶⁵. EKK Regresyonu' nun dezavantajı normal dağılım koşulu sağlanmadığında ve aşırı değerlerin olması durumunda tahmin edilen katsayıların gerçeği yansıtmamasıdır. Tablo 2 ' de EKK ile LAD Regresyonu karşılaştırılmıştır.

Tablo 2

EKK Regresyonu İle LAD Regresyonu' nun Karşılaştırılması

EKK REGRESYONU	LAD REGRESYONU
$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-2}}$	$\hat{\tau} = \frac{\sqrt{m} [e_{(k_2)} - e_{(k_1)}]}{4}$
$ t = \frac{ \hat{\beta}_1 }{S_{\hat{\beta}_1}}$	$ t = \frac{ \hat{\beta}_1 }{S_{\hat{\beta}_1}}$
$S_{\hat{\beta}_1} = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}}$	$S_{\hat{\beta}_1} = \frac{\hat{\tau}}{\sqrt{\sum (X_i - \bar{X})^2}}$

LAD Regresyonu' nda amaç, artıkların mutlak değerlerinin toplamını minimize etmektir. Hataların dağılımı Laplace dağılımı' na uyuyorsa, LAD Tahmincileri En Çok Benzerlik Tahminciler ile aynı sonuçları verirler.⁶⁶ Hatalar simetrik dağılmadığında ya

⁶⁵ Jan Kmenta, *Elements of Econometrics*, 2. Basım,, Maxwell Macmillan International Editions, 1971.

⁶⁶ Gilbert Bassett, And Roger Koenker, "Asymptotic Theory of Least Absolute Error Regression", *Journal of the American Statistical Association*, Vol.73, No. 363 (Sep., 1978), s. 618.

da örnek hacminin çok büyük olduğu durumlarda LAD Regresyonu' nun kullanılması uygundur.

LAD Testi' nde $|t|$ ' nin hesaplanması EKK' da hesaplanması ile aynıdır; ancak $\hat{\beta}_1$ ve $S_{\hat{\beta}_1}^2$ ' nin hesaplanması farklıdır. $S_{\hat{\beta}_1}^2$ formülündeki tek farklılık $\hat{\sigma}$ yerine $\hat{\tau}$ kullanılmasıdır.

LAD Regresyon' da τ/σ , regresyon doğrusunun eğimini tahmin etmede hangi yöntemin daha iyi olduğunu belirler. τ ve σ ' nin ikisi de tesadüfi hataların büyüklüğünün ölçüsüdür. $\tau, 1/2\theta$ ' ya eşittir. θ : Medyan noktasında hata dağılımının olasılık yoğunluğudur. (LAD Regresyonda bazen tahmin ortalama yerine medyanla yapılır, hata dağılımı simetrikse ortalama ve medyan çakışır.) σ büyükse hatalar geniş yayılmıştır, medyan etrafındaki olasılık yoğunluğu düşüktür, bundan dolayı θ küçük, τ büyüktür.

σ büyük olduğunda τ büyük, küçük olduğunda τ küçük olur; ancak τ/σ ' nin gerçek oranı hata dağılımının şekline bağlıdır. Hatalar normal dağılımlıysa büyük örnekte EKK Regresyon tahminleri LAD Regresyon tahminlerinden daha iyi sonuç verirler. Hatalar Laplace Dağılımı' na sahipse $\tau/\sigma = 0,707 < 1$ olur.

Yazılan formüller büyük örnek varsayımıyla geçerlidir. m de yaklaşık olarak n ' e eşittir. $\hat{\epsilon}_i$ artıkları da gerçek hatalarla (ϵ_i) aynı özelliklere sahiptir. k_1 ve k_2 ' nin $\frac{(n+1)}{2} \mp \sqrt{n}$, e yakın tamsayılar olduğu durumda τ , $\sqrt{n} [e_{(k_2)} - e_{(k_1)}] / 4$, e yaklaştırılmaktadır.

LAD' deki τ , EKK' deki σ ' ya benzemekteydi. EKK Regresyonunun standart sapması σc formuna sahiptir, LAD Regresyonunun standart sapması da yaklaşık olarak τc ' ye eşittir. c ikisinde de aynıdır. Regresyon modelinde açıklayıcı değişken olmaması durumunda aynı anakütleden $e_1, e_2, \dots, e_n$ gibi n bağımsız gözlemden oluşan tesadüfi örnek seçildiğinde, EKK Regresyon tahmini örnek ortalaması $\bar{e}$ ' ye eşittir, çünkü $a = \bar{e}$, $\sum (e_i - a)^2$ ' yi yani en küçük sapma kareleri toplamını veren değer olmaktadır.

Standart sapması $\sigma/\sqrt{n}$ ' dir. LAD Regresyon tahmini örnek medyanı $\tilde{e}$ ' dir, çünkü $a = \tilde{e}$, en küçük mutlak sapma toplamlarını $\sum |e_i - a|$ minimize eden a ' nın değeridir. Standart sapması yaklaşık olarak $\tau/\sqrt{n}$ ' dir.

Merkezi Limit Teoremi' ne göre, büyük örnek hacimlerinde (n) örnek ortalamasının dağılımı normale yaklaşırken, örnek meydanının dağılımı da normale yaklaşır. Örnek medyanının beklenen değeri anakütle medyanına yaklaşır. v anakütle hatalarının medyanını gösterebilir, büyük n için, $\bar{e}$, v ortalama ve $\tau/\sqrt{n}$ standart sapmayla yaklaşık olarak normal dağılıma sahiptir. $\hat{\tau}$, τ ' nun tahmini değeridir, % 95 güven aralığıyla v, yaklaşık olarak $\tilde{e} \pm 2\hat{\tau}/\sqrt{n}$ olur. Sadece v ya da güven aralığı bizim için önemli değildir, uzunluğu $4\tau/\sqrt{n}$ önemlidir. V için $e_{(k_1)}$ ' den $e_{(k_2)}$ ' ye güven aralığı, % 95 güven aralığıyla gösterilebilir. Büyük örnek hacimleri için iki aralık da benzerdir. $4\tau/\sqrt{n} \approx e_{(k_2)} - e_{(k_1)}$. Bunu gerçekten eşit yapabilmek için τ ' yu seçerken, e $\hat{e}$ ile, n de m= n-2 ile değiştirilir ve,

$$\hat{\tau} = \frac{\sqrt{m} [e_{(k_2)} - e_{(k_1)}]}{4}$$

formülü elde edilir. Sadece sıfıra eşit olmayan artıkların, küçük örnekler için iyi sonuç verdikleri bulunmuştur. EKK Regresyonu ile LAD Regresyonu'nun Karşılaştırılması Tablo 1' de özetlenmiştir.

M Regresyon' da amaç, herhangi $g(e)$ fonksiyonu için $\sum g(y_i - x_i'\hat{\beta})$ ' yı minimize etmektir. $g(e) = e^2$ olduğunda EKK tahmincileri M tahmincilerine eşittir.

M Regresyon' da hesaplanan $F_M = \frac{STR_{k,l1} - STR_{k,siz}}{(p-q)\hat{\lambda}}$ formülü EKK' de kullanılan F test istatistiğine çok benzemektedir. e_i^* ve $\rho(e)$ tanımlanırken, 1.5 yerine ∞ konursa; bu durumda F_M formülü F_{EKK} ' ya dönüşür. 1.5, ∞ ile değiştirildiğinde tüm i' ler için $e_i^* = e_i$ ve m=n (truncation olmamış e_i ' lerin sayısı= gözlem sayısı)

olacaktır. $\hat{\lambda}$ da $\hat{\sigma}_{EKK}^2$ ' na eşit olacaktır. $\rho(e) = e^2$ olacak, $STR = SSE$ ve $F_M = F_{EKK}$ olacaktır.

$$\rho(e) = \begin{cases} -k \leq e \leq k & \text{ise } e^2 \\ e < -k \text{ ya da } k < e & \text{ise } 2k|e| - k^2 \end{cases}$$

Tanımlamasında, k için başka değerler de seçilebilir. k ' nın büyük bir değeri için birçok artık $g(e) = e^2$ olacak, M -tahminleri EKK tahminlerine yaklaşık bir değer olacaktır. k ' nın küçük bir değeri için birçok artık $g(e) = 2k|e| - k^2$ olacak, $\sum (2k|e| - k^2)$ ' yi minimize etmek $\sum |e_i|$ ' yi minimize etmekle aynı olacağından M -tahminleri LAD tahminlerine yaklaşık bir değer olacaktır. Regresyon katsayılarının en iyi tahminlerini veren k değeri tesadüfi hataların anakütle dağılımına bağlıdır. Dağılımda çok sayıda aşırı değer varsa k ' nın küçük değeri, az sayıda aşırı değer varsa, normal dağılımda olduğu gibi, k ' nın büyük değeri en iyi sonucu verir. İki durum da göz önüne alındığında $k = 1,5\hat{\sigma}$ alınır.

EKK Yöntemi' nde, hataların normal dağılıma sahip olduğu varsayıldığı için bu yöntem parametrik bir yöntemdir. M Tahminleri için, hata terimlerinin normal dağılıma sahip olması gerekmez; ancak normale yakın olması durumunda iyi sonuçlar elde edilir.

Tablo 3

EKK Regresyonu ile Nonparametrik Regresyon' un Karşılaştırılması

EKK REGRESYONU	NONPARAMETRİK REGRESYON
$t_{EKK} = \frac{\hat{\beta}_{EKK}}{S_{\hat{\beta}_{EKK}}}$	$t_{NP} = \frac{U}{SD(U)}$
$\hat{\beta}_{EKK} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$	$U = \sum (r_i - \bar{r})X_i$
$S_{\hat{\beta}_{LS}} = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}}$	$\bar{r} = \frac{1}{2}(n+1)$ $\sum (r_i - \bar{r}) = 0$ $U = \sum (X_i - \bar{X})(r_i - \bar{r})$
$t_{LS} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\hat{\sigma} \sqrt{\sum (X_i - \bar{X})^2}}$	$t_{NP} = \frac{\sum (X_i - \bar{X})(r_i - \bar{r})}{\sqrt{\frac{n(n+1)}{12}} \sqrt{\sum (X_i - \bar{X})^2}}$

Tablo 3'te EKK Regresyonu ile Nonparametrik Regresyon'un karşılaştırması özetlenmiştir. Görüldüğü gibi; nonparametrik test istatistiği, t_{EKK} formülünde Y_i ' nin yerine r_i , $\hat{\sigma}$ ' nin yerine $\sqrt{n(n+1)/12}$ gelmiş halidir. $\hat{\sigma}$, $SD(Y_i)$ ' nin tahminidir, $\sqrt{n(n+1)/12}$, $SD(r_i)$ ' ye eşit olarak gösterilebilir. (β ' nin testi için EKK ile nonparametrik sıraya dayalı yöntemler arasındaki bu benzerlik β ' nin tahmini için geliştirilemez. Yani $\sum (X_i - \bar{X})(r_i - \bar{r}) / \sum (X_i - \bar{X})^2$ ' nin β ' nin iyi bir tahmini olduğu önerilemez. X_i ile r_i arasındaki korelasyon katsayısını $\sqrt{n-1}$ ' le çarparak t_{NP} ' yi belirlemek başka bir yoldur.)

M Regresyon ve Nonparametrik Regresyon hatalar normal dağılsa da dağılmasa da iyi sonuçlar verir, her ikisi de aşırı değerlerin olması durumunda kullanılır.

2.7 Aşırı Değerlerin Belirlenmesi

Daha önce de belirtildiği gibi, aşırı değerler diğer gözlem değerlerinden uzakta bulunan ve onlarla tutarlılık göstermeyen değerlerdir. Bunlar diğer gözlem değerleriyle karşılaştırıldıklarında çok büyük veya çok küçük değerlerdir⁶⁷. Bu değerler daha çok gözlem değerlerinin oluşturduğu frekans dağılımı aşırı basık olduğunda ortaya çıkarlar. Aşırı değerler seri içinde bulundukları yere göre analiz sonuçlarını etkileyebilecekleri gibi, analiz sonuçları üzerinde her hangi bir etkileri de olmayabilir. Bu nedenle bir gözlem değeri aşırı değer ise bunun nedeninin araştırılması gerekir.

Artığın pozitif veya negatif aşırı bir değere sahip olup olmadığının belirlenmesi “aşırı değerin belirlenmesi” anlamına gelmektedir.

2.7.1 Kutu Çizimleri

Aşırı değerler artıkların tahmini değerlere göre grafiği çizilerek gözlenebileceği gibi, histogram veya kutu çizimleri yapılarak, özel artıklar hesaplanarak ya da etki ölçüleri hesaplanarak da belirlenebilir.

Geleneksel tanıma göre; kartillerarası aralığın 1,5 katından fazla, üçüncü kartilin üzerinde ya da ilk kartilin altında kalan değerler aşırı değer olarak tanımlanır⁶⁸. Aşırı değerleri belirleyebilmek için verinin kutu çizimi yapılır. Bunun için kartiller arası fark,

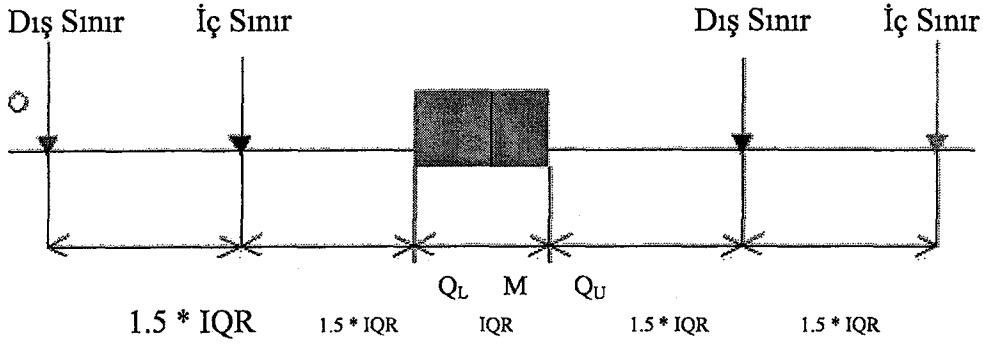
$$IQR = Q_U - Q_L$$

olarak hesaplanır. Burada Q_L ve Q_U en düşük ve en yüksek kartilleri ifade etmektedir. IQR genişlikli Q_L ve Q_U kutuları çizilir. Bu kutuların içine M medyan noktasına denk gelen yere dikey olarak bir çizgi çizilir. Q_L ’den $1,5 IQR$ uzaklıkta olan iç sınır ve $3 IQR$

⁶⁷ Lawrence, Arthur, s. 195

⁶⁸ “Outlier”, <http://mathworld.wolfram.com/Outlier.html> (15 Haziran 2005).

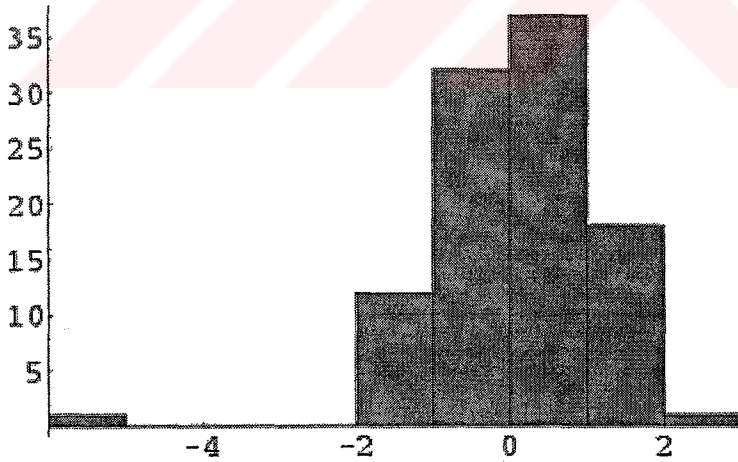
uzaklıkta olan dış sınır çizilir. İç sınır ve dış sınır arasında kalan gözlem değerleri “şüpheli aşırı değerler” olarak adlandırılır. Sınırlar dışında kalan gözlem değerleri de “çok şüpheli aşırı değerler” olarak adlandırılır⁶⁹.



Şekil 2. Kutu Çizimi

2.7.2 Histogram

Frekans serilerinde olduğu gibi sınıflı serilerin grafikleri de bir koordinat sistemi üzerine çizilebilir ve çizilen bu frekans serilerinin grafiklerine “histogram” adı verilir. Aşırı değerler çoğu zaman histogramla⁷⁰ da belirlenebilirler.



Şekil 3. Histogram

Şekil 3’ te görüldüğü gibi en sol tarafta yeralan nokta aşırı değerdir.

⁶⁹ “Methods of Detecting Outlier”, <http://www.netnam.vn/unescocourse/statistics/37.htm> (15 Haziran 2005).

⁷⁰ Turanlı, Güriş, s. 41.

2.7.3 Standartlaştırılmış Artıklar

Standart sapmasına bölünen artıklar “standartlaştırılmış artıklar” dır. Doğrusal regresyon varsayımlarına dayanmaktadır. Standartlaştırılmış artıkların, sıfır ortalama, 1 varyansa sahip olarak normal dağılıma uyması beklenir. Aşırı değerleri belirleyebilmek için bu artıklar da kullanılabilir,

$$\text{Standartlaştırılmış artıklar} = \sqrt{\frac{\hat{Y}_i - Y_i}{S_e}}$$

olarak hesaplanırlar, hesaplanan değer 2’ den büyük ya da -2’ den küçükse, bu durumda artıklar aşırı değerlidir. Örnek hacminin küçük olması durumunda aşırı değerler ortalama tahminine etki ediyorsa, mutlaka belirlenmeleri gerekmektedir⁷¹.

2.7.4 Studentlaştırılmış Artıklar

Aşırı değerleri belirleyebilmek için *Studentlaştırılmış Artıklar* da kullanılabilir: Bu artıklar t-dağılımı gibi (n-p-1) serbestlik derecesiyle dağılırlar ve bağımsız değildirler. Burada p açıklayıcı değişken sayısıdır. Bu nedenle istatistiksel olarak anlamlı olup olmadıklarının belirlenmesi için studentlaştırılmış artıkların kullanılmasını önermiştir⁷².

Artığın kovaryans matrisi,

$$\text{var}(e) = \sigma^2(1 - H)$$

şeklinde gösterilir. Burada,

$$H = X(X'X)^{-1}X'$$

olarak tanımlanan Hat matristir. i.’ci artığın varyansı ise,

⁷¹Robert A. Yaffee, Ph. D., “Robust Regression Modelling with STATA Lecture Notes”, www.bsos.umd.edu/gvpt/uslaner/robustregression.pdf (10 Haziran 2005), s. 33.

$$\text{var}(e_i) = \sigma^2(1 - h_i)$$

olarak hesaplanır. Burada h_i = leverage* istatistiğidir ve $X(X'X)^{-1}X'$ matrisinin i.'ci diagonalı** ile ifade edilir. $0 \leq h_i \leq 1$ aralığında yer alır. X de, x_{ij} açıklayıcı değişkenlerinin ilk sütunu 1 olan, $n \times (p+1)$ matrisidir.

Böylelikle e_i^s Studentlaştırılmış artık,

$$e_i^s = \frac{e_i}{\sqrt{s_{(i)}^2(1 - h_i)}}$$

olarak hesaplanır. Burada $s_{(i)}^2$ = i.'ci gözlemin atılarak kalan gözlem değerleriyle hesaplanan standart sapmadır.

Studentlaştırılmış artığın t-değeri, o gözlem değerinin aşırı değer olarak anlamlı olup olmadığını gösterecektir. Büyük örnekler için standartlaştırılmış ve studentlaştırılmış artıklar yaklaşık olarak aynı sonucu verirler⁷⁴.

2.7.5 Hat Matrisin Köşegenleri

Hat Matris bağımsız değişkenlerdeki aşırı değerler anlamına gelen leverage noktalarının belirlenmesinde kullanılır. Şu iki kavram birbirleriyle ilişkilidir; studentlaştırılmış artıklarla regresyon doğrusundan büyük sapmalar belirlenebilir. Genelde regresyon modelinin artıkları bağımsız ve özdeş dağılmadığı için artıkları kendi standart sapmaları ile ağırlıklandırmak önerilmektedir, bu işlem studentlaştırmaktır. Bu fikirden hareketle Hat Matris hesaplanır. Hat matris Y 'nin tahmini değerlere dönüştürülmesini sağlar. Hat matrisin diagonalı gözlem değerlerinin aşırı değer (leverage) olup olmadığını belirler.

* Açıklayıcı değişkenlerden birinin gözlem değerinde aşırı değerlerin olması "leverage nokta" olarak adlandırılır. Aşırı değer (Outlier) terimi genellikle bağımlı değişkenin gözlemlerinde aşırı değerlerin olması durumudur. (Birkess, Dodge, s. 20)

** Diagonal köşegen demektir. Diagonal matris de köşegenlerinde değerler olan, onun dışındaki tüm elemanları sıfır olan matristir.

⁷⁴ BIOST 515, "Lecture 7, Linear Regression Diagnostics", <http://courses.washington.edu/b515/l7.pdf> (5 Haziran), s. 9.

Hat matris regresyonun şu formülasyonundan gelir;

$$\hat{Y} = X\beta = X'(X'X)^{-1}X'Y$$

burada,

$$X'(X'X)^{-1}X'Y = \text{Hat matris} \rightarrow H$$

olarak belirlenebilir. Bu ifade,

$$\hat{Y} = H.Y$$

olarak gösterilebilir.

Hat matrisin diagonal bileşen değerleri ile aşırı değerlerin etkileri ölçülür. Aşırı değer $1/n$ ve 1 değerleriyle sınırlandırılır. 1' e yakın olan değer daha aşırı değerdir. Hat matrisin trace⁷⁵, i modeldeki değişken sayısıdır. Leverage $> 2p/n$ ise, Belsley (1980)' e göre yüksek aşırı değer (leverage)' dir.

2.7.6 Etki Ölçüleri

Bir gözlem değerinin regresyon modeli üzerindeki etkisini belirleyebilmek için, o gözlem değeri modelden atılarak meydana gelen değişiklikler analiz edilir. Bunun için etki ölçüleri geliştirilmiştir. Yaygın olarak kullanılan etki ölçüleri Cook'un D değeri, DFFIT ölçüsü ve DFBETAS ölçüsüdür.

2.7.6.1 Cook'un D Değeri

Gözlem değerlerinden biri silindiğinde, o gözlem değerinin regresyon modelindeki tahmin edilen tüm katsayılarla olan etkisini belirlemek için Cook' un D değeri hesaplanır. Cook'un D ölçüsü,

$$\text{Cook's } D_i = \left(\frac{1}{p} \right) \left(\frac{h_i}{1-h_i} \right) \left[\frac{e_i^2}{S^2(1-h_i)} \right]$$

⁷⁵ Trace; matrisin izi yani köşegenlerinin toplamıdır.

olarak hesaplanır. Burada p açıklayıcı değişken sayısı, h leverage istatistiği ve S^2 ise standart hatadır. $(p, n-p)$ serbestlik dereceli F dağılımının % 50' sini aşan D değerlerinin aşırı olduğunu ileri sürmüştür⁷⁶ Genelde $4/n$ değeri ile karşılaştırılarak karar verilmektedir. Cook'un D değeri $4/n$ değerinden büyük olursa aşırı değer olduğu kararına varılır.

2.7.6.2 DFBETAS

Aşırı gözlem değeri silinerek, o gözlem değerinin parametre tahmini üzerine olan etkisinin büyüklüğünü araştırmak amacı ile kullanılır. DFBETAS ölçüsü,

$$DFbeta_j = \frac{b_j - b_{(i)j}u_j}{\sqrt{\sum u_j^2(1-h_j)}}$$

formülü ile hesaplanır. Burada u_j ; aşırı değer atıldıktan sonra kalan gözlem değerleri ile hesaplanan artıklardır. Hesaplanan değer $2/\sqrt{n}$ ile karşılaştırılarak karar verilir. DFBETAS değeri $2/\sqrt{n}$ değerinden büyük olursa aşırı değer olduğu kararına varılır.

2.7.6.3 DFFITS

β' nın tahmini için i ' ci gözlem kullanılmadığında $\hat{Y}$ de meydana gelen değişikliği ölçmek için kullanılan DFFITS ölçüsü,

$$DFFITS_i = \frac{\hat{Y}_i - \hat{Y}_{ii}}{s_{(i)}\sqrt{h_{ii}}}$$

olarak hesaplanır. Burada $\hat{Y}_{ii}$, i ' ci gözlem değeri için tahmin edilen ortalamadır. Hesaplanan değer $2\sqrt{p/n}$ ile karşılaştırılarak karar verilir. DFFITS değeri $2\sqrt{p/n}$ inden büyük olursa aşırı değer olduğu kararına varılır.

Etki ölçülerinin formül ve karar kriterleri Tablo 4'te verilmiştir.

⁷⁶ John O. Rawlings, **Applied Regression Analysis: A Research Tool**, USA: Springer-Verlag New York, 1998, s. 362.

Tablo 4

Etki Ölçülerinin Formül ve Karar Kriterleri

Etki Ölçüsü	Formül	Karar Kriteri
Cook'un D Değeri	$Cook's D_i = \left(\frac{1}{p} \right) \left(\frac{h_i}{1-h_i} \right) \left[\frac{e_i^2}{S^2(1-h_i)} \right]$	$4/n$
DFBETAS	$DFbeta_j = \frac{b_j - b_{(i)j} u_j}{\sqrt{\sum u_j^2 (1-h_j)}}$	$2/\sqrt{n}$
DFFITS	$DFFITS_i = \frac{\hat{Y}_i - \hat{Y}_{ii}}{s_{(i)} \sqrt{h_{ii}}}$	$2\sqrt{p/n}$

Sonuç olarak bu üç etki ölçüsü karşılaştırılarak özetlendiğinde, aşırı değerlerin belirlenmesinde kullanılan bu ölçülerden Cook'un D değeri ve DFBETAS, $\hat{\beta}$ üzerindeki etkileri ölçerken, DFFITS $\hat{Y}_i$ üzerindeki etkileri ölçmektedir.

ÜÇÜNCÜ BÖLÜM

KANTİL REGRESYON

Alternatif regresyon yöntemlerinden biri de Kantil Regresyon'dur. Bu bölümde Kantil Regresyon ile ilgili açıklamalar yapmadan önce bölenlerden söz edilecektir.

3.1 Bölenler

Serileri iki, dört, on ve yüz eşit parçaya ayıran değerler genel olarak bölenler olarak adlandırılmaktadır. Seriyi iki eşit parçaya bölmek için hesaplanan değerlere medyan, dört eşit parçaya bölmek için hesaplanan değerlere kartil, on eşit parçaya bölmek için hesaplanan değerlere desil ve yüz eşit parçaya bölmek için hesaplanan değerlere santil adı verilmektedir. Bir seri için 3 kartil, 9 desil ve 99 santil hesaplanmaktadır. Seriyi dört eşit parçaya bölen kartiller birinci kartil, ikinci kartil ve üçüncü kartil olarak isimlendirilirler⁷⁷.

Seri birimlerinin $n/4$ 'ünün değeri birinci kartilden, $n/2$ 'sinin değeri ikinci kartilden ve $3n/4$ 'ünün değeri üçüncü kartilden küçük olacaktır. Birinci kartil $Q_{1/4}$, ikinci kartil $Q_{2/4}$, üçüncü kartil ise $Q_{3/4}$ şeklinde ifade edilmektedir. Serinin ikinci kartili aynı zamanda medyanıdır.

Basit serilerde kartiller medyana benzer şekilde hesaplanmaktadır. Yani seriyi ilk birimden başlayarak $n/4$ son birimden başlayarak $3n/4$ şeklinde bölen değer birinci kartil, ilk birimden başlayarak $3n/4$, son birimden başlayarak $n/4$ şeklinde bölen değer ise üçüncü kartil olmaktadır.

⁷⁷ Turanlı, Giriş, s. 100.

Frekans serilerinde kartillerin bulunabilmesi için medyanda olduğu gibi artan birikimli frekansların hesaplanması gerekir. Bu frekanslar hesaplandıktan sonra birinci kartil için seri birimlerinin $n/4$ değeri bulunarak bu değerin artan birikimli frekanslar sütununda hangi aralıkta yer aldığı belirlenir. Bu şekilde belirlenen aralığın büyük değeri karşısındaki X değeri serinin birinci kartil değerini verecektir. üçüncü kartilin hesabı için de aynı işlemler seri birimlerinin $3n/4$ değeri bulunarak yapılmaktadır.

Basit seride birinci kartil $(N+2)/4$ ' üncü terim, 3. kartil ise $(3N+2)/4$ ' üncü terimdir. $(N+2)/4$ ve $(3N+2)/4$ sayıları tam ise veya buçukla bitiyorsa medyanda olduğu gibi davranılır. Buna karşılık, tam veya buçukla bitmeyen sayılar için buçuktan küçük kusurlar atılır tam sayıya dönüştürülür. Mesela $3,25 \approx 3$ ve $3,75 \approx 4$ gibi.⁷⁸

3.2 Anakütlenin Modellenmesi

Herhangi bir dağılıma sahip örneğin yapısını göstermenin dört yolu vardır. Bunlar; kümülatif dağılım fonksiyonu, olasılık yoğunluk fonksiyonu, kantil fonksiyonu ve kantil yoğunluk fonksiyonudur.

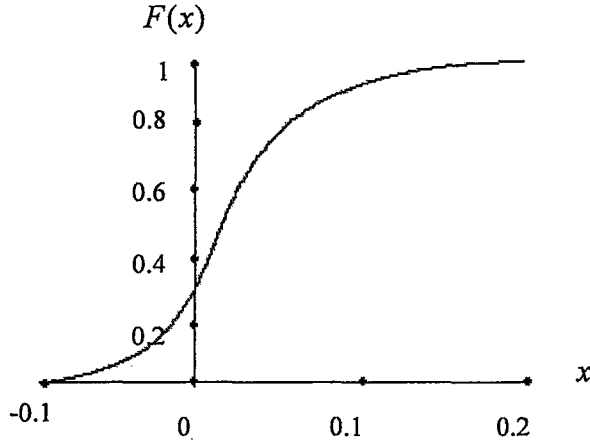
3.2.1 Kümülatif Dağılım Fonksiyonu

Kümülatif Dağılım Fonksiyonu $F(x)$ ' le gösterilir; verilen x değerinden küçük ya da eşit olan X değişkeninin olasılığı olarak tanımlanır ve

$$F(x) = \text{Olasılık}(\text{tesadüfi değişken } X \leq x)$$

olarak ifade edilebilir. Azalan fonksiyon olan kümülatif dağılım fonksiyonunun şekli,

⁷⁸ Prof.Dr. Özer Serper, Uygulamalı İstatistik I, 3. Basım, İstanbul: Filiz Kiabeve, 1996, s. 107.



Şekil 4. Kümülatif Dağılım Fonksiyonu

olacaktır. Şekilden de görülebileceği gibi, fonksiyon sol taraftan “0” olasılık değerinden başlamakta, sağ tarafta “1” olasılığına kadar devam etmektedir. Herhangi bir olasılığın (θ) örnekten elde edilen değerinin x ’ e göre çizimini verir. Kümülatif Dağılım Fonksiyonu (CDF) dağılımın tek yönlü tanımlamasını yapar⁷⁹.

3.2.2 Olasılık Yoğunluk Fonksiyonu

Bir değişkenin alabileceği değerlerle bu değerleri alma olasılıkları arasındaki bağıntıyı gösteren fonksiyona “Olasılık Fonksiyonu (PDF)” denir⁸⁰. $f(x)$ ’ le gösterilen olasılık yoğunluk fonksiyonu,

$$f(x)dx = \text{Olasılık}(x \leq X \text{ tesadüfi değişkeni} \leq x + dx),$$

olarak tanımlanabilir. Burada dx , x ’ in sonsuza doğru küçük aralığıdır. $f(x)$ eğrisinin altındaki alan, herhangi gözlenen değer toplam olasılığı, 1 olmalıdır. Kümülatif Dağılım Fonksiyonu ve Olasılık Fonksiyonu arasındaki ilişki,

$$f(x)dx = F(x + dx) - F(x) = dF(x)$$

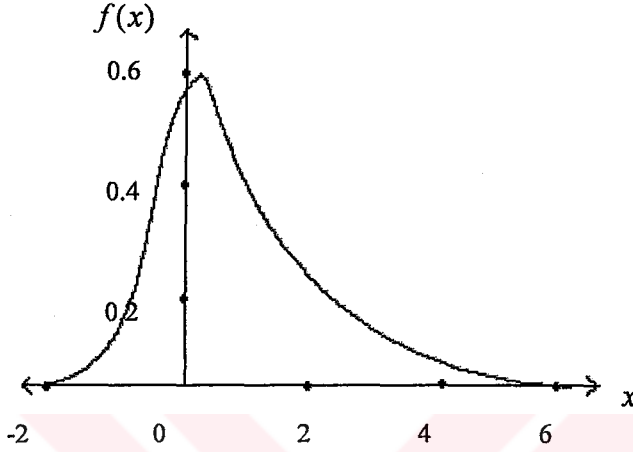
olacaktır. Olasılık Fonksiyonu, Kümülatif Dağılım Fonksiyonunun türevine eşittir ve

⁷⁹ Warren G. Gilchrist, *Statistical Modelling with Quantile Functions*, Florida: Chapman & Hall/CRC, 2000, s. 10.

⁸⁰ Doç. Dr. Bedriye Saraçoğlu, Y. Doç. Dr. Ferhan Çevik, *Matematiksel İstatistik*, Ankara: Gazi Büro Kitabevi, 1995, s. 67.

$$f(x) = dF / dx$$

olarak elde edilir ve



Şekil 5. Olasılık Yoğunluk Fonksiyonu

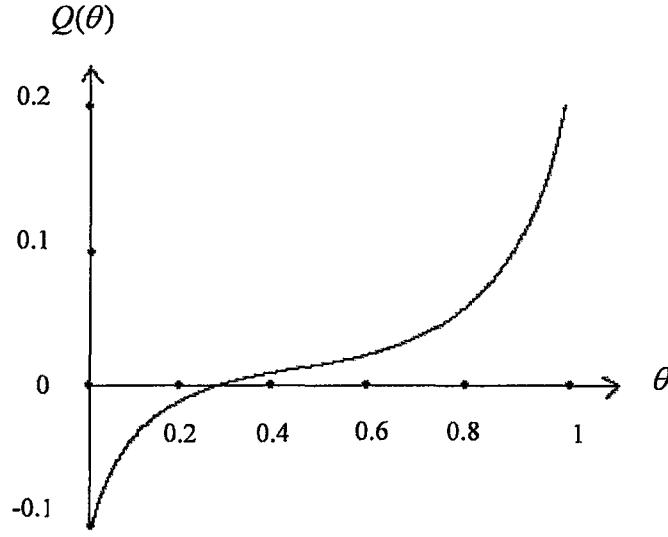
fonksiyonun çizimi Şekil 5 ‘ teki gibi olacaktır.

3.2.3 Kantil Fonksiyonu

Kantil Fonksiyonu QF , $Q(\theta)$ ile gösterilir. Dağılımı tarif etmenin üçüncü yoludur.

Kantil değeri değişkenin dağılımında yer alan ve dağılımı, kendisinden büyük olanlar ve kendisinden küçük olanlar diye ikiye bölen herhangi bir değerdir. Şöyle ki, değerlerin $\% \theta$ ’ sı, θ ’ cı kantilden daha küçüktür. (θ olasılık değerini ifade etmektedir.)

$$x_\theta = (X \leq x_\theta) \text{ olasılığı için } x' \text{ in değeridir} = \theta.$$



Şekil 6. Kantil Fonksiyonu

x_θ 'nin değeri, anakütlenin θ .' cı kantili olarak adlandırılır. $x_\theta = Q(\theta)$ fonksiyonu, θ 'cı kantili, θ 'nın bir fonksiyonu olarak ifade edilir ve kantil fonksiyon olarak adlandırılır. QF ve CDF, herhangi (x, θ) çifti için $x = Q(\theta)$ ve $\theta = F(x)$ şeklinde yazılabilir. Bu fonksiyonlar birbirlerinin tersine eşittir ve sürekli artan fonksiyonlardır. Böylelikle $Q(\theta) = F^{-1}(\theta)$ ve $F(x) = Q^{-1}(x)$ şeklinde de gösterilebilir⁸¹.

$Q(\theta)$ kantil fonsiyonuysa, θ 'nın tüm olasılıkları için, $0 \leq \theta \leq 1$, kantil değerlerini verir. Medyan da $Q(0,5)$ ' tir. Benzer şekilde $Q(1/4)$ ve $Q(3/4)$ kartillerdir. Hesaplamalar için normal dağılım tablosundan yararlanılır, örneğin 1,96 gibi bir değer 0,975 olasılığını alır. $N(\theta)$ standart normal dağılım için kantil fonksiyonuysa $N(0,975)$ değeri 1,96' ya eşittir, böylelikle kullanılan normal tablolar standart normal dağılım için kantil fonksiyon tablolarıdır.

⁸¹ Gilchrist, s. 13.

Dağılımları modelleyebilmek için kantil fonksiyon kullanılabilir. x verilmişken y ' nin θ ' cı kantili,

$$Q_y(\theta/x) = \beta x + \eta S(\theta)$$

olarak gösterilir. $\eta S(\theta)$ artık terimi ifade etmektedir. $S(\theta)$ simetrik olması gerekmeyen kantil fonksiyonudur, η ölçek parametresidir. “ y ” nin x üzerindeki kantil regresyon fonksiyonu” ya da “şartlı kantil fonksiyonu” olarak adlandırılır⁸².

3.2.4 Kantil Yoğunluk Fonksiyonu

Dağılımları modelleyebilmek için CDF’ nun türevini alarak PDF elde edildiği gibi, QF ’ in de türevi alınarak Kantil Yoğunluk Fonksiyonu (QDF) belirlenebilir ve,

$$q(\theta) = dQ(\theta) / d\theta$$

olarak gösterilir. $Q(\theta)$ azalmayan* bir fonksiyon olduğu için eğimi $q(\theta)$ negatif değildir, her zaman $0 \leq \theta \leq 1$ birim aralığında yer alır, olasılık yoğunluk fonksiyonu $f(x)$ ise sonsuz tanım aralığında yer alır.

Serinin mod değerinin olasılığı $p - \text{mod} \geq 0,5$ ise, dağılım sola çarpıktır ve $q(\theta)$ kantil yoğunluk fonksiyonu $q(\theta) \leq q(1-\theta)$ durumunu sağlar, $0 \leq \theta \leq 0,5$ ’ tir. Kantil fonksiyonu da $Q(\theta) + Q(1-\theta) \leq 2Q(0,5)$ durumunu sağlar ve,

$$\text{ortalama} \leq \text{medyan} \leq \text{mod}$$

sıralaması sağlanır.

⁸² Emanuel Parzen, “Nonparametric Statistical Data Modeling”, Journal of the American Statistical Association, Vol. 74, No. 365 (Mar., 1979), s. 105.

* Bir $f(x)$ fonksiyonu azalmayan ise ,tüm $b > a$ için, $f(b) \geq f(a)$ olur, $a, b \in I$, tersi durumda bir $f(x)$ fonksiyonu artmayan ise tüm $b > a$ için, $f(b) \leq f(a)$ olur, $a, b \in I$.
(<http://mathworld.wolfram.com/NondecreasingFunction.html>)

için genelleştirelim. Burada x_i' , bağımsız değişken vektörüdür. e_i bağımsız, sıfır etrafında simetrik ve F dağılımına sahiptir. Bu durumda θ ' cı kantil regresyon ($0 < \theta < 1$),

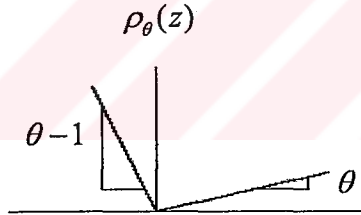
$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i: y_i \geq x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{i: y_i < x_i' \beta} (1 - \theta) |y_i - x_i' \beta| \right\}$$

$$\min_b \frac{1}{n} \sum_{i=1}^n \rho_{\theta}(y_i - x_i' \beta)$$

minimizasyon ile tahmin edilir ve y ' nin θ ' cı kantili olarak da adlandırılabilir. Kantil regresyonun bu şekildeki gösterimi doğrusal programlama gösterimidir. Burada I , karakteristik fonksiyondur, ρ_{θ} ise Check Fonksiyonudur ve,

$$\rho_{\theta}(z) = z(\theta - I(z \leq 0))$$

olarak tanımlanır. Bu fonksiyon,



Şekil 7. ρ_{θ} Fonksiyonu

Şekil 7' deki gibi gösterilebilir⁸⁵. $\theta=0,5$ olması durumunda kantil regresyon amaç fonksiyonu LAD amaç fonksiyonuna ve l_1 regresyona eşittir⁸⁶. Kantil regresyon amaç fonksiyonu mutlak sapmaların ağırlıklandırılmış toplamıdır.

⁸⁵ Roger Koenker, And Kevin F. Hallock, "Quantile Regression an Introduction", The Journal of Economic Perspectives, Vol.15, No.4 (Autumn 2001), s. 143

⁸⁶ S. Lee, Lecture Notes for MECT1, "Quantile Regression", <http://www.homepages.ucl.ac.uk/~uctplso/Teaching/MECT/lecture8.pdf> (10 Aralık 2004), s. 1.

Benzer şekilde serinin mod değerinin olasılığı $p - \text{mod} \leq 0,5$ ise, dağılım sağa çarpıktır ve $q(\theta)$ kantil yoğunluk fonksiyonu $q(\theta) \geq q(1-\theta)$ durumunu sağlar, $0 \leq \theta \leq 0,5$ ' tir. Kantil fonksiyonu da $Q(\theta) + Q(1-\theta) \geq 2Q(0,5)$ durumunu sağlar ve,

$$\text{ortalama} \geq \text{medyan} \geq \text{mod}$$

sıralaması sağlanır⁸³.

3.3 Kantil Regresyon

Ücretlerdeki ve gelirdeki eşitsizlik gibi dağılımın bozulduğu konuların incelenmesinde ekonomide yaygın olarak kullanılan Kantil Regresyon Modelleri şartlı ortalama fonksiyonları ve şartlı kantil fonksiyonları için tahmin yapılmasında kullanılır. Kantil Regresyon Medyan Regresyon'un belirlenen kantiller için genelleştirilmiş halidir. Bu regresyon modelleri uç değerlere ve eğikliğe En Küçük Kareler Yöntemi'nden daha az hassastır.

Kantil Regresyon Modeli aslında bir yerleşim modelidir. Basit yerleştirme modeli,

$$Y_i = \beta + e_i$$

olarak ifade edildiğinde, burada yer alan Y_i , simetrik F dağılım fonksiyonuna sahip, bağımsız, özdeş dağılımlı, β medyanlı tesadüfi değişkendir. Bu modelde θ .' cı örnek kantili,

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i: y_i \geq \beta} \theta |y_i - \beta| + \sum_{i: y_i < \beta} (1-\theta) |y_i - \beta| \right\}$$

ifadesinin minimizasyonu ile elde edilir⁸⁴. Bunu doğrusal regresyon modeli,

$$y_i = x_i' \beta + e_i$$

⁸³ J. Tiago de Oliveira, Benjamin Epstein, *Some Recent Advances in Statistics*, London: Academic Press, 1982, s. 35.

⁸⁴ Judge, s. 834.

θ .' cı kantil regresyon gözlem değerlerinin işaretlerine dayalı olarak,

$$\theta \min_{\beta} \frac{1}{n} \sum_{i=1}^n (\theta - 1/2 + 1/2 \operatorname{sgn}(y_i - x_i' \beta))(y_i - x_i' \beta)$$

şeklinde tahmin edilir. Burada $\operatorname{sgn}(a)$, a 'nın işaretidir ve a pozitif ise 1, negatif veya sıfır ise -1 değerini alır. Tahminlerin bu şekilde, yani gözlem değerlerinin büyüklüğü yerine gözlem değerlerinin işaretlerine dayalı olması, Kantil Regresyon'un robust bir yöntem olmasını sağlamaktadır⁸⁷. Minimizasyon için birinci mertebe koşulunun sağlanması gerekir. Birinci Mertebe Koşulu'nun⁸⁸ $K \times 1$ vektörü,

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^n (\theta - 1/2 + 1/2 \operatorname{sgn}(y_i - x_i' \beta))x_i = 0$$

olarak gösterilir. Bu ifade, Birinci Mertebe Koşulu Genelleştirilmiş Momentler Yöntemi(GMM)'ne uyan bir moment fonksiyonudur. Moment fonksiyonu,

$$\psi(x_i, y_i, \beta) = (\theta - 1/2 + 1/2 \operatorname{sgn}(y_i - x_i' \beta))x_i$$

olarak tanımlanabilir. $\psi(\cdot)$ 'nin moment fonksiyon olarak geçerli olabilmesi için belirli düzenleme şartları (C.R.C.)⁸⁹ altında,

$$E[\psi(x_i, y_i, \beta_\theta)] = 0$$

⁸⁷ Indermit S. Gill, **Crafting Labor Policy: Techniques and Lessons from Latin America**, Washington: World Bank, 2002, s. 249.

⁸⁸ $g(x, y) = c$ kısıtı altında $\max_{x,y} f(x, y)$ problemi çözülmek istensin, f artan bir fonksiyondur dolayısıyla, $f(x, y) = k$ ise, $k \leq k$ olur.

$1 = f_1(x^*, y^*)/g_1(x^*, y^*) = f_2(x^*, y^*)/g_2(x^*, y^*)$ tanımlanarak $((x^*, y^*))$ çözümü ifade etmektedir.) problem iki denklem halinde şöyle yazılabilir:

$f_1'(x^*, y^*) - \lambda g_1'(x^*, y^*) = 0$

$f_2'(x^*, y^*) - \lambda g_2'(x^*, y^*) = 0.$

Kısıt da modele eklenir,

$c - g(x^*, y^*) = 0$

İlk iki denklem Lagrange Fonksiyonunun türevi şeklinde gösterilebilir,

$L(x, y) = f(x, y) - \lambda(g(x, y) - c)$

Bu birinci mertebe koşulu olarak adlandırılır.

(<http://www.chass.utoronto.ca/~osborne/MathTutorial/MEN.HTM>) (10 Mayıs 2005)

⁸⁹ David Pollard, "Chapter 2: Classical Regularity Conditions",

<http://www.stat.yale.edu/~pollard/Asymptopia/Classical.pdf>, (5 Mayıs 2005)

olması gerekir. Genelleştirilmiş Momentler Yöntemi kullanılarak elde edilen parametre tahmincileri tutarlı ve asimtotik olarak normal olacaktır.

Belirli düzenleme şartları altında,

$$\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \xrightarrow{L} N(0, \Lambda_\theta)$$

olarak gösterilebilir. Burada,

$$\Lambda_\theta = \theta(1-\theta) \left(E \left[f_{u_\theta}(0/x_i) x_i x_i' \right] \right)^{-1} E \left[x_i x_i' \right] \left(E \left[f_{u_\theta}(0/x_i) x_i x_i' \right] \right)^{-1}$$

olarak tanımlanır⁹⁰. Olasılık değeri “1” olduğunda ve $f_{u_\theta}(0/x) = f_{u_\theta}(0)$ ise, yani hata teriminin yoğunluğu sıfır etrafındaysa ve x ’ ten bağımsızsa, Λ_θ ,

$$\Lambda_\theta = \frac{\theta(1-\theta)}{f_{u_\theta}^2(0)} \left(E \left[x_i x_i' \right] \right)^{-1}$$

şeklinde sadeleştirilebilir⁹¹. $f_{u_\theta}(. / x)$ x ’ ten bağımsız olduğunda, tüm kantillerin parametre vektörleri sadece kesim noktalarında farklılık gösterir.

Kantil katsayılarını yorumlayabilmek için, y ’ nin k açıklayıcı değişkenine göre şartlı kantilinin kısmi türevi alınmaktadır. Türev alındığında,

$$\partial Quant_\theta(y_i / x_i) / \partial x_{ik}$$

olacaktır. Bu türev, x ’ in k ’ ci değerindeki marjinal değişime göre, θ ’ cı şartlı kantildeki marjinal değişimi vermektedir.

⁹⁰ Asimtotik dağılımın genel gösterimi şu şekilde yapılabilir.

$$E \left[f_{u_\theta}(0/x_i) x_i x_i' \right] = \partial E \left[\psi(x_i, y_i, \beta_\theta) \right] / \partial \beta_\theta'$$

$$\theta(1-\theta) E \left[x_i x_i' \right] = E \left[\psi(x_i, y_i, \beta_\theta) \psi(x_i, y_i, \beta_\theta)' \right]$$

⁹¹ Roger Koenker, Gilbert Bassett, “Regression Quantiles”, *Econometrica*, Vol.46, No.1, (Jan., 1978), s.48.

3.4 Kantil Regresyonun Özellikleri

1) EKK ve Medyan Regresyon y' nin şartlı dağılımının ortası hakkında bilgi vermekte, Kantil Regresyon ise farklı kantil değerleri için y' nin x' e göre şartlı dağılımının tümü hakkında bilgi vermektedir.

2) Kantil Regresyon' da; $\min_b \frac{1}{n} \sum_{i=1}^n \rho_\theta(y_i - x_i' b)$ minimizasyonu, doğrusal programlama (LP) gösterimidir, bu durum tahmini kolaylaştırır.

3) Kantiller monoton dönüşümlere olanak verirler⁹². Herhangi $h(\cdot)$ monoton fonksiyonu için $Q_{h(y)/x}(\tau/x) = h(\theta_{y/x}(\tau/x))$ olur*.

4) Kantiller y' deki aşırı değerlere karşı kararlıdır (robust).

5) Hata terimi normal dağılmadığında, kantil regresyon tahmincileri EKK Tahmincilerinden çok daha etkin olabilir.

6) Kantil Regresyon değişen varyansın belirlenmesine imkan verir.

7) Kantil Regresyon amaç fonksiyonu için tahmin edilen katsayı vektörü bağımlı değişkendeki aşırı değerlere duyarlı değildir ve yerleşimin robust bir ölçüsüdür.

8) Farklı kantillerde farklı sonuçların ortaya çıkması, bağımlı değişkenin şartlı dağılımının farklı noktalarındaki açıklayıcı değişkenlerdeki değişikliklere farklı tepki vermesi olarak yorumlanabilir.

9) L-tahmincileri, kantil tahmincilerin doğrusal kombinasyonuna dayanır⁹³. İlk bölümde yeralan L-tahmincileri, $Y_{(1)}, Y_{(2)}, \dots, Y_{(T)}$ sıra istatistiklerinin doğrusal

⁹² S. Lee, s. 2.

* Monoton dönüşüm, bir serinin logaritmik şartlı medyanının, o serinin şartlı medyanının logaritmasına eşit olmasıdır.

⁹³ Stephen Portnoy, And Roger Koenker, "Adaptive L Estimation for Linear Models", *The Annals of Statistics*, Vol.17, No.1 (Mar., 1989), s. 362.

kombinasyonları olan tahmincilerdi. Bu tahminciler doğrusal model için genelleştirilmiş, medyan regresyon da diğer kantiller için geliştirilmiştir.

3.5 Kantil Regresyonun Doğrusal Programlama Gösterimi

Kantil Regresyon' un doğrusal programlama gösterimi sonlu sayıdaki simpleks iterasyonlarla tahminin belirleneceğini ifade eder. İterasyon sayısı doğrusal programlama algoritmasına göre küçüktür⁹⁴. EKK Regresyonu' ndan farklı olarak parametre vektör tahmini aşırı değerlere karşı robust 'tır.

Daha öncede söz ettiğimiz gibi y_i ' nin θ_i ' cı kantili,

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i: y_i \geq x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{i: y_i < x_i' \beta} (1 - \theta) |y_i - x_i' \beta| \right\} = \min_{\beta} \frac{1}{n} \sum_{i=1}^n \rho_{\theta}(u_{\theta_i})$$

ifadesinin minimizasyonu ile elde edilmekteydi. Bunu doğrusal programlama gösterimi olarak ifade edebilmek için y_i sadece pozitif elemanların bir fonksiyonu olarak,

$$y_i = \sum_{j=1}^K x_{ij} \beta_{\theta_j} + u_{\theta_i} = \sum_{j=1}^K x_{ij} (\beta_{\theta_j}^1 - \beta_{\theta_j}^2) + (\varepsilon_{\theta_i} - v_{\theta_i})$$

$$\beta_{\theta_j}^1 \geq 0, \quad \beta_{\theta_j}^2 \geq 0 \quad (j=1, \dots, K) \quad \text{ve} \quad \varepsilon_{\theta_i} \geq 0, \quad v_{\theta_i} \geq 0 \quad (i=1, \dots, n)$$

şeklinde yazılabilir. İlk denklem matris gösterimiyle yazıldığında doğrusal programlamanın primal problemine* dönüşür ve

$$Az = y \quad \text{kısıtı altında} \quad \min_i c'z$$

⁹⁴ I.Barrodale, "L₁ Approximation and the Analyses of Data", Applied Statistics, Vol.17, No.1(1968),s.51.

* Her doğrusal programlama probleminin ilişkili olduğu bir ikiz problemi vardır. Herhangi bir doğrusal programlama problemi veya modeli primal veya asıl problem veya model olarak adlandırılırken diğerine yani ikizine dual veya ikilik adı verilir. Problemlerin birbirine olan yakın ilişkisi göz önüne alındığında hangi problemin primal hangisinin dual olarak adlandırıldığı fazla bir şey fark ettirmez. Çünkü birinin optimal çözümü aynı zamanda ötekinin optimal çözümüdür.

olarak ifade edilir. Burada $A = (X, -X, I_n, -I_n)$, $y = (y_1, \dots, y_n)'$, $z = (\beta^1, \beta^2, u, v)'$, $c = (0', 0', \theta I', (1-\theta)I')$, $X = (x_1, \dots, x_n)'$, I_n ; n boyutlu birim matristir. $0'$; sıfırların $K \times 1$ vektörüdür, I ; ise 1'lerin $n \times 1$ vektörüdür. Doğrusal programlamanın dual problemi yaklaşık olarak daha önce,

$$\frac{1}{n} \min_{\beta} \sum_{i=1}^n (\theta - 1/2 + 1/2 \operatorname{sgn}(y_i - x_i' \hat{\beta}_{\theta})) x_i = 0$$

gösteriminde belirlenen birinci mertebe koşuluyla aynıdır ve

$$w'A \leq c' \quad \text{kasıtı altında} \quad \max_w w'y$$

olarak gösterilir. Duallik Teoremi; X matrisinin tüm sütunlarının sıralı olması durumunda hem primal hem de dual problemlerin çözümünün mümkün olduğunu ileri sürmektedir. Doğrusal programlamanın denge teoremi ancak bu durumda çözümün optimal olduğunu ifade eder.

Doğrusal programlama problemini çözmek için birçok algoritma öne sürülmüştür. Bunlardan en çok ilgi göreni $\theta = 1/2$ olan medyan regresyon için geliştirilmiştir⁹⁵. Bu algoritma küçük değişikliklerle herhangi bir kantil regresyonuna da uyarlanabilir⁹⁶. Söz konusu algoritmanın en büyük avantajı, bilinen diğer algoritmalara göre anlamlı bir şekilde simpleks dönüşümlerin sayısını azaltmasıdır.

$\hat{\beta}(\theta, y + X\gamma, X) = \hat{\beta}(\theta, y, X) + \gamma$, $\gamma \in R^K$ eşvaryans özelliği ile doğrusal programlama algoritması hesaplanır. $\hat{\beta}_{\theta}$ 'nin ilk değeri $\hat{\beta}_{\theta}^0$ biliniyorsa, gözlemleri düzlemin sağ tarafına yerleştirerek hesaplama süresini kısaltmak için kullanılabilir. $y_R = y - X\hat{\beta}_{\theta}^0$ olsun, $\hat{\beta}_{\theta}^R$, y_R 'nin x 'e göre kantil regresyon tahminidir. $\hat{\beta}(\theta, y + X\gamma, X) = \hat{\beta}(\theta, y, X) + \gamma$, $\gamma \in R^K$ eşvaryans özelliğini kullanarak,

⁹⁵ I. Barrodale, And F.D.K. Roberts, "An Improved Algorithm for Discrete L_1 Linear Approximation", *SIAM Journal on Numerical Analysis*, Vol.10, No.5 (Oct., 1973), s. 839.

⁹⁶ Roger W. Koenker, And Vasco D' Orey, "Algorithm AS 229: Computing Regression Quantiles", *Applied Statistics*, Vol.36, No.3 (1987), s. 383.

$\hat{\beta}_\theta = \hat{\beta}_\theta^R + \hat{\beta}_\theta^0$ bulunur. $\hat{\beta}_\theta^R$ ve $\hat{\beta}_\theta^0$ ı belirlemek, doğrudan $\hat{\beta}_\theta$ ' yı belirlemekten çok daha hızlıdır.

Mümkün ilk değer, sabitin, $\hat{e}_1, \dots, \hat{e}_n$ artıklarının $[n\theta]$.' cı sıra istatistiğiyle değiştirildiği, düzeltilmiş EKK tahminidir. Alternatif ilk değer, gözlem değerlerinin çok büyük olması durumunda gözlem değerlerinin bir kısmına uygulanan kantil regresyonla belirlenebilir.

3.5.1 Asimtotik Kovaryans Matris Tahmini

Kantil Regresyon modellerinde kovaryans matrisinin tahmini özellikle sabit varyans ve simetriklik varsayımlarının incelenmesi nedeni ile önem taşımaktadır. $\hat{\beta}_\theta$ için asimtotik kovaryans matris formülleri $f_{u_\theta}(0/x)$ ' in iki alternatif varsayımı atında,

$$\Lambda_\theta = \theta(1-\theta) \left(E \left[f_{u_\theta}(0/x_i) x_i x_i' \right] \right)^{-1} E \left[x_i x_i' \right] \left(E \left[f_{u_\theta}(0/x_i) x_i x_i' \right] \right)^{-1}$$

$$\Lambda_\theta = \frac{\theta(1-\theta)}{f_{u_\theta}^2(0)} \left(E \left[x_i x_i' \right] \right)^{-1}$$

denklemlerinden elde edilir. Λ_θ ' nın tahmini araştırmacının tercihinine göre keyfi olarak belirlenir, her yaklaşımın avantajları ve dezavantajları vardır.

Kovaryans matrisinin farklı tahminçiler kullanılarak tahmin edilmesi mümkündür. Bu tahminçilerden en çok kullanılanları Sıra İstatistiği Tahminçisi, Bootstrap Tahminçisi ve Kernel Tahminçisidir.

3.5.1.1 Sıra İstatistiği Tahminçisi

Kantil regresyonda kovaryans matrisinin tahmininde kullanılan sıra istatistiği tahminçisi, $f_{u_\theta}(0/x) = f_{u_\theta}(0)$ olması durumunda geçerlidir. Bu varsayım altında asimtotik kovaryans matrisi,

$$\Lambda_\theta = \sigma_\theta^2 E \left[x_i x_i' \right]^{-1}$$

$$\Lambda_{\theta} = \frac{\theta(1-\theta)}{f_{u_{\theta}}^2(0)}$$

olarak sadeleştirilebilir. $\hat{E}(xx') = 1/N \sum_{i=1}^n x_i x_i'$ ile tahmin edilir. İlk terim $\sigma_{\theta}^2, \hat{u}_{\theta_1}, \dots, \hat{u}_{\theta_n}$, in $[n\theta]$.' cı sıralı istatistiğinden oluşan bir güven aralığından belirlenir. Genelde, bir tesadüfi değişkenin θ .' cı kantili için kesin güven aralığı hesaplanabilir $Y \sim F_Y(\cdot)$.

Özellikle,

$$\Pr(y_{(j)} \leq \xi_{\theta} \leq y_{(k)}) = \Pr(y_{(j)} \leq \xi_{\theta}) - \Pr(y_{(k)} < \xi_{\theta})$$

$y_{(j)}$ ve $y_{(k)}, y_1, \dots, y_n$ ' in j.' ci ve k.' cı sıra istatistikleridir.

$$\Pr(y_{(j)} \leq \xi_{\theta}) = \Pr(j \text{ ya da daha fazla gözlem} \leq \xi_{\theta}) = \sum_{i=j}^n \binom{n}{i} \theta^i (1-\theta)^{n-i}.$$

Benzer şekilde,

$$\Pr(y_{(j)} \leq \xi_{\theta}) = \sum_{i=k}^n \binom{n}{i} \theta^i (1-\theta)^{n-i}$$

yazılabilir. İkinci ve üçüncü denklem birinci denklemde yerine konulursa,

$$\Pr(y_{(j)} \leq \xi_{\theta} \leq y_{(k)}) = \sum_{i=j}^{k-1} \binom{n}{i} \theta^i (1-\theta)^{n-i}$$

olur. ξ_{θ} için $(1-\alpha)$ düzeyinde simetrik bir güven aralığı oluşturmak daha doğru olur⁹⁷.

$j = [n\theta - l], \quad k = (n\theta + l), \quad X \sim B(n, \theta)$ olsun.

$$\Pr(y_{[n\theta-l]} \leq \xi_{\theta} \leq y_{[n\theta+l]}) = \Pr([n\theta - l] \leq X < [n\theta + l])$$

⁹⁷ Moshe Buchinsky, "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research", *The Journal of Human Resources*, Vol.33, No.1 (Winter, 1998), s. 88.

$$\approx \left(\left| \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}} \right| \leq \frac{1}{\sqrt{n\theta(1-\theta)}} \right)$$

$X - n\theta / \sqrt{n\theta(1-\theta)} \xrightarrow{L} N(0,1)$ olduğundan, en son denklemdeki olasılık

$$(1-\alpha) \text{ alınır, } l = Z_{1-\alpha/2} \sqrt{n\theta(1-\theta)} \text{ olacaktır. } \Pr(y_{(j)} \leq \xi_\theta \leq y_{(k)}) = \sum_{i=j}^{k-1} \binom{n}{i} \theta^i (1-\theta)^{n-i},$$

teki kesin güven aralığının uzunluğunu bu asimtotik normal güven aralığı ile birleştirilmesi ile σ_θ^2 'nin tahmini,

$$\hat{\sigma}_\theta^2 = \frac{n(y_{(n\theta+l)} - y_{(n\theta-l)})^2}{4Z_{1-\alpha/2}^2}$$

olacaktır. Aslında $\hat{f}_{u_\theta}(0) = 2Z_{1-\alpha/2} \sqrt{\theta(1-\theta)} / \sqrt{n} (y_{(n\theta+l)} - y_{(n\theta-l)})$ ile $f_{u_\theta}(0)$ tahmincisi belirlenebilir. Bu kantil regresyonun kovaryans matris tahmini için kullanılabilir.

3.5.1.2 Bootstrap Tahmincileri

Bootstrap Yöntemi, $\hat{\beta}_\theta$ 'nin asimtotik kovaryans matrisinin şekline göre üç farklı yolla hesaplanır⁹⁸. Bunlar Design Matris Bootstrap Tahmincisi, Hata Bootstrap Tahmincisi ve Sigma Tahmincisidir.

Genel şartlar altında asimtotik matrisin tutarlı tahmincisini sağlayan “design matris bootstrap tahmincisi” için bootstrap örnekleri x ve y , F_{nxy} 'nin deneysel bileşik dağılımından çekilir. F_{nxy} dağılımından tesadüfi olarak çekilmiş örnek $(y_i^*, x_i^*), i = 1, \dots, n$ olduğunda,

$$y_i = x_i' \beta_\theta + u_{\theta i}$$

için $Quant_\theta(y_i / x_i) = x_i' \beta_\theta$ olacaktır. Burada, $Quant_\theta(y_i / x_i)$ y_i 'nin şartlı kantilidir. Bu ilişkiden yararlanarak,

⁹⁸ B. Efron, “Bootstrap Methods: Another Look at the Jackknife” The Annals of Statistics, Vol.7, No.1 (Jan., 1979), s.1.

$$y^* = x^* \beta_\theta + u_\theta^*, \quad y^* = (y_1^*, \dots, y_n^*)' \text{ ve } X^* = (x_1^*, \dots, x_n^*)'$$

şeklinde yazılabilir.

$\hat{\beta}_\theta^*$, y^* , x^* üzerindeki kantil regresyonundan belirlenen bootstrap tahminini gösterir, $\hat{\beta}_{\theta 1}^*, \dots, \hat{\beta}_{\theta B}^*$ bootstrap tahminlerini elde etmek için B kere tekrarlanabilir. Λ_θ 'nin bootstrap tahmincisi şu şekilde verilir;

$$\hat{\Lambda}_\theta^{DMB} = \frac{n}{B} \sum_{j=1}^B (\hat{\beta}_{\theta j}^* - \bar{\beta}_\theta^*)'$$

$\bar{\beta}_\theta^* = 1/B \sum_{j=1}^B \hat{\beta}_{\theta j}^*$, $\bar{\beta}_\theta^*$ yerine esas değer olarak $\hat{\beta}_\theta$ da kullanılabilir. Bu daha önce;

$\Lambda_\theta = \theta(1-\theta) \left(E[f_{u_\theta}(0/x_i)x_i x_i'] \right)^{-1} E[x_i x_i'] \left(E[f_{u_\theta}(0/x_i)x_i x_i'] \right)^{-1}$ ' de verilen $\hat{\beta}_\theta$ 'nin asimtotik kovaryansının sürekli tahmincisidir. $\sqrt{n}(\hat{\beta}_\theta^* - \hat{\beta}_\theta)$ 'nin şartlı dağılımı $\sqrt{n}(\hat{\beta}_\theta^* - \beta_\theta)$ 'nin şartlı olmayan dağılımına zayıf olarak yaklaşır⁹⁹.

İkinci tahminci olan, bağımsızlık varsayımı altında sürekli tahminciyi veren “hata bootstrap tahmincisi” için x, F_{nx} ve $u_\theta, F_{\hat{u}_\theta}$ 'nin deneysel bileşik dağılımından çekilir*.

Bağımsızlık varsayımı altında F_{nx} ve $F_{\hat{u}_\theta}$ marjinal dağılımlarından yeniden örnekleme yaparak bootstrap tahmin yöntemi uygulanabilir. $F_{\hat{u}_\theta}$ dağılımından tesadüfi olarak çekilen n birimlik örnek $u_\theta^* = (u_{\theta 1}^*, \dots, u_{\theta n}^*)'$ ve F_{nx} dağılımından tesadüfi olarak çekilen örnek $X^* = (x_1^*, \dots, x_n^*)'$ olarak belirlendiğinde, kantil regresyon modeli,

⁹⁹ Peter J. Bickel, And David A. Freedman, “Some Asymptotic Theory for the Bootstrap”, The Annals of the Statistics, Vol. 9, No.6 (Nov., 1981), s. 1196.

* $F_{\hat{u}_\theta}$ 'nin deneysel bileşik dağılımı $u_{\theta i} = y_i - x_i' \hat{\beta}_\theta$, ($i = 1, \dots, n$)'e dayanır.

$$y^* = x^* \hat{\beta}_\theta + u_\theta^*$$

olacaktır. Bu kantil regresyoun bootstrap tahmircisi $\hat{\beta}_\theta^*$ ile çözmeye olanak sağlar. Design Matris Bootstrapping Yöntemi' nde olduğu gibi, B bootstrap tahmircilerini elde edebilmek için $\hat{\beta}_{\theta_j}^* (j = 1, \dots, B)$, B kere tekrarlanır. Λ_θ ' nın tahmircisi ,

$$\hat{\Lambda}_\theta^{DMB} = \frac{n}{B} \sum_{j=1}^B (\hat{\beta}_{\theta_j}^* - \bar{\beta}_\theta^*)(\hat{\beta}_{\theta_j}^* - \bar{\beta}_\theta^*)'$$

olarak design matris bootstrap yöntemiyle aynı şekilde belirlenir.

Hata Bootstrap Tahmircisi, bağımsızlık varsayımı altında Λ_θ ' nın sürekli tahmircisidir. Bağımsızlık varsayımı geçerli olmağında, yeniden örnekleme u_θ ile x arasında varolabilecek bir ilişkiyi yok etmekte ve bu yöntemi geçersiz hale getirmektedir. Design matris tahmircisi yerine, hata bootstrap tahmircisi kullanmanın avantajı yoktur, ikisi de bağımsızlık varsayımı altında aynı hesaplama süresiyle elde edilir. Küçük örneklerde bağımsızlık varsayımı sağlanıyorsa hata bootstrap tahmircisinin kullanılması daha iyi olabilir.

Üçüncü bootstrap tahmircisi de “Sigma Tahmircisi” dir. Bu tahmircide, kovaryans matrisinin sadece bir kısmı bootstrap tekniğiyle tahmin edilir, $\hat{\sigma}_\theta^2$. Bu tahmirci de bağımsızlık varsayımına ve daha önce verilen,

$$\Lambda_\theta = \frac{\theta(1-\theta)}{f_{u_\theta}^2(0)} \left(E[x_i x_i'] \right)^{-1}$$

olarak tanımlanan asimtotik kovaryans matrisinin şekline dayanır. Bu yöntem, bootstrap yöntemi kullanarak, Λ_θ gösterimini σ_θ^2 ' nin parametrik olmayan tahminiyle birleştirir. σ_θ^2 ' nin bir tahmircisi,

$$\hat{\sigma}_{\theta_D}^2 = \frac{n}{B} \sum_{j=1}^B (q_{\theta_j}^* - \hat{q}_\theta^*)^2$$

olarak belirlenir. Burada $\hat{q}^* = 1/B \sum_{j=1}^B \hat{q}_{\theta_j}^*$ olacaktır.

3.5.1.3 Kernel Tahmincisi

Kovaryans matrisinin elde edilmesinde kullanılan diğer bir tahminci Kernel tahmincisidir. Kernel tahmincisinin hesaplanmasında kernel fonksiyonunun ve bant genişliğinin belirlenmesi önemlidir.

$E[f_{u_\theta}(0/x)xx']$ için bir Kernel Tahmincisi,

$$\hat{E}[f_{u_\theta}(0/x)xx'] = (c_n^n)^{-1} \sum_{i=1}^n k(\hat{u}_{\theta_i}) x_i x_i'$$

olarak hesaplanır¹⁰⁰. Burada, $k(\cdot)$ kernel fonksiyonudur, $c_n = o_p(1)$ Kernel bant genişliğidir. $E[xx']$ terimi,

$$\hat{E}[xx'] = \sum_{i=1}^n x_i x_i' / n$$

olarak tahmin edilir.

Kernel bant genişliği seçiminde veriye bağlı olarak problem ortaya çıkar. c_n 'in optimal seçimine olanak sağlayan direkt bir yöntem yoktur.

$$\hat{E}[f_{u_\theta}(0/x)xx'] = (c_n^n)^{-1} \sum_{i=1}^n k(\hat{u}_{\theta_i}) x_i x_i'$$

gösterimindeki $\hat{f}_{u_\theta}(0) = (c_n n)^{-1} \sum_{i=1}^n k(\hat{u}_{\theta_i})$, $f_{u_\theta}(0)$ yoğunluğunun tahminidir bu da c_n 'in optimal seçimini zorlaştırır. Bu probleme karşı, birçok sayıda çapraz geçerlilik (cross-validation) yöntemi vardır (ör: EKK, Log Likelihood,...). Bunun için öncelikle

¹⁰⁰ J. L. Powell, "Censored Regression Quantiles", Journal of Econometrics, Vol.32, No.1, s. 143.

bir çapraz geçerlilik yöntemi ile c_n belirlenir, sonra ilk denklemdeki $E[f_{u_\theta}(0/x)xx']$ 'yi tahmin etmek için optimal olarak seçilen c_n kullanılır.

Bağımsızlık varsayımı altında $f_u(o)$ 'ın da tahmin edilmesi gereklidir. Yukarıda tahmin edilen ilk adıma uygun olarak Λ_θ kovaryans matrisinin tahmini,

$$\Lambda_\theta = \frac{\theta(1-\theta)}{f_{u_\theta}^2(0)} \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1}$$

olacaktır.

3.6 Kantil Regresyonda Sabit Varyansın İncelenmesi

Sabit varyansın geçerli olup olmadığı, kantil regresyonlar tahmin edilerek analiz edilebilir. Standart sabit varyanslı regresyon modelinde regresyon doğrusu y' nin x' e göre şartlı beklenen değeridir. Hata terimlerinin dağılımı sabit varyanslı ise kantiller her zaman medyandan aynı uzaklıkta yani regresyon doğrusuna paralel olacaktır.

Kantil regresyonda sabit varyans geçerli olduğunda, açıklayıcı değişkenlerle bulunacak katsayıların eşit; fakat sabit terim farklı olacaktır. Bu durumda EKK ile bulunacak doğrusal regresyon modeli ile medyan regresyon modeli aynı olacaktır. $f_{u_\theta}(0/x) = f_{u_\theta}(0)$ olduğunda, herhangi iki kantil parametre vektörleri β_{θ_1} ve β_{θ_2} sadece kesim noktalarında farklılık gösterir, eğim katsayıları aynıdır¹⁰¹.

Sabit varyansın incelenmesi için hesaplanan test istatistiği için kovaryans matrisi kullanılmaktadır. Kovaryans matrisinin non-local alternatif hipotezler altında geçerli olması ve local alternatif hipotezler altında geçerli olmasına göre farklı görüşler bulunmaktadır*.

¹⁰¹ Moshe Buchinsky, "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research", s. 106.

* Buchinsky (1998) ile Koenker and Bassett(1982)'de kullanılan teknik arasındaki temel fark; Buchinsky (1998)'de test istatistiği için kullanılan kovaryans matrisi non-local alternatif hipotezler altında geçerliken, Koenker and Bassett(1982)'de sadece local alternatif hipotezler altında geçerlidir. Koenker

Sabit varyansın incelenmesinde minimum uzaklık yaklaşımı kullanılabilir. İlk olarak sabit varyans kısıtı altında eğim katsayıları, kısıtlı katsayı vektörü, β^R , ye göre,

$$Q(\beta^R) = (\hat{\beta}_\theta - R.\beta^R)' A^{-1} (\hat{\beta}_\theta - R.\beta^R)$$

minimizasyonu ile elde edilir. Burada A ağırlık matrisidir ve $A \rightarrow \psi$, pozitif belirli(definite) matristir. $\hat{\beta}_\theta = (\hat{\beta}_{\theta 1}, \dots, \hat{\beta}_{\theta p})$, p sayıdaki kantil regresyon tahminlerinin kısıtsız vektörüdür. $\psi = \Lambda_\theta$ ise, β_θ^R , nin tahmini optimal minimum uzaklık tahmincisinin tahminidir. $\psi \neq \Lambda_\theta$ ise minimum uzaklık tahmincisinin asimtotik kovaryans matrisi,

$$\Lambda_\theta^R = (R' \psi^{-1} R)^{-1} R' \psi^{-1} \Lambda_\theta \psi^{-1} R (R' \psi^{-1} R)^{-1}$$

olacaktır. β_θ^R kısıtlı parametrelerin $(p + K - 1) \times 1$ vektörüdür.

Kısıtlama matrisi R,

$$R' = (R_1, \dots, R_p)$$

olacaktır. Burada,

$$R_j = \begin{pmatrix} e_j & o_m \\ o_v & I_{k-1} \end{pmatrix}$$

' dir.

e_j , 1 haricinde sıfırların j.'ci değerdeki $p \times 1$ vektörüdür,

o_v , sıfırların $(K-1) \times 1$ vektörüdür,

o_m , sıfırların $p \times (K-1)$ matrisidir,

and Bassett(1982) kantil parametre vektörleri üzerinde belli yapıya sahip çoklu değişen varyans modelini önermişlerdir.

I_{K-1} , birim matristir.

Alternatif kantil regresyonlardaki $(\beta_{\theta_1}, \dots, \beta_{\theta_p})$ kesim noktalarının eşit olmasına gerek yoktur¹⁰².

Optimal minimum uzaklık (MD) Tahmincisinin $\hat{\beta}_\theta^R$ ' nin asimtotik dağılımı,

$$\sqrt{n}(\hat{\beta}_\theta^R - \beta_\theta^R) \xrightarrow{L} N(0, \Lambda_\theta^R)$$

olacaktır. Burada; $\Lambda_\theta^R = (R' \Lambda_\theta^{-1} R)^{-1}$ ' dir. $\Lambda_\theta = \psi$ için, eğim katsayılarının eşit olduğunu savunan sıfır hipotezi altında;

$$n(\hat{\beta}_\theta - R\hat{\beta}_\theta^R)' A^{-1}(\hat{\beta}_\theta - R\hat{\beta}_\theta^R) \xrightarrow{L} \chi^2_{(pK-p-K+1)}$$

olacaktır.

3.7 Kantil Regresyonda Simetrinin İncelenmesi

Hata terimlerinin dağılımı simetrikse, koşullu ortalama veya regresyon fonksiyonu medyanda veya % 50 kantilde olacaktır. Simetrinin incelenmesi için minimum uzaklık yaklaşımı kullanılmaktadır. $\hat{\beta}_\theta$ ve $\hat{\beta}_{(1-\theta)}$, θ ' cı ve $(1-\theta)$ ' cı kantil regresyonlar için tahmin edilmiş parametre vektörlerini göstermektedir. Simetri; $\hat{\beta}_{0.50} = 0.5(\beta_\theta + \beta_{(1-\theta)})$ olmasını ifade eder. $\beta_{0.50}$; medyan parametre vektörüdür. ($\theta=0.50$)

$\beta_\theta, \dots, \beta_{\theta_p}, \theta_1, \dots, \theta_p$ kantillerindeki parametre vektörleridir, p de tek bir sayıdır. Simetrinin test edilebilmesi için, $\theta_{1+(p-1)/2} = 0.50$ ' in $(j=0, \dots, (p-1)/2-1)$ ve

¹⁰² Moshe Buchinsky, "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research", s. 106.

$\theta_{1+(p-1)/2} = 0.50$ ' ci kantil için $\beta_{\theta_{j+(p-1)/2}}$ ' nin bilinmesi gerekmektedir. Simetriklik varsayımı gereği $\beta_{\theta_{p-j}} = 2\beta_{0.50} - 2\beta_{\theta_{j+1}}$ olmaktadır¹⁰³.

$\hat{\beta}_{\theta}^U = (\hat{\beta}_{\theta_1}', \dots, \hat{\beta}_{\theta_p}')'$; p kısıtsız kantil tahminlerinin (stacked) yığın vektörüdür.

$\beta_{\theta}^R = (\beta_1', \dots, \beta_{1+(p-1)/2}')'$ kısıtlı anakütle parametresidir. β_{θ}^R ' nin tahmini,

$$\min_b (\hat{\beta}_{\theta}^U - Rb)' A^{-1} (\hat{\beta}_{\theta}^U - Rb)$$

olacaktır.

Kısıtlama matrisi R,

$$R = (R_1' R_2')' . R_1$$

olarak gösterilir. Burada R_1 , $(1 + (p-1)/2)K$ boyutlarında birim matristir ve,

$$R_2 = \begin{pmatrix} -I_K & O_K & O_K & \dots & O_K & O_K \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O_K & O_K & O_K & \dots & -I_K & 2I_K \end{pmatrix}$$

olarak gösterilir ve $K(p-1)/2 \times K(1 + (p-1)/2)$ ' lik bir matristir.

Test istatistiğinin asimtotik dağılımı, sıfır hipotezi altında, $A \rightarrow \Lambda_{\theta}$, $n \rightarrow \infty$ için,

$$n(\hat{\beta}_{\theta}^U - R\hat{\beta}_{\theta}^R)' A^{-1} (\hat{\beta}_{\theta}^U - R\hat{\beta}_{\theta}^R) \xrightarrow{L} \chi^2_{((p-1)K/2)}$$

olacaktır. $\hat{\beta}_{\theta}^R$, minimum uzaklık (MD) tahminini göstermektedir. Dikkat edilmelidir ki testlerin sonucu sıfır hipotezinin red ya da reddedilmemesine dayanır, alternatif hipotezlerin gücünü ölçmeye dayanmaz.

¹⁰³ Whitney K. Newey, And James L. Powell, "Asymmetric Least Square Estimation and Testing", *Econometrica*, Vol. 55, No. 4 (Jul., 1987), s. 819

Kantil regresyon modelleri için sabit varyans ve simetriklik varsayımlarının incelenmesi için genelleştirilmiş momentler yöntemi de kullanılabilir. $pK \times 1$ boyutlu vektör olan moment fonksiyonu,

$$\psi(x, y, \beta_1, \dots, \beta_p)' = (\psi_1(x, y, \beta_1)', \dots, \psi_p(x, y, \beta_p)')$$

olarak tanımlandığında, simetri ve değişen varyansın sıfır hipotezi altında, genelleştirilmiş momentler yöntemi kullanılarak q ' dan pK ' ya kadar tahmin edilecek parametre vardır.

β_θ^R için $\hat{\beta}_\theta^R$ tahmincileri,

$$\min_b \left(\frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, b) \right)' A^{-1} \left(\frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, b) \right)$$

GMM gösteriminin minimizasyonu ile belirlenir¹⁰⁴.

$$A \xrightarrow{P} E[\psi(x, y, \hat{\beta}_\theta) \psi(x, y, \hat{\beta}_\theta)']$$

olmasını sağlayan A seçildiğinde, $n \rightarrow \infty$ ' ken, etkin bir tahminci belirlenebilir.

$$\psi(x, y, \hat{\beta}_\theta); \psi(x, y, \beta_1, \dots, \beta_p)' = (\psi_1(x, y, \beta_1)', \dots, \psi_p(x, y, \beta_p)')$$

olarak tanımlandığında, sıfır hipotezi altında (simetri ya da değişen varyans için);

$$n \left(\frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, \hat{\beta}_\theta^R) \right)' A^{-1} \left(\frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, \hat{\beta}_\theta^R) \right) \xrightarrow{L} \chi^2_{(pk-Q)}, n \rightarrow \infty$$

olacaktır. Bu yaklaşımın dezavantajı;

$$\min_b \left(\frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, b) \right)' A^{-1} \left(\frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, b) \right)$$

¹⁰⁴ Lars Peter Hansen, "Large Sample Properties of Generalized Methods of Moments Estimators", *Econometrica*, Vol. 50, No.4 (Jul., 1982), s. 1029.

gösteriminin çözümünün doğrusal programlama ile gösterilememesidir ve bu nedenle çözümünün zor olmasıdır.

3.8 Otoregresif Kantil Modeller

Klasik doğrusal regersyon modelinde otokorelasyon sözkonusu olduğunda herhangi bir hata terimi s devre önceki hata teriminden etkilenmektedir¹⁰⁵. Kantil regresyonda da bu durum,

$$Q(\theta; x_{t-1}) = \alpha x_{t-1} + \eta S(\theta), \quad t = 1, 2, \dots, t-1, t, \dots -1 < \alpha < 1, \eta > 0.$$

olarak gösterilir. Bu model “birinci mertebeden otoregresif kantil model” olarak adlandırılır. Sadece bağımsız değişkenin bir dönem gecikmeli değeri denklemin sağında yer aldığı için “birinci mertebeden” ifadesi kullanılmıştır. İkinci mertebeden modelde eşitliğin sağ tarafında, $\theta x_{t-1} + \phi x_{t-2}$ yer alır¹⁰⁶. Bu model üzerinden işlem yapılmak istendiğinde,

$$x_t = \alpha x_{t-1} + \eta S(u_t), \quad t = 1, 2, \dots, t-1, t, \dots, x_0 = 0.$$

modeli kullanılır.

$S(\theta)$ sıfır etrafında simetrik dağılıyorsa, x_t de sıfır ortalamaya sahip olacak ve otokorelasyonlu olacaktır. Gözlemler arasındaki otokorelasyon katsayısı h gecikme uzunluğu için, $\rho(h) = \alpha^h$ olarak gösterilir. $S(\theta)$ simetrik olmayan dağılıma sahipse, $S(\theta)$ ’ nın aşırı değerlerine karşılık x değerleri ani sıçrayış gösterecek böylelikle süreçte süreklilik olmayacaktır. Süreklilik olmaması genellikle otoregresif süreçte simetrik dağılıma sahip olmama sonucunda ortaya çıkar. Bu durumda farklı gözlemlerle çalışmak yerine dağılım şeklinin etkileri üzerine çalışılmalıdır.

¹⁰⁵ Güriş, Çağlayan, s. 401.

¹⁰⁶ Gilchrist, s. 260.

3.9 l_1 -Tahmincileri ile Kantil Regresyon

l_1 -Tahmincisi kantil regresyonun tüm ağırlıkların $\theta = 0.5$ ' te olduğu doğrusal bir fonksiyonudur. $\theta = 1/2$ olduğunda,

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i: y_i \geq x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{i: y_i < x_i' \beta} (1-\theta) |y_i - x_i' \beta| \right\}$$

denkleminin minimizasyonu, $\sum |y_i - x_i' \beta|$ ' nin minimizasyonu haline gelir. Bu $\hat{\beta}^*(0.5)$ tahmincisi, $\sum |y_i - x_i' \beta|^p$ ' yi minimize eden l_p tahmincisinin özel bir durumu olan l_1 -Tahmincisi olarak tanımlanır. Aynı zamanda LAV (en küçük mutlak değer) tahmincisi, LAR(en küçük mutlak artık) tahmincisi, LAE (en küçük mutlak hata) tahmincisi ve MAD(en küçük mutlak sapma) tahmincisi olarak da adlandırılabilir¹⁰⁷.

$\hat{\beta}^*(0.5)$ ' in istatistiksel özelliklerine göre, eğer hatalar ,

$$f(e_i) = (2\lambda)^{-1} \exp \left\{ -\frac{|e_i|}{\lambda} \right\}$$

yoğunluk fonksiyonuyla çift kuyruklu üssel dağılıma sahipse benzerlik fonksiyonunun maksimizasyonu, $\sum_{i=1}^T |e_i|$ ' nin minimizasyonuna eşittir ve böylelikle $\hat{\beta}^*(0.5)$ en çok benzerlik tahmincisi olacaktır¹⁰⁸. Normal dağılıma göre bu yoğunluk daha sivri ve daha geniş kuyrukludur; ancak geniş kuyruklu birçok dağılımın aksine sonlu varyansa sahiptir. Bundan dolayı en çok benzerlik tahmincisi istenen özellikleri sağlamaktadır. Hatalar yukarıdaki yoğunluk fonksiyonuna sahip olduğunda $\hat{\beta}^*(0.5)$, EKK' dan daha iyi sonuç verir¹⁰⁹.

¹⁰⁷ Judge, s. 836.

¹⁰⁸ Robert Blattberg, And Thomas Sargent, "Regression with Non-Gaussian Stable Disturbances", *Econometrica*, Vol.39, No.3 (May., 1971), s. 501.

¹⁰⁹ V. Kerry Smith, And Thomas W. Hall, "A comparison of Maximum Likelihood Versus BLUE Estimators", *The Review of Economics and Statistics*, Vol.54, No.2 (May., 1972), s. 186.

e_i herhangi bir dağılımdan geldiğinde, minimizasyon probleminin tek bir çözümü varsa tahminci sapmasızdır, birden fazla çözüm söz konusuysa sapmasızlık bir algoritma yardımı ile belirlenebilir¹¹⁰.

$\hat{\beta}^*(0.5)$ ' in sınırlayıcı dağılımı,

$$\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Omega \otimes Q^{-1})$$

gösterimindeki δ ' nın alt vektörü ile belirlenir. w_{ij} , $2[f(0)]^{-2}$ haline gelir ve,

$$\sqrt{T}[\hat{\beta}^*(0,5) - \beta] \xrightarrow{d} N\left(0, [2f(0)]^{-2} Q^{-1}\right)$$

olacaktır. $f(0)$ medyandaki yoğunluğun değeridir¹¹¹. $2[f(0)]^{-2}$, F dağılım fonksiyonlu örneğin medyanının asimtotik varyansıdır. Medyan, yerleştirme tahmincisi olarak ortalamadan daha üstündür. Bu nedenle, I_1 -Tahmincisi EKK Tahmincisinden çok daha etkin olacaktır. Cauchy, 2-kuyruklu üssel dağılım ve aşırı değerlerin etkili olduğu diğer hata dağılımları bu sınıfa girmektedir.

Parametre(β) ile ilgili hipotezleri test etmek için Wald İstatistiği kullanılır. Bunun formülasyonu,

$$\sqrt{T}[\hat{\beta}^*(0,5) - \beta] \xrightarrow{d} N\left(0, [2f(0)]^{-2} Q^{-1}\right)$$

olacaktır. Bu da $f(0)$ için sürekli bir tahminci bulunmasına olanak sağlar. Bunun gibi bir tahminci,

$$\hat{f}(0) = \frac{r - s}{T(\hat{e}_{(r)} - \hat{e}_{(s)})}$$

¹¹⁰ R. L. Sielken, And H. O. Hartley, "Two Linear Programming Algorithms for Unbiased Estimation of Linear Models", *Journal of the American Statistical Association*, Vol. 68, No. 343 (Sep., 1973), s. 639.

¹¹¹ Koenker and Bassett, Regression Quantiles, s. 35.

'dir. Burada $r = \lceil T/2 \rceil + v$, $s = \lceil T/2 \rceil - v$ dir. $\lceil . \rceil$ tamsayı kısmını ifade eder, v bir tamsayıdır. $\hat{e}_{(1)}, \hat{e}_{(2)}, \dots, \hat{e}_{(T)}$ artıkları sıralı l_1 artıklarıdır. En iyi v ' nin seçimi açık değildir. Tahminci aslında fonksiyonun düzgünlüğüne(smoothness) ve gözlem sayısına bağlıdır. Bu tahminci K/T oranı büyük olduğunda iyi sonuç vermez.

Wald Testi' ne ek olarak Benzerlik Oranı (LR) ve Lagrange Çarpanı Testleri' ne benzeyen, aynı zamanda l_1 tahmin yöntemlerine dayanan testler de kullanılmaktadır. β üzerindeki doğrusal kısıtlamaların testi için LR istatistiği,

$$\lambda_{LR} = 2[2F(0)]^{-1} (S_R - S_U)$$

olacaktır. Burada S_U ve S_R , kısıtsız ve kısıtlı modeldeki artıkların mutlak değerlerinin toplamını göstermektedir.

$$y_i = x_{i1}'\beta_1 + x_{i2}'\beta_2 + e_i$$

modelinde $\beta_2 = 0$ hipotezi için LM test istatistiği,

$$\lambda_{LM} = g' \Delta g$$

olarak hesaplanır. Burada $g = \sum x_{i2} \text{sgn}(y_i - x_{i1}'\bar{\beta}_1)'$ dir. $\bar{\beta}_1, \beta_1$ için kısıtlı l_1 tahmincisidir. Δ de $(X'X)^{-1}$ in ikinci diagonal sütunudur. Sıfır hipotezi altında her iki istatistik de kısıtlı $\chi^2_{(j)}$ dağılımına sahiptir, burada j kısıt sayısıdır. LM Testi $f(0)$ ' ın tahminini gerektirmediginden daha avantajlıdır. Beklenildiği gibi heavy-tailed dağılımlar için l_1 tahminine dayalı testler EKK artıklarına dayalı testlerden çok daha güçlüdür¹¹².

¹¹² Gilbert Bassett, And Roger Koenker, "An Empirical Quantile Function for Linear Models with iid Errors", Journal of the American Statistical Association, Vol.77, No.378 (Jun., 1982), s.407.

3.10 Gastwirth ve Tukey Tahmincileri ile Kantil Regresyon

Kantil regresyonun daha genel fonksiyonları olan ve basit yerleştirme modeli için önerilen L-tahmincilerinin bazılarına benzeyen tahminciler elde etmek mümkündür. Tahminci,

$$\hat{\beta}(\pi) = \sum_{i=1}^M \pi(\theta_i) \hat{\beta}^*(\theta_i)$$

olarak belirlendiğinde, $\pi = (\pi(\theta_1), \pi(\theta_2), \dots, \pi(\theta_M))'$ dir ve simetrik ağırlıklıdır. $\pi(\theta_i)$ ve e_i ' nin dağılımının her ikisi de simetriktir, $\hat{\beta}(\pi), \beta$ ' nın tahmini olarak alınabilir. Sonuçta,

$$\sqrt{T}(\hat{\beta}(\pi) - \beta) \xrightarrow{d} N(0, \pi' \Omega \pi Q^{-1})$$

haline gelir. Ağırlıklar yukarıda belirtildiği gibi Gastwirth' ün önerdiği

$$\pi' = \left(\pi\left(\frac{1}{3}\right), \pi\left(\frac{1}{2}\right), \pi\left(\frac{2}{3}\right) \right) = (0.3, 0.4, 0.3)$$

şeklindedir, ya da Tukey' in önerdiği,

$$\pi' = \left(\pi\left(\frac{1}{4}\right), \pi\left(\frac{1}{2}\right), \pi\left(\frac{3}{4}\right) \right) = (0.25, 0.5, 0.25)$$

üçlü ortalamadır (trimean). β ile ilgili hipotezlerin testinin formülasyonunda $\sqrt{T}(\hat{\beta}(\pi) - \beta) \xrightarrow{d} N(0, \pi' \Omega \pi Q^{-1})$ kullanılabilir. Q yerine $X'X/T$ yazılır, ve Ω elemanları için sürekli bir tahminci kullanılır, bunun için $f(\xi_\theta)$ için bir tahminciye ihtiyaç vardır. Bu tahminci,

$$\hat{f}(\xi_\theta) = \frac{r-s}{T(\hat{e}_{(r)} - \hat{e}_{(s)})}$$

olacaktır. Burada $r = [T\theta] + v$, $s = [T\theta] - v$ dir. $\hat{e}_{(1)}, \hat{e}_{(2)}, \dots, \hat{e}_{(T)}$ artıkları sıralı l_1 artıklarıdır.

3.11 Trimmed EKK Tahmincileri ile Kantil Regresyon

Basit yerleştirme modelinin kesilmiş ortalamasına benzeyen bir yöntem de trimmed EKK tahmincisidir*. Bu tahminciyi belirlemek için ilk olarak, $\hat{\beta}^*(\alpha)$ ve $\hat{\beta}^*(1-\alpha)$ hesaplanır, burada α arzu edilen trimming oranıdır ($0 < \alpha < 0.5$). Daha sonra $y_i - x_i'\hat{\beta}^*(\alpha) \leq 0$ ve $y_i - x_i'\hat{\beta}^*(1-\alpha) \geq 0$ olan gözlemler atılır ve kalan gözlemlere EKK Yöntemi uygulanır. Bu tahminci $\bar{\beta}_\alpha$ ile gösterilir ve uygun koşullar altında,

$$\sqrt{T}[\bar{\beta}_\alpha - \beta] \xrightarrow{d} N(0, \sigma^2(\alpha, F)Q^{-1})$$

olarak gösterilebilir. $\sigma^2(\alpha, F)$; F dağılımlı anakütlenin α trimmed ortalamasının asimtotik varyansını göstermektedir. $\sigma^2(\alpha, F)$ için,

$$\hat{\sigma}^2(\alpha, F) = \frac{1}{(1-2\alpha)^2} \left(\frac{S}{T-K} + \alpha(c_1^2 + c_2^2) - \alpha^2(c_1^2 + c_2^2) \right)$$

olarak sürekli bir tahminci geliştirilmiştir¹¹³. Burada S , trimmed örnekten elde edilen artık kareleri toplamını göstermektedir. $c_1 = \bar{x}'(\hat{\beta}^*(\alpha) - \bar{\beta}_\alpha)$, $c_2 = \bar{x}'(\hat{\beta}^*(1-\alpha) - \bar{\beta}_\alpha)$ olarak hesaplanır, $\bar{x}$ açıklayıcı değişken gözlemlerinin ortalamalarını içeren K -vektörüdür. Trimming EKK ya da l_1 -tahmincilerinde olduğu gibi başlangıç tahminine dayalı olduğu durumlarda, $\bar{\beta}_\alpha$ 'nın trimmed tahmincilerine göre daha tercih edilebilir olduğu ortaya konulmuştur. Trimmed EKK tahmincisi asimtotik olarak Huber' in M-

* Koenker ve Bassett (1978) tarafından ortaya atılmış daha sonra Ruppert ve Carroll (1980) tarafından geliştirilmiştir.

¹¹³ David Ruppert, And Raymond J. Carroll, "Trimmed Least Square Estimation in the Linear Model", *Journal of the American Statistical Association*, Vol.75, No. 372 (Dec., 1980), s. 828.

tahmincisine eşittir ve sabit varyanslıdır. l_1 -tahmincisi $\alpha = 1/2$ olduğu durumda trimmed EKK tahmincisinin kısıtlı bir hali olarak gösterilebilir.



DÖRDÜNCÜ BÖLÜM

UYGULAMA

Bu bölüme kadar Alternatif Regresyon Modelleri ile ilgili teorik açıklamalar yapıldı. Söz konusu alternatif modellerin tahmininin, testlerinin ve yorumlanmalarının yapılışı ve bunlardan hangisinin daha iyi sonuç verdiğinin belirlenmesi ve sonuçların karşılaştırılması için bir uygulama yaptık. Uygulamaya konu olarak büyüme seçildi ve OECD ülkeleri ile ilgili modeller kuruldu. Konu olarak farklı ülkelerin büyümelerinin seçilme nedeni ülkeden ülkeye farklılıklar gösteren büyüme oranlarında oluşacak aşırı değerlerin olmasının beklenmesidir.

4.1 İktisadi Büyüme Kavramı

Kavram olarak büyüme, bir değişkendeki oransal artışı ifade etmektedir. İktisadi büyüme ise en basit anlamı ile milli gelirdeki artış oranıdır. Milli gelir artışı enflasyon nedeni ile farklı boyutlarda olabileceğinden enflasyondan arındırılmış reel milli gelirdeki artış oranı ele alınır. Herhangi iki periyot arası oransal değişim oranı iktisadi büyüme oranı olabilirse de, yıllık değişim incelenmektedir. Bazı durumlarda iktisadi büyüme negatif olabilir, yani küçülme söz konusu olabilir.

Milli gelir, ülkedeki toplam üretimin parasal ölçümü ile elde edilen büyüklüktür. Artışı veya büyümesi çeşitli faktörlere bağlıdır. Ülkedeki sermaye birikiminin, işgücünün artışı, yatırımlar gibi faktörler milli geliri büyütürken, teknolojik gelişme gibi faktörler de milli gelir artışına neden olmaktadır.

Ülkede sermaye birikimi genel olarak tasarruflardan kaynaklanmaktadır. Bu nedenle tasarrufların artışının veya tasarruf oranının yüksek olmasının büyümeye olumlu katkısı olacaktır. Ayrıca ülkede sermaye artışı, yurt dışından alınacak borçlardan

da kaynaklanıyor olabilir. Alınacak borçlarla yatırım yapılarak, kaynaklar üretime dönüştürülerek milli gelir artışı yaratılabilir.

İşgücünün artışı insan gücü gerektiren sektörler açısından önemlidir. Yeterli sayıda yetişmiş işgücü daha fazla üretim yapılmasını sağlayabilir. İşgücü artışı nüfus artışının bir sonucu olarak da karşımıza çıkmaktadır. Nüfusun artışı işgücünü arttırırken, tüketici sayısını, işçi sayısını...vs. arttıracaktır. Daha çok insana yatırım yapılmasını gerektirecektir.

Bir ülkede yatırımların artışı yatırımlar tamamlandığında üretimin artacağını göstergesi olması açısından önemlidir. Yatırımların çokluğunun veya büyüklüğünün ölçüsü olarak milli gelire veya gayri safi milli hasılaya oranları kullanılabilir. Yatırımların GSMH' ye oranındaki artış büyümeyi arttıracaktır.

Teknolojik gelişme özellikle daha uzun dönemde büyümeye neden olmaktadır. Teknolojik gelişme ile yeni ürünler geliştirildiğinden bunların üretimi ve tüketimi milli geliri arttırıcı etki yaratacaktır. Ayrıca teknolojik gelişme ile verimlilik artışı olacağından üretim artacak ve milli gelir büyüyecektir.

Büyümenin incelenmesinde yıpranma payının da dikkate alınması gerekecektir. Üretimde kullanılan araçlar zaman içinde eskiyecek, yıpranacak ve yenilenmeleri gerekecektir. Yıpranma hem yeniden yerine konmayı, hem de mevcudun bakımını ifade etmektedir.

İktisadi büyüme tüm iktisatçılar için önemli olmuş, farklı iktisadi görüşlere göre farklı büyüme teorileri geliştirilmiştir. Burada bunlardan Neoklasik Büyüme Teorisi ile bu teoriden farkları belirtilerek yeni iktisadi büyüme teorisi olan içsel büyüme teorisi özetlenecektir.

4.1.1 Neoklasik Büyüme Teorisi

Pekçok Neoklasik iktisatçının katkısı ile Neoklasik Büyüme Teorisi geliştirilmiştir¹¹⁴. Solow' un öncü çalışmasına (1956) dayanan neoklasik büyüme teorisi, birçok iktisatçının 1960' lı Literatürde Solow Modeli olarak da adlandırılan Neoklasik Büyüme Modeli' nin temel varsayımları; kapalı bir ekonomi, rekabetçi piyasalar, rasyonel davranan bireyler, üretim faktörleri sermaye ve işgücünün her biri için ölçeğe göre azalan getiriye, üretim fonksiyonu için sabit getiriye öngören bir üretim teknolojisi olarak özetlenebilir. Nüfus ve işgücündeki artışa, teknolojik değişme gibi modelde dışsal olarak yerverilmekte ve beşeri sermayedeki üretkenlik veya verimlilik değişimleri dikkate alınmamaktadır. Bu varsayımlar altında kurulan model, fert başına sermayenin yine fert başına üretim veya tüketim ile aynı oranda artış gösterdiği dengeli büyüme modelidir. Denge sözkonusu olduğunda fert başına gelir ve tüketimdeki artış oranı teknolojik gelişme hızıyla eşit olmaktadır. Diğer bir ifadeyle, modelde dışsal bir değişken olan teknoloji, fert başına gelirdeki artışı sağlayan tek faktördür ve denge durumundaki büyüme hızı tasarruf eğiliminden bağımsız olarak ortaya çıkmaktadır. Nüfus artışı ile teknolojik değişmeyi dışsal kabul eden modelde kamunun, uygulayacağı politikaların bir rolü yoktur.

Solow' un büyüme modeli, aynı teknoloji parametreleri ve seçeneklerine sahip ülkelerin, sadece servetlerinin başlangıç düzeylerinde farklılık gösterdiklerini, zamanla aynı sürekli-sabit kişi başı gelir düzeyine yaklaşımları gerektiğini ileri sürmektedir¹¹⁵. Bu sonuç literatürde şartsız β yaklaşımı (unconditional β convergence) olarak bilinir. Ülkeler mikroekonomik şartlarda farklılık gösterdiğinde, sonuç olarak farklı düzeyde sürekli-sabit kişi başı gelir düzeyine sahip olacaklardır. Solow Büyüme Modeli, farklılıklar kontrol altına alındıktan sonra, fakir ülkelerin zengin ülkelere daha hızlı

¹¹⁴ I. Ehrlich, "The Problem of Development: Introduction", *Journal of Political Economy*, (Oct., 1990), Vol.98, No.5, s. 1.

¹¹⁵ R. M. Solow, "A Contribution to the Theory of Economic Growth," *The Quarterly Journal of Economics*, 70, 1956, s. 65.
David Begg, Stanley Fischer, Rudiger Dornbusch, *Economics*, 7. Basım, The McGraw-Hill Companies, 2003, s. 431.

büyüyebileceğini öne sürer. Modelin bu öngörüsü şartlı β yakınsaması olarak bilinir¹¹⁶.

4.1.2 İçsel Büyüme Teorisi

İktisat literatüründe içsel büyüme teorisinin temellerinin Romer (1986) ve Lucas'ın (1988) çalışmalarına dayandığı konusunda görüş birliği bulunmaktadır. Bu çalışmalar büyümenin, ekonomik sistemin kendi dinamikleri içinde, bir takım faktörlerin etkileşimiyle içsel olarak gerçekleştiğini ileri sürer. Bu bakımdan, büyümeyi, tanımlanan model ve dolayısıyla ekonomik sistem dışındaki etkenlere bağlayan neoklasik büyüme yaklaşımından önemli ölçüde farklıdır. Yapılan analizlerde amaç, artık terimin büyüme açısından hesaplanması olmayıp, bu terimi etkileyen faktörleri ve bu çerçevede ülkeler arasında artık terimin farklılaşmasına yol açan özel kesim ile kamu kesimi tercihlerini incelemektir¹¹⁷.

İçsel büyüme teorisi, iktisadi büyümeyi piyasa mekanizması içinde faaliyet gösteren ekonomik güçlerin içsel olarak belirlediğini varsayar. Bu teori büyümenin itici gücünü tanımlar ve bunun birikimini sağlayan etkenler ile büyüme sürecinin işleyişini açıklar. İçsel Büyüme Modelleri, büyümenin itici gücü olarak tanımladıkları faktörler itibarıyla üç ana grupta incelenebilir¹¹⁸.

i. nüfus artışı ve beşeri sermaye birikimini birer karar değişkeni olarak ele alan modeller.

ii. teknolojik değişmeyi, dışsal ve özerk bilimsel buluşlar yerine, piyasa güçlerinin yönlendirdiği girişimci kararlarına bağlayan modeller.

iii. büyüme sürecinde kamunun rolünü bağımsız bir değişken olarak dikkate alan modeller.

¹¹⁶ Marcelo Mello, Alvaro Novo, "The New Empirics of Economic Growth: Quantile Regression Estimation of Growth Equations", preliminary draft, http://www.econ.pucrio.br/PDF/seminario/marcelo_melo.PDF (12 Aralık 2004), s.2.

¹¹⁷ Begg, Fischer, And Dornbusch, s. 431-433.

¹¹⁸ I. Ehrlich, s. 1-11.

Yakınsama Testleri sadece gelir yakınsamasını tayin etmek için değil, büyüme teorilerini birbirinden ayırmak için de kullanılırlar. Gelir yakınsaması, büyüme modelinin neoklasik büyüme modeli olduğunu gösterir. bazı endojen büyüme teorileri başlangıç gelir katsayısının işaretinin pozitif olduğunu öngörür. Bu durumda, kişi başı gelirden uzaklaşma sözkonusudur, yani başlangıç geliri yükseldiğinde büyüme oranı da yükselir. Başlangıç gelir katsayısının tahmininin işareti pozitifse üzerinde çalışılan model endojen büyüme modelidir. Bu nedenle başlangıç gelirinin katsayısının işaretinin belirlenmesi önemlidir.

Büyüme modelleri ile ilgili uygulamalarda istatistiksel güçlüklerle karşılaşılmaktadır. Şartlı yakınsamanın sözkonusu olduğu uygulamalarda politika değişkenlerinin katsayılarının tüm ülkeler için aynı olacak şekilde kısıtlandığı ortalama regresyon tahmin yöntemi yani En Küçük Kareler Yöntemi kullanılır. Bu durumda politika değişkenlerinden birinde meydana gelen değişikliğin zengin ülke gayri safi yurt içi hasılası üzerinde meydana getirdiği etki ile, fakir ülke gayri safi yurt içi hasılası üzerinde meydana getirdiği etki aynı olacak, bu sonuç da gerçeği yansıtmayacaktır. Bu nedenle politika değişken katsayılarının ülkenin gelişmişlik derecesiyle ilgili olan bir ölçüye bağlı olması gerekmektedir.

4.2 Uygulamada Kullanılan Veri ve Değişkenler

Uygulamada 27 OECD ülkesinin 1985 ve 2000 yıllarına ait yatay kesit verileri kullanılarak bir büyüme modeli tahmin edilmeye çalışılmıştır. Veriler http://pwt.econ.upenn.edu/php_site/pwt61_form.php sitesinden elde edilmiştir. Büyüme artış oranlarını açıklamak için kullanılacak model, tasarruf, nüfus artışı, teknolojik gelişme, büyüme oranı değişkenlerinin yer aldığı Solow Modeli' ne benzeyen; reel gayri safi yurtiçi hasıla büyüme oranı, gayri safi yurtiçi hasıla, yatırımların gayri safi yurtiçi hasılaya oranı, kişi başı reel gayri safi milli hasılanın büyüme oranı, teknolojik büyüme oranı, yıpranma oranı değişkenlerinin yer aldığı Mankiw, Romer ve Weil tarafından geliştirilen modeldir¹¹⁹. Model,

¹¹⁹ G. Mankiw, D. Romer ve D. Weil, "A Contribution to the Empirics of Economic Growth," Quarterly Journal of Economics, Vol.152 (May., 1992), s. 407..

$$\ln(Y_{00}/Y_{85}) = \alpha_0 + \alpha_1 \ln Y_{85} + \alpha_2 \ln(I/GSYIH) + \alpha_3 \ln(n+g+\delta) + \varepsilon_i$$

' dir. Burada,

(Y_{00}/Y_{85}) :2000 yılı reel gayri safi yurtiçi hasılasının 1985 yılı reel gayri safi yurt içi hasılasına oranı olan büyüme oranıdır.

Y_{00} :2000 yılı çalışan kişi başına reel gayri safi yurtiçi hasıladır (1=1\$)

Y_{85} :1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıladır (1=1\$)

$(I/GSYIH)$:Reel yatırımların reel gayri safi yurtiçi hasılaya oranıdır.

$$((I_{85}/GSYIH_{85})/(I_{00}/GSYIH_{00}))$$

$(n+g+\delta)$: n; kişi başı reel gayri safi milli hasılanın büyüme oranı, g; teknolojik büyüme oranı, δ ; yıpranma oranıdır.

Görüldüğü gibi model tam olarak Solow Modeli değildir. Bazı model denemelerinde Solow Modeli'nde yer alan tasarruflara yerverilmiştir. Bu modellerde değerlerini belirleyemediğimiz teknolojik büyüme oranı yerine tasarruf artış oranı alınmıştır.

Uygulamada çok sayıda model denenmiş, bunlar iktisadi ve istatistiksel olarak test edilmiştir. Denenen herbir model önce En Küçük Kareler Yöntemi ile doğrusal olarak, daha sonra tam logaritmik olarak tahmin edilmiştir. Söz konusu modellerin testinden sonra aynı modeller robust ve kantil regresyon yöntemleri ile de tahmin edilmişlerdir. Burada tahmin edilen modellerden iyi sonuç veren iki tanesine yerverilmiştir.

4.2.1 Birinci Model Denemesi

2000 yılı reel gayri safi yurtiçi hasılasının 1985 yılı reel gayri safi yurt içi hasılasına oranının bağımlı değişken olduğu regresyon modelinde, teknolojik büyüme oranının verisi bulunamadığından, teknolojik büyüme oranı ve yıpranma oranının toplamı $(g + \delta)$ sabit 0,05 alınarak,

$$(Y_{00} / Y_{85}) = \beta_0 + \beta_1 Y_{85} + \beta_2 (I / GSYIH) + \beta_3 (n + g + \delta) + \varepsilon_i$$

modeli,

$$(Y_{00} / Y_{85}) = 1,808940 - 0,000008 Y_{85} - 0,018582 (I / GSYIH) - 0,018126 (n + g + \delta)$$

olarak En Küçük Kareler Yöntemi ile tahmin edilmiş ve tablodaki sonuçlar elde edilmiştir.

Bağımlı Değişken : (Y_{00} / Y_{85}) Gözlem sayısı :27				
Değişkenler	Katsayılar	Std. Hata	t	$P > t $
Sabit	1.87867	0.7399173	2.54	0,018
Y_{85}	-0.000008	0.0000067	-1,20	0,242
$(I/GSYIH)$	-.0189789	12.48577	-0,29	0,773
$(n + g + \delta)$	-3.040444	0,123016	-0,24	0,810
$R^2 = 0.0743$ $\bar{R}^2 = -0.0464$ $F_{0,005/2;4-1;27-4} = 0.62$ Prob (F istatistiği)=0.6119 $\sum e_i^2 = 2.6495$				

Tahmin edilen katsayıların istatistiksel olarak anlamlılığının test edilmesi için t Testi yapıldığında hipotezler,

$$H_0 : \beta_i = 0, \hat{\beta}_i \text{ istatistiksel olarak anlamsızdır.}$$

$$H_1 : \beta_i \neq 0, \hat{\beta}_i \text{ istatistiksel olarak anlamlıdır.}$$

olarak kurulmuştur. Katsayıların hesaplanan t değerlerinin $t_{\alpha:0.05/2,27-4}=2,069$ tablo değeriyle karşılaştırıldığında, sadece sabit katsayının istatistiksel olarak anlamlı olduğu, diğer katsayıların anlamsız olduğu görülmüştür.

Modelin genel olarak anlamlılığı için F testi yapıldığında hipotezler,

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0, \text{ model genel olarak anlamsızdır.}$$

$$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq 0, \text{ model genel olarak anlamlıdır.}$$

olarak kurulmuştur. Hesaplanan F değeri= 0.62 , $F_{0,005/2;4-1;27-4}=3,03$ tablo değerinden küçük olduğu için model genel olarak anlamsızdır. Modelin belirlilik katsayısı (R^2) oldukça düşüktür.

Normal dağılıma uygunluk için eğiklik ve basıklık incelemiş,

Normallik için Skewness/Kurtosis Testi				
Değişken	Pr(Skewness)	Pr(Kurtosis)	Adj $\chi^2_{(2)}$	Prob> χ^2
Artıklar	0.001	0.018	13.20	0.0014

olasılık değeri 0,05' ten küçük olduğundan normal dağılım olmadığına karar verilmiştir.

Modelde aşırı değerlerin belirlenmesi için DFFITS, COOKSD testleri ile modelde üç bağımsız değişken olduğundan DFBETA1, DFBETA2, DFBETA3 Testleri yapılmıştır. Test sonuçları aşağıdaki tabloda görülmektedir.

	DFFITS	COOKSD	DFBETA1	DFBETA2	DFBETA3
1	0.1009373	0.0026616	-0.0049576	0.0892421	0.0389013
2	0.0000753	0.0000000	0.0000224	-0.0000206	-0.0000229
3	0.0035981	0.0000034	0.0020508	-0.0016386	-0.0016680
4	-0.0751985	0.0014745	-0.0374585	0.0051145	-0.0456432
5	-0.3159573	0.0251278	-0.2542409	0.0449787	0.0051421
6	-0.0224405	0.0001316	-0.0064112	-0.0016732	0.0118550
7	-0.1006243	0.0026323	-0.0118543	0.0531286	0.0531888
8	-0.0046367	0.0000056	-0.0008501	0.0020806	0.0020398
9	-0.1045073	0.0028387	-0.0490748	0.0625096	0.0316460
10	-0.0338355	0.0002989	0.0043888	-0.0083447	0.0153697
11	-0.2088578	0.0110197	0.0355546	0.0702581	0.0807337
12	-0.4730334	0.0568907	0.2148743	0.1262310	0.3421061
13	*0.7264320	0.1034624	-0.1428434	-0.3750436	-0.2169161
14	-0.1277767	0.0042262	-0.0391930	0.0459747	-0.0670820
15	-0.0288073	0.0002169	-0.0019530	-0.0241213	0.0083188
16	-0.0823568	0.0017669	0.0360473	-0.0465704	0.0295098
17	*1.1504070	*0.2630250	-0.9310331	0.0850847	0.3689570
18	*1.7208340	*0.4854494	*1.1784980	-0.7055982	*0.8469921
19	-0.9566799	*0.2163482	0.2815562	0.2219862	-0.7811430
20	-0.0789350	0.0016219	-0.0518925	0.0145980	0.0021400
21	-0.3376985	0.0284688	-0.1180248	0.1986599	-0.1531194
22	0.0450316	0.0005297	0.0046526	0.0362032	-0.0033244
23	0.0480917	0.0006043	-0.0415565	0.0123770	-0.0135925
24	0.2040226	0.0107651	-0.1241963	0.0079938	-0.1238130
25	-0.1149320	0.0034141	-0.0389630	0.0164572	0.0453879
26	-0.7398002	0.1385180	*0.4666932	-0.0299697	-0.5270786
27	0.0550863	0.0007925	0.0439371	-0.0186063	0.0157902

Hesaplanan değerler;

DFFITS Testi için $2\sqrt{p/n} = 2\sqrt{3/27} = 0,666$ değeri ile,

COOKS D Testi için; $4/n=0,148$ değeri ile,

DFBETAS Testi için; $2/\sqrt{n} = 0,384$ değeri ile karşılaştırılmıştır.

Karşılaştırma yapılmış ve aşırı değer olduğu belirlenen noktalar * ile belirtilmiştir.

Yine seri değerlerinde aşırı değerlerin belirlenmesi için standartlaştırılmış ve studentlaştırılmış artıklar ile HAT Matrisin köşegen değerleri hesaplanmıştır. Hesaplanan değerler aşağıdaki tabloda görülmektedir.

	rstndrt	rstudnt	hat
1	0.1038946	0.1016347	0.4965571
2	0.0003455	0.0003379	0.0472720
3	0.0110628	0.0108196	0.0995774
4	-0.2318236	-0.2269933	0.0988937
5	-0.9217587	-0.9186248	0.1057846
6	-0.0905507	-0.0885761	0.0603136
7	-0.3501485	-0.3433684	0.0790868
8	-0.0187706	-0.0183581	0.0599652
9	-0.3573093	-0.3504293	0.0816750
10	-0.1479857	-0.1448018	0.0517738
11	-0.8771626	-0.8726015	0.0541846
12	-0.7913118	-0.7846733	0.2665490
13	2.3972020	2.7069380	0.0671787
14	-0.4705999	-0.4624878	0.0709181
15	-0.0428630	-0.0419225	0.3207374
16	-0.2744214	-0.2688299	0.0857996
17	2.3474550	2.6328150	0.1603165
18	2.9281250	3.6159870	0.1846567
19	-1.4826050	-1.5247040	0.2824841
20	-0.3045613	-0.2984692	0.0653702
21	-1.015794	-1.0165290	0.0993926
22	0.110440	0.1080417	0.1480087
23	0.090172	0.0882051	0.2291512
24	0.491435	0.4831766	0.1513178
25	-0.505699	-0.4973562	0.0506936
26	-0.8532056	-0.8479789	0.4321827
27	0.1339428	0.1310498	0.1501594

Hesaplanan deęerler;

Standartlařtırılmıř artıklar iin 2 deęeri ile,

Studentlařtırılmıř artıklar iin $t_{0,05,n-k} = t_{0,05,27-4} = 2,069$ deęeri ile,

HAT Matris elemanları iin $2p/n = 0,222$ deęeri ile karřılařtırılır.

Karřılařtırma yapılmıř ve ařırı deęer olduęu tespit edilen noktalar * ile belirtilmiřtir.

Yapılan testler doęrultusunda ařırı deęerlerin olduęu gzlemlenmiřtir. Deęiřkenlerin logaritması alınarak tam logaritmik model denenecektir.

Tam Logaritmik Model

Tahmin edilen doęrusal modelin deęiřkenlerinin logaritması alınarak,

$$\ln(Y_{00} / Y_{85}) = \alpha_0 + \alpha_1 \ln Y_{85} + \alpha_2 \ln(I / GSYIH) + \alpha_3 \ln(n + g + \delta) + \varepsilon_i$$

modeli,

$$\ln(Y_{00} / Y_{85}) = 1,5424 - 0,1709 \ln Y_{85} + 0,0246 \ln(I / GSYIH) - 0,1969 \ln(n + g + \delta)$$

olarak En Kk Kareler Yntemi ile tahmin edilmiř ve tablodaki sonular elde edilmiřtir.

Bağımlı Değişken : $\ln(Y_{00} / Y_{85})$ Gözlem sayısı :27				
Değişkenler	Katsayılar	Std. Hata	t	$P > t $
sabit	1.5424	1.6835	0.92	0.369
$\ln Y_{85}$	-0.1709	0.1064	-1.61	0.122
$\ln(L/GSYİH)$	0.0246	0.0369	0.67	0.512
$\ln(n + g + \delta)$	-0.1969	0.4413	-0.45	0.660
$R^2=0.1123$ $\bar{R}^2=-0.0035$ $F_{0,005/2;4-1;27-4}=0.94$ Prob (F istatistiği)=0.97 $\sum e_i^2 = 0,9739$				

Tahmin edilen katsayıların istatistiksel olarak anlamlılığının test edilmesi için t Testi yapıldığında hipotezler,

$$H_0 : \beta_i = 0, \hat{\beta}_i \text{ istatistiksel olarak anlamsızdır.}$$

$$H_1 : \beta_i \neq 0, \hat{\beta}_i \text{ istatistiksel olarak anlamlıdır.}$$

olarak kurulmuştur. Katsayıların hesaplanan t değerlerinin $t_{\alpha:0.05/2,27-4}=2,069$ tablo değeriyle karşılaştırıldığında, hiçbir katsayının istatistiksel olarak anlamlı olmadığı görülmüştür.

Modelin genel olarak anlamlılığı için F testi yapıldığında hipotezler,

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0, \text{ model genel olarak anlamsızdır.}$$

$$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq 0, \text{ model genel olarak anlamlıdır.}$$

olarak kurulmuştur. Hesaplanan F değeri= 0.94 , $F_{0,005/2;4-1;27-4}=3,03$ tablo değerinden küçük olduğu için model genel olarak anlamsızdır. Modelin belirlilik katsayısı (R^2) oldukça düşüktür.

Artıkların normal dağılımına uygunluğu incelemek için Skewness/Kurtosis Testleri yapılmıştır ve sonuçlar tabloda görülmektedir.

Normallik için Skewness/Kurtosis Testi				
Değişken	Pr(Skewness)	Pr(Kurtosis)	Adj $\chi^2_{(2)}$	Prob> χ^2
Artıklar	0.001	0.015	13.43	0.0012

olasılık değeri 0,05' ten küçük olduğunda artıkların normal dağılıma uymadığına karar verilmiştir.

Modelde aşırı değerlerin belirlenmesi amacı ile DFITS, COOKSD testleri ile modelde üç bağımsız değişken olmasından dolayı DFBETA1, DFBETA2, DFBETA3 Testleri yapılmış ve aşağıdaki sonuçlar elde edilmiştir.

	DFITS	COOKSD	DFBETA1	DFBETA2	DFBETA3
1	-0.1265918	0.0041786	-0.0099831	-0.0918427	-0.0760792
2	0.0279765	0.0002044	0.0082296	0.0023917	-0.0067617
3	0.0513379	0.0006879	0.0269666	-0.0164040	-0.0259078
4	-0.0713336	0.0013273	-0.0273340	-0.0189274	-0.0465888
5	-0.3957764	0.0378864	-0.2564307	-0.0751780	-0.0285505
6	-0.0220295	0.0001268	-0.0052012	-0.0071659	0.0096349
7	-0.0382411	0.0003819	-0.0082239	0.0217965	0.0211018
8	0.0425113	0.0004716	0.0104763	-0.0117832	-0.0193981
9	-0.0460089	0.0005527	-0.0220007	0.0285167	0.0145812
10	-0.0343936	0.0003089	0.0018744	-0.0160635	0.0119878
11	-0.1944878	0.0095365	0.0137153	0.0233929	0.0779772
12	-0.5330642	0.0731248	0.1290447	0.3345914	0.3704433
13	*0.7902945	0.1196770	-0.0062680	-0.5019772	-0.2607395
14	-0.0963980	0.0024146	-0.0313575	0.0062543	-0.0562910
15	-0.1688619	0.0074020	-0.0291922	-0.1099660	0.0639340
16	-0.1078717	0.0030244	0.0342838	-0.0680469	0.0319397
17	*1.0615710	*0.2380300	-0.8735893	0.3177269	0.3494794
18	*1.5086930	*0.3686125	*0.9205433	-0.4348806	*0.8538396
19	-1.2157410	*0.3552094	0.0426036	0.8350367	-0.6445289
20	-0.0794552	0.0016427	-0.0422214	-0.0149096	-0.0050630

21	-0.4213279	0.0446334	-0.1634924	0.3222633	-0.1222149
22	-0.0177674	0.0000825	-0.0030490	-0.0127803	0.0000486
23	-0.1228089	0.0039361	0.1092745	-0.0454315	0.0282573
24	0.2163744	0.0120928	-0.1266276	0.0408891	-0.1361437
25	-0.1273517	0.0041834	-0.0378499	-0.0257591	0.0403932
26	-1.0835250	*0.2912185	*0.7929688	-0.2246435	-0.6724802
27	0.0808150	0.0017026	0.0558299	-0.0057392	0.0327734

Hesaplanan değerler;

DFFITS Testi için $2\sqrt{p/n} = 2\sqrt{3/27} = 0,666$ değeri ile,

COOKS D Testi için; $4/n = 0,148$ değeri ile,

DFBETAS Testi için; $2/\sqrt{n} = 0,384$ değeri ile karşılaştırılmıştır.

Karşılaştırma yapılmış ve aşırı değer olduğu tespit edilen noktalar * ile belirtilmiştir.

Seri değerlerinde aşırı değerlerin tespiti için standartlaştırılmış ve studentlaştırılmış artıklar ile HAT Matrisin köşegen değerleri hesaplanmıştır.

	residstd	rstudt	hat
1	-0.2333880	-0.2285287	*0.2348027
2	0.1312272	0.1283908	0.0453286
3	0.1751521	0.1714165	0.0823125
4	-0.2119973	-0.2075403	0.1056545
5	-1.3097300	-1.3315590	0.0811732
6	-0.0833736	-0.0815533	0.0680049
7	-0.1279880	-0.1252193	0.0853086
8	0.1842247	0.1803083	0.0526602
9	-0.1544063	-0.1510907	0.0848586
10	-0.1376870	-0.1347161	0.0611919

11	-0.9019774	-0.8981796	0.0447877
12	-0.5951281	-0.5865806	*0.4523113
13	* 2.4774480	*2.8298150	0.0723512
14	-0.3655617	-0.3585696	0.0674036
15	-0.3951577	-0.3877904	0.1593908
16	-0.3578438	-0.3509564	0.0863184
17	* 2.1006410	*2.2853610	0.1774750
18	* 2.9578390	*3.6750250	0.1442250
19	-1.3606050	-1.3877180	*0.4342300
20	-0.3200198	-0.3136847	0.0602908
21	-0.9349009	-0.9322360	0.1696163
22	-0.0526296	-0.0514759	0.1064526
23	-0.1840064	-0.1800945	*0.3174092
24	0.5196524	0.5112401	0.1519152
25	-0.5489684	-0.5404541	0.0526045
26	-1.0823660	-1.0866090	*0.4985788
27	0.2430842	0.2380470	0.1033438

Hesaplanan değerler;

Standartlaştırılmış artıklar için 2 değeri ile,

Studentlaştırılmış artıklar için $t_{0,05,n-k} = t_{0,05,27-4} = 2,069$ değeri ile,

HAT Matris elemanları için $2p/n = 0,222$ değeri ile karşılaştırılır.

Karşılaştırma yapılmış ve aşırı değer olduğu tespit edilen noktalar * ile belirtilmiştir.

Yapılan testler doğrultusunda regresyon modelinin normal dağılıma uymadığı ve aşırı değerlerin olduğu gözlemlenmiştir, Bu nedenle Alternatif Regresyon Yöntemleri'nden M- Regresyon ve Kantil Regresyon ile model tahmin edilmiştir.Daha önce elde edilen EKK Yöntemi ile tahmin sonuçları da karşılaştırma yapabilmek için aşağıdaki tabloda verilmiştir.

Bağımlı Değişken : $\ln(Y_{00}/Y_{85})$							
Değişken	EKK	M	Kantil Regresyon				
			$\theta = 0.10$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.90$
Sabit	1.5424 (1.6835)	0.3385 (0.6989)	0.8832 (0.9521)	0.7259 (1.0310)	0.2421 (0.5655)	2.0202 (5.5187)	60379 (1,3527)
$\ln Y_{85}$	-0.1709 (0.1064)	-0.1529 (0.0442)	-0.1978 (0.0915)	-0.1781 (0.0833)	-0.1359 (0.0429)	-0.1673 (0.4178)	-0.1629 (0.0691)
$\ln(I/GSYIH)$	0.0246 (0.0369)	0.0451 (0.0153)	0.0395 (0.0170)	0.0424 (0.0359)	0.0448 (0.0157)	-0.0062 (0.1051)	-0.1252 (0.0491)
$\ln(n + g + \delta)$	-0.1969 (0.4413)	-0.5332 (0.1832)	-0.4823 (0.2230)	-0.4729 (0.2867)	-0.5072 (0.1756)	-0.0263 (1.4944)	1.3038 (0.3978)
R^2	0.1123						
Pseudo R^2			0.3162	0.2375	0.1517	0.1180	0.2739
F	0.94	9.08					

EKK Yöntemi sonuçlarına bakıldığında 1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla $\ln(Y_{85})$ değişkeninin, $\ln(I/GSYIH)$ değişkeninin ve n kişi başı reel gayri safi milli hasılanın büyüme oranı, g teknolojik büyüme oranı, δ yıpranma oranının toplamının logaritmasından oluşan $\ln(n + g + \delta)$ değişkeninin katsayılarının istatistiksel olarak anlamsız olduğu, R^2 'nin düşük olduğu, F Testi sonucunun anlamsız olduğu görülmektedir.

Yapılan normallik testi sonucuna göre artıklar normal dağılıma uymadığından ve modelde aşırı değerler olduğundan robust yöntemlerden biri olan M Regresyon ve Kantil Regresyon ile model yeniden tahmin edilmiştir.

M Regresyon sonuçlarına bakıldığında sabit katsayı dışında tüm katsayıların istatistiksel olarak anlamlı olduğu görülmektedir. F Testi sonucu da anlamlı çıkmıştır.

Yavaş büyüme oranına sahip ülkelerde (10. kantil için) 1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla $\ln(Y_{85})$ değişkeninin, reel yatırımların reel gayri safi yurtiçi hasılaya oranı olan $\ln(I/GSYIH)$ değişkeninin ve $\ln(n + g + \delta)$ değişkeninin katsayısının istatistiksel olarak anlamlı olduğu yani büyüme üzerinde etkisi olduğu, sadece sabit katsayının istatistiksel olarak anlamsız olduğu görülmektedir. Orta düzeyde büyüme oranına sahip ülkelerde (50. kantil) ise sabit katsayı dışında tüm değişkenlerin katsayılarının istatistiksel olarak anlamlı olduğu görülmektedir. 50.

kantilde hesaplanan sonuçlar aynı zamanda Medyan Regresyon sonuçlarıdır. Hızlı büyüme oranına sahip ülkelerde (90. kantil için) ise 1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla $\ln(Y_{85})$, reel yatırımların reel gayri safi yurtiçi hasılaya oranı $\ln(I/GSYIH)$, n kişi başı reel gayri safi milli hasılanın büyüme oranı, g teknolojik büyüme oranı, δ yıpranma oranının toplamının logaritmasından oluşan $\ln(n+g+\delta)$ değişkenlerinin büyüme üzerinde anlamlı etkisi olduğu görülmekte ve sabit katsayının da istatistiksel olarak anlamlı çıktığı görülmektedir.

1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla $\ln(Y_{85})$ değişkeninin yani başlangıç gelirinin katsayısı tüm yöntemlerde negatif çıkmıştır, bu da şartlı yakınsamanın bir göstergesidir.

Yavaş ve hızlı büyüme oranına sahip ülkelerde, büyüme oranını etkileyen değişkenlerin farklı olduğu gözlemlenmiştir. Sabit varyans olsaydı her kantil regresyonda bağımsız değişken katsayıları aynı, sadece sabit değişkenin katsayısı farklı olurdu, EKK Yöntemi ile bulunan sonuçlar medyan regresyon ile aynı sonuçları verirdi.

4.2.2 İkinci Model Denemesi

İlk modelde n kişi başı reel gayri safi milli hasılanın büyüme oranı, g teknolojik büyüme oranı, δ yıpranma oranı değişkenlerinin toplamı ile elde edilen $(n+g+\delta)$ değişkeni hesaplanırken $(g+\delta)$ sabit bir değer olarak 0,05 alınmıştı, ikinci modelde $(g+\delta)$ hesaplanırken, g teknolojik büyüme oranı yerine s tasarruf oranları, δ yıpranma oranı da sabit bir değer olarak 0,10 alınmıştır. Değişken hesaplamalarında böyle bir dönüşüm yapıldıktan sonra,

$$(Y_{00}/Y_{85}) = \beta_0 + \beta_1 Y_{85} + \beta_2 (I/GSYIH) + \beta_3 (n+s+\delta) + \varepsilon_i \text{ modeli;}$$

$$(Y_{00}/Y_{85}) = 0,899254 - 0,000014 Y_{85} + 0,030472 (I/GSYIH) + 8,244571 (n+s+\delta)$$

olarak En Küçük Kareler Yöntemi ile tahmin edilmiş ve tablodaki sonuçlar elde edilmiştir.

Bağımlı Değişken : (Y_{00} / Y_{85}) Gözlem sayısı :27				
Değişkenler	Katsayılar	Std. Hata	t	$P > t $
sabit	0.899254	0.267963	3.36	0,003
Y_{85}	-0.000014	0,000005	-2.53	0,019
(I/GSYIH)	0.030472	0.051497	0,59	0,56
$(n+s+\delta)$	8.244571	2.0792	3,97	0,001
$R^2=0.4488$ $\bar{R}^2=0.3769$ $F_{0,005/2;4-1;27-4}=6.24$ Prob (F istatistiği)=0.0029 $\sum e_i^2 = 1,5778$				

Tahmin edilen katsayıların istatistiksel olarak anlamlılığının test edilmesi için t Testi yapıldığında hipotezler,

$$H_0 : \beta_i = 0, \hat{\beta}_i \text{ istatistiksel olarak anlamsızdır.}$$

$$H_1 : \beta_i \neq 0, \hat{\beta}_i \text{ istatistiksel olarak anlamlıdır.}$$

olarak kurulmuştur. Katsayıların hesaplanan t değerlerinin $t_{\alpha:0.05/2,27-4}=2,069$ tablo değeriyle karşılaştırıldığında, sadece (I/GSYIH) değişkeninin katsayısının istatistiksel olarak anlamlı olmadığı, diğer tüm değişkenlerin katsayılarının istatistiksel olarak anlamlı olduğu görülmüştür..

Modelin genel olarak anlamlılığı için F testi yapıldığında hipotezler,

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0, \text{ model genel olarak anlamsızdır.}$$

$$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq 0, \text{ model genel olarak anlamlıdır.}$$

olarak kurulmuştur. Hesaplanan F değeri= 6,24 , $F_{0,005/2;4-1;27-4}=3,03$ tablo değerinden büyük olduğu için model genel olarak anlamlıdır. Modelin belirlilik katsayısı (R^2) 0,4488' dir.

Modelde değişen varyans Breusch-Pagan Testi ile incelenmiş sonuçlar tabloda verilmiştir.

Breusch-Pagan / Cook-Weisberg Testi	
H_0 :	Artıklar sabit varyansa sahiptir.
H_1 :	Artıklar değişen varyansa sahiptir.
χ^2 =	15.93
Prob (χ^2) =	0.0001

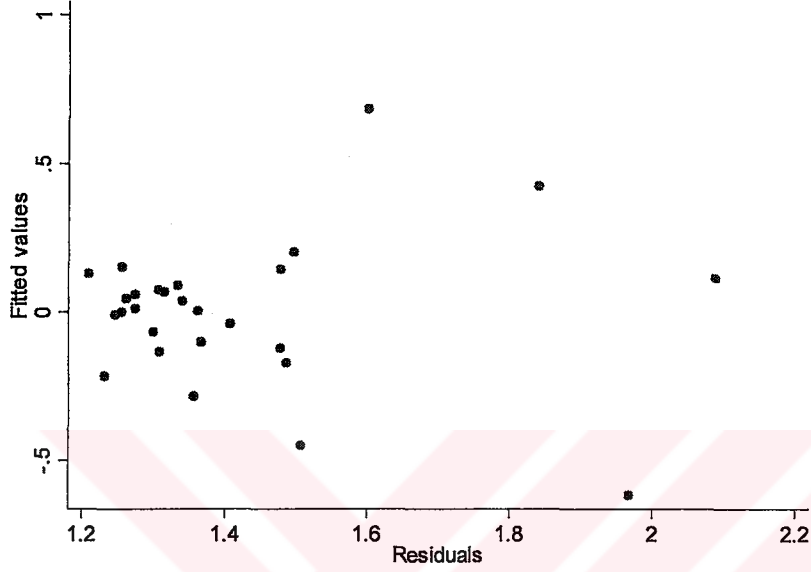
olasılık değeri 0,05' ten küçük olduğundan, değişen varyansın olduğuna karar verilmiştir.

Otokorelasyonun varlığı artıkların korelogramı ile incelenmiş,

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.022	-0.022	0.015	0.902
		2 -0.266	-0.359	2.224	0.329
		3 -0.127	-0.193	2.752	0.431
		4 0.016	-0.086	2.760	0.599
		5 0.092	0.032	3.058	0.691
		6 -0.088	-0.179	3.348	0.764
		7 0.330	0.509	7.618	0.367
		8 -0.182	-0.485	8.995	0.343
		9 -0.256	-0.321	11.849	0.222
		10 0.156	0.158	12.977	0.225
		11 0.020	-0.805	12.997	0.294

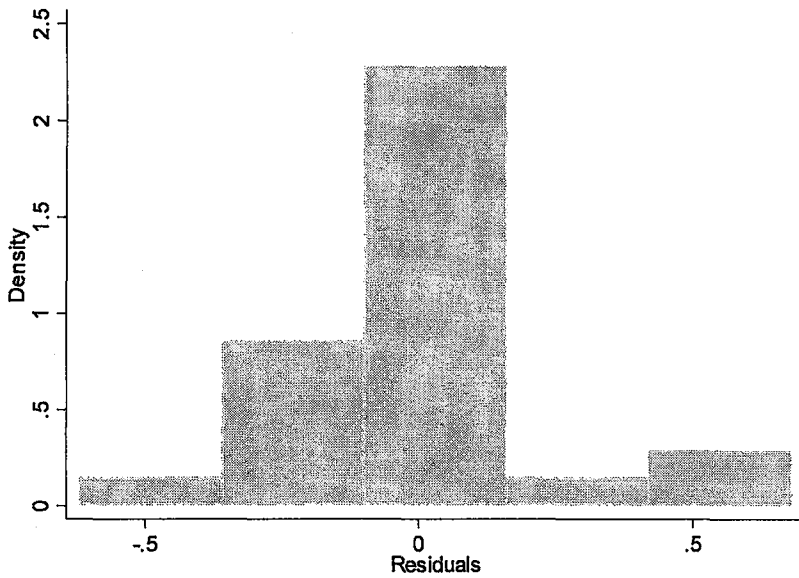
0,05' ten küçük olasılık değeri olmadığından otokorelasyonun olmadığına karar verilmiştir.

Modelin tahmini deęerleri ile artıklarının grafięi incelendięinde, aşırı deęerli gözlemlerin olduęu görölmektedir; ancak kesin karar verilebilmesi için dięer aşırı deęer testleri yapılmalıdır.



Artıkların histogram grafięi incelendięinde yine aşırı deęerlerden şüphelenilmektedir.

(bin=5, start=-0.61805975, width=0.25989705)



Modelde aşırı değerlerin belirlenmesi için DFFITS, COOKSD testleri ile modelde üç bağımsız değişken olmasından dolayı DFBETA1, DFBETA2, DFBETA3 Testleri yapılmıştır. Test sonuçları aşağıdaki tabloda görülmektedir.

	DFFITS	COOKSD	DFBETA1	DFBETA2	DFBETA3
1	-0.8030844	*0.1627629	0.0877723	-0.7565600	-0.2016799
2	0.1523175	0.0059696	0.0595906	-0.0536160	-0.0855272
3	0.0783866	0.0016008	0.0509894	-0.0399023	-0.0138608
4	-0.1027788	0.0027412	-0.0619597	0.0111928	-0.0048598
5	-0.3102932	0.0243156	-0.2530428	0.0582434	0.0697545
6	-0.1404878	0.0050937	0.0017180	0.0576345	0.0769664
7	-0.1941810	0.0097457	0.1134622	0.0710932	0.0389882
8	*0.9077886	*0.1853660	-0.2980212	-0.0870259	*0.7454239
9	-0.0106024	0.0000294	-0.0046642	0.0052937	0.0058528
10	0.1302073	0.0044230	0.0109552	0.1099118	-0.0100484
11	0.0027180	0.0000019	-0.0009988	0.0014044	-0.0007601
12	*1.4005730	*0.3361816	-1.1792790	0.1150265	0.1346748
13	*1.1303810	*0.3251906	0.1321671	-0.0082107	*1.0079490
14	-0.6168654	0.0852813	0.2865390	0.2882808	0.0209557
15	-0.0019494	0.0000009	-0.0013357	0.0004871	0.0006727
16	0.0027180	0.0000019	-0.0009988	0.0014044	-0.0007601
17	*1.4005730	*0.3361816	-1.1792790	0.1150265	0.1346748
18	*1.1303810	*0.3251906	0.1321671	-0.0082107	*1.0079490
19	-0.6168654	0.0852813	0.2865390	0.2882808	0.0209557
20	-0.0019494	0.0000009	-0.0013357	0.0004871	0.0006727
21	-0.3428262	0.0289730	-0.1457195	0.2392950	0.0602220
22	0.0602039	0.0009464	0.0059220	0.0474508	0.0001704
23	0.3258313	0.0272842	-0.2599998	0.0708257	-0.0702320
24	0.2621137	0.0174456	-0.1899197	0.0167176	-0.0229908
25	-0.0601075	0.0009414	-0.0254995	0.0128632	0.0236810
26	-2.5079550	*1.0465580	*2.102481	-0.3734463	-1.5096250
27	0.2172138	0.0121793	0.183951	-0.0878482	-0.0526028

Hesaplanan değerler;

DFFITS Testi için $2\sqrt{p/n} = 2\sqrt{3/27} = 0,666$ değeri ile,

COOKS D Testi için; $4/n=0,148$ değeri ile,

DFBETAS Testi için; $2/\sqrt{n} = 0,384$ değeri ile karşılaştırılmıştır.

Karşılaştırma yapılmış ve aşırı değer olduğu tespit edilen noktalar * ile belirtilmiştir.

Yine seri değerlerinde aşırı değerlerin belirlenmesi için standartlaştırılmış ve studentlaştırılmış artıklar ile HAT Matrisin köşegen değerleri hesaplanmıştır. Hesaplanan değerler aşağıdaki tabloda görülmektedir.

	residstd	rstudt	hat
1	-0.8897675	-0.8855847	*0.4512611
2	0.5976626	0.5891180	0.0626602
3	0.2700791	0.2645624	0.0807019
4	-0.4052057	-0.3977211	0.0626000
5	-0.8808761	-0.8764245	0.1113856
6	-0.1612081	-0.1577538	0.0439385
7	0.2341726	0.2292988	0.0808906
8	0.3546657	0.3478223	0.0554670
9	0.0404025	0.0395158	0.0873529
10	0.2935591	0.2876458	0.0533892
11	-0.5376045	-0.5291226	0.0658537
12	-0.5052588	-0.4969183	0.1324727
13	1.7904070	1.8875200	0.1878541
14	-0.0383784	-0.0375360	0.0738878
15	0.2052526	0.2009251	*0.2957521
16	0.0093528	0.0091473	0.0811290
17	*2.8139800	*3.3986760	0.1451685
18	0.7771484	0.7702464	*0.6829146
19	-1.8104650	-1.9121580	0.0942619
20	-0.0070431	-0.0068883	0.0741514
21	-1.1430550	-1.1511030	0.0814725
22	0.1480930	0.1449069	0.1472032
23	0.6200228	0.6115264	0.2211191
24	0.8090808	0.8028036	0.0963318
25	-0.2656684	-0.2602284	0.0506494
26	-2.8909850	-3.5436760	*0.3337230
27	0.5329478	0.5244818	0.1464079

Hesaplanan değerler;

Standartlaştırılmış artıklar için 2 değeri ile,

Studentlaştırılmış artıklar için $t_{0,05,n-k} = t_{0,05,27-4} = 2,069$ değeri ile,

HAT Matris elemanları için $2p/n = 0,222$ değeri ile karşılaştırılır.

Karşılaştırma yapılmış ve aşırı değer olduğu tespit edilen noktalar * ile belirtilmiştir.

Yapılan testler doğrultusunda aşırı değerlerin olduğu gözlemlenmiştir. Değişkenlerin logaritması alınarak tam logaritmik model denenecektir.

Tam Logaritmik Model

Tahmin edilen doğrusal modelin değişkelerinin logaritması alınarak,

$$\ln(Y_{00}/Y_{85}) = \alpha_0 + \alpha_1 \ln Y_{85} + \alpha_2 \ln(I/GSYIH) + \alpha_3 \ln(n+g+\delta) + \varepsilon_i$$

modeli,

$$\ln(Y_{00}/Y_{85}) = 4,1655 - 0,2328 \ln Y_{85} + 0,0510 \ln(I/GSYIH) - 0,6407(n+s+\delta)$$

olarak En Küçük Kareler Yöntemi ile tahmin edilmiş ve tablodaki sonuçlar elde edilmiştir.

Bağımlı Değişken : $\ln(Y_{00}/Y_{85})$ Gözlem sayısı :27				
Değişkenler	Katsayılar	Std. Hata	t	$P > t $
sabit	4.1655	1.0757	3.87	0.001
$\ln Y_{85}$	-0.2328	0.0883	-2.64	0.015
$\ln(I/GSYIH)$	0.0510	0.0305	1.67	0.108
$\ln(n+s+\delta)$	0.6407	0.1842	3.48	0.002
$R^2=0.4133$ $\bar{R}^2=0,3367$ $F_{0,005/2;4-1;27-4}=5.40$ Prob (F istatistiği)=0.0058 $\sum e_i^2 = 0,643702$				

Tahmin edilen katsayıların istatistiksel olarak anlamlılığının test edilmesi için t Testi yapıldığında hipotezler,

$$H_0 : \beta_i = 0, \hat{\beta}_i \text{ istatistiksel olarak anlamsızdır.}$$

$$H_1 : \beta_i \neq 0, \hat{\beta}_i \text{ istatistiksel olarak anlamlıdır.}$$

olarak kurulmuştur. Katsayıların hesaplanan t değerlerinin $t_{\alpha:0,05/2,27-4}=2,069$ tablo değeriyle karşılaştırıldığında, $\ln(I/GSYIH)$ değişkeninin katsayısının istatistiksel olarak anlamlı olmadığı onun dışındaki tüm katsayıların istatistiksel olarak anlamlı olduğu görülmüştür..

Modelin genel olarak anlamlılığı için F testi yapıldığında hipotezler,

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0, \text{ model genel olarak anlamsızdır.}$$

$$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq 0, \text{ model genel olarak anlamlıdır.}$$

olarak kurulmuştur. Hesaplanan F değeri= 5,40 , $F_{0,005/2;4-1;27-4}=3,03$ tablo değerinden büyük olduğu için model genel olarak anlamlıdır. Modelin belirlilik katsayısı (R^2) 0,4133' tür. Çoklu doğrusal bağıllık olabilirse de $\ln(I/GSYIH)$ tek taraflı bakıldığında 0,10 hata payı ile anlamlı olabilir.

Modelde değişen varyans Breusch-Pagan Testi ile incelenmiş sonuçlar tabloda verilmiştir.

Breusch-Pagan / Cook-Weisberg Testi	
H_0 : Artıklar sabir varyansa sahiptir.	
H_1 : Artıklar değişen varyansa sahiptir.	
χ^2	= 21.04
Prob (χ^2)	= 0.000

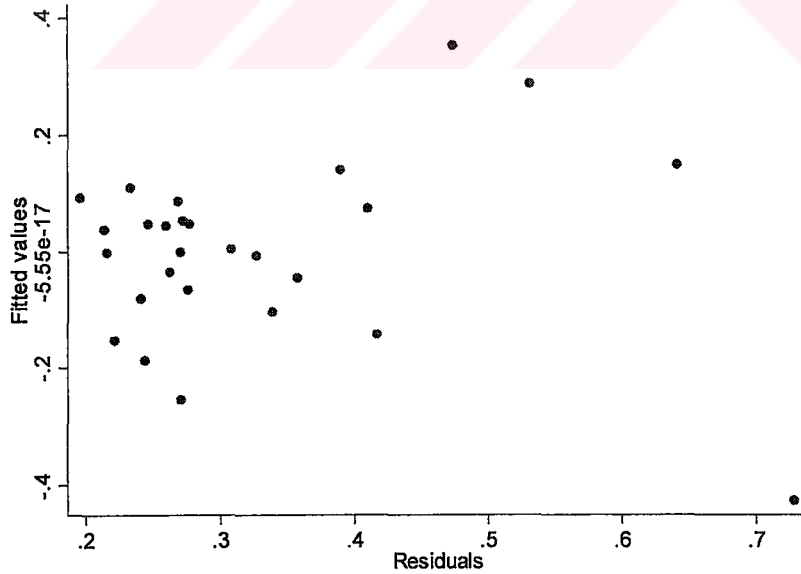
olasılık değeri 0,05' ten küçük olduğundan, değişen varyansın olduğuna karar verilmiştir.

Otokorelasyonun varlığı artıkların korelogramı ile incelenmiş,

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.130	0.131	0.507	0.477
		2	-0.298	-0.479	3.287	0.193
		3	-0.182	-0.101	4.364	0.225
		4	0.102	0.113	4.717	0.318
		5	0.223	0.205	6.480	0.262
		6	0.045	0.009	6.556	0.364
		7	0.131	0.378	7.223	0.406
		8	-0.288	-0.706	10.652	0.222
		9	-0.260	-0.267	13.603	0.137
		10	0.178	0.192	15.062	0.130
		11	0.064	-0.500	15.259	0.171

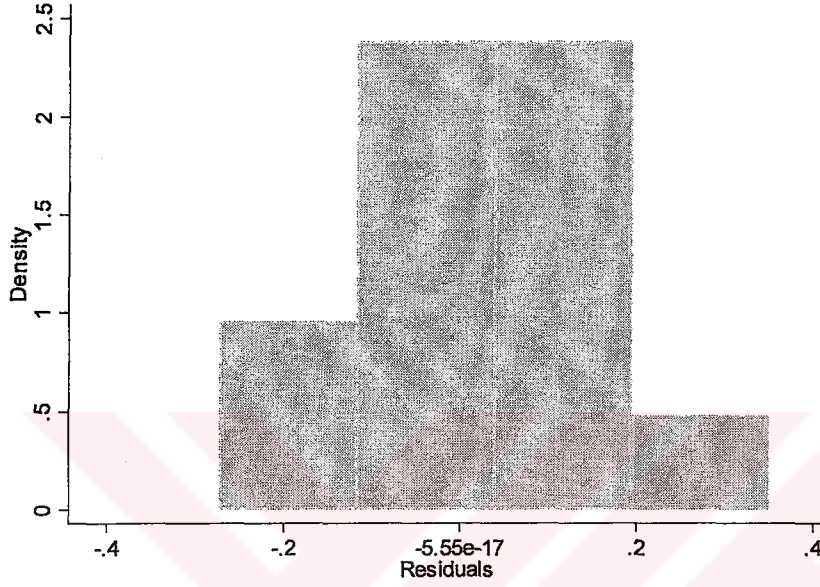
0,05' ten küçük olasılık değeri olmadığından otokorelasyonun olmadığına karar verilmiştir.

Modelin tahmini değerleri ile artıklarının grafiği incelendiğinde, aşırı değerli gözlemlerin olduğu görülmektedir; ancak kesin karar verilebilmesi için diğer aşırı değer testleri yapılmalıdır.



Artıkların histogram grafiği incelendiğinde yine aşırı değerlerden şüphelenilmektedir.

(bin=5, start=-0.42820632, width=0.15587718)



Modelde aşırı değerlerin belirlenmesi için DFFITS, COOKSD testleri ile modelde üç bağımsız değişken olduğundan DFBETA1, DFBETA2, DFBETA3 Testleri yapılmıştır. Test sonuçları aşağıdaki tabloda görülmektedir.

	DFFITS	COOKSD	DFBETA1	DFBETA2	DFBETA3
1	-0.4203047	0.0443893	-0.0108810	-0.3467722	-0.1329300
2	0.1699278	0.0073972	0.0610935	-0.0042556	-0.0928592
3	0.0728733	0.0013829	0.0434738	-0.0208619	0.0003401
4	-0.1702262	0.0074355	-0.0702126	-0.0490979	-0.0498714
5	-0.4896070	0.0557446	-0.3141535	-0.0836910	0.0142333
6	-0.0701394	0.0012813	-0.0134161	-0.0326859	-0.0197253
7	0.1782325	0.0081834	0.0578135	-0.1083867	-0.1041834
8	0.1128936	0.0032916	0.0363278	-0.0336521	-0.0349943
9	0.0699329	0.0012752	0.0373416	-0.0447820	-0.0240912
10	0.0851729	0.0018876	0.0038076	0.0314566	-0.0397405
11	-0.1278995	0.0042284	-0.0089784	0.0255868	0.0808994
12	0.1792618	0.0083639	-0.0444448	-0.1373380	-0.0558430
13	*1.0139360	*0.2261328	-0.1724358	-0.1650242	*0.8335320
14	-0.0034868	0.0000032	-0.0015206	0.0009460	0.0019764
15	-0.0038684	0.0000039	-0.0008937	-0.0025423	0.0010268

16	0.0062416	0.0000102	-0.0014129	0.0034604	-0.0023267
17	*1.1156130	*0.2510466	-0.9735112	0.3251731	0.1345086
18	*1.5294460	*0.5641139	0.1826086	0.0397675	*1.3889480
19	-0.9568058	*0.2199152	0.0078070	*0.8639424	0.1585577
20	-0.0557776	0.0008114	-0.0307593	-0.0070476	0.0113272
21	-0.4349056	0.0472765	-0.1829595	0.3690042	0.0609840
22	-0.0159998	0.0000669	-0.0027343	-0.0112585	0.0002243
23	0.3665170	0.0346727	-0.2942233	0.1188659	-0.1024666
24	0.2783613	0.0195775	-0.1977074	0.0887345	-0.0289597
25	-0.0930921	0.0022486	-0.0327670	-0.0166891	0.0260071
26	-4.6305060	*2.6309510	*3.9955200	-1.2041540	-2.692305
27	0.0880013	0.0020166	0.0660557	-0.0130947	-0.005754

Hesaplanan değerler;

DFFITS Testi için $2\sqrt{p/n} = 2\sqrt{3/27} = 0,666$ değeri ile,

COOKS D Testi için; $4/n = 0,148$ değeri ile,

DFBETAS Testi için; $2/\sqrt{n} = 0,384$ değeri ile karşılaştırılmıştır.

Karşılaştırma yapılmış ve aşırı değer olduğu tespit edilen noktalar * ile belirtilmiştir.

Yine seri değerlerinde aşırı değerlerin belirlenmesi için standartlaştırılmış ve studentlaştırılmış artıklar ile HAT Matrisin köşegen değerleri hesaplanmıştır. Hesaplanan değerler aşağıdaki tabloda görülmektedir.

	residstd	rstudt	Hat
1	-0.9422172	-0.9398232	0.1666687
2	0.6757563	0.6675629	0.0608525
3	0.2909121	0.2850426	0.0613509
4	-0.6473199	-0.6389383	0.0662756
5	-1.5924840	-1.6511670	0.0808191
6	-0.2840672	-0.2783119	0.0597197
7	0.5748539	0.5663012	0.0901279
8	0.5217945	0.5133728	0.0461278
9	0.2319306	0.2270983	0.0866143
10	0.3198828	0.3135498	0.0687180

11	-0.5030237	-0.4946956	0.0626557
12	0.3093719	0.3032032	*0.2590117
13	1.9088120	2.0349880	0.1988813
14	-0.0134716	-0.0131756	0.0654493
15	-0.0095322	-0.0093226	0.1468901
16	0.0201140	0.0196721	0.0914603
17	*2.2911820	*2.5507380	0.1605747
18	1.3335110	1.3577420	*0.5592614
19	-1.3640820	-1.3915750	*0.3209996
20	-0.2204055	-0.2157889	0.0626288
21	-1.0021430	-1.0022400	0.1584604
22	-0.0473892	-0.0463498	0.1064728
23	0.5353702	0.5268957	*0.3260917
24	0.8750781	0.8704564	0.0927765
25	-0.4077717	-0.4002581	0.0513176
26	-3.4931530	-4.9860850	*0.4630749
27	0.2914636	0.2855849	0.0867187

Hesaplanan değerler;

Standartlaştırılmış artıklar için 2 değeri ile,

Studentlaştırılmış artıklar için $t_{0,05,n-k} = t_{0,05,27-4} = 2,069$ değeri ile,

HAT Matris elemanları için $2p/n = 0,222$ değeri ile karşılaştırılır.

Karşılaştırma yapılmış ve aşırı değer olduğu tespit edilen noktalar * ile belirtilmiştir.

Yapılan testler doğrultusunda regresyon modelinin normal dağılıma uymadığı ve aşırı değerlerin olduğu gözlenmiştir, değişen varyans da vardır, bu modelle çalışmaya devam edilecektir. Bu nedenle Alternatif Regresyon Yöntemleri'nden M-Regresyon ve Kantil Regresyon ile model tahmin edilmiştir. Daha önce elde edilen EKK Yöntemi ile tahmin sonuçlarına da karşılaştırma yapabilmek aşağıdaki tabloda yerverilmiştir.

Bağımlı Değişken : $\ln(Y_{00}/Y_{85})$							
Değişken	EKK	M	Kantil Regresyon				
			$\theta = 0.10$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.90$
Sabit	4.1655 (1.0757)	7.4013 (1.1164)	1.4832 (1.3074)	1.4181 (0.6376)	5.6908 (1.4041)	6.0414 (1.4551)	8.0588 (0.3984)
$\ln Y_{85}$	-0.2328 (0.0883)	-0.4694 (0.0875)	-0.0978 (0.1262)	-0.1351 (0.0549)	-0.3283 (0.1099)	-0.3757 (0.1174)	-0.5238 (0.0281)
$\ln(I/GSYIH)$	0.0510 (0.0305)	0.0734 (0.0243)	0.0561 (0.0232)	0.0606 (0.0278)	0.0592 (0.0394)	0.0275 (0.0272)	0.0711 (0.0079)
$\ln(n+s+\delta)$	0.6407 (0.1842)	0.9806 (0.1619)	0.1214 0.1302	-0.0949 (0.1309)	0.8661 (0.1805)	0.7816 (0.2087)	0.9636 0.0738
R^2	0.4133						
Pseudo R^2			0.2774	0.1664	0.1748	0.4562	0.6449
F	5.40	15.35					
iterasyon sayısı		6	5	8	4	9	6

EKK Yöntemi sonuçlarına bakıldığında 1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla $\ln(Y_{85})$ değişkeninin ve n kişi başı reel gayri safi milli hasılanın büyüme oranı, s tasarruf oranı, δ yıpranma oranının toplamının logaritmasından oluşan $\ln(n+s+\delta)$ değişkeninin katsayıları ve sabit katsayı anlamlı, F anlamlı, R^2 yüksektir.

Modelde değişen varyans incelenmiş ve değişen varyans olduğuna karar verilmiştir, bundan dolayı robust yöntemlerden biri olan M Regresyon Yöntemi ve Kantil Regresyon Yöntemi uygulanmıştır.

M Regresyon sonuçlarına bakıldığında tüm katsayıların istatistiksel olarak anlamlı olduğu görülmektedir. F Testi sonucu da anlamlı çıkmıştır.

Yavaş büyüme oranına sahip ülkelerde (10. kantil için) sadece reel yatırımların reel gayri safi yurtiçi hasılaya oranı olan $\ln(I/GSYIH)$ değişkeninin büyüme üzerinde etkisi olduğu, onun dışındaki katsayıların istatistiksel olarak anlamsız olduğu görülmektedir. Orta düzeyde büyüme oranına sahip ülkelerde (50. kantil) ise yavaş büyüme oranına sahip ülkelerin aksine sadece reel yatırımların reel gayri safi yurtiçi hasılaya oranı olan $\ln(I/GSYIH)$ değişkeninin büyüme üzerine etkisinin olmadığı, diğer tüm katsayıların istatistiksel olarak anlamlı olduğu görülmektedir. 50. kantilde hesaplanan sonuçlar aynı zamanda Medyan Regresyon sonuçlarıdır. Hızlı büyüme

oranına sahip ülkelerde (90. kantil için) ise 1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla $\ln(Y_{85})$, reel yatırımların reel gayri safi yurtiçi hasılaya oranı $\ln(I/GSYIH)$, n kişi başı reel gayri safi milli hasılanın büyüme oranı, s teknolojik büyüme oranı, δ yıpranma oranının toplamının logaritmasından oluşan $(n + s + \delta)$ değişkenlerin büyüme üzerinde anlamlı etkisi olduğu görülmektedir.

1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla $\ln(Y_{85})$ değişkeninin yani başlangıç gelirinin katsayısı tüm yöntemlerde negatif çıkmıştır, bu da şartlı yakınsamanın bir göstergesidir.

Yavaş ve hızlı büyüme oranına sahip ülkelerde, büyüme oranını etkileyen değişkenlerin farklı olduğu gözlemlenmiştir. Sabit varyans olsaydı her kantil regresyonda bağımsız değişken katsayıları aynı, sadece sabit değişkenin katsayısı farklı olurdu, EKK Yöntemi ile bulunan sonuçlar medyan regresyon ile aynı sonuçları verirdi.

4.2.3 İki Modelin Karşılaştırılması

2000 yılı reel gayri safi yurtiçi hasılasının 1985 yılı reel gayri safi yurt içi hasılasına oranının bağımlı değişken olduğu, teknolojik büyüme oranının verisi bulunamadığından, teknolojik büyüme oranı ve yıpranma oranının toplamı $(g + \delta)$ sabit 0,05 alınarak tahmin edildiği birinci modelin EKK Yöntemi sonuçlarına bakıldığında 1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla $\ln(Y_{85})$ değişkeninin, $\ln(I/GSYIH)$ değişkeninin ve n kişi başı reel gayri safi milli hasılanın büyüme oranı, g teknolojik büyüme oranı, δ yıpranma oranının toplamının logaritmasından oluşan $\ln(n + g + \delta)$ değişkeninin katsayılarının ve sabit katsayının istatistiksel olarak anlamsız olduğu, R^2 'nin düşük olduğu, F Testi sonucunun anlamsız olduğu görülmektedir.

$(g + \delta)$ hesaplanırken, g teknolojik büyüme oranı yerine s tasarruf oranları, δ yıpranma oranı da sabit bir değer olarak 0,10 alınan ikinci modelin EKK Yöntemi sonuçlarına bakıldığında 1985 yılı çalışan kişi başına reel gayri safi yurtiçi hasıla (Y_{85}) değişkeninin ve n kişi başı reel gayri safi milli hasılanın büyüme oranı, s teknolojik

büyüme oranı, δ yıpranma oranının toplamının logaritmasından oluşan $\ln(n+s+\delta)$ değişkeninin katsayıları anlamlı, F anlamlı, R^2 yüksektir.

Bu sonuçlara bakıldığında R^2 ' sinin yüksek olması ve katsayılarının daha anlamlı olması bakımından ikinci modelin ilk modele göre daha üstün olduğu söylenebilir.

Modeller logaritmik olarak tahmin edildiğinde, yapılan normallik testi sonucuna göre, iki modelde de artıklar normal dağılıma uymadığından ve modellerde aşırı değerler olduğundan robust yöntemlerden biri olan M Regresyon ve Kantil Regresyon ile modeller yeniden tahmin edilmiştir.

Robust M Regresyon sonuçlarına bakıldığında birinci modelde sadece sabit katsayı istatistiksel olarak anlamsızken, ikinci modelde tüm katsayıların anlamlı olduğu görülmektedir.

10. Kantil sonuçlarını her iki model için karşılaştırdığımızda; ilk modelin sadece sabit katsayısı anlamsızken, ikinci modelde ise sadece (I/GSYIH) değişkeninin katsayısı anlamlıdır. Birinci modelin Pseudo R^2 ' si 0,3162, ikinci modelin ise 0,2774' tür. Yavaş büyüme oranına sahip ülkelerde büyüme oranını etkileyen bağımsız değişkenler farklılık göstermektedir. 50. kantil sonuçlarına baktığımızda, birinci modelde sadece (I/GSYIH) değişkeninin katsayısı anlamsızken, ikinci modelde de sadece (I/GSYIH) değişkeninin katsayısı anlamsızdır. Birinci modelin Pseudo R^2 ' si 0,2375, ikinci modelin ise 0,1748' dir. İkinci model için, orta düzeyde büyüme oranına sahip ülkelerde (50. kantil) düşük büyüme oranına sahip ülkelerin aksine sadece reel yatırımların reel gayri safi yurtiçi hasılaya oranı olan (I/GSYIH) değişkeninin büyüme üzerine etkisinin olmadığı sonucuna varılabilir. 90. kantil sonuçlarına baktığımızda, birinci modelde tüm katsayılar anlamlıyken, ikinci modelde de tüm katsayıların anlamlı olduğu görülmektedir. Birinci modelin Pseudo R^2 ' si 0,2739, ikinci modelin ise 0,6449' dur. Yine her iki model için de hızlı büyüme oranına sahip ülkelerde (90. kantil için) ise tüm bağımsız değişkenlerin büyüme üzerine etkisinin olduğu sonucuna varılabilir. Pseudo R^2 ' ler karşılaştırıldığında ise yüksek kantillerde ikinci modelin daha iyi sonuç verdiği görülmektedir

SONUÇ

Değişkenler arasındaki ilişkiyi daha iyi açıklayabilmek için Alternatif Regresyon Modelleri anakütlenin veya incelenen verinin yapısına da bağlı olarak kullanılmaktadır. Bu çalışma Alternatif Regresyon Modelleri' ni karşılaştırmak amacı ile yapılmıştır. Sonuçların karşılaştırılması için büyüme ile ilgili bir uygulama yapılmıştır. Elde edilen sonuçlar aşağıda özetlenmiştir.

Klasik tahmin teorisi normal dağılım temeline dayanmaktadır. Robust yöntemler gözlem değerlerini tamamen kabul etmekle tamamen dışlamak arasında geçiş oluştururlar. Bu nedenle aşırı değerleri dışlayan yöntemlere göre daha üstündürler. Robust yöntemler klasik yöntemlere göre daha üstündürler ve klasik yöntemlerin aşırı gözlem değerlerine duyarlılığına karşın robust yöntemler duyarsızdır.

Dağılımın normal olmadığı durumlarda robust tahmin yöntemleri kullanılır. Normallik varsayımının sağlanmaması, dağılımların simetrik olmaması ve aşırı gözlem değerlerinin olması robust tahminlerin yapılmasına neden olmaktadır.

M tahmincileri En Çok Benzerlik Tahmini' ne dayalı, R Tahmincileri rank testlerine dayalı tahmincilerdir. W Tahmincileri merkezdeki gözlem değerine olan uzaklığa dayanan ağırlıklı tahmincilerken, L Tahmincileri sabit ağırlıklı sıra istatistiklerinin doğrusal kombinasyonlarıdır. Adaptive Tahminciler ise örneğin özelliğine göre seçilmektedir. L Tahminleri ağırlıkların aldıkları değerler yerine sadece gözlemlerin sırasına dayanması sebebiyle W Tahmincilerinden farklılık gösterir.

En Küçük Kareler Regresyonu' nda artıkların kareleri toplamı minimize edilirken, artıkların kareleri yerine farklı değerlerin minimizasyonu söz konusu olabilir. Bu durumda Alternatif Regresyon Modelleri olarak adlandırılan farklı regresyon modelleri kullanılabilir.

M Regresyon robust bir yöntemdir. Bir istatistiksel yöntemin robust olarak adlandırılabilmesi için varsayımlar geçerli olmasa bile iyi sonuçlar vermelidir. En Küçük Kareler ve Medyan Regresyon Y' nin X' e göre şartlı dağılımının tümü hakkında bilgi vermektir. Kantiller bağımlı değişkendeki aşırı değerlere karşı robust'tırlar.

Kantil Regresyon deęiřen varyansın belirlenmesine imkan verir. Farklı kantillerde farklı sonuçların ortaya ıkması, bağımlı deęiřkenin řartlı daęılımının farklı noktalarındaki açıklayıcı deęiřkenlerdeki deęiřikliklere farklı tepki vermesi olarak yorumlanabilir.

Uygulamada büyüme ile ilgili ok sayıda model denenmiřtir. Bu modeller doęrusal ve tam logaritmik model olarak tahmin edilmiř gerekli testler yapılmıřtır. Kullanılan veride ařırı deęerler olduęu belirlenmiřtir. Modellerde Robust ve Kantil Regresyon Modelleri daha iyi sonuç vermiřtir.

Uygulama sonucunda büyüme hızının farklı büyüklüklerde farklı deęiřkenler tarafından etkilendięi belirlenmiřtir. dięer bir ifade ile yavař ve hızlı büyüme oranına sahip lkelerin büyüme oranlarını etkileyen deęiřkenler farklıdır.



Çalışmada Kullanılan Veriler

	Y_{00} / Y_{85}	Y_{85}	$(I/GSYIH)(n + g + \delta)^*(n + s + \delta)^{**}$		
Avustralya	1,314438	39232,2	4,372967	6,3478	0,119449
Avusturya	1,409201	35900,29	0,802186	5,52	0,099357
Belçika	1,383467	41021,85	0,50546	5,314	0,115889
Kanada	1,264426	41358,95	1,066708	6,216	0,120616
İsviçre	1,015831	46673,35	1,11609	5,685	0,112905
Danimarka	1,366036	37172,65	1,257476	5,336	0,117913
İspanya	1,333994	33068,25	0,288509	5,307	0,098743
Finlandiya	1,42544	34194,12	0,507451	5,413	0,107049
Fransa	1,285342	38228,05	0,279398	5,48	0,107241
Birleşik Krallık	1,382677	32291,96	1,330648	5,402	0,097577
Yunanistan	1,172863	30048,98	0,568035	5,154	0,093825
Macaristan	1,355758	18780,66	0,057651	4,691	0,100891
İrlanda	2,26366	28738,15	0,241948	5,499	0,160455
İzlanda	1,238497	36379,03	0,569271	6,086	0,099852
İtalya	1,308579	41227,47	3,754925	5,183	0,097888
Japonya	1,365508	28368,3	1,810333	5,383	0,096085
Kore Cum.	2,283225	16139,49	0,739998	6,036	0,108938
Lüksemburg	2,203431	46806,78	0,38146	6,294	0,219609
Meksika	1,057284	23256,09	0,030348	6,89	0,111916
Hollanda	1,254809	41624,05	0,989137	5,68	0,107918
Yeni Zelanda	1,070185	36778,42	0,108073	6,107	0,115427
Norveç	1,377329	39229,23	2,806457	5,573	0,107579
Polonya	1,622732	12473,5	1,211002	5,304	0,081142
Portekiz	1,698621	20609,57	0,868276	5,04541	0,103104
İsveç	1,232813	36869,37	0,974388	5,45432	0,105521
Türkiye	1,350116	11262,77	0,517304	6,962	0,146202
Amerika	1,339934	48164,1	0,678201	6,014	0,114259

* İlk model için kullanılan değişkendir, $g + \delta$ sabit 0,05 alınmıştır.

** İkinci model için kullanılan değişkendir, g yerine s tasarruflar alınmış, δ yerine ise 0,10 alınmıştır

KAYNAKÇA

- Adichie, J. N. "*Estimates of Regression Parameters Based on Rank Tests*", **The Annals of Mathematical Statistics**. Vol 38, No.3 (Jun., 1967), s. 894.
- Adichie, J. N. "*Rank Tests of Sub-Hypotheses in the General Linear Regression*", **The Annals of Statistics** . Vol 6, No.5 (Sep., 1978), s. 1012.
- Atkinson, A.C. "*Influential Observations, High Leverage Points and Outliers in Linear Regression*", **Statistical Science**. Vol.1, No.3, (Aug., 1986), s.397.
- Barrodale, I. " *L_1 Approximation and the Analyses of Data*", **Applied Statistics**. Vol.17, No.1 (1968), s. 51.
- Barrodale, I. and F.D.K. Roberts. "*An Improved Algorithm for Discrete L_1 Linear Approximation*", **SIAM Journal on Numerical Analysis**. Vol.10, No.5 (Oct., 1973), s. 839.
- Bassett, Gilbert and Roger Koenker. "*Asymptotic Theory of Least Absolute Error Regression*", **Journal of the American Statistical Association**. Vol.73, No. 363 (Sep., 1978), s. 618.
- _____. "*An Empirical Quantile Function for Linear Models with iid Errors*", **Journal of the American Statistical Association**. Vol.77, No.378 (Jun., 1982), s.407.
- Begg, David, Stanley Fischer, Rudiger Dornbusch. **Economics**. 7. Basım, The McGraw-Hill Companies, 2003.
- Bickel, Peter J. And David A. Freedman. "*Some Asymptotic Theory for the Bootstrap*", **The Annals of the Statistics**. Vol. 9, No.6 (Nov., 1981), s. 1196.
- Bickel, P.J. And J.L.Hodges Jr., "*The Asymptotic Theory of Galton's Test and Related Simple Estimate of Location*", **The Annals of Mathematical Statistics**. Vol.38, No.1, (Feb., 1967),s.73.

BIOST 515, "Lecture 7, Linear Regression Diagnostics".
<http://courses.washington.edu/b515/l7.pdf> (5 Haziran), s. 9.

Birkes, David and Yadolah Dodge. **Alternative Methods of Regression**. New York: John Wiley,1993.

Blattberg, Robert and Thomas Sargent. "Regression with Non-Gaussian Stable Disturbances", **Econometrica**. Vol.39, No.3 (May., 1971), s. 501.

Buchinsky, Moshe. "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research", **The Journal of Human Resources**. Vol.33, No.1 (Winter, 1998), s. 101.

Büyüklü, Ali Hakan. "Robust Tahmin Edicilerin Kullanılma Sebebi ve Sonuçları", **IV.Ulusal Ekonometri ve İstatistik Sempozyumu Bildirileri**. M.Ü., Antalya- 14-16 Mayıs 1999, s.655.

Conover, William Jay. **Practical Nonparametric Statistics**. Newyork:John Wiley & Sons Inc., 1971.

Crow , Edwin L. And M.M. Siddiqui. "Robust Estimation of Location", **Journal of the American Statistical Association**. Vol.62, No.318,(Jun., 1967), s.353.

Çağlayan, Ebru. "Yarıparametrik Regresyon Modelleri ile Yaşam Boyu Sürekli Gelir Hipotezinin Türkiye Uygulaması", MÜ Sosyal Bilimler Enstitüsü (Doktora Tezi), İstanbul 2002.

Dallal, Gerard E. "Regression Diagnostics". <http://www.tufts.edu/~gdallal/diagnose.htm> (11 Kasım 2004).

Draper, David. "Rank-Based Robust Analysis of Linear Models", **Statistical Science**. Vol.3, No.2 (May., 1988), s.239.

Efron , B. "Bootstrap Methods: Another Look at the Jackknife" **The Annals of Statistics**. Vol.7, No.1 (Jan., 1979), s.1.

- Efromovich, Sam. **Nonparametric Curve Estimation: Methods, Theory & Applications**. Secaucus, USA: Springer Verlag New York, 1999.
- Ehrlich, I. "*The Problem of Development: Introduction*", **Journal of Political Economy**. (Oct., 1990), Vol.98, No.5, s. 1.
- Finch, S.J. **Testing for Normality**. New York: Marcel Dekker Incorporated, 2001.
- Fox, John. "Robust Regression", Appendix to An R and S-PLUS Companion to Applied Regression, January 2002. <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-robust-regression.pdf> (11 Kasım 2004), s.1
- Freund, John E. **Mathematical Statistics**. 5. Edition, New Jersey :Prentice-Hall International Editions, 1999.
- Gastwirth, Joseph L. "On Robust Procedures", **Journal of the American Statistical Association**. Vol.61, No.316, (Dec., 1966), s.929.
- Gilchrist, Warren G. **Statistical Modelling with Quantile Functions**. Florida: Chapman & Hall/CRC, 2000.
- Gill, Indermit S. **Crafting Labor Policy: Techniques and Lessons from Latin America**. Washington: World Bank, 2002.
- Güriş, Selahattin ve Ebru Çağlayan. **Ekonometri Temel Kavramlar**. 2. Basım, İstanbul: Der Yayınları, 2005.
- Hampel, Frank R. "*The Influence Curve and its Role in Robust Estimation*", **Journal of the American Statistical Association**. Vol.69, No 346, June 1974, s.383.
- Hansen, Lars Peter. " *Large Sample Properties of Generalized Methods of Moments Estimators*", **Econometrica**. Vol. 50, No.4 (Jul., 1982), s. 1029.
- Hoaglin, David C., Frederick Mosteller ve John W. Tukey. **Understanding Robust and Exploratory Data Analysis**. USA: John Wiley & Sons, 1983.

- Hogg, Robert V. "*More Light on the Kurtosis and Related Statistics*", **Journal of the American Statistical Association**. Vol.67, No.338, (June 1972), s. 422.
- _____. "*Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory*", **Journal of the American Statistical Association**. Vol.69, No.348, (Dec. 1974),s.909.
- Huber, Peter J. **Robust Statistics**. New York: John Wiley & Sons, 1981.
- _____. "*Robust Estimation of a Location Parameter*", **The Annals of Mathematical Statistics**. Vol.35, No.1,(Mar., 1964), s.73.
- _____. "*The 1972 Wald Lecture Robust Statistics*", **The Annals of Mathematical Statistics**. Vol.43, No.4, (Aug. 1972), s. 1041.
- Jaeckel, Louis A. "*Some Flexible Estimates of Location*", **The Annals of Mathematical Statistics**. Vol.42, No.5, (Oct., 1971), s. 1540.
- Jolliffe, Ian T. **Principal Component Analysis**. New York :Springer-Verlag, 2002.
- Judge, George G. ve Diğerleri. **The Theory an Practice of Econometrics**. 2.Baskı, Canada: John Wiley, 1985.
- Jung, Sin-Ho. "*Quasi-Likelihood for Median Regression Models*", **Journal of the American Statistical Association**. Vol.91, No.433 (Mar., 1996), s. 251.
- Kmenta, Jan. **Elements of Econometrics**. 2. Basım,, Maxwell Macmillan International Editions, 1971.
- Koenker, R. "*The Trimmed Mean in the Linear Model*", **The Annals of Statistics**. Vol.15, No.1(Mar.1987), s.39.
- Koenker, Roger and Gilbert Bassett. "*Regression Quantiles*", **Econometrica**. Vol.46, No.1, (Jan., 1978), s.33
- Koenker, Roger and Kevin F. Hallock. "*Quantile Regression an Introduction*", **The Journal of Economic Perspectives**. Vol.15, No.4 (Autumn 2001), s. 143

- Koenker, Roger W. And Vasco D' Orey. "*Algorithm AS 229: Computing Regression Quantiles*", **Applied Statistics**. Vol.36, No.3 (1987), s. 383.
- Lawrence, Kenneth D. **Robust Regression**. New York and Basel: Marcel Dekker, Inc.,1990.
- Lee, S. Lecture Notes for MECT1, "Quantile Regression".
<http://www.homepages.ucl.ac.uk/~uctplso/Teaching/MECT/lecture8.pdf> (10 Aralık 2004),s.1.
- Mankiw, G., D. Romer and D. Weil. "A Contribution to the Empirics of Economic Growth," **Quarterly Journal of Economics**. Vol.152 (May., 1992), s. 407..
- Mello, Marcelo and Alvaro Novo. "*The New Empirics of Economic Growth: Quantile Regression Estimation of Growth Equations*", preliminary draft.
http://www.econ.pucrio.br/PDF/seminario/marcelo_melo.PDF (12 Aralık 2004).
- "Methods of Detecting Outlier". <http://www.netnam.vn/unescocourse/statistics/37.htm> (15 Haziran 2005).
- Mosteller, Frederick. "*On Some Useful Inefficient Statistics*",**Annals of Mathematical Statistics**. Vol.17, No.4,(Dec.1946), s.377.
- Newbold, Paul. **İşletme ve İktisat için İstatistik**. Çev. Ümit Şenesen, Literatür Yayınları, No.44, İstanbul, 2000.
- Newey, Whitney K. And James L. Powell. "*Asymmetric Least Square Estimation and Testing*", **Econometrica**. Vol. 55, No. 4 (Jul., 1987), s. 819
- Oliveira, J. Tiago de and Benjamin Epstein. **Some Recent Advances in Statistics**. London: Academic Press, 1982.
- Parzen, Emanuel. "*Nonparametric Statistical Data Modeling*", **Journal of the American Statistical Association**. Vol. 74, No. 365 (Mar., 1979), s. 105.
- Pollard, David. "Chapter 2: Classical Regularity Conditions"
<http://www.stat.yale.edu/~pollard/Asymptopia/Classical.pdf>, (5 Mayıs 2005)

- Portnoy, Stephen and Roger Koenker. "*Adaptive L Estimation for Linear Models*", **The Annals of Statistics**. Vol.17, No.1 (Mar., 1989), s. 362.
- Powell, J. L. "*Censored Regression Quantiles*", **Journal of Econometrics**. Vol.32, No.1, s. 143.
- Rao, C Radhakrishna and Helge Toutenburg. **Linear Models: Least Squares and Alternatives**. 2. Basım, USA: Springer-Verlag New York, 1999.
- Rawlings, John O. **Applied Regression Analysis: A Research Tool**. USA: Springer-Verlag New York, 1998.
- Rousseeuw, Peter J. "*Least Median of Squares Regression*", **Journal of the American Statistical Association**. Vol.79, No.388, (Dec., 1984), 871.
- Rousseeuw, Peter J., Cristophe Croux. "*Alternatives to the Median Absolute Deviation*", **Journal of the American Statistical Association**. Vol.88, No.424, (Dec., 1993), s.1273.
- Ruppert, David and Raymond J. Carroll. "*Trimmed Least Square Estimation in the Linear Model*", **Journal of the American Statistical Association**. Vol.75, No. 372 (Dec., 1980), S. 828.
- Saraçoğlu, Doç. Dr. Bedriye ve Y. Doç. Dr. Ferhan Çevik. **Matematiksel İstatistik**. Ankara: Gazi Büro Kitabevi, 1995.
- Schrader, Ronald M. And Thomas P. Hettmansperger, "*Robust Analyses of Variance Based Upon a Likelihood Criterion*", **Biometrika**. Vol 67, No.1, (Apr., 1980), s. 93.
- Serper, Prof.Dr. Özer. **Uygulamalı İstatistik I**. 3. Basım, İstanbul: Filiz Kitabevi, 1996.
- Sielken, R. L. And H. O. Hartley. "*Two Linear Programming Algorithms for Unbiased Estimation of Linear Models*", **Journal of the American Statistical Association**. Vol. 68, No. 343 (Sep., 1973), s. 639.

Smith, V. Kerry and Thomas W. Hall. "*A comparison of Maximum Likelihood Versus BLUE Estimators*", **The Review of Economics and Statistics**. Vol.54, No.2 (May., 1972), s. 186.

Solow, R. M. "*A Contribution to the Theory of Economic Growth*,". **The Quarterly Journal of Economics**, 70, 1956, s. 65 Ying, Z., S.H. Jung, and L.J.Wei. "*Survival Analysis with Median Regression Models*", **Journal of the American Statistical Association**. Vol 90, no 429,(Mar., 1995), s.178.

Turanlı, Münevver ve Selahattin Güriş. **Temel İstatistik**. 2. Basım, İstanbul: Der Yayınları, 2005.

"Outlier", <http://mathworld.wolfram.com/Outlier.html> (15 Haziran 2005).

Yaffee, Robert A. "Robust Regression Modelling with STATA Lecture Notes".www.bsos.umd.edu/gvpt/uslaner/robustregression.pdf (10 Haziran 2005)