T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ PROGRAMI
YÜKSEK LİSANS TEZİ

TÜRKİYE'DE GELİR FARKLILIĞININ KANTİL REGRESYON MODELİ İLE İNCELENMESİ: 2002-2010 YILLARI KARŞILAŞTIRMASI

Muhammed Hanifi VAN

Danışman
Prof. Dr. Mehmet Vedat PAZARLIOĞLU

İZMİR-2013

YÜKSEK LİSANS
TEZ/ PROJE ONAY SAYFASI

Üniversite	: Dokuz Eylül Üniversitesi
Enstitü	: Sosyal Bilimler Enstitüsü
Adı ve Soyadı	: Muhammed Hanifi VAN
Tez Başlığı	: Türkiye"de Gelir Farklılığının Kantil Regresyon Modeliyle İncelenmesi :2002-2010 Yılları Karşılaştırması
Savunma Tarihi	: 28.06.2013
Danışmanı	: Prof.Dr.Mehmet Vedat PAZARLIOĞLU

JÜRİ ÜYELERİ

Ünvanı, Adı, Soyadı	**Üniversitesi**	**İmza**
Prof.Dr.Mehmet Vedat PAZARLIOĞLU	DOKUZ EYLÜL ÜNİVERSİTESİ	
Prof.Dr.Şenay ÜÇDOĞRUK	DOKUZ EYLÜL ÜNİVERSİTESİ	
Doç.Dr.Hakan AY	DOKUZ EYLÜL ÜNİVERSİTESİ	

Oybirliği (✓)
Oy Çokluğu ( )

Muhammed Hanifi VAN tarafından hazırlanmış ve sunulmuş **"Türkiye"de Gelir Farklılığının Kantil Regresyon Modeliyle İncelenmesi :2002-2010 Yılları Karşılaştırması"** başlıklı Tezi( ) / Projesi( ) kabul edilmiştir.

Prof.Dr. Utku UTKULU
Enstitü Müdürü

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum "**Türkiye'de Gelir Farklılığının Kantil Regresyon Modeli ile İncelenmesi: 2002-2010 Yılları Karşılaştırması**" adlı çalışmanın, tarafımdan, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

..../..../.....

Muhammed Hanifi VAN

ÖZET

Yüksek Lisan Tezi

Türkiye'de Gelir Farklılığının Kantil Regresyon Modeli İle İncelenmesi: 2002-2010 Yılları Karşılaştırması

Muhammed Hanifi VAN

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Ekonometri Anabilim Dalı

Ekonometri Programı

Modelin fonksiyonel yapısının doğru seçilmesi ve seçilen modelin temel varsayımlarının sağlanması ekonometride çok önemli konulardır. Genel olarak, değişkenler arasındaki ilişki parametrik olmayan yöntemlerle incelenmektedir. Bu yöntemlerin içinde en çok bilinen ve en sık kullanılan yöntem En Küçük Kareler yöntemidir. Fakat bu yöntem kullanılan veri setinin yapısı ve normallik varsayımının sağlanmaması gibi nedenlerden dolayı iyi sonuçlar vermemektedir. Bu durumlarda alternatif regresyon yöntemlerini kullanmamız gerekecektir.

Bu çalışmanın amacı son yıllarda yaygınlaşan Kantil regresyon yöntemi ile En Küçük kareler yöntemini kıyaslayarak üstünlüklerini ortaya koymaktır. İlk bölümde En küçük mutlak sapmalar, M regresyon gibi alternatif yöntemler tanıtılmıştır. İkinci bölümde ise normallik varsayımı gerektirmeyen ve kullanıcıya ortalamadan farklı olarak veri setini bir çok faklı kantillere bölen Kantil regresyon yöntemi tanıtılmıştır. Son bölümde ise Uygulama olarak Türkiye'de gelir farklılığının Kantil regresyon modeli ile 2002- 2010 yılları karşılaştırması yapılmıştır.

Anahtar Kelimeler: En Küçük Kareler (EKK), Kantil Regresyon , En Küçük Mutlak Sapmalar (LAD), M Regresyon.

ABSTRACT

Master's Thesis

Analiysing Income Diversty in Turkey by Quantile Regression Model: Comarsion of 2002 and 2010 Years

Muhammed Hanifi VAN

Dokuz Eylül University

Graduate School of Social Sciences

Department of Econometrics

Econometrics Program

Selecting the correct functional structure of model and providing the basic assumptions of the selected model are very important subjects in econometrics. Generally, relationship among variables is examined using parametric methods. Ordinary Least Squares is most known and frequently used method among these methods. But this method not gives good results because of reasons such as used data set structure and not provided the assumption of normality. In such cases, we will need to use alternative regression methods.

The purposes of this study compare the Quantile regression method which is popularized recent years with Ordinary Least Squares and reveals the advantages of Quantile regression method. In the first chapter, alternative methods such as least absolute deviations and m regression have been introduced. In the second chapter, Quantile regression which is not require the assumption of normality and for user unlike averages dividing the data set to the many different quantiles has been introduced. In the last chapter as an empirical part, we compare the income differences in Turkey between 2002 and 2010 with using Quantile regression model.

Keywords: Ordinary Least Squares (OLS), Quantile Regression, Least Absolute Deviations(LAD), M Regression.

TÜRKİYE'DE GELİR FARKLILIĞININ KANTİL REGRESYON MODELİ İLE İNCELENMESİ: 2002-2010 YILLARI KARŞILAŞTIRMASI

İÇİNDEKİLER

KISALTMALAR

QR	Kantil Regresyon
QRM	Kantil Regresyon Modeli
LRM	Liner Regresyon Modeli
LAD	En Küçük Mutlak Sapmalar
EKK	En Küçük Kareler
MAD	Mutlak Sapmalar Medyanı
SST	Kareler Toplamı
SSE	Hata Kareler Toplamı
SSR	Regresyona Bağlı Kareler Toplamı

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

GİRİŞ

Ekonometri, iktisat teorisinden hareketle ele alınan değişkenler arasındaki ilişkiyi ve bu ilişkinin derecesini incelemektedir. Değişkenler arasındaki söz konusu ilişki incelemek üzere ekonometrik olarak yapılan çalışmalarda yaygın olarak En küçük Kareler Yöntemi kullanılmaktadır. Bu yöntemin kullanılabilmesi için temel varsayımların sağlanması en önemli şart olarak ortaya çıkmaktadır. Özellikle normallik varsayımın karşılanmaması yapılan ekonometrik analizin güvenirliliği olumsuz etkilenecektir. Asimetrik dağılımlara sahip verilerde genellikle normallik varsayımı sağlanamamaktadır ve bu varsayımın sağlanamaması durumunda da farklı bir yöntem kullanmak daha sağlıklı sonuçların elde edilmesini sağlayabilecektir.

Bu çalışmada en küçük kareler yönteminden daha esnek varsayımlara sahip olan Kantil regresyon yöntemi tanıtılmıştır. Kantil regresyonun iki yönden diğer yöntemlere göre avantaja sahiptir. Bunlardan ilki bu yöntemin özellikle simetrik olmayan verilerde daha iyi sonuç vermesidir. Diğer önemli bir avantajı da veri setini belli kantillere ayırarak sapan gözlemlerin etkisini ortadan kaldırmak ve bağımlı değişkeni ortalama olarak açıklamak yerine alt, üst ve orta gibi 19 farklı grup kadar açıklayabilmesidir.

Bu çalışmanın amacı Türkiye'deki gelir farklılığını 2002 ile 2010 yıllarına ait TÜİK'in yapmış olduğu kapsamlı hanehalkı veri anketini inceleyerek, hem bu söz konusu yıllar itibari ile hem de bu on yıllık süre içerisinde gelişmişlik düzeyini göstermek bakımından bu yılları bir biri ile kıyaslayarak farklı gelir gruplarındaki, farkı sosyo-kültürel yapıya sahip gelirleri karşılaştırmaktır.

Tez çalışması üç bölümden oluşmaktadır. Birinci bölümde, en çok tercih edilen regresyon yöntemleri olan En Küçük Kareler ile Kantil Regresyonun anlaşılabilmesi adına koşullu ortalama yerine medyanı dikkate alan yöntemlerden olan Edgeworth Çoklu medyan, En Küçük Mutlu Sapmalar Regresyonu ile M Regresyon tanıtılmıştır. Daha sonra bu doğrusal regresyonlar ile Robust regresyonların bir birlerine karşı avantaj ve dezavantajları karşılaştırılmıştır. İkinci bölümde ise yine Kantil Regresyonun anlaşılabilmesi için ilk olarak dağılım fonksiyonları ve kantil dağılım fonksiyonu tanıtılmıştır. Gerekli temel tanımlar yapıldıktan sonra kantil regresyon yönteminin nasıl çalıştığından bahsedilmiştir. Son

bölümde, hanehalkındaki bireylerin gelirini etkileyen etmenleri incelemek üzere yaş, deneyimi ifade etmesi bakımından yaş değişkeninin karesi, bireyin cinsiyeti ve medeni durumu gibi değişkenler ele alınmıştır. Bağımlı değişken olarak ele alınan gelir değişkenin normal dağılmaması üzerine EKK ile birlikte Kantil Regresyon modeli farklı gelir gruplarını açıklayabilecek şekilde 5 ayrı kantille tahminleme yapılmıştır. Bu kurulan tüm modeller ayrıca 2002 yılı ile 2010 yılı karşılaştırılmış ve bu iki yıl arasındaki farklılıklar ortaya konulmuştur.

BİRİNCİ BÖLÜM
REGRESYON MODELLERİ

Çok sayıda alternatif regresyon bulunmakla birlikte bunlar arasında en yaygın olarak kullanılanı en küçük kareler regresyonudur. Tahmin edilen parametre sayısından az olduğu yani serbestlik derecesi fazla ise, veri seti uygun ise, (aykırı gözlem ve farklı varyansın olmaması) Regresyon denkleminin fonksiyonel formu iyi tahminlenmiş ise, kayıp veri yok ise en küçük kareler regresyonu şu durumlarda iyi çalışmaktadır[1].

Veri setleri sıklıkla, değişken sayısının çokluğunun yol açtığı düşük serbestlik derecesi, eksik veri, değişkenler arasındaki güçlü eşdoğrusallık, farklı varyans, doğrusal olmayan karmaşık ilişki, aykırı gözlem ve gözlem sayısının azlığı gibi analiz yapmayı zorlaştıran birkaç özelliğe sahip olabilmektedir. En azından bu problemlerin bazılarının üstesinden gelinebilir. Örneğin eş doğrusallık faktör analiziyle kaldırılabilir ya da bazı dönüşümler yardımıyla yok edilebilir. Diğer problemler eksik veri gibi ve serbestlik derecesi kolaylıkla ortadan kaldırılamayacak ve bazı teknikler bu problemlerden etkilenen veri setleri için uygun olmayabilir[2]. Bu ve benzer gerekçelerden dolayı alternatif regresyon modellerini kullanmak daha doğru bir yol olacaktır.

Bu bölümde tüm alternatif regresyon modellerini anlatmak yerine kantil regresyonun anlaşılmasında yardımcı olabileceği düşünülen en küçük kareler, robust regresyon, en küçük mutlak sapmalar, M regresyon ve Edgeworth Çoklu Medyan gibi alternatif regresyonlar anlatılacaktır.

[1] Lionel C. Briand ve diğerleri, "A Pattern Recognition Approach for Software Engineering Data Analysis", **İEEE Transactions on Software Engineering**, Cilt:18, 1992, ss. 931-932.
[2] Andrew R. Gray ve Stephen G. MacDonell,"A Comparison of Alternatives to Regression Analysis Model Building Techniques to Develop Predictive Equations for Software Metrics"**, The Information and Software Technology**, Cilt:39, 2005, ss. 425-437.

1.1. DOĞRUSAL REGRESYON MODELİ

Regresyon, bağımlı değişken olarak ifade edilen Y ve bağımsız değişken olarak kullanılan X değişkeni ya da değişkenleri arasındaki ortalama ilişkinin matematiksel bir fonksiyonla açıklanması olarak tanımlanabilir[3].

Genel bir doğrusal regresyon modeli

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i \qquad [1.1]$$

Şeklinde ifade edilir. Burada Y_i bağımlı değişkeni, X'ler ise bağımsız değişkeni ifade etmektedir. Modelde yer alan β_0 sabit parametre iken β_i'ler kısmi regresyon parametreleridir.

Diğer taraftan bağımsız değişken ya da değişkenler tarafından bağımlı değişkenler tam olarak açıklanamadığından modele hata terimi de eklemek gerekmektedir. Hata terimi sadece modelin şekli nedeni ile yer almamaktadır[4]. Hata teriminin modelde yer alması, modelde yer alması gereken değişkenlere yer verilmemesi ya da modelde yer almaması gereken bir değişkenin modelde yer alması ve bunlara ek olarak ölçme hataları gibi nedenlerle zaruri hale gelmektedir[5].

Bir regresyon modelinde bağımsız değişkenlerin bağımlı değişkenleri açıklayabilmesi için temel varsayımların sağlanması gerekmektedir. Bu varsayımlar genel olarak hata terimi ile ilgili varsayımlardır, ilk olarak hata terimin rassal dağılımı varsayımı, ikinci olarak hata terimi (u) ile açıklayıcı değişkenler arasındaki varsayım ve son olarak açıklayıcı değişkenlerin kendi aralarındaki ilişkileri ile ilgili varsayımlardır[6]. Bu varsayımlar,

1) $E(\varepsilon_i) = 0$

2) $V(\varepsilon_i) = \sigma^2$

3) $Cov(\varepsilon_i ; \varepsilon_j) = 0$

4) $\varepsilon \sim N(0, \sigma^2)$

---

[3] Şahin Akkaya ve M.Vedat Pazarlıoğlu, **Ekonometri 1**, Anadolu Matbaacılık, İzmir, 2000, s.47.
[4] Russel Davidson ve James Gordon Mackinnon, **Econometric Theory and Methods**, Oxford University Press, 2004, s.101.
[5] Selahattin Güriş ve Ebru Çağlayan, **Ekonometri Temel Kavramlar**, İstanbul: Der Yayınları, 2005, s.87.
[6] Bavos Abraham ve Johannes Ledolter, **Statistical Methods for Forecasting,** Jhon Wiley, Sons,inc, Hoboken, New Jersey,2005, s. 9.

X_i bağımsız değişkenlerinin koşulu altında y_i bağımlı değişkeninin koşullu dağılımı olur ve bunun varsayımları:

1) Koşullu beklenen değer $E(Y_i / X_i) = f(X_i; \beta)$, bağımsız x_i değişkenlerine ve β parametresine, varyans $V(y_i / x_i) = \sigma^2$ bağımsız x_i' lere ve zamana bağlıdır.
2) Bağımlı y_i değişkeni ve y_{i-k} değişkeni arasında farklı zaman aralıkları için korelasyon yoktur.

$$Cov(y_i; y_{i-k}) = E\left[y_i - f(x_i; \beta)\right]\left[y_{i-k} - f(x_{i-k}; \beta)\right] = 0 \qquad [1.2]$$

3) X_i koşulunda Y_i, $f(x_i; \beta)$ beklenen değerli σ^2 varyanslı normal dağılıma sahiptir. $N(f(x_i; \beta); \sigma^2)$ şeklinde gösterilir.

Bu varsayımlar, koşullu dağılan y_i' nin , x_i bağımsız değişkenlerinin bir fonksiyonu olduğunu gösterir[7].

Basit doğrusal regresyon modelinin parametrelerinin örnekten tahmini için tahmincilerin belirlenmesi gerekmektedir. Bağımlı değişken Y_i**,** hata terimi ε_i**'**inin doğrusal fonksiyonudur. Temel varsayımlar nedeni ile ε_i normal dağıldığından Y_i de normal dağılacaktır. Dağılımın ortalaması $\left(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}\right)$ve varyansı σ^2 olmaktadır[8].

$$Y_i \sim N\left(\beta_0 + \beta_1 X_i, \sigma^2\right) \qquad [1.3]$$

Bu nedenle β_0 ve β_1**'**in tahmini Y_i**'**inin ortalamasının tahminidir. Bu tahmincilerin belirlenmesinde en küçük kareler, en çok benzerlik yöntemleri gibi yöntemler kullanılabilir[9].

1.1.1. Robust Regresyon

Parametrik testler bir dağılama bağlı olarak bir veya daha fazla anakütle parameteresini dikkate alan varsayımları ele alır, parametrik olmayan testler ise

[7] Jack Johnston ve John Dinardo, **Econometric Methods**, Mcgraw Hill, Madison, 1997, ss. 6-7.
[8] Ramu Ramanathan, **Introductory Econometrics With Applications**, The Dryden Press, San Diego, 1998, ss.90-92.
[9] Güriş ve Çağlayan, ss, 92-93.

anakütle parametresi hakkında herhangi bir varsayımda bulunmaz. Parametrik ve parametrik olmayan testler veriler tarafından temsil edilen ölçüm düzeylerine dayanır[10].

Parametrik ve parametrik olmayan yöntemler, parametrenin kullanılıp kullanılmasına bağlıdır. Parametre bir ana kütlenin tanımlayıcı sayısal ölçüsüdür[11]. Bunlar oran, ortalama, standart sapma ve varyans gibi ölçülerdir.

Parametrik olmayan regresyon klasik küme teorisi temeline dayanır ve olasılık teorisinin genel yapısını kullanır. Parametrik regresyonda ise gözlem değerlerinden tahmin değerlerinin sapmasının regresyon modelinden ya da diğer rassal ölçüm hatalarından kaynaklandığı varsayılır. En küçük kareler yönteminden önemli farkı, parametrik olmayan regresyonun yalnızca hata terimlerinin sürekliliği gibi genel varsayımları yapmasıdır. Bazı parametrik olmayan regresyon yöntemleri hata terimleri ya da regresyon fonksiyonlarında her hangi bir varsayım ortaya koymamaktadır. Etkili gözlemler regresyon katsayılarının tahminine ve tahmin sonuçlarına önemli ölçüde tesir etmektedir. Aykırı değerlerin dikkatle incelenmesi araştırmalar açısından oldukça önem arz etmektedir. Aykırı değerlerin varlığı regresyon modelinin iyi bir tahmin yapmasını zorlaştırır. Etkili gözlemlerin varlığı durumunda, aykırı değerlerinden en küçük kareler yöntemine göre daha az etkilenen bir regresyon uygunluk yöntemi kullanmalıdır. Literatürde pek çok yöntem önerilmiştir[12].

Bunlardan biri Robust M-tahmin edicisidir. M-tahmin edici regresyon, iteratif olarak çözülür. Regresyon parametrelerinin parametrik olmayan tahmin edicileri tüm örneklem büyüklükleri ve normal olmayan hata modelleri için en küçük kareler tahmin edicilerine göre daha küçük standart hataya sahiptirler. Parametrik olmayan regresyon istatistiksel model kurma için bir alternatif olarak kullanılır. Özellikle gözlem sayısının az olduğu, çoklu bağlantı ve aykırı değerlerin varlığı durumunda regresyon problemi doğrusal olmayan hedef programlama problemi olarak

[10] David J. Sheskin, **Handbook of Parametric and Nonparametric Statistical Procedeures**, CRC Pres Company, Washington D.C., 2004, s. 126.
[11] Fikret İkiz ve diğerleri, **İstatistiğe Giriş**, Barış Yayınları, 2002, s.263.
[12] Kamile Şanlı, "Parametrik Olmayan Robust Regresyon", http://www.yad.org.tr/oturum2.pdf, (07.06.2012), s.1

modellenecek ve parametrik olmayan robust regresyon yöntemiyle çözümleme yapılması tercih edilecektir[13].

Parametreleri elde etmek için kullanılan en yaygın yöntemlerden biri olan En Küçük Kareler Yöntemi bazı varsayımlara dayanmaktadır. Parametrelerin tahminin de hataların normal dağılıma sahip olduğu varsayımı gerçekleştiğinde EKK'ler yöntemi en iyi çözüm olduğu Gauss tarafından gösterilmiştir[14].

Normallik varsayımına duyarsız olan yaklaşımlara robust yaklaşımları denilmektedir. Robust regresyon, parametrik modellerin varsayımlarının karşılanmadığı durumlarda tahminlerin kararlılığını artırmak için dizayn edilmiş istatiksel yöntemlerin ortak bir sınıfıdır. Robust regresyon yöntemi, büyük hataların ağırlıklarını azaltarak bu hataların etkisini düşürmektedir. Aykırı gözlem ve etkili gözlemlerin tespit edilmesi için kullanılan yöntemler robust regresyonun bir parçası olarak ele alınabilecektir. Herhangi bir varsayıma bağlı olmayan özellikle de normallik varsayımına karşı duyarsız olan yaklaşımlar genel olarak "robust" (sağlam) olarak ifade edilmektedir[15].

Regresyon analizinde en küçük kareler yöntemi, gözlem değerleri, değişkenler ve hata terimi hakkında birtakım varsayımların sağlandığı durumlarda geçerlilik kazanır. Bu varsayımlar geçerli olmadıkça yapılmış olan hesaplamaların ve elde edilmiş olan regresyon denklemlerinin istatistikî bir değeri olamaz. Çünkü varsayımların bozulmalarının bu değerler üzerine çok önemli etkileri olabilmektedir. Varsayımların tutmaması elde edilen modelin, populasyonu iyi temsil etmediğini göstermektedir. Buna bağlı olarak elde edilen regresyon denkleminden yapılacak tahminlerin hatalı olma ihtimali yüksek olur. Hata terimi normal dağılış göstermemesi durumunda farklı teknik olan robust rergresyon tekniğini kullanmak zorunlu hale gelecektir[16].

---

[13] Kamile Şanlı, Yöneylem Araştırması Derneği, "YA/EM Doktora Öğrencileri Kolokyumu", 16.07.2011,http://www.yad.org.tr/oturum2.pdf

[14] Julian J. Faraway, **Extending the Linear Model With R: Generalized Linear, Mixed Effects and Nonparametric Regression Models**. Boca Raton: Chapman & Hall. CRC, 2005, ss. 229.

[15] Latif Öztürk. **Doğrusal Regresyonda Sağlam Kestirim Yöntemleri ve Karşılaştırılmaları**, (Yayınlamış Doktora Tezi), Mimar Sinan Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 2003.

[16] Chen Haifeng ve Meer Peter, "Robust Resgression With Projection Based M- Estimators", **Proceedings of the Ninth IEEE International Conference on Computer**, 2003 s. 1.

1.1.2. Edgeworth Çoklu Medyan

Boscovich tarafından basit iki değişkenli regresyon modelinde en iyi doğruyu elde etmek üzere minimum mutlak hata toplamını ilk kez formüle etmiş ve uygulamıştır[17].

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_{i1}| \qquad [1.4]$$

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1}) = 0 \qquad [1.5]$$

Boscovich, veri merkezinden geçen bir doğruyu kısıtlamayı önermiştir. Daha sonra β_1'in optimal seçimini hesaplamaya dayanan bir geometrik algoritma geliştirmiştir.

$$Z(\beta_1) = \sum_{n=1}^{N} |\beta_1 (x_n - \bar{x}) - (y_n - \bar{y})| \qquad [1.6]$$

Aslında bu metod çok yeni bir metod değildir. 1760'lı yıllarda Rudjer Boscovich bir enlem derecesinin boyunu beş gözlem kullanarak y_i enlem boylarını θ_i bunların farklı açılarını ifade etmek üzere aşağıda minimizasyon problemiyle dünyanın elipsliğinin tahminlemesini önermiştir.

$$\min |y_i - \alpha - \beta \sin^2 \theta_i| \qquad [1.7]$$

Ortalama hata kısıtına bağlı olarak $n^{-1}\sum (y_i - \hat{\alpha} - \hat{\beta} \sin^2 \theta_i)$ sıfıra eşitlenir. Daha sonra Laplace bu problemin çözümünü ağırlıklandırılmış medyan hesaplamasıyla çözülebileceğini göstermiştir. Edgeworth metod ya da çoklu (plural) medyan metot, 18. yy sonlarına kadar Legendre ve Gauss tarafından bulunan, üstün saltanatı devam eden en küçük kareler yaklaşımına doğrudan rakip olarak Boscovich metod egemenliğine son vermeyi amaçlamıştır.

Edgeworth, denklem [1.5] deki katsayıları kısıtlamaya gerek kalmadan genel bir uygulama sunmuştur. Burada mutlak hata yaklaşımıyla medyan kısıtını tartışarak, hataların sıfır ortalama kısıtını azaltmayı önermiştir. Laplace' in hatalara ilişkin bu teorisinin cazibesi, (çoklu) plural medyan hataların Gaussian kurallara uygun

[17] Elvezio Ronchetti, **Statistical Data Analysis Based on The L Norm and Realted Methods, Bounded Influence in Regression**, North Holland, 1987, ss.65-80.

hatalardan ziyade üst kuyrukların olduğu ve daha çok uygunsuz gözlemler olduğu zaman en küçük kareler tahmincilerinden daha doğru sonuçlar vermesindendir[18].

Edgeworth, regresyon parametrelerinin seçimi için basit bir model geliştirmiştir. Parametrelerin optimal değerlerini göstermek için Laplace'in prosedürünü kullanmıştır[19]. Bazı uygunsuz şartlarda çoklu medyan içeriğini Laplace ölçeğini biraz genişletebilmiş ama ortalamadan ziyade medyanın üstünlüğü iddiasını destekleyecek her hangi bir matematiksel desteği sağlayamamıştır. Laplace regresyondaki gibi, çoklu ve tek değişkenli medyan için optimizasyon problemi arasındaki benzerlik kuşkusuz tüm kalıntılarda daha zorlayıcı olacaktır[20].

1.1.3. En Küçük Mutlak Sapmalar (LAD) Regresyonu

En küçük mutlak sapma (LAD) 1757 yılında Roger Joseph Boscovich tarafından en küçük kareler metodundan 50 yıl önce tanıtılmıştır. En küçük kareler metodunda parametreler hata kareler toplamını $\sum_{i=1}^{n} \hat{e}_i^2$ en küçük yapacak değerden seçilir. En küçük mutlak sapma metodunda ise parametreler hataların mutlak değer toplamını $\sum_{i=1}^{n} |\hat{e}_i|$ en küçük yapacak olan olası değerden seçer. Yani en küçük mutlak sapma tahmincileri α ve β aşağıdaki denklemin minimizasyonu sağlanan a ve b değerleridir[21].

$$\sum |y_i - (a + bx_i)| \qquad [1.8]$$

Burada $y_i - (a + bx_i)$ farkı $\hat{Y} = (a + bx_i)$ doğrusundan, gözlemlerin koordinat düzleminde meydana getirdiği (x_i, y_i) ikililerden sapması olarak adlandırılır. LAD tahmin kavramı en küçük kareler kavramından çok daha zor değildir. Gelişen bilgisayar teknolojisiyle şimdiki tahmin hesaplamaları, LAD metodu daha karmaşık

[18] Robert V. Hawley ve Neal C. Gallagher, "On Edgeworth's Method for Minimum Absolute Error Linear Regression", **İEEE Transaction on Signal Processing**, Cilt:42, 1994, ss. 2045-2054.
[19] Bijan Bidabad," L_1 Norm Computational Algorithms,18.12.2012", http://www.bidabad.com/doc/l1-article6.pdf
[20] Roger Koenker ve Galton, Edgeworth, "Frisch and Prospect for Kantile Regression in Econometrics", **Journal of Econometrics**, Cilt:95, 2000, ss.347-374.
[21] Peter J. Rousseeuw, "Least Median of Squares Regression", **Journal of the American Statistical Association**, Cilt:79, 1984, ss.871-880.

hale gelmiştir. LAD metodu için bir formülasyon yoktur, bunun yerine hesaplamalara dayanan algoritmalar mevcuttur[22].

Algoritmada amaç en küçük toplam mutlak sapma yönüne sahip en iyi doğruyu elde etmektir. Algoritmanın temeli verilen her (x_0, y_0) noktaları için tüm doğrular arasından en iyi doğrudan geçen noktaları bulma prosedürüdür. Bu prosedür iki veri noktası arasından geçen LAD regresyonu doğrusu ile birlikte kullanılır. Bu yüzden algoritma (x_1, y_1) noktalarından biri ile başlar ve içinden geçebileceği en iyi doğruyu bulur. Bu doğru diğer veri noktalarından da geçebilir. Daha sonraki noktalar (x_2, y_2) şeklinde gösterilir ve bu noktaların geçebilecekleri en iyi doğru bulunur. Daha sonra aynı işlemler (x_3, y_3) noktaları ve onun geçebileceği en iyi doğru için de devam ettirilir. Sonuçta en son doğru bir önceki doğrulara benzer şekilde elde edilebilecektir. Hangi noktalardan geçtiklerine bakılmaksızın tüm doğrular arasında bu en iyi doğru olacaktır. Yani bu LAD regresyon doğrusu olacaktır[23].

Şimdi verilen (x_0, y_0) noktalarından geçen tüm doğrular arasından en iyi doğruyu bulmak için bir prosedür tanımlanmalıdır. Verilen her bir (x_i, y_i) noktaları için (x_0, y_0) ve (x_i, y_i) noktalarının içinden geçen doğrunun eğimi $(x_i - y_i)/(x_0 - y_0)$ olarak hesaplanır. Eğer bazı "i" ler için $x_i = x_0$ ise eğim tanımlanamaz ise bazı noktalar önemsenmeyebilir. Veri noktaları şöyle gösterilebilir. $(y_1 - y_o)/(x_1 - x_0) \le (y_2 - y_0)/(x_2 - x_0) \le \ldots \le (y_n - y_o)/(x_n - x_0)$. sonra $T = \sum |x_i - x_0|$ elde edilir.

Aşağıdaki koşulların karşılanmasıyla k endeksi elde edilir.

$$|x_1 - x_0| + \ldots + |x_{k-1} - x_0| < \frac{1}{2}T \qquad [1.9]$$

$$|x_1 - x_0| + \ldots + |x_{k-1} - x_0| + |x_k - x_o| > \frac{1}{2}T \qquad [1.10]$$

[22] Brian S. Cade ve Jon D. Richards. "Permutation Test For Least Absolute Deviation Regression", **International Biometric Society**, Cilt:52, 1996, ss.886-902.
[23] Steven P. Ellis, "Instability of Least Squares, Least Absolute Deviation and Least Median of Squares Linear Regression", **Statistical Science**, Cilt:13, 1998, ss. 337-344.

(x_0, y_0) noktaları arasından geçen en iyi doğru $\hat{Y} = \alpha^* + \beta^* X$ doğrusudur ve buradan da katsayılar şöyle elde edilecektir.

$$\beta^* = \frac{y_k - y_0}{x_k - x_0} \qquad [1.11]$$

$$\alpha^* = y_0 - \beta^* x_0 \qquad [1.12]$$

$\sum \left| y_i - (a + bx_i) \right|$ denkleminin değeri mutlak sapma toplamıdır ve varsayılan doğru her veri noktasından geçmemektedir. Eğer doğru bir miktar yukarı çıkarsa ε kadar hata oluşur, ε kadar mutlak sapma artış ya da azalış gösterir. Doğrunun yukarı ya da aşağıya kaymasına göre $(x_0 - y_0)$ noktalarından geçen en iyi doğruyu bulmak için işlenen süreç aşağıdaki gibi açıklanmıştır[24].

Eğer doğru sadece bir veri noktasından geçerse, ikinci bir noktayla karşılaşıncaya kadar saat yönünde ya da saat yönün tersine kendi veri noktasına bağlı olarak döndürülebilir. Böylece mutlak sapma bir nokta için sıfır olarak kalacaktır ve diğer mutlak sapmaların her biri çeşitli miktarlara göre hem azalabilir hem de artabilir. Toplam mutlak sapmanın artış ya da azalışına göre saat yününde ya da saat yönün tersine döndürme, tersi bir etki gösterecektir. Bu yolla ya da aksi yönde döndürme, toplam mutlak sapmayı azaltacak ya da en azından artırmayacaktır. Toplam mutlak sapmayı minimize etmek için en azından iki veri noktasından geçen doğrulara gereksinim duyulur.

Algoritmanın Gerekçeleri: (x_0, y_0) noktalarından geçen en iyi doğruyu bulma prosedürü aşağıdaki gibi gösterilebilir. $(x_0 - y_0)$ noktalarından geçen tüm doğrular arasından $\sum \left| y_i - (a + bx_i) \right|$ denklemini minimize eden bir doğru bulmak istenir. $(x_0 - y_0)$ noktaları arasından geçen $\hat{Y} = a + bX$ doğrusu için $y_0 = a + bx_0$ olur, bunun sonucu olarak $a = y_0 - bx_0$ olacaktır.

[24] Peter Bloomfield ve William L. Steiger, **Least Absolute Deviations Theory, Applications, and Algorithms**, Boston, 1983, ss.152-172.

$y_i - (a + bx_i)$ daki sapma, $(y_i - y_0) - b(x_i - x_0)$ olarak yazılabilir. Bu yüzden b'nin bir fonksiyonu olarak $\sum\left|(y_i - y_0) - b(x_i - x_0)\right|$ denklemini mimimize eden b değeri bulunmak istenir. Bir fonksiyonun minimumunu bulmak için kullanılan ortak teknik o fonksiyonun türevini almaktır. Fakat, |t| mutlak değer fonksiyonu t=0 olduğunda türevi alınamaz. Bu da $(y_i - y_0) - b(x_i - x_0)$ da b değeri haricindeki tüm b değerlerinde türev alınabilir yani $b_i = (y_i - y_0)/(x_i - x_0)$ değeri haricinde. (1.9) eşitsizliği (1.7)'ün türev alınabilir koşulunu $b > \beta^*$ için pozitif ve $b < \beta^*$ değeri için negatif koşulunu göstermektedir.

(x_k, y_k) noktası ile (x_0, y_0) noktasından geçen en iyi doğru dikkate alınır. Bunu doğrulamak için $y_k = \alpha^* + \beta^*$ olduğunu göstermek için denklem (1.9) daki α^* ve β^* tanımları kullanılabilir.

LAD tahmincileri $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2 \ldots \hat{\beta}_k$ olarak ele alındığında mutlak hata değerleri toplamını $\sum|\hat{e}_i|$ en küçük yapan değerlerden seçilir. Yani bu tahminciler b_0, b_1, b_2...b_k değerlerinin minimizasyonu ile olmaktadır.

$$\sum\left|y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3})\right| \qquad [1.13]$$

Minimizasyonu sağlamak için herhangi bir formülasyon bulunmamasına karşın, söz konusu tahmincilerin elde edilmesinde kullanılabilecek algoritmalar tanımlanabilir. Algoritma, ele alınan veri setinin hiçbir bozulma ya da benzersiz olmama problemini içermediğini varsaymaktadır.

Algoritmanın işleyişini kolayca açıklamak için vektör notasyonu kullanılacaktır.

$$x_i = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix} \qquad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \text{ ve}$$

Bundan sonra (1.12)'deki mutlak sapma toplamı aşağıdaki gibi yazılabilir

$$\sum|y_i - b' x_i| \qquad [1.14]$$

Buradaki amaç (1.13)'nin minimizasyonunu sağlayan **b** vektörünü bulmak olacaktır.

LAD regresyonundaki algoritmalar genel olarak şu şekilde sıralanabilir: Öncelikle herhangi bir doğrudan başlanır, sonra iyi doğru bulunur daha sonra daha iyi bir doğru elde edilir ta ki en iyi doğruya ulaşıncaya kadar devam edilir. Çoklu LAD regresyonunda benzer iterasyonlar devam ettirilir. Başlangıç **b** vektörü ile yapılır sonra daha iyi vektör bulununcaya kadar devam ettirilir en iyi vektör $\hat{\beta}$ elde edilinceye kadar devam edilir. Her bir adımda b tahmincilerinin bir vektörü olur sonra ilk olarak uygun yöne sahip **d** vektörü bulunur ve $b^* = b + td$ için bulunan t değeri en iyisi olacaktır.

"d" yönünde en iyi tahminciler vektörünün bulunmasında (1.15) değerini minimize eden bir sürece ihtiyaç duyulur.

$$\sum \left| y_i - (b + td)' x_i \right| \qquad [1.15]$$

Eğer $z_i = y_i - b' x_i$ ve $w_i = d' x_i$ yazılırsa, sonra süreç (2.4)'ü minimize eden t değeri elde edilebilir.

$$\sum \left| z_i - t w_i \right| \qquad [1.16]$$

Bu denklem (1.13)'i minimize eden b değerinin bulunmasında karşılaşılan probleme benzemektedir. z_i / w_i oranı alınarak artan değere göre sıralamaya koyulur. Bu sıralamaya göre w ve z değerleri yeniden endekslenerek k değerleri elde edilir.

$$\left|w_1\right| + \left|w_2\right| + \ldots + \left|w_{k-1}\right| < \frac{1}{2} T \qquad [1.17]$$

$$\left|w_1\right| + \left|w_2\right| + \ldots + \left|w_{k-1}\right| + \left|w_k\right| > \frac{1}{2} T \qquad [1.18]$$

Burada $T = \sum \left| w_i \right|$ t'nin minimize değeri z_k / w_k olmaktadır.

Uygun yönün bulunması: Her bir adımda, algoritma d_1, d_2, d_3, d_4'e kadar dört yönsel vektör olarak ele alınabilir. Genel durumlarda p açıklayıcı değişken ve onunla ilgili p+1 vektör ele alınır. "d_j" pozitif yönünün yanı sıra "$-d_j$" negatif yönde türev alınır ve her bir d_j vektöründen dolayı sekiz farklı yön sunulur.

Bu sekiz yön arasından en umut verici yön (1.15) eşitliğini t=0 noktasına en çabuk yaklaştırarak (en dik şekilde) düşürendir. (1.15) eşitliğinin ne kadar hızlı

düşürüleceğini belirlemek amacıyla t = 0 noktasında eşitliğin sağ taraf türevi hesaplanır. (2.4)'teki notasyon terimlerine göre t = 0 noktasındaki sağ taraf türevi $W_- + W_0 - W_+$'dır. Burada W_-, z_1 / w_1 in negatif olduğu i indisleri için $|w_i|$'lerin toplamıdır, W_0 $z_i = 0$ için $|w_i|$'lerin toplamıdır ve W_+ ise z_i / w_1'in pozitif olduğu $|w_i|$ 'lerin toplamıdır.

Sekiz yönün her biri için bu türev hesaplanır ve türevi en negatif olan en uygun yön olarak seçilir. Eğer bütün türevler pozitif ise mevcut **b** vektörü en iyi $\hat{\boldsymbol{\beta}}$ vektörüdür ve algoritma sonlandırılır.

LAD regresyonunda katsayı tahminleri ele alındığında ilk olarak tahminciler α,β ve sonra da $\hat{e}_i = y_i - \left(\hat{\alpha} + \hat{\beta} x_i\right)$ hesaplanır. Serbestlik derecesi n-2 alınır ve sıfır dışındaki hatalar büyükten küçüğe $\hat{e}_{(m)} > \hat{e}_{(2)} > \hat{e}_{(1)}$ doğru sıralanır.

k_1 değeri $\dfrac{m+1}{2-\sqrt{m}}$ değerine en yakın tamsayı değeridir.

k_2 değeri ise $\dfrac{m+1}{2+\sqrt{m}}$ değerine en yakın tamsayı değeridir.

Daha sonra $\hat{\tau}$ değeri hesaplanır.

$$\hat{\tau} = \frac{\sqrt{m}\left[\hat{e}_{(k2)} - \hat{e}_{(k1)}\right]}{4} \qquad [1.19]$$

$$S_{(\hat{\beta})} = \frac{\hat{\tau}}{\sqrt{\sum \left(x_i - \overline{x}\right)^2}} \qquad [1.20]$$

Test istatistiği ise

$$|t| = \frac{\left|\hat{\beta}\right|}{S_{(\hat{\beta})}} \qquad [1.21]$$

Testin p-değeri Prob[T≥|t|] olasılığı olarak hesaplanır. Burada T n-2 serbestlik dereceli t dağılımının bir rassal değişkenini göstermektedir[25].

[25] David Birkes ve Yadolah Dodge, **Alternative Methods of Regression,** John Willey & Sons, Canada, 1993 s. 85.

1.1.4. M Regresyon

M-regresyon ve Doğrusal Mutlak Sapma (LAD) regresyonu robust istatistiğinin bir parçasıdır. İstatistiksel modelin varsayımları sağlanmadığında dahi makul ölçüde ve iyi performans gösteren bir istatistiksel prosedür robust olarak ifade edilmiştir. Eğer eldeki veriler doğrusal regresyon modeline uygunsa en küçük kareler (EKK) tahminleri ve testi iyi sonuçlar verecektir ancak rassal hatalar için normallik varsayımı geçersiz ise robust durumdan bahsedilemeyecektir. M-regresyon 1964 'te Peter Huber tarafından tanıtılan M-tahmin düşüncesine dayandırılarak bu varsayıma karşı güçlülüğü kazandırmak amacıyla geliştirilmiştir[26].

En küçük kareler tahmininde $\hat{\alpha}$ ve $\hat{\beta}$ 'nın seçimi $\sum e_i^2$ 'yi olabildiğince küçük yapacak şekilde seçilir. En küçük mutlak sapma tahmininde ise bu seçim $\sum |\hat{e}_i|$ mümkün olabildiğince küçük yapılmaya çalışılarak yerine getirilir. M-tahminde bu düşünce genelleştirilir ve $\hat{\alpha}$ ve $\hat{\beta}$ 'nın seçimi $\sum \rho(\hat{e}_i)$ ifadesini mümkün olduğunca küçültecek şekilde gerçekleştirilir. Burada $\rho(e)$, e 'nin bir fonksiyonudur. En küçük kareler ve en küçük mutlak sapma tahmini, $\rho(e) = e^2$ ve $\rho(e) = |e|$ olarak M-tahminin özel durumları olarak görülebilir [27].

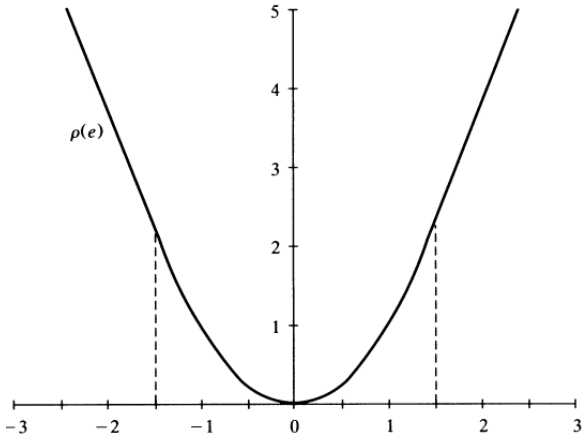
M-tahmin: M-tahmin, e^2 ve $|e|$ arasında bir uzlaşı sağlamak amacıyla $\rho(e)$ fonksiyonunu kullanır. LAD tahminlerinin EKK tahminlerine göre ana avantajı sapan gözlemlere çok duyarlı değillerdir. Ancak sapan gözlem bulunmadığında EKK tahminleri daha doğru olabilecektir. İki metodun avantajları, "e" sıfıra yakın olduğunda $\rho(e)$ 'nin "e^2 " olarak alınması ve "e" sıfırdan uzak olduğunda ise $\rho(e)$ 'nin $|e|$ olarak alınmasıyla birleştirebilir. Daha spesifik olarak aşağıdaki eşitlik kullanılır[28].

[26] Norman R. Draper ve Harry Smith, **Applied Regression Analysis,** John & Sons, Canada, 1998. ss. 567-568
[27] Brikes ve Dodge, s. 86.
[28] Birkes ve Dodge, s. 87.

$$\rho(e)=\begin{Bmatrix} e^2,\ eğer -k \le e \le k \\ 2k|e|-k^2,\ eğer\, e<-k\, veya\, k<e \end{Bmatrix} \qquad [1.22]$$

Huber 'in tavsiyesi izlenilerek $k=1.5\hat{\sigma}$ olarak alınabilir. Burada $\hat{\sigma}$, rassal hatalar populasyonunun standart sapması σ'nın tahminidir. $\rho(e)$'yi sürekli bir fonksiyon yapmak amacıyla $|e|$ yerine $2k|e|-k^2$ kullanılmıştır. Bu fonksiyonun grafiği Şekil 1 'de gösterilmiştir.

Şekil 1: $\boldsymbol{k=1.5\hat{\sigma}}$ iken Huber 'in M-tahmini tanımında kullanılan $\boldsymbol{\rho(e)}$ fonksiyonunun grafiği.



Kaynak: Norman R. Draper ve Harry Smith, **Applied Regression Analysis,** John & Sons, Canada, 1998. ss. 567-568

σ'yı tahminlemek için $\hat{\sigma}=1.483MAD$ kullanılır. Burada MAD, mutlak sapmaların $|\hat{e}_i|$ medyanıdır. 1.483 çarpanının seçilmesinin nedeni $\hat{\sigma}$'yı rassal hataların dağılımının normal olması durumunda σ'nın iyi bir tahminleyicisi olarak elde etmektir.

Huber' in M-tahminleri $(\hat{\alpha}\ ve\ \hat{\beta})$ aşağıdaki fonksiyonu minimize eden a ve b değerleridir.

$$\sum \rho\left(y_i-(a+bx_i)\right) \qquad [1.23]$$

"a" ve "b", ρ fonksiyonunun argümanı olarak (1.22)'de açık bir şekilde görülmesinin yanı sıra ρ tanımından dolaylı olarak görülmektedir. ρ fonksiyonu $k = 1.5\hat{\sigma}$'ı içermektedir ve $\hat{\sigma}$ sapmalardan $\left(y_i - \left(a + bx_i\right)\right)$ aracılığıyla hesaplanmaktadır. Şimdi, sırada (1.23) denklemini minimize edecek bir algoritmanın geliştirilmesine ihtiyaç vardır[29].

Algoritmayı başlatırken $\hat{\alpha}$ *ve* $\hat{\beta}$ 'nın başlangıç tahminleri olarak en küçük kareler tahminleri alınmaktadır. Bu değerlerin sapmaları ve σ 'nın tahminini hesaplamak için kullanılacaktır. Ardından $\hat{\alpha}$ *ve* $\hat{\beta}$'nın geliştirilmiş tahminleri elde edilecektir. Bu geliştirilmiş tahminler yeni sapmaları ve σ'nın geliştirilmiş bir tahminini hesaplamada kullanılacaktır. Daha sonra yeni sapmalar ve yeni $\hat{\sigma}$, $\hat{\alpha}$ *ve* $\hat{\beta}$'nın geliştirilmiş tahminleri elde etmede kullanılacaktır. Geliştirilmiş tahminler önceki tahminlerle aynı (veya en azından yaklaşık olarak aynı) oluncaya kadar algoritmanın bu şekilde iterasyona devam edilir.

Daha spesifik olarak, algoritmanın herhangi bir adımı için $\hat{\alpha}$ *ve* $\hat{\beta}$'nın o anki tahminleri a^0 ve b^0 olsun. $\left(y_i - \left(a + bx_i\right)\right)$ sapmaları ve bu sapmalardan $\hat{\sigma}^0 = 1.483MAD$ hesaplanır. Ardından, büyük sapma değerlerinden kurtulmak amacıyla y değerlerinin düzeltilmesi yapılır. Mevcut tahminlenen regresyon doğrusundan y_i'nin sapmaları $e_i^0 = \left(y_i - \left(a^0 + b^0 x_i\right)\right)$ şeklinde hesaplanır. Böylece $y_i = a^0 + b^0 x_i + e_i^0$'dır. Sonra şu tanım yapılır: $y_i^* = a^0 + b^0 x_i + e_i^*$. Burada e_i^*, e_i^0 değerinin kırpılmasıyla elde edilen düzeltilmiş sapmadır. Bu sapma düzeltmesiyle mutlak değer olarak $1.5\hat{\sigma}^0$'den daha büyük sapmaya izin verilmez. Başka bir deyişle, eğer e_i^0 sapması $-1.5\hat{\sigma}^0$ ve $1.5\hat{\sigma}^0$ arasında ise $e_i^* = e_i^0$ (böylece $y_i^* = y_i$)'dır. Eğer e_i^0 sapması $-1.5\hat{\sigma}^0$'dan küçükse $e_i^* = -1.5\hat{\sigma}^0$'dır ve e_i^0 sapması $1.5\hat{\sigma}^0$'dan büyükse $e_i^* = -1.5\hat{\sigma}^0$'dır. Ardından $y_1^*, y_2^* ... y_n^*$ düzeltilmiş verisinden en küçük kareler tahmini yapılarak $\hat{\alpha}$ *ve* $\hat{\beta}$'nın geliştirilmiş tahminleri elde edilmiş olunur[30].

[29] Birkes ve Dodge, s.347.
[30] Birkes ve Dodge, s.349.

Algoritma akla yatkın görünse de (1.23)'u nasıl minimize edeceği açık olmayabilir. $\hat{\sigma}$ 'nın sabit tutulup (1.23)'un a ve b 'e göre türevlerinin alınarak 0 'a eşitlenmesiyle minimizasyon yapılır. Bu işlem, aşağıdaki a ve b 'den oluşan iki bilinmeyenli iki denklemin elde edilmesine neden olacaktır.

$$\sum x_i \rho'\left(y_i - \left(a + bx_i\right)\right) = 0 \qquad [1.24]$$

Dikkat edilirse $\rho'(e)$ türevi $1.5\hat{\sigma}$ 'ya eşit veya daha küçük tüm e değerleri için $-3\hat{\sigma}$ değerine ve $1.5\hat{\sigma}$ 'ya eşit veya daha büyük tüm e değerleri için $3\hat{\sigma}$ değerine sahiptir. Böylece eğer $e_i = y_i - \left(a + bx_i\right)$ sapmaları kırpılmış e_i^* sapmalarıyla yer değiştirirse (1.24)'ün çözümü aynı kalacaktır. Burada e_i, $-1.5\hat{\sigma}$ ve $1.5\hat{\sigma}$ arasında ise $e_i^* = e_i$ 'dır ve e_i, $-1.5\hat{\sigma}$ 'dan küçükse $e_i^* = -1.5\hat{\sigma}$ olarak alınır ve e_i, $1.5\hat{\sigma}$ 'dan büyükse $e_i^* = -1.5\hat{\sigma}$ 'a eşittir. Başka bir deyişle, (1.24)'ün çözümünü değiştirmeksizin düzeltilmiş $y_i^* = a^0 + b^0 x_i + e_i^*$ değerleri y_i değerleriyle yer değiştirebilir. Böylece (1.23)'ü minimize eden a ve b değerlerinin değiştirilmesine gerek kalmaz. Düzeltmenin bir sonucu olarak $\rho\left(y_i^* - \left(a_i + b_i x_i\right)\right) = \left[y_i^* - \left(a_i + b_i x_i\right)\right]^2$ 'dir. $\sum\left[y_i^* - \left(a_i + b_i x_i\right)\right]^2$ ifadesinin minimize edilmesi tanım gereği düzeltilmiş veriden elde edilen en küçük kareler tahminlerini verir.

Çoklu regresyonda M-tahminin bulunması prosedürü, basit regresyonda tanımlanan prosedürün doğrudan genelleştirilmiş halidir. $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p$ Huber M-tahmini değerleri aşağıdaki fonksiyonu minimize eden $b_0, b_1, ..., b_p$ değerleridir.

$$\sum \rho\left(y_i - \left(b_0 + b_1 x_{i1} + ... + b_p x_{ip}\right)\right) \qquad [1.25]$$

Burada $\rho(e)$, (1.22)'de tanımlanan fonksiyondur. Aşağıdaki vektör notasyonunun kullanılması uygun olacaktır.

$$b = \begin{bmatrix} \\ b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} \text{ ve } x_i = \begin{bmatrix} \\ 1 \\ x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix}$$

Huber M-tahmininin $\hat{\beta}$ vektörü $\sum \rho\left(y_i - b'x_i\right)$ fonksiyonunu minimize eden **b** olarak tanımlanır.

β ile gösterilen regresyon katsayılar vektörü ilk olarak en küçük kareler tahmin vektörü tarafından tahminlenir. β 'nın bu başlangıç tahmini sapmaları ve σ 'nın bir tahminini hesaplamak için kullanılır. Böylece β 'nın geliştirilmiş bir tahmini elde edilir. Algoritmaya bu şekilde geliştirilmiş β tahmininin önceki adımda yapılan geliştirilmiş β tahminiyle aynı (veya en azından yaklaşık olarak aynı) oluncaya kadar devam edilir.

Daha spesifik olmak gerekirse, algoritmanın herhangi bir adımında o anki geliştirilmiş β tahmini b^0 olsun. $y_i\left(b^0\right)' x_i$ ifadesinden sapmalar hesaplanır ve bu sapmalar kullanılarak da $\hat{\sigma} = 1.483MAD$ bulunur. Ardından büyük sapma değerlerinden kurtulmak için y değerlerinin düzeltmesi yapılır. Mevcut tahminlenen regresyon doğrusundan y_i değerlerinin sapmaları $e_i^0 = y_i\left(b^0\right)' x_i$ denklemiyle hesaplanır. Böylece $y_i = y_i\left(b^0\right)' x_i + e_i^0$ olur. Mutlak değerce $1.5\hat{\sigma}$ 'dan daha büyük olan sapmaya sahip olan e_i^0 'lar kırpılarak düzeltilmiş e_i^* sapması elde edilir ve $y_i^* = y_i\left(b^0\right)' x_i + e_i^*$ değerleri hesaplanır. Artık $y_1^*, ..., y_n^*$ düzeltilmiş verisinden en küçük kareler tahmini yapılarak β 'nın geliştirilmiş tahmini elde edilir.

$Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + e$ genel doğrusal regresyon modelinde $\beta_{q+1} = ... = \beta_p = 0$ testinin en küçük kareler test istatistiği hatırlanırsa aşağıdaki gibidir.

$$F_{LS} = \frac{SSR_{sınırlandırılmış} - SSR_{sınırlandırılmamış}}{(p-q)\hat{\sigma}^2_{LS}} \qquad [1.26]$$

Burada, SSR artık kareler toplamına karşılık gelmektedir ve $SSR = \sum \hat{e}_i^2$ ve $\hat{\sigma}^2_{LS} = \sum \hat{e}_i^2 / (n-p-1)$'dir. $SSR_{kısıtlanmış}$ ve $SSR_{kısıtlanmamış}$'daki artıklar sırasıyla $Y = \beta_0 + \beta_1 X_1 + ... + \beta_q X_q + e$ düşürülmüş modeline ve $Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + e$ tüm modele en küçük kareler metodunun uygulanmasıyla hesaplanır. $\hat{\sigma}^2_{LS}$ hesaplanırken tüm modelden elde edilen artıklar kullanılır.

Benzer bir test istatistiği M-regresyonda aşağıdaki gibi kullanılır.

$$F_M = \frac{STR_{kısıtlanmış} - STR_{kısıtlanmamış}}{(p-q)\hat{\lambda}} \qquad [1.27]$$

Burada, STR dönüştürülmüş artıkların toplamına karşılık gelmektedir, $STR = \sum \rho(\hat{e}_i)$ ve $\hat{\lambda} = (n/m)\sum e_i^2 / (n-p-1)$'dir. m tamsayısı kırpma işleminin gerek duyulmadığı $(|\hat{e}_i| \le 1.5\hat{\sigma})\hat{e}_i$ artıklarının sayısını göstermektedir. $STR_{kısıtlanmış}$ ve $STR_{kısıtlanmamış}$'daki artıklar sırasıyla düşürülmüş ve tüm modele uygulanan M-regresyon prosedüründen hesaplanmıştır. Dönüştürülmüş modelin tahminleme prosedürü daha önce anlatılanlardan biraz daha farklıdır. σ'nın tahmini iterasyonlu bir şekilde bulunmaz. Bunun yerine düşürülmüş modelin regresyon katsayılarının M-tahmin vektörünü elde etmek için yapılan tüm iterasyonlar boyunca tüm modelden elde edilmiş $\hat{\sigma}$ tahmini değişikliğe uğratılmadan kullanılır. $\hat{\lambda}$ hesaplanırken tüm modelden elde edilen artıklardan faydalanılır.

Testin yaklaşık p-değeri en küçük kareler testinde olduğu gibi hesaplanır. $prob[F \ge F_M]$ ifadesinde F, p-q ve n-p-1 serbestlik dereceli F dağılımına sahip rassal bir değişkeni göstermektedir. F_M formülasyonu F_{LS} formülasyonuna oldukça benzerdir. Aslında e_i^* ve $\rho(e)$ tanımında kullanılan 1.5 katsayısını ∞ ifadesiyle yer değiştirirsek F_M tam olarak F_{LS}'e eşit olur. Dikkat edilirse eğer 1.5 katsayısı ∞ ifadesiyle yer değiştirirse tüm i değerleri için $e_i^* = \hat{e}_i$ ve m=n olur ve böylece $\hat{\lambda}$ ile $\hat{\sigma}^2_{LS}$ çakışır. Ayrıca $\rho(e) = e^2$ olur böylece STR=SSR ve F_M ile F_{LS} çakışır.

Yukarıda çoklu doğrusal regresyon modelinde $\beta_{q+1} = ... = \beta_p = 0$ 'ın nasıl test edileceği tanımlanmıştır. $\beta = 0$'ın test edilmesi ise p=1 ve q=0 olduğu özel bir duruma karşılık gelmektedir. Test istatistiği F_M (1.27) formülasyonundaki gibidir. Yaklaşık p-değeri $prob\left[F \geq F_M\right]$ olarak hesaplanabilir. Burada F, 1 ve n-2 serbestlik dereceli F dağılımına sahip rassal bir değişkeni göstermektedir. Ayrıca t ifadesi n-2 serbestlik dereceli t dağılımlı bir rassal değişkeni gösterir ve $\left|t_M\right| = \sqrt{F_M}$ iken $\Pr ob\left[\left|t\right| \geq \left|t_M\right|\right]$ olasılığından yaklaşık p-değeri hesaplanabilir.

İKİNCİ BÖLÜM

KANTİL REGRESYON

Kantil fonksiyonu tek değişkenli dağılımların karşılaştırılması ve tanımlanması için yeterlidir. Fakat bağımsız değişkenler ile bir bağımlı değişkeni arasındaki ilişkiyi modellediğimiz zaman, kantil regresyon modeli (QRM) ve kantil fonksiyonu için bir regresyon model türünü de tanımlamak gerekli hale gelmektedir. Birden fazla bağımsız değişken verildiğinde doğrusal regresyon modeli koşullu ortalama fonksiyonunu belirlerken, kantil regresyon modelleri ise koşullu kantil fonksiyonlarını belirler.

Referans noktası olarak doğrusal regresyon modelini kullanarak kantil regresyon modeli ve tahmini tanıtılacaktır. Kantil regresyon ile doğrusal regresyon modellerinin temel model kurulumu, LRM için en küçük kareler tahmini ve QRM için benzer tahminleme yaklaşımları, iki tür modelin özelliklerinin arasındaki karşılaştırma bu bölümde yapılmak istenmektedir.

Kantil ve sansürlü regresyon modellerinin faydalı özellikleri şöyle özetlenebilir.

Bu modeller bağımlı değişkenin tüm koşullu dağılımlarını tanımlamak için kullanılabilir.

Kantil regresyon modeli kullanılan doğrusal algoritma programlamasının kolayca üstesinden gelebilecek, problem tahmini yapan doğrusal programlara sahiptir.

1- Kantil regresyon amaç fonksiyonu robust yer ölçüsü veren ağırlıklandırılmış mutlak sapma toplamıdır ve bunun tahminlenmiş katsayı vektörü bağımlı değişkendeki sapan gözlemlere karşı duyarlı değildir.

2- Hata terimi normal dağılmadığında, kantil regresyon tahmincileri en küçük kareler tahmincilerinden daha etkin olabilirler.

Farklı kantillerde çözümler farklı parametre vektörlerini sağlar; farklı tahminler, dağılımın farklı noktalarındaki değişim için bağımlı değişkenin tepkisindeki farklılık olarak yorumlanabilir[31].

[31] Buchinsky Moshe, **The Theory And Practice Of Kantile Regression. Pablished Doctoral Dissertation**. Cambridge, Massachusetts: Graduate Faculty of Harvar University,1991.

2.1. KANTİL KAVRAMI

Küçükten büyüğe doğru sıralanmış bir seride seriyi, iki parçaya bölen değere Medyan, dört eşit parçaya bölen değerlere Kartil, on eşit parçaya bölen değerlere Desil, yüz eşit parçaya bölen değerlerde Persantil adı verilmektedir.

Bu şekilde seriyi eşit parçalara ayıran ölçülere genel olarak Kantil adı verilmektedir. Medyan aynı zamanda ikinci kartil, aynı zamanda beşinci desil ve ellinci pörsentile eşit olacaktır. Küçükten büyüğe sıralı bir seride (N) adet eleman varsa (i) inci kantilin serideki sıra numarası genel olarak $K_i = X_{(iN+k/2)/k}$ ile bulunabilir. Burada k kantilin türü ile ilgili sabit bir sayı olup, k=2 için medyan, k=4 için kartil, k=10 için desil, k=100 için persantil değerleri elde edilir[32].

X , F dağılım fonksiyonuna sahip rastgele değişken ve P, (0,1) aralığında bir reel sayı olduğu düşünüldüğünde

$$P\left(X \le x_p\right) \ge p \qquad [2.1]$$

$$P\left(X \ge x_p\right) \ge 1-p \qquad [2.2]$$

eşitsizliklerini sağlayan x_p değerine X 'in (ya da dağılımın) p.kantili adı verilmektedir. F dağılım fonksiyonu kullanarak p. kantil ise

$$F\left(x_p^-\right) \le p \le F\left(x_p\right) \qquad [2.3]$$

Eşitsizliğini sağlayan değer olarak tanımlanabilir, burada $p = 0.5$ için $x_{0.5}$ değeri dağılımın medyanı, ve $p = 0.25$ ve $p = 0.75$ için $x_{0.25}$ ve $x_{0.75}$ değerleri sırasıyla dağılımın sırasıyla 1. ve 3. çeyreklikleri (quartile) olarak adlandırılır[33].

[32] Murat Karagöz, **İstatistik Yöntemleri**. Bursa: Ekin Yayın Dağıtım,2009, s.62.
[33] İrem Altındağ, **Kantil Regresyon ve Bir Uygulama**, (Yayınlanmış Yüksek Lisans Tezi) Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya.

2.2. ÖRNEK KANTİLİNİN ÖRNEKLEM DAĞILIMI

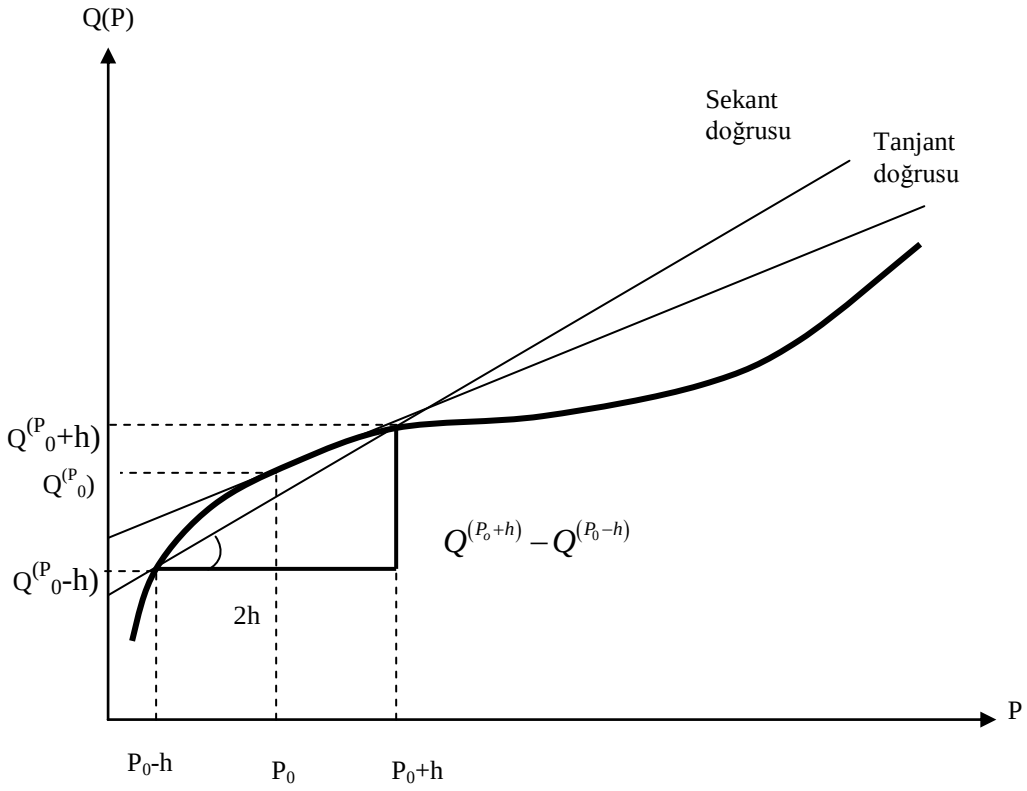
Büyük örneklerde örnek kantillerinin nasıl davrandığı önemli bir noktadır. $f = F'$ olasılık yoğunluk fonksiyonlu ve $Q^{(p)}$ kantil fonksiyonlu bir dağılımdan elde edilen $y_1, y_2 \ldots y_n$ geniş bir örnek için $Q^{(p)}$'nin dağılımı yaklaşık olarak normal, $Q^{(p)}$ ortalama ve $\frac{p(1-p)}{n}.\frac{1}{f\left(Q^{(p)}\right)^2}$ varyanslıdır. Özellikle bu örnek dağılımının varyansı kantilde ölçülen olasılık yoğunluk fonksiyonu tarafından tam olarak tanımlanır. Kantilde yoğunluluğa bağlılık basit bir açıklamaya sahiptir. Eğer daha yüksek yoğunluklu veriler varsa kantiller daha az değişkendir. Bunun tersi durumunda daha az yoğunluluk varsa kantiller daha çok değişkenlik gösterecektir.

Kantil örnekleme değişkenliğini tahminlemek için olasılık yoğunluk fonksiyonu bilinmeyen bir tahminleme yolu gerekirken yukarıdaki yaklaşım kullanılır. Bu tahminleme için standart bir yaklaşım şekil 2 ile gösterilmiştir. P noktasındaki $Q^{(p)}$ fonksiyonuna teğet olan doğrunun eğimi P'ye göre kantil fonksiyonun türevidir. Ya da benzer biçimde denklem 2.4'ün olasılık yoğunluk fonksiyonun tersidir.

$$\frac{d}{dp}Q^{(p)} = \frac{1}{f\left(Q^{(p)}\right)} \qquad [2.4]$$

Bu terim küçük h değerleri için $\left(p-h, \hat{Q}^{(p-h)}\right)$ ve $\left(p+h, \hat{Q}^{(p+h)}\right)$noktalarına doğru sekant doğrusun eğimi olan $\frac{1}{2h}\left(\hat{Q}^{(p+h)} - \hat{Q}^{(p-h)}\right)$ gibi farklı oranlar tarafından tahminlenebilir.

Şekil 2: Kantil Fonksiyonun Eğiminin Tahminlenmesi



Kaynak: Hao, Lingxin ve Daniel Q. Naiman, **Quantile Regression**, Sage Publications, 2007, s.12.

Şekil 2 deki p noktasında (tanjant çizgisinin eğimi) $Q^{(p)}$ fonksiyonunun türevi $\left(Q^{(p_0+h)} - Q^{(p_0-h)}\right) / 2h$ olarak gösterilir.

2.3. OLASILIK FONKSİYONU

Rassal bir değişkenin alabileceği değerlerle bu değişkenin söz konusu değerleri alma olasılıkları arasındaki ilişkiyi, bağlantıyı gösteren fonksiyona olasılık fonksiyonu denir. Rassal bir değişken olan X, kesikli bir değişkense buna kesikli olasılık fonksiyonu denir.

$$f\left(x_i\right) = \Pr\left\{X = x_i\right\} = p_i \qquad \text{i=1,2,...,n} \qquad [2.5]$$

X'in x değerini alma olasılığına p_i denir. "a" ve "b" gibi iki sabit sayı arasındaki aralıkta, sürekli rassal bir değişkenin değer alma olasılığına sürekli olasılık fonksiyonu denir. Bu olaslılık aşağıdaki gibi gösterilebilir.

$$\int_{-\infty}^{\infty} f(x)dx \qquad [2.6]$$

Ya da

$$\Pr\{a \le X \le b\} = \int_{a}^{b} f(x)dx \qquad [2.7]$$

Rassal bir değişken olan x verildiğinde, X'in ele alınan bu değişkene eşit ya da küçük çıkma olasılığını veren fonksiyona F(x) ile gösterilen dağılım fonksiyonu denir.

$F(x) = \Pr\{X \le x\}$ ve bu değer 0 ile 1 arasında olacaktır.

Olasılık fonksiyonu ile dağılım fonksiyonu arasındaki ilişki şu şekilde gösterilebilir. Olasılık fonksiyonu ile dağılım fonksiyonu arasındaki bağlantı yani birinden diğerini geçiş şu şekil de sağlanabilir[34].

$$F(x) = \int_{-\infty}^{x} f(x)d(x) \qquad [2.8]$$

Dağılım fonksiyonundan yoğunluk fonksiyonuna geçiş ise

$f(x) = \frac{d}{dx}\left(F(x)\right)$ olacaktır.

2.4. KANTİL FONKSİYONU VE KANTİL YOĞUNLUK FONKSİYONU

Kantil fonksiyonu hakkında sistematik olarak ilk defa makalesinde bahseden Parzen, Kantil fonksiyonunu Q(p) olarak ifade etmiştir. Buradaki p değeri 0 ile 1 arasında bir değer almaktadır.

$$Q(p) = F^{-1}(p) = \inf\{x : F(x) \ge p\} \qquad [2.9]$$

x_p değeri anakütlenin p'ninci kantili olarak adlandırılır[35].

$$x_p = \left(X \le x_p\right) = p \qquad [2.10]$$

[34] Bedriye Saraçoğlu ve Ferhan Çevik, **Matematiksel İstatistik Olasılık ve Önemli Dağılımlar**, Gazi Büro Kitabevi, Ankara, 1995, s.85.

[35] Emanuel Parzen, "Kantile Probability and Statistical Data Modeling", **Statistical Science**, Cilt:19, 2004, ss. 652-662.

Bunda temel özellik $-\infty < x < \infty$ ve $0 < p < 1$ olarak ifade edilmiş ve eğer $Q(p) \leq x$ ise $F(x) \geq p$ olacaktır. Sonuç olarak X, $Q(p)$ için özdeş dağılır ve aşağıdaki gibi gösterilebilir[36].

$$\Pr\left[Q(p) \leq x\right] = \Pr\left[p \leq F(x)\right] = F(x) \qquad [2.11]$$

Kantil yoğunluk fonksiyonu da benzer özelliklere sahip olarak, kantil dağılım fonksiyonunun türevi alınarak elde edilebilir.

$$q(p) = \frac{dQ(p)}{dp} \qquad [2.12]$$

Q(p) azalmayan bir fonksiyona sahip ve q(p) negatif olmayan ve $0 \leq p \leq 1$ değerlerini içermektedir[37].

2.5. DOĞRUSAL REGRESYON MODELLERİ VE ONUN YETERSİZLİKLERİ

Doğrusal regresyon modeli sosyal bilim araştırmalarında yaygın olarak kullanılan standart istatiksel metottur. Fakat bağımlı değişkeninin tüm koşullu dağılım özelliklerini açıklamaksızın bir bağımlı değişkeninin koşullu ortalamasına yoğunlaşmaktadır. Aksine kantil regresyon modelleri bağımlı değişkeninin tam koşullu dağılım özelliklerinin analizini kolaylaştırmaktadır. QRM VE LRM çeşitli yönlerde bir birine benzerler, her ikisi de parametreleri bilinmeyen, doğrusal sürekli bağımlı değişkenleri ile ilgilenmektedirler; fakat kantil regresyon modeli ve doğrusal regresyon modelleri farklı nicel modeller ve hata terimine ait farklı varsayımlara güvenmektedirler. Bu benzerlikleri ve farklılıkları en iyi şekilde anlamak için başlangıç noktası olarak LRM sergilenecek sonra QRM tanıtılacaktır. Açıklanması kolay olması açısından tek bağımsız değişken ele alınacaktır.

LRM ile standart doğrusal regresyon modeli ifade edilmektedir.

---

[36] Parzen, ss. 105-121.

[37] Warren Gilchrist, **Statistical Modelling with Quantile Functions,** Cherman & Hall/Crc Press, London, 2000, ss. 11-13.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon \qquad [2.13]$$

Burada ε_i, özdeş, bağımsız ve sıfır ortalamalı ve σ^2 varyanslı normal dağılımlıdır. $\beta_0 + \beta_1 x$ fonksiyonu x veriyken y'nin koşullu ortalamasına karşılık verilere uygun olarak gösterilebilir bu da $E[y \mid x]$ gösterilir ve burada x bağımsız değişkenine karşılık olarak y değerlerinin popülasyondaki ortalamasını gösterilmektedir.

LRM nin önemli özelliği farklı varyans varsayımı, koşullu varayans Var(y|x) bağımsız değişkenin tüm değerleri için sabit bir σ^2 varsayımıdır. Farklı varyans ihmal edildiğinde koşullu ölçek ve koşullu ortalamanın kendiliğinden modellenmesi için LRM nin kendiliğinden düzenlemesi olasıdır. Örneğin denklem 2.13 te düzenlenmiş model $y_i = \beta_0 + \beta_1 x_i + e^{\gamma}\varepsilon_i$ olacaktır. Burada γ bilinmeyen ilave bir parametredir ve şöyle yazılabilir $Var(y \mid x) = \sigma^2 e^{\gamma}$.

LRM nin üçüncü farklı özelliği normallik varsayımıdır. Çünkü LRM veriler için olası en iyi uyumu en küçük kareler sağlar, sadece tanımlayıcı amaç için normallik varsayımlarını kullanmaksızın LRM yi kullanırız. Fakat, sosyal bilim araştırmalarında, LRM bağımsız değişkenlerin bağımlı değişken üzerindeki önemli etkinin olup olmadığının testi öncelikle kullanılır. Hipotez testleri parametre tahminini aşmakta ve tahmin edicinin örnek değişkenliğinin tanımlanmasını gerektirir.

Hesaplan "p" değerleri normal dağılıma ya da büyük örnek durumda güvenilir olacaktır. Bu koşullardan birinin ihlal edilmesi "p" değerlerinin yanlı olmasına neden olabilir ve bu da hipotez testlerinin geçersiz olmasına yol açacaktır. Liner regresyon modelinde verilerin büyük bir çoğunu takip etmeyen sapan gözlemler, tahmin edilen regresyon doğrusu üzerinde gereksiz bir etki yapma eğilimindeler. LRM de alışılagelmiş pratik yöntem sapan gözlemleri tespit etmek ve onları yok etmektir.

2.6. KOŞULLU MEDYAN VE KANTİL REGRESYON MODELİ

Medyan regresyon modellemesi 18. yüzyılın ortalarında Boscovich tarafında önerilmiştir ve sonrasında Edgeworth ve Laplace tarafından geliştirilmiştir. Medyan regresyon modeli LRM' nin koşullu ortalama tahminin zararlarını söylemektedir.

Medyan regresyon koşullu medyana göre bir bağımsız değişkenin etkisini tahminler, böylece çarpık bir dağılımda merkezi yeri ifade eder.

Hem şekil ve hem de merkezi konumdaki değişimleri modellemek için Koenker ve Bassett medyan regresyondan daha genel bir formu kantile regresyon modelini (QRM) önermişlerdir. QRM koşullu dağılımda çeşitli kantillerdeki bir açıklayıcı değişkenin potansiyel fark etkilerini tahminler, örneğin 0.5'inci kantilden 0.95'ci kantile kadar ki 19 eşit uzaklıktaki kantillerin sıklığını tahminlemektedir. Medyan ile medyan dışındaki kantiller, bu tahminlenmiş 19 regresyon çizgisinin konum değişikliğinin (medyan çizgisinin yerindeki kayma, değişme) yanı sıra, hem ölçek hem de daha karmaşık şekil kaymasını (medyan dışındaki çizgiler) elde eder.

Aşağıdaki (2.14) denkleminde LRM ile ilgili olarak QRM aşağıdaki gibi gösterilebilir.

$$yi = \beta_0^{(p)} + \beta_1^{(p)} x_i + \beta_2^{(p)} x_i \qquad [2.14]$$

Burada 0<p<1 p'ninci kantile sahip popülâsyonun özelliğini göstermektedir. LRM tekrar hatırlandığında, x_i verildiğinde y'nin koşullu ortalaması $E(y_i|x_i)=\beta_0+\beta_1 x_i$ dir ve bu eşitlik hata teriminin beklenen değerinin sıfır olmasını gerektirmektedir. Sonuç olarak, QRM ile ilgili olarak x_i verildiğinde p'ninci koşullu kantilleri $Q^{(p)}\left(yi \mid x_i\right) = \beta_0^{(p)} + \beta_1^{(p)} x_i + \beta_2^{(p)} x_i$ olarak belirtilir. Koşullu p'ninci kantil x_i bağımsız değişkenin spesifik değeri, spesifik kantil parametreleri $\beta_0^{(p)}$ ve $\beta_1^{(p)}$ tarafından tanımlanmaktadır. LRM de olduğu gibi QRM'de de hata terimi eşdeğer olarak tanımlanmıştır. Çünkü $\beta_0^{(p)} + \beta_1^{(p)}$ bir sabit ve $Q^{(p)}\left(yi \mid x_i\right) = \beta_0^{(p)} + \beta_1^{(p)} x_i + \beta_2^{(p)} x_i + Q^{(p)}\left(\varepsilon_i\right) = \beta_0^{(p)} + \beta_1^{(p)}$ dir. Böylece QRM eşdeğer formülasyonu hata teriminin p'ninci kantili sıfır olmayı gerektirir.

İlgilenilen p. kantilin farklı değerleri içinde bu durum geçerli olacaktır. Denklem 2.14 te p yerine q yazıldığında $yi = \beta_0^{(q)} + \beta_1^{(q)} x_i + \beta_2^{(q)} x_i$ olur bu da $\varepsilon_i^{(p)} - \varepsilon_i^{(q)} = \left(\beta_0^{(q)} - \beta_0^{(p)}\right) + x_i\left(\beta_1^{(q)} - \beta_1^{(p)}\right)$ haline gelir. Verilen x_i sabit değeri tarafından iki sabit hata terim oluşur. Diğer bir deyişle, $\varepsilon_i^{(p)}$ ve $\varepsilon_i^{(q)}$ dağılımı birinden diğerine kaymadır. QRM de göz önüne alınması gereken önemli bir hususta

i=1,2,…,n kadar $\varepsilon_i^{(p)}$'nin bağımsız ve özdeş dağıldığıdır; bu durum kısaca "iid" olarak söylenir. Bu durumda $\varepsilon_i^{(p)}$'nin q'nüncü kantili "i" ye değil "q" ve "p" ye bağlı olan $c_{p,q}$ sabitidir. Denklem 3.5 kullanılarak, q'nüncü koşullu fonksiyonu $Q^{(q)}(yi \mid x_i) = Q^{(p)}(yi \mid x_i) + c_{p,q}$ olarak gösterebiliriz. Sonuç olarak "iid" durumu, koşullu kantil fonksiyonu β_1'in ortak değerini alan $\beta_1^{(p)}$ eğimi ile birinden diğerine basit bir kayma ya da değişmedir. Yani diğer bir değişle, "iid" varsayımı bağımlı değişkeninde şekil kaymasının olmadığını söyler. 2.13 denklemi, bir denklemle sadece bir koşullu ortalamaya sahip denklem 2.14 teki LRM'ye benzemediğini ortaya koymakta ve QRM ise sayısız koşullu kantile sahip olabilir. Bu sayısız denklem denklem 2.14 formunda gösterilebilir. Örneğin QRM 19 kantille belirlense 19 denklem, 19 koşullu kantillerin $\left(\beta_1^{0.05}, \beta_1^{0.10}, ..., \beta_1^{0.95}\right)$ her birinde x_i için 19 katsayı elde edilir. Bu kantiller eşit uzaklıkta olmayabilirler, fakat pratik olarak eşit uzaklıkta olmak onların yorumlamasını kolaylaştırmaktadır

Bu sonuçlar LRM'nin koşullu ortalamasında oldukça farklılık arz etmektedir. Koşullu kantiller şekil ve yer kaymasının özetini kullanabilen bir dağılım olarak tanımlanır.

2.7. KANTİL REGRESYON TAHMİNİ

Şimdiye kadarki regresyon modellerinin en belirgin ortak noktası ortalamaya dayalı tahmin yapmalarıydı. Yani bu regresyon modelleri bağımsız değişkenlerin bir fonksiyonu olarak bağımsız değişkenlerin verilen değerlerini amaç fonksiyonunu koşullu dağılımın ortalamasını modellemeye yoğunlaşmıştır. Fakat ortalama, ilgilenilen koşullu dağılımların özelliğinde sadece bunlardan bir tanesidir. Kantil regresyon modeli ile bağımsız değişkenlerin değişen değerleri ile birlikte koşullu dağılımdaki değişikliğin diğer özelliklerinin nasıl olduğun kıyaslanacaktır[38].

En küçük kareler tahmincisi toplam hata kareleri minimize etmek için parametrelerin bu değerlerini alarak parametre tahmincileri $\hat{\beta}_0, \hat{\beta}_1$ için çözer.

[38] Mark S. Handcock, **Relative Distribution Methods in the Social Sciences**, Springer-Verlag, New York, 1999, s.213.

$$\min \sum_i \left(y_i - \left(\beta_0 + \beta_1 x_i \right) \right)^2 \qquad [2.15]$$

Eğer LRM varsayımları yerine getirilmiş ise bağımlı değişkenin tahmini $\hat{\beta}_0$, $\hat{\beta}_1$ değerleri örnek hacmi sonsuza giderken anakitlenin koşullu ortalaması E(y | x) e yakalaşacaktır. Denklem 2.15 te minimizasyon ifadesi tahmin doğrusu $y = \hat{\beta}_0 + \hat{\beta}_1 x$ ve veri noktaları (x_i, y_i) arasındaki dikey uzaklığın karesinin toplamıdır.

Minimizasyon problemini elde etmek için aşağıdaki adımlar elde edilir.

a) $\hat{\beta}_0$, $\hat{\beta}_1$'e göre denklem 2.15'in kısmı türevi alınır.

b) Elde edilen her bir kısım sıfıra eşitlenir.

c) İki bilinmeyene ait iki denklem çözülür

Sonrasında aşağıdaki tahminciler elde edilir.

$$\hat{\beta}_1 = \frac{\sum_i^n \left(x_i - \bar{x} \right)\left(y_i - \bar{y} \right)}{\sum_i^n \left(x_i - \bar{x} \right)^2} \qquad [2.16]$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad [2.17]$$

QR tahmin edicisinin LR tahmin edicisinden önemli bir farkı şudur ki, QR'de doğrudan noktanın uzaklığı ağırlıklandırılmış dikey uzaklık (karesiz) toplamı kullanılarak ölçülür, burada ağırlık doğrunun üstündeki noktalar için p ve doğrunun altındaki noktalar için 1-p olarak alınır. Bu p oranı için her bir seçenek, örneğin p=0.10, 0.25, 0.50, farklı bir uyumlu koşullu kantil fonksiyonuna yol açar. Amaç olası her bir p değeri için arzu edilen özellikler ile bir tahmin edici bulmaktır.

Somutlaştırmak için, ilk önce medyan regresyon için tahmin ediciyi ele alınacak. 1. bölümde, y'nin medyanı E|y-m|'in minimizasyon değeri olarak nasıl görülebileceği tanımlanmıştır. Benzer bir durum için medyan regresyon durumunda mutlak hata toplamını minimizasyon için seçilebilir(gözlem değerinden uyum değerine mutlak uzaklık).

Tahmin edici denklem 2.12 β_i'lere göre minimize edilse:

$$\sum_i \left| y_i - \beta_0 - \beta_1 x_i \right| \qquad [2.18]$$

Uygun model varsayımları altında, örnek hacmi sonsuza gittikçe, popülasyon düzeyinde x verildiğinde y'nin koşullu medyanı elde edilir.

Denklem 2.18 minimizasyon ifadesinde, medyan regresyon doğrusu olarak adlandırılan çözüm sonucunda, regresyon doğrusunun altında kalanın yarısı ile regresyon doğrusun altında kalan verilerin diğer yarısındaki bu veriler regresyon doğrusundan geçmek zorundadır. Yani, kabaca hataların yarısı pozitif diğer yarısı da negatiftir.

Şekil 3'ün sağ panelinde, sekiz adet hipotetik veri noktası (x_i,y_i) ve bu noktalarla bir birine bağlı 28 doğru(8(8-1)/2=28) gösterilmektedir. Kesiki doğru tahmini medyan doğrusudur, yani bu doğru tüm veri noktalarından mutlak dikey uzaklık toplamı minimizasyonudur. Altı gözlem noktası medyan regresyon doğrusunun üzerine değil de bu doğrunun altı veya üstünde kalmıştır. (x, y) düzlemindeki her doğru (β_0, β_1) eğim ve sabitin seçimi $y = \beta_0 + \beta_1 x$ ten elde edilir, bu yüzden (β_0, β_1) düzlemindeki noktalar ve (x, y) düzlemindeki noktalar arasında bir uyuma sahiptir. Şekil 3 teki sağ panel, (β_0, β_1) düzlemindeki grafik sol paneldeki her bir doğruya karşılık gelen noktayı içerdiğini göstermektedir. Koyu daire sol paneldeki medyan regresyon doğrusuna karşılık gelen sağ paneli göstermektedir[39].

Ek olarak, eğer sabit ve eğim (β_0, β_1) ile bir doğru verilen (x_i, y_i) noktalarından geçerse ve sonra $y = \beta_0 + \beta_1 x$, ve bu yüzden (β_0, β_1), $\beta_1 = \left(\frac{y_i}{x_i}\right) - \left(\frac{1}{x_i}\right)\beta_0$ doğrusunun üzerine düşer.

Sekiz doğru şekil 3'de sol paneldeki sekiz noktaya karşılık gelen sağ paneldeki sekiz doğruyu göstermektedir. Bu doğrular (β_0, β_1) düzlemini poligan(çok kenarlı) bölgelere böler. Bir benzer örnek olarak şekil 3 teki gölgeli alan gösterilebilir. Bu alanın herhangi birinde, bu noktalar (x ,y) düzlemindeki doğruların bir kümesine karşılık gelmekte, ve bunlar tam olarak benzer şekilde veri setini ikiye bölmekteler (bunun anlamı bir veri doğrusunun altındaki ve üstündeki noktalar benzerdir).

[39] Lingxin Hao ve Daniel Q. Naiman, **Quantile Regression**, Sage Publications, 2007, s.36.

Sonuç olarak, şekil 4'te minimizasyonu araştıran (β_0, β_1) fonksiyonu her alanda doğrusaldır, yani bu fonksiyon çok düzlemli yüzey formlu grafik konvekstir. Bu örnekte şekil 4 te iki farklı açıdan konveks olarak çizilmiştir. Tepe noktası, kenarlar, noktaların yansıtılan yüzeylerinin görünümü, doğru parçası ve alan sırasıyla (β_0, β_1) düzleminde şekil 3'ün sağ panelinde gösterilmektedir.

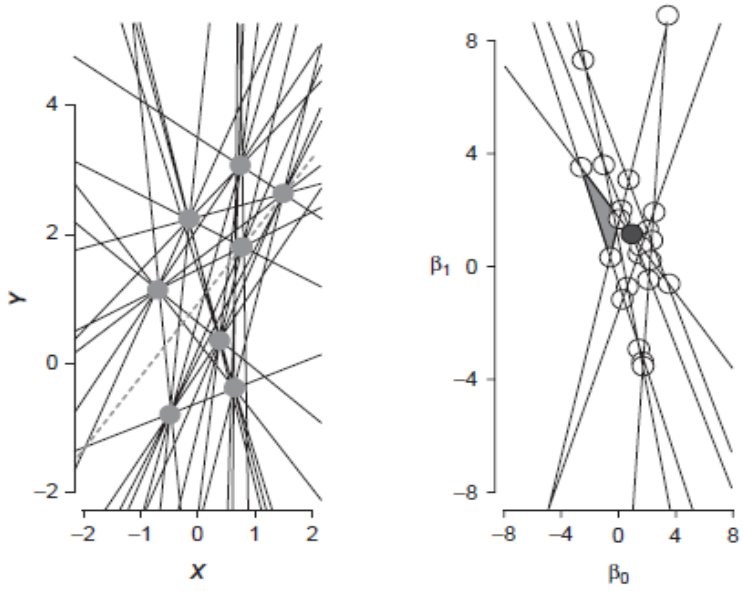
Her bir köşeye karşılık veri noktalarıyla ilişkili bir doğruya karşılık gelir. Yüzeyde İki tepe noktasına bağlı her bir kenar bir doğru parçasına karşılık gelmektedir, burada ilk doğru ile tanımlanan veri noktalarından bir tanesi bir başka veri noktası yerine de kullanılmış ve diğer noktalar onların altında ya da üstünde kalmıştır.

Denklem 2.18. deki mutlak uzaklık toplamının minimizasyonu için çözüm yolu, bunlardan biri medyan regresyon katsayılarına $(\hat{\beta}_0, \hat{\beta}_1)$ neden olan, doğrusal programlama problemleri çözümleri için dış nokta çözüm yollarına dayanabilir.

Bir köşeye karşılık gelen (β_0, β_1) noktalarının herhangi birinden başlanabilir. Minimizasyon çok düzlemli yüzeyin kenarı boyunca bir açıdan diğer açıya olan hareket tekrarlanarak elde edilir, seçilen bir açıdan minimuma varınca kadar bu yönde en dik tepeden inmektedir. Önceki paragraftaki tanımlama kullanılarak, eş veri noktaları tarafından tanımlanan doğrudan doğruya tekrarlı bir şekilde hareket edilir, seçilen biriyle şuan ki iki tanesinden biri yeni veri noktasının takası her bir adımda karar verildiğinde denklem 2.18. en küçük değer yapmaya yol açacaktır[40].

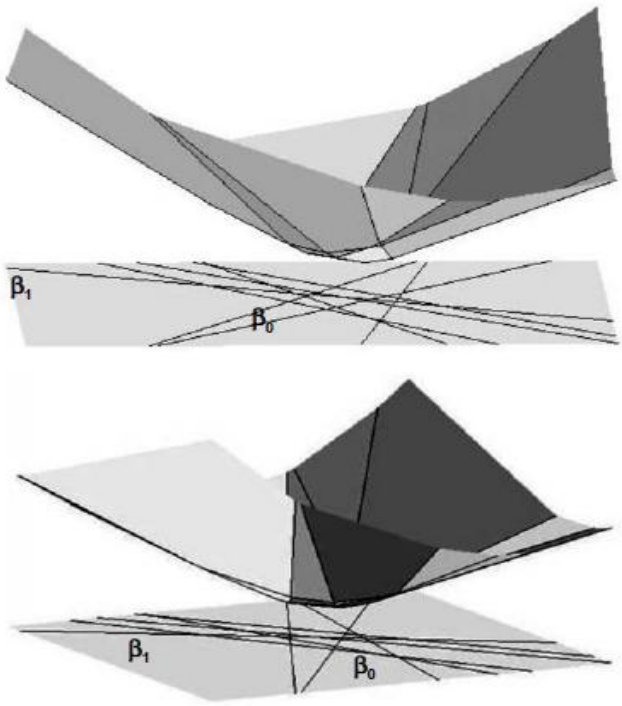
[40] Hao ve Naiman, s.37.

Şekil 3: Doğru ve Noktaların İkili Gösterimi



Kaynak: Hao, Lingxin ve Daniel Q. Naiman, **Quantile Regression**, Sage Publications, 2007, s.35.

Şekil 4: Poledral Yüzeylerin Gösterimi



Kaynak: Hao, Lingxin ve Daniel Q. Naiman, **Quantile Regression**, Sage Publications, 2007, s.36.

Mutlak hata toplamı minimumuna en alttaki köşe noktaları yüzeyinin altındaki (β_0, β_1) yüzeyindeki noktalarda ulaşılır.

Medyan regresyon tahmin edicisi p'ninci kantil regresyon tahmin edicisi içinde genelleştirilebilir.Tek değişkenli bir dağılım olan $y_1,\ldots,y_n$'nin p'nin kantil dağılımı örnek noktalarından ağırlıklandırılmış uzaklık toplamı minimum olan q değeridir. Buradaki q'nün üstündeki 1-p ağırlıklandırmasını alırken q'nün altındaki değerler p ağırlıklandırmasını alacaktır[41]. Benzer bir şekilde, y_i gerçek değeri ile $\hat{y}_i = \hat{\beta}_0^{(p)} + \hat{\beta}_1^{(p)}$ tahmin değerleri arasındaki uzaklığın tartılı toplamı minimzasyon değeri olan $\hat{\beta}_0^{(p)}$, $\hat{\beta}_1^{(p)}$ p'ninci kantil regresyon tahmin edicileri olarak tanımlanır. Burada $\hat{y}_i$ tahmin değeri gözlem değeri y_i'nin altında tahminlenmiş ise buradaki tartı 1-p değilse p kullanılır. Diğer bir deyişle $y_i - \hat{y}_i$ ağırlıklandırılmış toplam hatayı mininimzasyonunu araştırırız. Burada pozitif hatalar p ağırlıklandırmasını alırken negatif hatalar ise 1-p ağırlıklandırmayı alacaktır. P'ninci kantil regeresyon tahmin edicileri $\hat{\beta}_0^{(p)}$, $\hat{\beta}_1^{(p)}$ minimizasyon için aşağıdaki gibi seçilir[42].

$$\sum_{i=1}^{n} d_p\left(y_i, \hat{y}_i\right) = p \sum_{y_i > \beta_0^{(p)} + \beta_1^{(p)} x_i} \left| y_i - \beta_0^{(p)} + \beta_1^{(p)} x_i \right| + (1-p) \sum_{y_i < \beta_0^{(p)} + \beta_1^{(p)} x_i} \left| y_i - \beta_0^{(p)} + \beta_1^{(p)} x_i \right| \qquad [2.19]$$

Denklem 2.18 ten farklı olarak negatif hatalar pozitif hatalar gibi benzer önemde ifade edilir ve denklem 2.19 pozitif ve negatif hatalar farklı işaretlerle ağırlıklandırılır. Denklem 2.19'daki ilk toplam doğrunun yukarısındaki noktalar için $y = \beta_0^{(p)} + \beta_1^{(p)} x$ doğrusundaki veri noktalarının dikey uzaklık toplamıdır. İkinci toplam ise aşağıdaki tüm veri noktalarının toplamıdır.

Her bir kantil regresyon için katsayıların tahmininin tüm verileriyle ağırlıklandırıldığı bilinen en yaygın hatadır. Kantil regresyon katsayıları $\hat{\beta}_0^{(p)}$, $\hat{\beta}_1^{(p)}$ hesaplamaları için bir çözüm yolu medyan regresyon katsayılarınki gibi

[41] Roger Kuenker ve Vasco D'Orey, "Computing Regression Quantiles", **Applied Statistics**, Cilt:46, 1987, ss. 383-393.
[42] Hao ve Naiman, s. 37.

geliştirilebilir. “p’ninci” kantil regresyon tahmin edicisi medyan regresyon tahmin edicinin benzer özellikteki bir durumuna sahiptir. $y = \hat{\beta}_0^{(p)} + \hat{\beta}_1^{(p)}$ tahmin doğrusu altında kalan veri noktalarının oranı “p” ve yukarıda kalan oran ise 1-p olacaktır.

Örneğin, 0.10 cu kantil regresyon doğrusu için katsayıları tahmin etiğimiz zaman, doğrunun altındaki gözlemler ağırlığın 0.90’ını vermekte ve yukarısında doğru ise ağırlığın 0.10’unu alır. Sonuç olarak tahmin doğrusunun yukarısında kalan (x_i, y_i) veri noktalarının % 90’nı pozitif hatalara yol açar ve doğrunun altında kalan %10 ise negatif hatalara yol açar. Diğer taraftan, 0.90’ıncı kantile regresyon katsayılarını tahmin etmek için doğrunun altında kalan noktalara 0.10 ağırlığı verilmekte ve geriye kalanlar ise 0.90 olacaktır. Sonuç olarak gözlemlerin %90’ı negatif geriye kalanların %10’u da pozitif hatalara sahip olacaktır.

β_0 ’a göre yönsel türevle ilgili basit bir ispat regresyon doğrusunun altında ve üstündeki veri noktaları içinde benzer bir sonuç göstermektedir. Bir kantil minimizasyon probleminin çözümü olarak ela alınabilir. Burada ilk olarak medyan, 0,5’inci kantille başlanacak.

Minimizasyon problemini harekete geçirmek için ilk olarak bilinen μ ortalama ve y dağılımını dikkate alacağız. $(Y-\mu)^2$ sapma kareler kullanılarak verilen Y veri noktasının μ değerinden ne kadar uzak olduğu ve sonra $E\left[(Y-\mu)^2\right]$ ortalama mutlak sapma ile ortalamada Y’nin μ den ne kadar uzak olduğu ölçülebilir. Bir dağılımın merkezinin nasıl tanımlandığı hakkındaki bir yol olarak, minimize edilmiş Y’den ortalama sapma karelerdeki μ noktası için araştırılabilir.

$$E\left[(Y-\mu)^2\right] = E[Y]^2 - 2E[Y]\mu + \mu^2$$

$$= (\mu - E[Y])^2 + \left(E\left[Y^2\right] - (E[Y])^2\right)$$

$$= (\mu - E[Y])^2 + Var(Y) \qquad [2.20]$$

Burada 2.terim olan Var(Y) sabit olduğundan ilk terim olan $(\mu - E[Y])^2$ (2.20) eşitliğinde minimize olmuştur. $\mu = E[Y]$ olarak ele alınması ilk terimi sıfır (2.20) eşitliğini minimize eder ve μ ’nün diğer değerleri ilk terimi pozitif yaparken (2.20) eşitliğini minimizasyondan ayırır.

Benzer olarak n büyüklüğünde bir örneğin örnek ortalaması minimizasyon probleminin çözümü olarak görülebilir. $\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2$: Ortalama kareler uzaklığı minimizasyonunu μ noktasında araştırırız.

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2=\frac{1}{n}\sum_{i=1}^{n}(\mu-\bar{y})^2+\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2=(\mu-\bar{y})^2+s_y^2 \qquad [2.21]$$

Denklem (2.21) te $\bar{y}$ örnek ortalaması s_y^2 örnek varyansını göstermektedir. Bu minimizasyon probleminin çözümü için ilk terimin küçük olma durumunda μ 'nün değeri $\mu=\bar{y}$ olarak alınır.

Somutlaştırmak için aşağıdaki 9 değeri bir gözlem olarak ele alalım: 0.23, 0.87, 1.36, 1.49, 1.89, 2.69, 3.10, 3.82 ve 5.25 değeri. Verilen bir μ noktasından örnek noktalarının ortalama kareler uzaklığının bir çizimi şekil 2.7 de verilmiştir. Dikkat edildiğinde fonksiyonun minimumu konveks ile düzleştirilmiş paraboldur. . Kareler uzaklığının kullanılması yerine mutlak uzaklık $|Y-m|$ ile Y'nin m'den ne kadar uzak olduğunu ölçebiliriz ve ortalama mutlak uzaklık $E|Y-m|$ ile popülasyondaki m den ortalama uzaklığı ölçebiliriz. Yine $E|Y-m|$ minimizasyonu ile m değeri için çözebiliriz. Görüleceği üzere $|Y-m|$ fonksiyonu konvekstir. Böylece minimizasyon çözümü için burada m noktasına göre türev sıfır bulunur ya da karşı işaretteki 2 yönsel türevle bulunur. Medyan dağılımı çözümdür.

Benzer olarak örnek düzeyini çalışacaktır. $\frac{1}{n}\sum_{i=1}^{n}|y-m|$ ile örnek noktalarının m'den ortamla mutlak uzaklığı tanımlanır. Yukarıdaki aynı dokuz örnek için bu fonksiyonun bir çizimi şekil 6 da verilmiştir. Bu fonksiyon 7 şeklindeki fonksiyonun çizimi ile kıyaslandığında şekil 6 görünürde parabolik ve konveks kalır. Şekil 6'daki fonksiyon parça parça doğrusaldır ve her örnek noktasında eğim değişmektedir. Fonksiyonun minimum değeri şekilde gösterilmiş olan örneğin 1.89 olan medyan değeri ile çakışmaktadır. Bu daha genel bir olayın özel bir durumudur. Her örnek için $f(m)=|y_i-m|$ ile tanımlı fonksiyon $f_i(m)=|y_i-m|/n$ "V" şeklindeki fonksiyonunun tanımıdır. (y_i=1.49 veri noktası ile ilgili olarak f_i fonksiyonu için

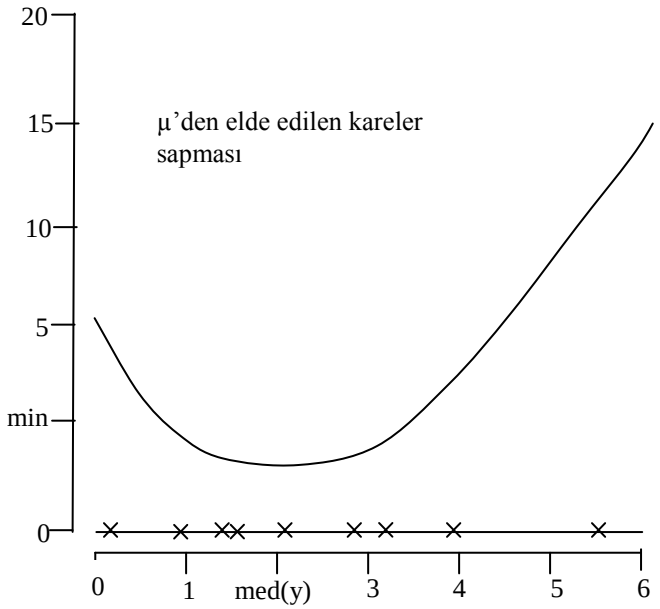
şekil 7'ye bakınız. f_i fonksiyonu m=y_i m>y_i için $\frac{1}{n}$ ve m<y_i için $-\frac{1}{n}$'nin türevi alındığında minimium(sıfır) değeri alır. m=y_i türev alınamadığında pozitif yönde $\frac{1}{n}$ ve negatif yönde $-\frac{1}{n}$'den yönsel türevi alınır. Bu fonksiyonun toplamı alındığında m'de f'in yönsel türevi pozitif yönde (s-r)/n ve negatif yönde (r-s)/n dir. Burada "s" m'in sağ tarafındaki veri noktalarının sayısı ve "r" ise m'in sol tarafında kalan veri noktalarının sayısını göstermektedir. f'in minimizasyonu m örnek medyanı olarak alındığında, m'in sağ ve solundaki veri noktalarının sayısı eşit olduğu durumda gerçekleşir. Medyanın gösteriminde olduğu gibi diğer kantiller için de genişletilebilir. Her $p \in (0,1)$ için, verilen bir "q" için Y'den uzaklık, mutlak uzaklık ile ölçülür fakat Y'nin q'nün sağ ve solunda olmasına bağlı olarak farklı ağırlıklar uygulanır. Verilen q değeri için Y'den uzaklık aşağıdaki gibi gösterilir.

$$d_p(Y,q) = \begin{cases} (1-p)|Y<q & Y<q| \\ p|Y-q & Y \le q| \end{cases} \qquad [2.22]$$

$Y : E\left[d_p(Y,q)\right]$ de ortalama uzaklığı minimum olan q değerini ararız. P. Kantil q olduğunda minimizasyon sağlanır.

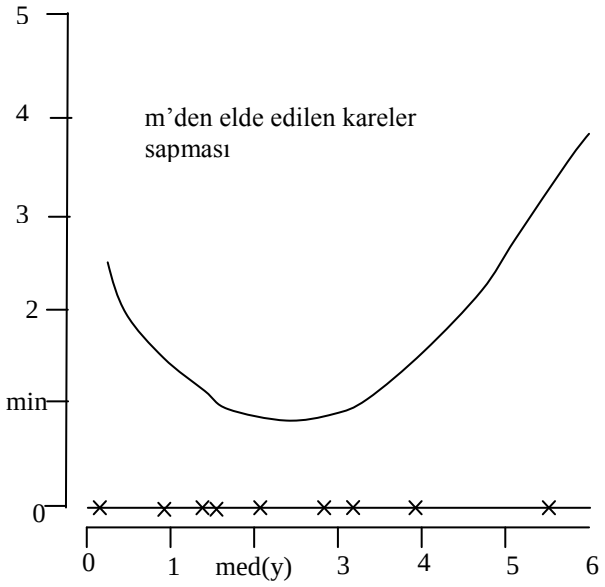
$$\frac{1}{n}\sum_{i=1}^{n} d_p(y_i,q) = \frac{1-p}{n}\sum_{y_i<q}|y_i-q| + \frac{p}{n}\sum_{y_i>q}|y_i-q| \qquad [2.23]$$

Şekil 5: Ortalama için Minimizasyon Problemi



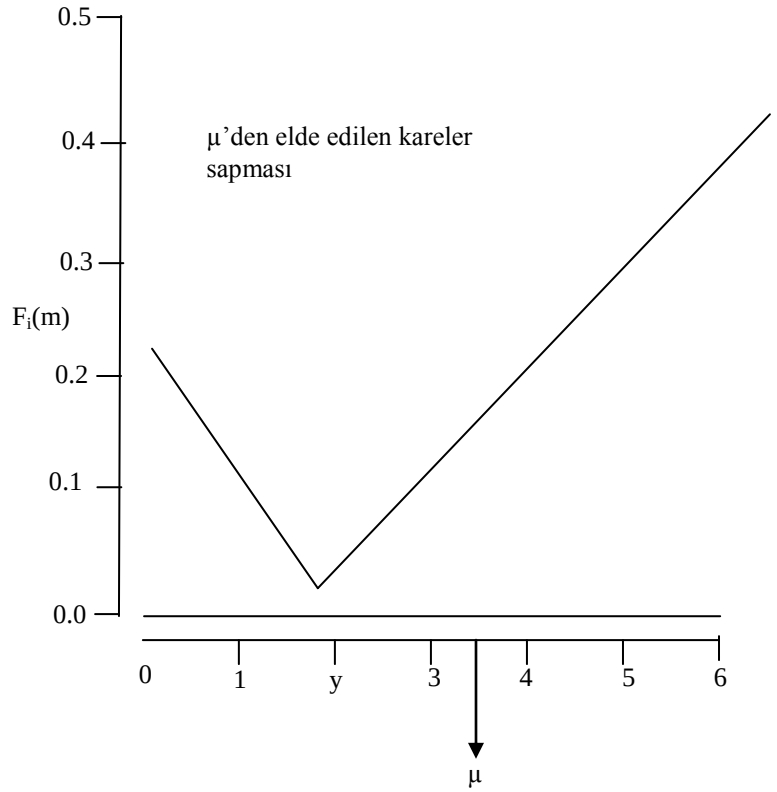
Kaynak: Hao, Lingxin ve Daniel Q. Naiman, **Quantile Regression**, Sage Publications, 2007, s.18.

Şekil 6: Medyan için minimizasyon problemi



Kaynak: Hao, Lingxin ve Daniel Q. Naiman. **Quantile Regression**, Sage Publications, 2007, s.18.

Şekil 7: V Şekil Fonksiyonu ile Minimizasyon Problemi Gösterimi



Kaynak: Hao, Lingxin ve Daniel Q. Naiman, **Quantile Regression**, Sage Publications, 2007, s.19.

2.9. KANTİLLERİN ÖZELLİKLERİ: RANK VE OPTİMİZASYON

Her rassal X değeri dağılım fonksiyonu tarafından tanımlanmıştır.

$$F(x) = P(X \le x), \qquad [2.24]$$

$0 < \tau < 1$ her değerine karşılık

$F^{-1}(\tau) = \inf\{x : F(x) \ge \tau\}$ X'in τ.kantilidir. Medyan yani F-1(1/2), burada merkezi bir rol oynamaktadır.

Kantiller, tüm bunları takip eden temel basit bir optimizasyondan kaynaklanmaktadır. Basit bir teorik karar problemi düşünüldüğünde: Bir nokta

tahmini F dağılım fonksiyonlu rassal bir değişken gerektirir. Eğer azalma aşağıdaki şekilde gösterilen parçalı doğrusal fonksiyon tarafından tanımlanmışsa bazı $\tau \in (0,1)$ için

$$\rho_\tau(u) = u(\tau - I(u<0)) \qquad [2.25]$$

Beklenen azalmayı bulmak için $\hat{x}$ bulunur[43]. Bu minimizasyonu araştırmak için

$$E\rho_\tau(X-\hat{x}) = (\tau-1)\int_{-\infty}^{\hat{x}}(X-\hat{x})dF(X) + \tau\int_{\hat{x}}^{\infty}(X-\hat{x})dF(X) \qquad [2.26]$$

$\hat{x}$ e göre türev alınır.

$$\frac{\partial E(\rho_\tau(X-\hat{x}))}{\partial\hat{x}} = (\tau-1)\frac{\partial\int_{-\infty}^{\hat{x}}(X-\hat{x})\partial F(X)}{\partial\hat{x}} + \tau\frac{\partial\int_{\hat{x}}^{\infty}(X-\hat{x})\partial F(X)}{\partial\hat{x}}$$

$$= \int_{-\infty}^{\hat{x}}\partial F(X) - \tau\int_{\hat{x}}^{\infty}\partial F(X)$$

$$= (1-\tau)F(\hat{x}) - \tau(1-F(\hat{x}))$$

$$= F(\hat{x}) - \tau$$

Burada beklenen azalmanın minimizasyonun $\{x : F(x) = \tau\}$ her bir elemanı için F durağan olduğundan sonuç sıfıra eşit olacaktır. Tek bir çözüm olduğunda $\hat{x} = F^{-1}(\tau)$ aksi takdirde τ. kantil aralığından en küçüğü seçilmek zorundadır. Asimetrik doğrusal azalmalar için nokta tahmincisinin kantilleri vermesi doğaldır. Simetrik durumda azalan mutlak değerin medyanı sağladığı iyi bilinmektedir. Azalma doğrusal ve asimetrik olduğunda, iki marjinal azalmadan düz olanını tahmin noktası olarak tercih edilmesi daha muhtemeldir. Yani örneğin alt tahmin (underestimate) aşırı tahminden 3 kat daha fazla değere sahipse, eşitliği sağlamak

[43] Thomas Shelburne Ferguson, **Matematical Statistics: A Desicion Theoretic Approach**, Academic Press, New York, 1967, s. 51.

için seçilen $\hat{x}$ yani P(X≤ $\hat{x}$) üç kat daha büyük olacaktır P(X≥ $\hat{x}$) den. Yani, seçilen $\hat{x}$ F'in yüzde yetmiş beşi olacaktır[44].

Dağılım fonksiyonu tarafından yerine konulduğunda

$$F_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \le \hat{x}) \qquad [2.27]$$

Beklenen azalan değeri minimize etmek için $\hat{x}$ seçilebilir.

$$\int \rho_\tau(x-\hat{x})\, dF_n(x) = n^{-1}\sum_{i=1}^{n} \rho_\tau(x-\hat{x}) \qquad [2.28]$$

Böylece τ . Örnek kantili sağlanır[45].

Kantillerin temel özelliklerinden birisi eşit varyans özelliğidir. Böyle bir durumda, eğer rassal bir değişken için sabit bir dönüşüm h(örneğin üstel ya da logaritmik fonksiyon) uygulanırsa, kantiller kantil fonksiyonu için benzer bir dönüşüm uygulayarak elde edilebilirler. Diğer bir deyişle, eğer Y'nin p'ninci kantili q ise h(Y)'nin p'ninci kantili de h(q) olacaktır. Benzer bir durum örnek kantilleri içinde yapılabilir.

Örnek kantillerin diğer bir temel özelliği aykırı değerlerin etkisi için duyarsızlıkları ile ilgilidir. Bu özellikli kantil regresyon, kantil temelli faydalı içeriğin çoğunda ve kantillerin yapım aşamasındaki yardımda benzer özelliğe sahiptir. Örnek medyanı "m" örnek veriler $x_1, x_2 \ldots x_n$ olarak verildiğinde medyanın altındaki ve üstündeki x veri değerlerinin değiştirilmesiyle örnek yeniden düzenlenebilir. Örnek için bazı düzenlemeler örnek medyanını vermede etki sahibi olmayabilir.[46] Benzer özellik p'ninci kantil için de alınır. Buna karşılık örnek ortalaması için bu durum: bazı $x_i + \Delta$ değerleri için her örnek x_i değeri değişimi $\frac{\Delta}{n}$ ile örnek ortalama değişir. Bazı özgün verilerin etkisi örnek ortalama için sınırlandırılamaz ama örnek kantili için sınırlandırılabilir.

Bir bağımlı değişken analizinde, araştırmacılar en iyi modeli elde etmek ya da yorumlamaya kolaylık olması açısından sık sık ölçeği dönüştürürler. Tahmin ve

[44] Roger Koenker, **Quantile Regression,** Cambridge University Press, New York, 2005, s. 6.
[45] Koenker, s.6.

modelin eşit varyans özelliği durum (situations) olarak adlandırılır, eğer veriler dönüştürülürse model ya da tahminde de benzer dönüşümlere uğrayacaktır. Bağımlı değişkenini dönüştürdüğümüz zaman, eşit varyans özelliği bilgisi model tahminini yeniden yorumlamamıza yardımcı olur.

Bağımlı değişkenin her doğrusal dönüşümü için yani, y' ye bir sabit eklenmesi ya da bir sabitle çarpılmasıyla LRM'nin koşullu ortalaması tam olarak dönüştürülebilir. Bunu şu şekilde yazabiliriz[47]:

$$E\left(c+ay \mid x\right)=c+aE\left(y \mid x\right) \qquad [2.29]$$

$$Q^{(p)}\left(c+ay \mid x\right)=c+a\left(Q^{(p)}\left[y \mid x\right]\right) \qquad [2.30]$$

Burada koşul a pozitif bir değerdir. Eğer a negatif ise $Q^{(p)}\left(c+ay \mid x\right)=c+a\left(Q^{(1-p)}\left[y \mid x\right]\right)$ olur. Nonliner dönüşümlerde sık sık ortaya çıkan bu durum arzu edilmektedir. Logaritmik dönüşümü sağa çarpık bir dağılımda sık kullanılan bir durumdur. Diğer dönüşümler diğer dönüşümler ise daha iyi bir model elde etmek ya da daha fazla normallik sağlamak için göz önünde bulundurulur.

Logaritmik dönüşümü yakın bir terimdeki bir bağımsız değişkenin etkisini modellemek içinde kullanılır(yüzdelik değişimler). Diğer bir değişle bir değişkenin etkisi çarpan ölçekte toplama ölçekten daha fazla görülür.

2.9.1. Robustnes (Dayanıklılık, Sağlamlılık)

Y değişkeni ile ilgili model varsayımlarının ihmal edilmesi uç değerlerin duyarsızlaştırılması olarak ifade edilir. Doğrusal regresyon modelleri tahminleri uç değerler için hassas olabilirler. Koşullu ortalama ve ortalamadaki bozulmaları net bir şekilde görülebilir. Buna karşın, uç değerlerin yok edilmesi çalışmaları sosyal bilimlerin çoğunda pek tercih edilmemektedir.

Doğrusal regresyon modellerinin aksine kantil regresyon modelleri uç değerler için duyarlı değildir.

[47] Hao ve Naiman, s.47.

2.9.2. Kantil Regresyon ve Çıkarsama

Bundan sonraki aşamada istatiksel çıkarsama, kantil regresyon modelinden elde edilen katsayı tahminleri için güven aralığı ve standart hata gibi özellikler incelenecektir. İlk olarak doğrusal regresyon model çıkarsaması gözden geçirilecek, hipotez testleri ve güven aralığı yapılarında kullanılan kantillerin asimtotik dağılımları sonsuz örnekler tarışılacaktır. Sonra da QRM katsayıları hakkındaki çıkarsamalar için izin verilen bootstrap prosedürü anlatılcaktır. Bootstrap prosedürü asimptotik olarak daha tercih edilebilirdir. Çünkü çarpıklıktaki değişim ve yapısal ölçeğin standart hataları için çözüm karmaşıktır ve varsayımlar tatmin edici olsa bile asimptotik prosedür için varsayımlar daima tutulamazlar. Bootstrap prosedürü her tahminin güven aralığı standart hatalarını sağlamada esneklik sağlar.

2.9.3. QRM'de Standart Hata ve Güven Aralığı

QRM de Katsayılar $\beta^{(p)}$ için çıkarsama yapılmak istenildiğinde $Q^{(p)}\left(y_i \mid x^i\right)=\sum_{j=1}^{k}\beta_j^{(p)}x_j^{(i)}$ daha önce bu model $y_i=\sum_{j=1}^{k}\beta_j^{(p)}x_j^{(i)}+\varepsilon_i^{(p)}$, burada $\varepsilon_i^{(p)}$ p. Kantilrde ortak dağılama sahip sıfır değerine eşittir.

$\beta_j^{(p)}$ için sonuç çıkarma LRM' deki yapıya benzer olarak $\hat{\beta}_j^{(p)}$ nin standart hatanın $s_{\hat{\beta}_j^{(p)}}$ ölçümlerine dayanan hipotez testleri ya da güven aralığı formunda olabilecektir. Bu standart hata, standart normal dağılıma sahip $\left(\hat{\beta}_j^{(p)}-\beta_j^{(p)}\right)/s_{\hat{\beta}_j^{(p)}}$ niteliğinde ve asimptotik özelliklere sahip olacaktır. QRM için standart hatalar bağımsız özdeş dağılıma (i.i.d) varsayımı altında önemli derecede basittirler. Bu durumda $\beta^{(p)}$ nin asimptotik kovaryans matrisi

$$\sum_{\hat{\beta}^{(p)}}=\frac{p(1-p)}{n}\frac{1}{f_{\varepsilon^{(p)}}(0)^2}\left(X'X\right)^{-1} \qquad [2.31]$$

olarak ele alınır. Denklem 2.31 de görülen $f_{\varepsilon^{(p)}}(0)$ terimi hata dağılımının p. kantilinde hesaplanan hata terimi $\varepsilon^{(p)}$'nin olasılık yoğunluk fonksiyonudur. LRM' dekine benzer olarak, kovaryans matrisi $\left(X^tX\right)^{-1}$ matrisinin çoklu ölçeğidir. Fakat

QRM de çarpan $\frac{p(1-p)}{n}\frac{1}{f_{\varepsilon^{(p)}}(0)^2}$, tek değişkenli bir örnek $\varepsilon_1^{(p)},\varepsilon_2^{(p)},...,\varepsilon_n^{(p)}$ ye dayanan bir örnek kantilnin asimptotik varyansıdır. Olasılık yoğunluk terimi olarak görülen bu ifade tek değişkenli bir durumdakine benzer olarak tanımlanma ihtiyacı duyulan bir bilinmeyendir. $\frac{1}{f_{\varepsilon^{(p)}}} = \frac{d}{dp}Q^{(p)}\left(\varepsilon^{(p)}\right)$ değeri $\frac{1}{h}\left(\hat{Q}^{(p)}\left(p+h\right)-\hat{Q}^{(p)}\left(p-h\right)\right)$ farklı bir oran kullanılarak tahmin edilebilir. Buradaki örnek kantiller $\hat{Q}\left(p\pm h\right)$, QRM model uyumu için $\hat{\varepsilon}_i^{(p)} = y_i\sum_{j=1}^{k}\hat{\beta}_j^{(p)}x_j^{(i)}, i=1,2,...,n$ hatalara dayanır.

Eğer h İİD (bağımsız özdeş dağılım) durumuna sahip değilse bu durumda baş edilmesi zor kompleks bir yapı haline gelecektir. Bu durumda $\varepsilon_i^{(p)}$ ortak dağılımdan sahip olunmayacak fakat, fakat bu kantiller p. kantil de sıfır olabilecektir. Bu aynı dağılıma sahip olmayanlar ele alındığında, X'X matrisinin bir ağırlıklandırması olan $D_1 ve D_0$ versiyonunu tanıtma ihtiyacı doğacaktır. Bu dağılım kovaryans matris ve doğru katsayılar bileşeni ile ortalamaya sahip bu form: $\sum_{\hat{\beta}(p)} = \frac{p(1-p)}{n}D_1^{-1}D_0D_1^{-1}$

Burada $D_0 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x^{(i)t}x^{(i)}$ ve $D_1 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}w_ix^{(i)t}x^{(i)}$ ve $x^{(i)}$ 1xk boyutlu x matrisinin i.ninci satırıdır. D_0 ve D_1 matrisleri kxk boyutlu matrislerdir. Ağırlık olarak $w_i = f_{\varepsilon_i^{(p)}}(0)$ olasılık yoğunluk fonksiyonunda $\varepsilon_i^{(p)}$'nin sıfırdır. D_1 değeri $\tilde{X}^t\tilde{X}$ olarak ifade edilmiş ve burda $\tilde{X}$ ise X in $\sqrt{w_i}$ değeri ile çarpını göstermektedir. İ.İ.D koşulları altında $\hat{\beta}^{(p)}$ nin asimptotik dağılımı ilgilenilen kantil de olasılık yoğunluk fonksiyonu hesaplanarak görülebilir. Fakat her bir kantil için hata özdeş dağılmadığından bu terimleri her bir "i" inci değer için farklı bir ağırlıklandırmaya neden olacaktır. Yoğunluk fonksiyonu bilinmediği için denklem 2.32 de gösterilen wi ağırlığı bilinmek mecburi hale gelecektir. Her türlü metot için $\hat{\beta}^{(p)}$ için kovaryans matrisi $\sum\frac{p(1-p)}{n}\hat{D}_1^{-1}\hat{D}_0\hat{D}_1^{-1}$ olarak tahmin edilebilir.

$$\hat{D}_0 = \frac{1}{n}\sum_{i}^{n}x^{(i)t}x^{(i)} \quad ve \quad \hat{D}_1 = \frac{1}{n}\sum_{i}^{n}\hat{w}_ix^{(i)t}x^{(i)} \qquad [2.32]$$

Özdeş katsayı tahmincisi $\hat{\beta}^{(p)}$ için standart hata tahmini kovaryans matrisinin diyagonal elamanlarının karekökü alınarak elde edilebilir. İ.İ.D durumunda kantil regresyon katsayıları için güven aralığı ve bağımsız değişkenlerin hipotez testleri elde edilebilir.

2.10. KANTİLE REGRESYON BOOTSRAP METOD

Bootsrap metodu popülasyonun n hacimli bir örnekten hesaplanan parametre tahminin örnek dağılımı için elde edilen Monte-Carlo metodudur. Basit Monte-Carlo simülasyonu örnek dağılımına yakın olmak için kullanıldığında, popülasyon dağılımının bilindiği varsayılır, n hacimli örnekler dağılımdan elde edilir ve her bir örnek parametreleri hesaplamak için kullanılır. Tahmini parametreleri hesaplanan bu deneysel dağılımlar arzulanan örnek dağılımlarına yaklaşmak, benzemek için kullanılır. Özellikle, tahmini standart hata parametre tahminleri örneğinin standart hatası kullanılarak tahminlenebilir.

Bootstrap yaklaşımının sıradan Monte-Carlo simülasyondan farkı 1979 yılında Efron tarafından tanıtılmıştır. Hipotetik bir örnek dağılımından elde etmek yerine, n hacimlik örnekler yerine gerçek bir veri setinden elde edilir. 'M' ile tanımlanan yeni örneklerin sayısı genellikle güven aralığı için 500 ile 2000 arasında ve standart hata tahmini için 50 ile 200 arasındadır. Her bir yeni örneklem orijinal örneklemdeki elaman sayısına benzemesine dışarıda kalan diğerlerinde fazla bazı orijinal data noktaları içerecektir. Bu yüzden bu yeni örneklemin her biri rastsal olarak orijinal gözlemlerden ayrılacaktır.

Bootstrap yaklaşımını somut bir örnekte verecek olursak, bir popülasyonun $Q^{(0.25)}$ %25'inin tahmini dikkate alınırsa $y_1, y_2, ..., y_n$ bir örneklem için yüzde 25'ine dayanır. Bu tahminin standart hatası tahminlenmek istendiğinde. Bu yaklaşımla verilen $Q^{(p)}$'nin varyansı için büyük örnek tahmini kullanmaktır. Bu $Q^{(0.25)}$'in standart hatası için tahmin olarak $\sqrt{\frac{p(1-p)}{nf\left(Q^{(p)}\right)^2}} = \sqrt{\frac{(1/4)(3/4)}{nf\left(Q^{(p)}\right)^2}} = \frac{\sqrt{3}}{4\sqrt{nf\left(Q^{(0.25)}\right)}}$ verilir, f burada olasılık yoğunluk fonksiyonu olarak verilmiştir, ve bunu tahmin etmek gerecektir. Bootstrap yaklaşımı için bunu hesaplamak çok yönlüdür: n hacimli büyük bir örneklem yerine orijinal veri setinden elde edilir. Bu örneklemlerin her biri

bootstrap örneğini yerine kullanılır. M. bootstrap örneği $y_1, y_2, ..., y_n$ için $\hat{Q}_m^{(0.25)}$ değeri hesaplanır. Bu büyük sayılar M(50-200) kez tekrarlanması $\hat{Q}_m^{(0.25)}, m = 1, 2, ..., M$ bir örneğe yol açar. Daha sonra arzulanan standart hatayı elde etmek için $\hat{Q}_m^{(0.25)}, m = 1, 2, ..., M$ 'nin s_{boot} standart hatası kullanılır.

Bootstrap tahminleri arzu edilen popülâsyonun yüzde 25'i için tahmin güven aralığı da kullanılabilir. Bu yaklaşımların farklı türü bunun için kullanılabilir. Bir diğer alternatif ise örneklemden orijinal tahmini kullanmaktır, bunun standart hatası ve sboot $\hat{Q}^{(0.25)} \pm Z_{\alpha/2} S_{boot}$ % 100(1-α) ile güven aralığı tahminlenir.

Diğer bir alternatif bootstrap tahminlerinin deneysel kantillerinden yararlanmaktır. Bootstrap için %95 güven aralığı için, bootstrap tahminlerinin 0.25 ve 0.95 kantilleri aralığı alınabilir. Daha da özelleştirmek için, eğer en büyükten en küçüğe bootstrap tahminleri $\hat{Q}_1^{(0.25)}, \hat{Q}_2^{(0.25)}, ..., \hat{Q}_{1000}^{(0.25)}$ sıralanırsa sıra istatistiğini $\hat{Q}_1^{(0.25)}, \hat{Q}_2^{(0.25)}, ..., \hat{Q}_{1000}^{(0.25)}$ verir ve ayrıca güven aralığı $\left[\hat{Q}_{50}^{(0.25)}, \hat{Q}_{951}^{(0.25)}\right]$ olarak alınır. Benzer bir yapı arzu edilen her olasılık için güven aralığı mümkündür.

QRM de bu fikir genişletilerek, bağımlı ve bağımsız $(x_i, y_i), i = 1, 2, ..., n$ değişken örneklerinden meydana gelen $\beta^{(p)} = \beta_1^{(p)}, \beta_2^{(p)}, ..., \beta_k^{(p)}$ kantil regresyon parametre tahminlerine ait standart hatalar tahminlenmek istenir. X,Y bootstrap çifti yaklaşımı bu çiftlerin yeri değiştirilerek yapılan örnklemden n hacimli bootstrap örneklemi elde edilir ve bunlar mikro birimler denilmekte (burada x ve y ayrı birey olan verilerdir). Bir örnekteki çiftlerin aynı veri kopyası onların çeşitliliğine göre hesaplanır bu yüzden k defa görünen bir kopya k defadan fazla örnekte görünecektir.

Her bir bootstrap örneği bir parametre tahminine yol açar ve M bootstrap tahminlerinin standart hataları alınarak belli bir katsayı tahmini olan $\hat{\beta}_i^{(p)}$'nin standart hatası tahminlenir. Bootsrap tahminleri çeşitli şekillerde bir birinden ayrı kantil regresyon parametresi $\hat{\beta}_i^{(p)}$ için güven aralığı elde edilebilir. Bu metotlardan biri standart hata ve güven aralığı $\hat{\beta}_i^{(p)} \pm Z_{\alpha/2} s_{boot}$ kullanmaktır. Alternatif olarak örnek kantillerine dayalı güven aralığı temel alınabilir. Örneğin $\hat{\beta}_i^{(p)}$'nin %95 güven aralığı için bootstrap tahminleri $\hat{\beta}_m^{(p)}$ M den oluşan örneklemin 97.5 ci pörsentili için

2.5 ci pörsentildir. Çoklu kantil regresyon örneğin 19 eşit uzaklıktaki kantille (P=0.05…0.95) ortak olarak dikkate alınabilir. 19 modelin tüm olası kantil regresyon katsayıları arasındaki kovaryans tahmin edilebilir. Örneğin, model iki değişkeniyle $\hat{\beta}_2^{(p)}$ ve $\hat{\beta}_3^{(p)}$ katsayılarla ilişkili bir de sabit terimli olarak oluşturulduğunda, bu durumda 3X19=57 katsayılar tahminlenir ve bunlara bağlı olarak 57X57 kovaryans matrisi oluşmuş olacaktır. Bu matris her kantildeki $\left(Var\hat{\beta}_1^{(0.05)}\right)$ ve $\left(Var\hat{\beta}_1^{(0.5)}\right)$ her bir değişkenin katsayısı için sadece varyanslarını sağlamaz, aynı zamanda benzer değişkenler için $\left(Cov\hat{\beta}_1^{(0.05)}\right)$ ve $\left(Cov\hat{\beta}_1^{(0.5)}\right)$ farklı kantillerde kovaryansları da tahminleyebilir.

Varyans ve kovaryansın birlikte tahminlenmek, benzer değişkenlere ilişkin $\hat{\beta}_i^{(p)}$ ve $\hat{\beta}_i^{(q)}$ katsayılarına eşdeğer hipotez testleri yapılabilir, fakat p ve q arasındaki uzaklığa karşılıkta Wald testi kullanılmalıdır.

$$\text{Wald İstatistiği}=\frac{\left(\hat{\beta}_j^{(p)}-\hat{\beta}_j^{(q)}\right)}{\hat{\sigma}^2_{\hat{\beta}_j^{(p)}-\hat{\beta}_j^{(q)}}}$$

Paydadaki $\hat{\sigma}^2_{\hat{\beta}_j^{(p)}-\hat{\beta}_j^{(q)}}$ terimi $\hat{\beta}_j^{(p)}-\hat{\beta}_j^{(q)}$'nin varyans farklarını hesaplar bunlarda sağ taraftaki kovaryans ve varyansların tahminlerinin yerine konulmasıyla aşağıdaki denklem kullanılarak elde edilir.

$$Var\left(\hat{\beta}_j^{(p)}-\hat{\beta}_j^{(q)}\right)=Var\left(\hat{\beta}_j^{(p)}\right)+Var\left(\hat{\beta}_j^{(q)}\right)-2Cov\left(\hat{\beta}_j^{(p)},\hat{\beta}_j^{(q)}\right) \qquad [2.33]$$

Sıfır hipotezi altında Wald istatistiği bir serbestlik dereceli yaklaşık olarak χ^2 dağılımına sahiptir.

Daha genel bir ifadeyle, kantillere karşı çoklu kantilrin eşitliğini test edebiliriz. Örneğin, modelde iki değişkene ek olarak sabit terimin de var olduğunu varsayalım ve koşullu p'inci ve q'uncu kantil fonksiyonların birinde diğerine bir kayma olup olmadığını test etmek isteyelim; yani,

$H_0: \beta_2^{(p)}=\beta_2^{(q)}$ ve $\beta_3^{(p)}=\beta_3^{(q)}$ buna karşın

$H_a: \beta_2^{(p)}\neq\beta_2^{(q)}$ ya da $\beta_3^{(p)}\neq\beta_3^{(q)}$

Sabit terim ihlal edilmiş bu durumda. Wald istatistiğini test etmek için aşağıdaki gibi tanımlanabilir. İlk olarak, kovaryans tahminlemek için

$\sum_{\hat{\beta}^{(p)}-\hat{\beta}^{(q)}}=\begin{bmatrix}\hat{\sigma}_{11} & \hat{\sigma}_{12}\\ \hat{\sigma}_{21} & \hat{\sigma}_{22}\end{bmatrix}$ den $\hat{\beta}^{(p)}-\hat{\beta}^{(q)}$'nun $\sum\hat{\beta}^{(p)}-\hat{\beta}^{(q)}$ kovaryans matrisi tahminlemesi elde edilir ve burada girdiler tahminlenen vekil varyans ve kovaryanslar aşağıda gösterildiği gibi elde edilir.

$$\sigma_{11}=Var\left(\hat{\beta}_1^{(p)}-\hat{\beta}_1^{(q)}\right)=Var\left(\hat{\beta}_1^{(p)}\right)+Var\left(\hat{\beta}_1^{(q)}\right) -2Cov\left(\hat{\beta}_1^{(p)},\hat{\beta}_1^{(q)}\right) \quad [2.34]$$

$$\sigma_{12}=\sigma_{21}=Cov\left(\hat{\beta}_1^{(p)},\hat{\beta}_2^{(q)}\right)+Cov\left(\hat{\beta}_1^{(p)},\hat{\beta}_2^{(q)}\right)-Cov\left(\hat{\beta}_1^{(p)},\hat{\beta}_2^{(q)}\right) -Cov\left(\hat{\beta}_1^{(p)},\hat{\beta}_2^{(q)}\right) \quad [2.35]$$

$$\sigma_{22}=Var\left(\hat{\beta}_2^{(p)}-\hat{\beta}_2^{(q)}\right)=Var\left(\hat{\beta}_2^{(p)}\right)+Var\left(\hat{\beta}_2^{(q)}\right) -2Cov\left(\hat{\beta}_2^{(p)},\hat{\beta}_2^{(q)}\right) \quad [2.36]$$

Daha sonra test istatistiği de aşağıda gibi hesaplanır:

$$W=\begin{bmatrix}\hat{\beta}_1^{(p)}-\hat{\beta}_1^{(q)}\\ \hat{\beta}_2^{(p)}-\hat{\beta}_2^{(q)}\end{bmatrix}^t\sum\nolimits^{-1}_{\hat{\beta}^{(p)}-\hat{\beta}^{(q)}}\begin{bmatrix}\hat{\beta}_1^{(p)}-\hat{\beta}_1^{(q)}\\ \hat{\beta}_2^{(p)}-\hat{\beta}_2^{(q)}\end{bmatrix} \quad [2.37]$$

Burada sıfır hipotezi altında iki serbestlik dereceye sahip yaklaşık olarak χ^2 dağılımına sahiptir.

2.11. QRM'NİN UYUM İYİLİĞİ

Doğrusal regresyon modellerinde uyum iyiliğinin ölçümü R^2 tarafından yapılır.

$$R^2=\frac{\sum_i\left(\hat{y}_i-\bar{y}\right)^2}{\sum_i\left(y_i-\bar{y}\right)^2}=1-\frac{\sum_i\left(y_i-\hat{y}\right)^2}{\sum_i\left(y_i-\bar{y}\right)^2} \quad [2.38]$$

[2.41] eşitliğinin ikinci kısmındaki ifade gözlemlenmiş y değeri ile modelden elde edilen $\hat{y}_i$ tahmin değeri arasındaki uzaklığın kareler toplamıdır. Diğer bir yandan payda, gözlemlenen y_i değeri ile modelde sadece sabit terim bulunursa elde edilecek tahmin değeri arasındaki uzaklığın kareler toplamını ifade eder. R^2 değeri modeldeki bağımlı değişken tarafından açıklanan bağımsız değişkenlerin değişim

oranı olarak yorumlanır. Bu değer sıfır ile bir arasındadır ve büyük değerlerde modelin uyum iyiliği R kare tarafından daha iyi açıklanır.

R^2 İstatistiği kolaylıkla kantil regresyon modeli içinde geliştirilebilir. Doğrusal regresyon modeller hata kareler minimizasyonuna dayanırken kantil regresyon modeller ise ağırlıklandırılmış uzaklık toplamının minimizasyonuna dayanır. Bu uzaklık $\sum_{i=1}^{n} d_p\left(\hat{y}_i, y_i\right)$ farklı ağırlıklar kullanılır $y_i > \hat{y}_i$ ya da $y_i < \hat{y}_i$ olmasına bağlı olarak. Kenker ve Machado (1999) uyum iyiliği modeli için sadece sabit terimli model için ağırlıklandırılmış uzaklık toplamı kıyaslanmasını önermişlerdir. Tüm p'inci kantil regresyon model için ağırlıklandırılmış uzaklık toplamı olarak V1(p) verilmiş olsun, ve V1(p) sadece sabit terim içeren ağırlıklandırılmış uzaklık toplamı olarak verilmiştir. Tek değişkenli model için aşağıdaki gibi örneklendirilebilir.

$$V^1(p) = \sum_{i=1}^{n} d_p\left(\hat{y}_i, y_i\right)$$
$$= \sum_{y_i \geq \beta_0^{(p)} + \beta_1^{(p)} x_i} p\left|y_i - \beta_0^{(p)} + \beta_1^{(p)} x_i\right| \qquad [2.39]$$
$$+ \sum_{y_i \leq \beta_0^{(p)} + \beta_1^{(p)} x_i} (1-p)\left|y_i - \beta_0^{(p)} + \beta_1^{(p)} x_i\right|$$

ve

$$V^0(p) = \sum_{i=1}^{n} d_p\left(y_i, \hat{Q}^{(p)}\right) = \sum_{y_i > \bar{y}} p\left|y_i - \hat{Q}^{(p)}\right| + \sum_{y_i < \bar{y}} (1-p)\left|y_i - \hat{Q}^{(p)}\right| \qquad [2.40]$$

Sadece sabit terim içeren model için, uygun sabit terim $y_{1,} y_2 \ldots y_n$ örneklem için p'ninci qunatile $\hat{Q}^{(p)}$ örneğidir. Bunun uyum iyiliğide aşağıdaki gibi tanımlanır.

$$R(p) = 1 - \frac{V^1(p)}{V^0(p)} \qquad [2.41]$$

Bu eşitlikten dolayı $V^1(p)$ ve $V^0(p)$ değerleri negatif olamaz ve $R(p)$ en fazla bir değerini alabilecektir. Ek olarak ağırlıklı uzaklık toplamı tam uyumlu model için minimize edildiğinden $V^1(p)$ asla $V^0(p)$'dan büyük olamaz, bu yüzdende $R(p)$'nin alacağı değer aralığı 0 ile 1 arasında olacak ve bu değerin büyüklüğü daha uyumlu bir model olduğunu da gösterecektir. Yukarıdaki eşitlikte p de QRM'nin yerel olarak uyum iyiliği ölçülmüştür. QRM'nin global olarak

değerlendirilmesi için incelenen ortak $R(p)$'lerin tüm dağılımın ele alınması gerekmektedir[48].

Yukarıda tanımlanan $R(p)$ değeri sadece sabit terimli model ile her bağımsız değişkenli ayrıca sabit terimli modellerin karşılaştırılmasına izin verir. Bu Koenker ve Machado tarafından yuvalanmış model için tanıtılan bir karşılaştırılan uyum iyiliği kısıtlamasıdır. Daha geniş bir ifadeyle, verilen bir modeldeki iyileştirme daha sınırlı model göreli $R(p)$'nin ölçülmesi olabilir. Sonuç sayısı Göreli $R(p)$ değeri ile gösterilir. Daha az kısıtlı p'inci kantil regresyon modeli için ağırlıklı uzaklık toplamı olarak $V^2(p)$, daha fazla kısıtlı p'inci kantil regresyon modeli için ağırlıklı uzaklık toplamı olarak da $V^1(p)$ gösterilir. Göreli $R(p)$ değeri aşağıdaki gibi ifade edilebilir.[49]

$$Göreli\, R(p) = 1 - \frac{V^2(p)}{V^1(p)} \qquad [2.42]$$

(2.42) denklemi eski R^2'lerin benzeri gibi düşünülebilir. Buradaki göreli R^2 uyum iyiliğini kriteri olarak tanımlanabilmektedir.

[48] Roger Koenker ve Jose A.F. Machado, "Goodness of Fit and Related Inference Processes For Quantile Regression", **Journal of the American Statistical Association**, Cilt:94, 1999, ss. 1296-1310.
[49] Hao ve Naiman, s.52.

ÜÇÜNCÜ BÖLÜM
UYGULAMA

Bu bölümde daha önceki bölümlerde anlatılan ekonometrik modellerin uygulaması karşılaştırmalı olarak verilmiştir. Çalışmanın amacı, veriler ve değişkenler, yapılan ekonometrik analizlerle birlikte değerlendirme ve sonuç kısımlarına yer verilmiştir.

3.1. UYGULAMANIN AMACI

Ülkelerin gelişmişlik düzeyini belirleyen en önemli unsurlardan biri gelir dağılımıdır. Geçmişten günümüze devletler refah düzeylerini yani iktisadi ve sosyal gelişmelerini hızlandırmak için yoğun bir çaba içerisinde olmuşlardır. Refah düzeyini artırmanın önemli ilkelerinden biride insanların ekonomik anlamda ihtiyaçlarını karşılayabilecek gelire bağlı olmaktadır. Gelişmiş ve gelişmekte olan ülkelerin karşılaştıkları büyük problemlerden biri de gelir dağılımındaki farklılıklardır. Çünkü bir ülkede gelişmişlik düzeyi ülkenin gelirinin o ülkede yaşayan vatandaşlar tarafın eşit bir şekil paylaşılmasıyla ilgilidir[50].

Bu çalışmanın amacı Türkiye'deki gelir farklılığını 2002 ile 2010 yıllarını dikkate alarak, hem bu söz konusu yıllar itibari ile hem de bu on yıllık süre içerisinde gelişmişlik düzeyini göstermek bakımından bu yılları bir biri ile kıyaslayarak farklı gelir gruplarındaki, farkı sosyo-kültürel yapıya sahip gelirleri karşılaştırmak.

3.2. DEĞİŞKENLER VE VERİLER

Bu çalışmada kullanılan veriler, Türkiye İstatistik Kurumu (TUİK) tarafından; hanelerin yaşam düzeylerini, gelirini tüketimini sosyo-ekonomik durumlarını incelemek üzere 2002 yılından itibaren her yıl düzenli olarak hanekalkı bütçe anketleri kullanılmıştır. Hanedeki bireylerin ve bunlardan meydana gelen hanehalklarının tüketim yapılarını, gelir düzeylerini; sosyo-ekonomik gruplara

[50] Özlem Kiren Gürler ve Şenay Üçdoğruk, "Türkiye'de Cinsiyete Göre Gelir Farklılığının Ayrıştırma Yöntemiyle Uygulanması", **Journal of Yasar University**, Cilt:12, ss.571-589.

ayırarak, kır, kent ve bölgelere göre ortaya koyan TUİK. Bu çalışmayla tüketim alışkanlıkları, tüketim harcaması türleri ile mal ve hizmet harcamalarının çeşitliliği, hanehalkının sosyo-ekonomik özellikleri, hanehalkı fertlerinin çalışma durumları, hanehalkının toplam geliri, gelirin kaynaklar hakkındaki bilgileri sunmaktadır.

Türkiye İstatistik kurumu tarafından yapılan Hanehalkı Bütçe Anketi 1 Ocak – 31 Aralık tarihleri arasında bir yıl süreyle her ay değişen ve tabakalama yöntemi ile seçilen örnek hanehalkına uygulanmaktadır. Türkiye geneli, kentsel ve kırsal yerler ayrımında tüketim harcaması göstergeleri elde edilmektedir. 2002 yıllına ait veri setinde Türkiye geneli 9555 hanehalkından toplam fert sayısı 107610 ile çalışılmıştır. 2010 hanehalkı yine aynı tarihler arasında bir yıl süre ile her yıl değişen 1104, yıllık toplam 13248 örnek hanehalkına uygulanarak, Türkiye geneli, kentsel kırsal yerler ayrımında tüketim harcaması ve gelir göstergeleri ile çalışılmıştır.

TUİK'ten elde edilen bu veri setlerinden 2002 ve 2010 yıllarına ait fert kılavuzu kullanılmıştır. Çalışmada kullanılan veriler: Hanehalkı geliri, yaş, eğitim durumu, cinsiyet değişkenleridir. Türkiye'deki hanelerin beşeri sermayeleri incelenmek istendiğinden hem 2002 hem de 2010 fert veri setindeki çalışmayan bireyler veri setinden çıkarılmıştır. 2002 yılana ait geriye kalan gözlem sayısı 11833, 2010 yılına ait kalan gözlem sayısı ise 12990 dır.

3.3. TANIMLAYICI İSTATİSTİKLER

Tanımlayıcı istatistikler tablosuna bakıldığında (tablo 1) 2002 yılından 2010 yılına kadar ortalama gelirde ciddi bir artışın olduğu görülmektedir. 2002 yıllında ortalama toplam yıllık ortalama gelir yaklaşık 4600 lira iken, 2010 yılın ortalama yıllında fert geliri 12450 civarında olduğu görülmektedir. Türkiye'nin 2001 yılında yaşadığı ekonomik kriz sonucun bir sonraki yıllarda gelirin düşük olması beklentilere uygun olmaktadır. 2002 yılından itibaren Ekonomi otoritelerin uyguladığı ekonomik reformlarla örneklemdeki hanehalkı bireylerin gelirinde yaklaşık üç katlık bir artış olmuştur.

Bağımsız değişken incelendiğinde ortalama yaşın 35'ten 38'e çıktığı görülmektedir. Eğitim durumu incelendiğinde ise okuma yazma bilmeyenler ile diplomasız okuryazar ortalamasında çok az bir artış görülmektedir. 8 yıllık zorunlu

eğitime nedeniyle ilkokulların ortalamasında ise azalma olmuştur. Eğitimde dikkat çeken önemli unsurlardan biri de yüksek okul ve fakülte bitirenler ile yüksek lisans ve doktora ortalamasındaki artıştır.

Tablo 1: Tanımlayıcı İstatistikler Tablosu

Değişkenler	**2002**		**2010**	
	Ortalama	**Std. Sapma**	**Ortalama**	**Std. Sapma**
Bağımlı Değişkenler				
Toplam Yıllık Fert geliri	4608.831	7191.409	12456.73	16758.12
Logaritmik Toplam Yıllık Fert geliri	8.060175	1.176257	9.060581	1.188706
Yaş	35.83166	12.72117	38.35081	.0980124
YaşKare	1445.722	1028.018	1635.145	1043.38
Eğitim (bitirilen eğitim yılı)	-	-	-	-
Okur Yazar Değil	.062368	.2418329	.0632794	.2434743
Okur Yazar	.0475788	.2128823	.0540416	.2261084
İlk Okul	.4637877	.498708	.3919169	.4881971
Ortaokul ve Dengi	.1313276	.3377726	.1597383	.3663772
Lise ve Dengi	.1963154	.3972266	.1921478	.3940038
Lisans ve Önlisans	.0936364	.2913346	.1291763	.3354079
Lisans Üstü	.0049861	.0704387	.0096998	.0980124
Medeni Durum	-	-	-	-
Evli	.7479084	.4342319	.7578137	.428423
Bekar	.2520916	.4342319	.2421863	.428423
Cinsiyet	-	-	-	-
Kadın	.246345	.4309	.3137798	.4640459
Erkek	.753655	.4309	.6862202	.4640459

Gözlem sayısı (n) 2002 yılı için 1183, 2010 yılı içinse 12990 olarak elde edilmiştir.

3.4. KOŞULLU ORTALAMA VE MEDYAN REGRESYON KARŞILAŞTIRMASI

Bu aşamada EKK ve Medyan Regresyon bağımlı değişken olan toplam yılık gelir herhangi bir transformasyon yapılmadan ve gelirin logaritması alınarak karşılaştırılmıştır.

İzleyen tablo 2 de tüm katsayılar %1 anlamlılık düzeyinde anlamlıdır. İncelenen bu modelde koşullu ortalama bağımsız değişkenler ve gelirin merkezi yer ölçüsü arasındaki ilişkiyi sunmaktadır. Koşullu ortalama modelleri, EKK yöntemi gibi ortalamayı tahminlediğinden dolayı bu tür modeller gelir dağılımının üst kuyruklarını (sağa çarpık) yakalama eğilimindedirler. Örneğin yaş değişkenin katsayısı EKK yöntemine göre yaklaşık 1032 olarak elde edilmiştir. Medyan Regresyon yönteminde (Kantil regresyon q=0.50) bu katsayı yaklaşık 904 olarak tahminleşmiştir. Bu durum koşullu ortalamanın sapan gözlemlerden ve yüksek değerlerden nasıl etkilendiğini ortaya koyması bakımından ilgi çekicidir. Gelir değişkenin logaritmik dönüşümü yapılarak modeller tahmin edildiğinde de aynı durumun söz konusu olduğu görülmektedir.

Tablo 2: EKK ve Medyan Regresyon: 2010 yılı

Değişkenler	Bağımlı Değişken: Gelir				Bağımlı Değişken: log(Gelir)			
	EKK		MEDYAN		EKK		MEDYAN	
	Katsayılar	Standart Hata	Katsayılar	Standart Hata	Katsayılar	Standart Hata	Katsayılar	Standart Hata
Yaş	1031.866	58.88693	903.809	28.05944	.1323642	.0044837	.1136788	.0038838
Yaşkare	-8.64724	.7411019	-8.448947	.3473244	-.0012604	.0000552	-.0011084	.0000478
Eğitim	2439.26	50.47473	1852.289	23.67027	.1697248	.0036461	.1337492	.0031585
Sabit	-24841.34	1142.928	-19552.16	535.2932	5.164338	.0866637	6.019271	.0750765

Tüm değişkenler %1, %5 ve %10 anlamlılık düzeyinde anlamlıdır.

Tablo 3: Uyum İyiliği R^2 Karşılaştırılması

		Uyum iyiliği: R^2
EKK	**Gelir**	**0.1946**
Medyan Regresyon	**Gelir**	**0.1813**
EKK	**Log(gelir)**	**0.2529**
Medyan Regresyon	**Log(gelir)**	**0.1594**

3.5. NORMALLİK TESTİ

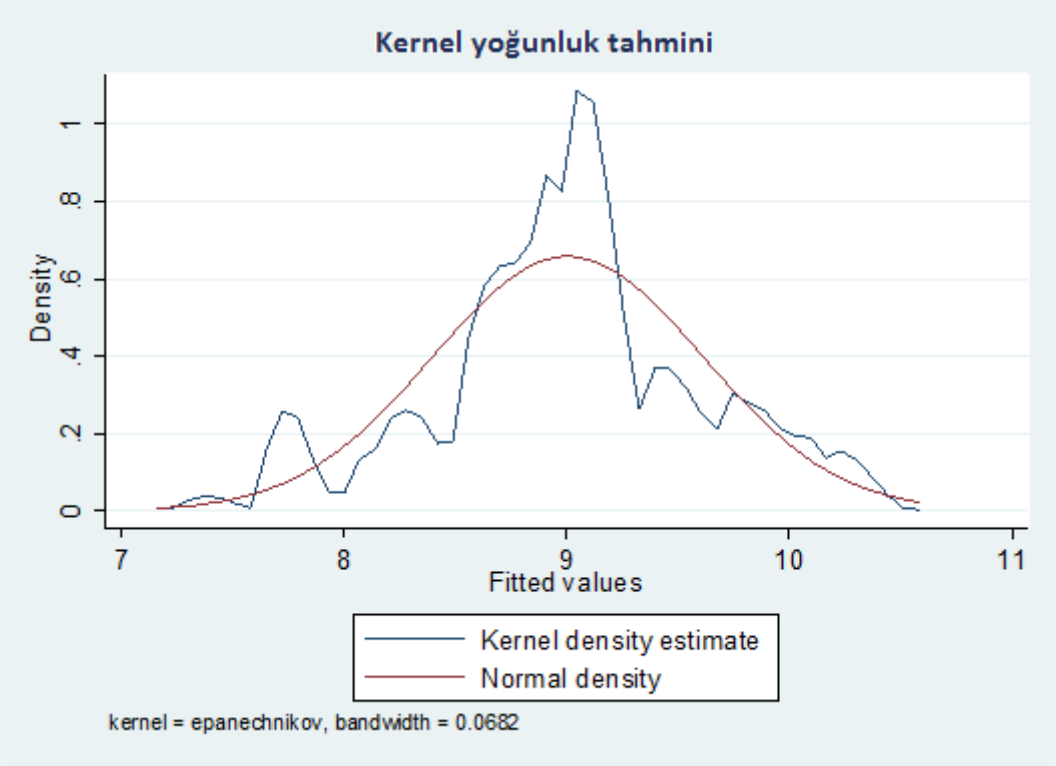
Modellere ilişkin normal dağılım varsayımın sağlanıp sağlanmadığı skewness/kurtosis normallik testi ile tablo 4 de test edilmiştir.

Tablo 4: Skewness/Kurtosis Normallik Testi:

Değişken	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	Prob>chi2
hata	1.3e+04 0.0000	0.0049	47.61	0.000

%5 anlamlılık düzeyinde hatalar normal dağılmamaktadır.

Şekil 6: Kernel Yoğunluk Fonksiyonu:



Yukarıdaki normal dağılım fonksiyonu ile kernel yoğunluk tahmini kıyaslamasına bakıldığında serilerin normal dağılama sahip olmadığı görülmektedir.

3.6. MEDYAN REGRESYON VE EN KÜÇÜK KARELER YÖNTEMİ

Herhangi bir logaritmik dönüşüm yapılmadan Tablo 5 ve Tablo 6 da 2002 ve 2010 yıllarını EKK VE QR (kantil regresyon) değerleri 0.50'inci kantil için modeller tahminlenmiştir. Tablo 5 te EKK yönetimi ile tahmin edilen modelde kadın katsayısı %1, %5, %10 anlamlılık düzeyinde anlamlı değilken diğer model ve değişkenler %1 anlamlılık düzeyinde anlamlıdır.

2002 yılına ait model incelendiğinde katsayı değerleri olarak medyan regresyona göre daha büyük katsayılar tahminlenmiştir. Parametrelerin işaret yönü her iki modelde de katsayılar beklentilere uygundur. Hanehalkı bireylerinin gelirine yaş ve kişinin aldığı eğitim pozitif bir etki yaparken, deneyimi ifade eden yaşın karesi ve bekâr ve kadın olması ise gelirine negatif bir etki yapmaktadır. Yani evli olan bireylerin bekâr olanlara göre gelirinin daha fazla olduğu, erklerinde kadınlara göre daha fazla gelire sahip olduğu görülmektedir.

Her iki model için çarpıcı olan eğitimin gelir üzerindeki etkisinin farklı eğitim düzeylerine göre daha fazla olmasıdır. Örneğin ortaokul mezunlarının ilkokul mezunlarına göre geliri 1120 TL daha fazla iken bu değer lise mezunların yaklaşık olarak iki kat daha fazladır. Fakat medyan regresyonda ortaokul mezunlarının geliri ilkokula göre 534 TL daha fazla iken lise mezunlarında bu fark 3 katına çıkmaktadır. 2-3 yıllık yüksek okul ve fakülte bitirenlerin ilkokul mezunlarına göre daha gelirleri daha da artmakta iken, özellikle yüksek lisans ve doktora yapmış olan bireylerin geliri arasında ciddi farklar oluşmuştur. Fakat bu büyük farklılık medyan regresyonda bu kadar belirgin değildir. Söz konusu durum gelirin dağılımıyla ilişkili olarak açıklanabilir. Her iki modelde de yaşın gelire etkisi pozitif olmasına rağmen deneyimi ifade eden yaşın karesinin gelire etkisi negatif çıkmıştır. Bu hanehalkı bireylerinin deneyimi artıkça gelirinin artmadığını ve deneyimden ziyade kişinin aldığı eğitimin daha çarpıcı etkiye bir etkiye sahip olduğunu gösterir.

2010 yılında tüm katsayılar %1 anlamlılık düzeyinde anlamlıdır. Katsayıların işaret yönleri 2002 yılı veri seti ile aynı iken, gelirin ortalama olarak daha fazla olmasına karşın eğitim düzeyleri arasındaki makas biraz daha kapanmıştır. Örneğin 2002 yılına ait modelde ilkokulu bitirenlere göre ile orta okulu bitirenler ile liseyi

bitiren bireyler arasındaki gelir farkı iki kattan fazla iken, 2010 yılında bu 0.5 kata kadar düşmüştür. Benzer durum diğer eğitim düzeyleri içinde geçerlidir.

Tablo 7 ve Tablo 8 ise bağımlı değişken olan fertlerin yıllık gelirlerine logaritmik dönüşüm yapılarak elde edilmiştir. Burada tüm katsayılar %1 anlamlılık düzeyinde anlamlı ve tüm değişkenlerin katsayılarının işaret yönü beklentilere uygundur. Logaritmik modelde EKK ile Medyan regresyon tahmin sonuçları arasında daha belirgin bir fark ortaya çıkmamaktadır. Her iki model karşılaştırmasında da benzer sonuçlar ortaya çıkmıştır.

Tablo 5: En Küçük Kareler ve Medyan Regresyon: 2002 Yılı

Değişkenler	2002			
	EKK		MEDYAN	
	Katsayılar	Standart Hata	Katsayılar	Standart Hata
Bağımlı değişken: gelir				
Yaş	363.9313	28.91381	207.2639	12.3486
Yaşkare	-2.98039	.3391868	-1.842425	.1449391
Eğitim (bitirilen eğitim yılı)				
Okur Yazar Değil	-2484.531	273.4703	-1754.368	117.1341
Okur Yazar	-1207.181	295.4263	-673.1972	126.1453
Ortaokul ve Dengi	1120.561	190.7955	534.9552	81.76207
Lise ve Dengi	2585.102	164.3804	1514.385	70.44336
Lisans ve Önlisans	5804.668	216.3931	4224.805	92.73915
Lisans Üstü	11605.62	852.5823	8822.435	362.4894
Medeni Durum				
Bekar	-279.6031	192.3779	-339.9007	82.21386
Cinsiyet				
Kadın	-2577.753	147.6821	-1880.827	63.1749
Sabit	-4460.771	603.6816	-1527.482	257.6511

Temel sınıflar: Her iki modelde 2002 yılına göre ilk okul mezunu evli erkek (sadece çalışan bireyler modele dahil edilmiştir).
Kadın değişkeni dışındaki tüm katsayılar %1, %5 ve % 10 anlamlılık düzeyinde anlamlıdır.

Tablo 6: En Küçük Kareler ve Medyan Regresyon: 2010 yılı

Değişkenler	2010			
	EKK		MEDYAN	
	Katsayılar	Standart Hata	Katsayılar	Standart Hata
Bağımlı değişken: gelir				
Yaş	4614.852	326.1896	3678.066	175.0882
Yaşkare	-206.7113	19.8607	-181.9702	10.66142
Eğitim (bitirilen eğitim yılı)				
Okur Yazar Değil	-4659.97	585.1367	-3882.214	314.1372
Okur Yazar	-770.2624	600.0485	-837	321.2989
Ortaokul ve Dengi	3809.817	420.71	2393.643	225.7931
Lise ve Dengi	6941.337	371.3513	5193.369	199.4157
Lisans ve Önlisans	16345.86	415.6167	13908.64	223.1289
Lisans Üstü	39640.96	1307.433	24333.64	699.4989
Medeni Durum				
Bekâr	-1852.548	389.4135	-978.0356	208.9748
Cinsiyet				
Kadın	-7247.721	289.1161	-6392	155.1845
Sabit	-10177.47	1331.815	-6612.072	714.7738

Temel sınıflar: Her iki modelde de 2002 yılına göre ilkokul mezunu evli erkek olarak alınmış (sadece çalışan bireyler modele dâhil edilmiştir). Tüm katsayılar %1, %5 ve % 10 anlamlılık düzeyinde anlamlıdır.

Tablo 7: En Küçük Kareler ve Medyan Regresyon: 2002 yılı

Değişkenler	2002			
	EKK		MEDYAN	
	Katsayılar	Standart Hata	Katsayılar	Standart Hata
Bağımlı değişken: log(gelir)				
Yaş	0.1114265	0.0047272	0.0917736	0.0038019
Yaşkare	-0.0010444	0.0000552	-0.0008451	0.0000444
Eğitim (bitirilen eğitim yılı)				
Okur Yazar Değil	-0.7251582	0.0495893	-0.7093859	0.039851
Okur Yazar	-0.442928	0.048624	-0.3859778	0.0390724
Ortaokul ve Dengi	0.1653842	0.0301038	0.1461504	0.0242094
Lise ve Dengi	0.5145341	0.0255245	0.4757754	0.0205268
Lisans ve Önlisans	0.9906715	0.0329026	0.8551431	0.0264494
Lisans Üstü	1.468038	0.1263087	1.383392	0.1007735
Medeni Durum				
Bekar	-0.1353503	0.0310074	-0.1579388	0.0249411
Cinsiyet				
Kadın	-0.7234245	0.0261954	-0.4765104	0.0210554
Sabit	5.516364	0.0989517	6.001448	0.0795828

Temel sınıflar: Her iki modelde de 2002 yılına göre ilkokul mezunu evli erkek olarak alınmış (sadece çalışan bireyler modele dâhil edilmiştir). Tüm katsayılar %1, %5 ve % 10 anlamlılık düzeyinde anlamlıdır.

Tablo 8: En Küçük Kareler ve Medyan Regresyon: 2010 Yılı

Değişkenler	2010			
	EKK		MEDYAN	
	Katsayılar	Standart Hata	Katsayılar	Standart Hata
Bağımlı değişken: log(gelir)				
Yaş	.5755552	.0233257	.4618984	.0210534
Yaşkare	-.0285574	.0014171	-.023274	.0012791
Eğitim (bitirilen eğitim yılı)				
Okur Yazar Değil	-.6641526	.0478276	-.6561642	.0431671
Okur Yazar	-.1577099	.0452782	-.2459203	.0408302
Ortaokul ve Dengi	.1862585	.0291173	.1679204	.0262843
Lise ve Dengi	.5525843	.0253679	.4378968	.022909
Lisans ve Önlisans	1.144031	.0281281	.929938	.025404
Lisans Üstü	1.662624	.0863255	1.36527	.0776958
Medeni Durum				
Bekar	-.0674866	.0273792	-.1037696	.0247226
Cinsiyet				
Kadın	-.857555	.0214963	-.4984355	.0194135
Sabit	6.502607	.0952936	7.171453	.0860091

Temel sınıflar: Her iki modelde 2010 yılına göre ilk okul mezunu evli erkek (sadece çalışan bireyler modele dahil edilmiştir).

Tüm katsayılar %1, %5 ve % 10 anlamlılık düzeyinde anlamlıdır.

3.7. BAĞIMSIZ DEĞİŞKEN KATSAYI GRAFİKLERİ

Kantil regresyon modellerinin Doğrusal regresyon modellerinin önemli bir ayırımı da bağımsız değişkenlere ait katsayı grafikleridir. Bu grafik logaritmik ve logaritmiksiz olmak üzere iki şekilde Tablo 8 ve Tablo 9 da verilmiştir. STATA 12 programında qrqreg kodu ile belirli 19 kantille (0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95) elde edilmiştir. Bu katsayı grafiklerini çizmek için yeterli büyük örneklem sağlamak için Boostrap yöntemiyle 500 tekrar sayısı seçilmiştir.

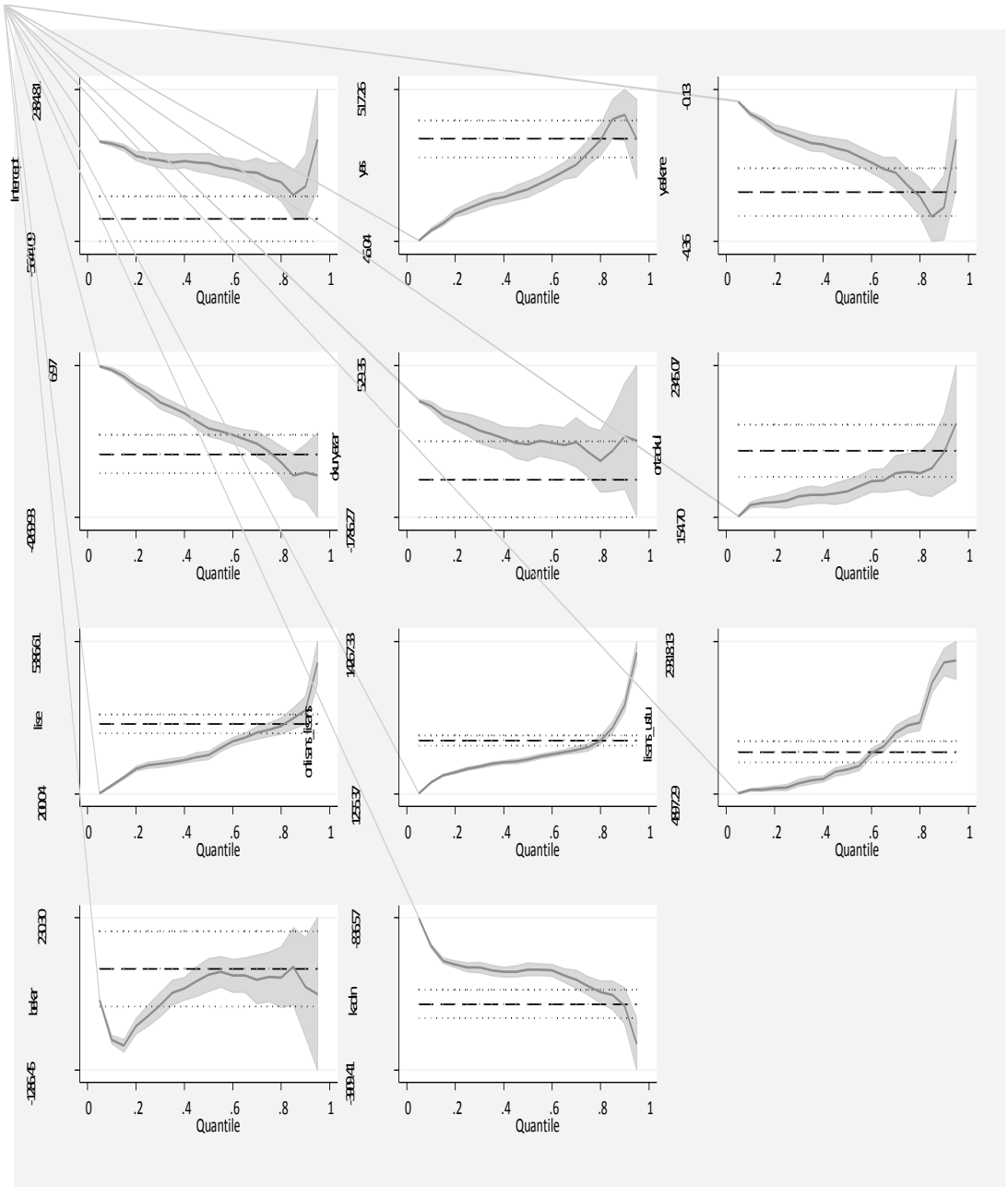
2002 ve 2010 veri setiyle elde edilen modellerin katsayıların işaret yönleri aynı olmasından dolayı her iki yılın katsayı grafikleri benzer olmaktadır. Katsayılar için yatay bir doğru, bağımlı değişkenin p değerine göre değişiklik olmadığını gösterir. Yani bağımlı değişken kantilleri üzerine bağımsız değişkenlerin sabit etkisi tüm kantiller için aynı olduğu anlamına gelmektedir. Diğer bir deyişle, böyle yatay bir doğru tüm bağımsız değişkenlerin merkezi yer ölçüsündeki değişim üzerinde pek bir etki yaratmadığı anlamına gelir. Eğer doğru sıfır değerleri arasında değilse, düm düz yatay bir doğru merkezi eğilim ve ölçek değişikliğini gösterir. Bu durumda merkezi eğilim ölçümü medyan tarafından gösterilir.

Pozitif bir medyan katsayısı sağa doğru merkezi konumun değiştiğini gösterirken sola doğru bir medyan ise sola doğru merkezi eğilimin değiştiğini gösterir. Dümdüz yukarıya doğru eğimli bir doğru ise pozitif ölçek değişimini ifade etmektedir. Kesikli noktalar EKK %95 güven aralığını gösterirken, bu ikisinin arasındaki kesikli çizgide EKK doğrusunu göstermektedir. Aşağıdaki tablolarda doğrunun etrafındaki gölgeli alan ise kantil regresyonun %95 güven aralığını göstermektedir.

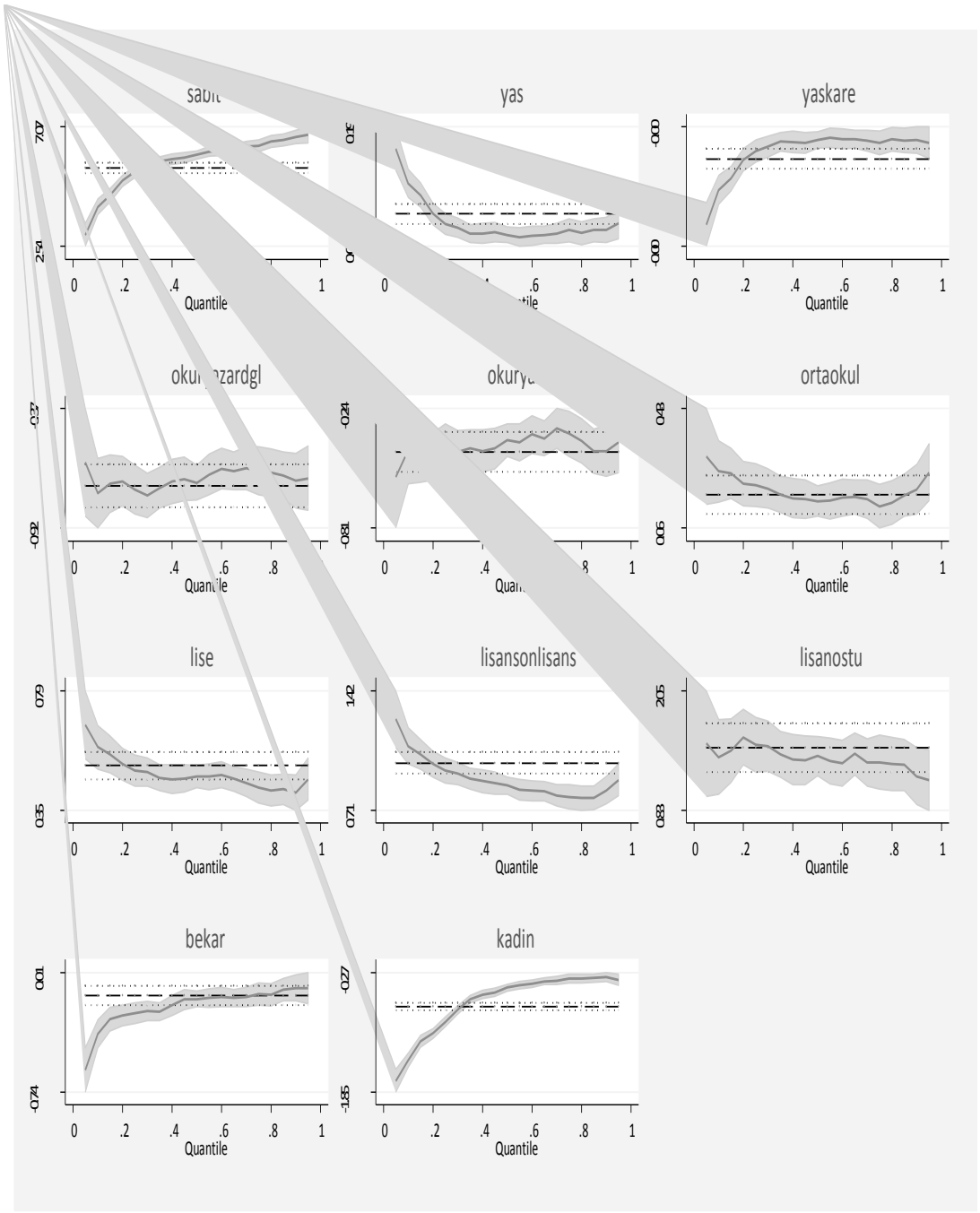
Tüm katsayı grafiklerinde doğru sıfır doğrusunun üstünde ya da altında olduğu için %5 anlam düzeyinde anlamlıdır. 2002 yılına ait tablo 8 de sol baştan ikinci grafik olan yaş değişkenin gelirin 0.5-0.85 arasındaki kantiller üzerine etkisi pozitif ve anlamlı iken 0.85'ten sonra bu etki negatif olmaya başlamıştır. Sabit terime ilişkin katsayı ise ilkokulu bitirmiş evli bir erkeğin gelirini göstermektedir ve katsayının gelir üzerine etkisi negatif olmaktadır. Sabit terim yani temel sınıf, deneyim (yaskare) ve kadınların gelir dağılımı üzerine etkisi negatif ve anlamlıdır.

Bu da bu üç katsayının gelir dağılımına etkisi sola çarpıktır. Yas, bekar ve diğer eğitim düzeylerindeki hanehalkı bireylerinin gelir dağılımı üzerindeki etkisi pozitif yani sağa çarpık ve anlamlıdır.

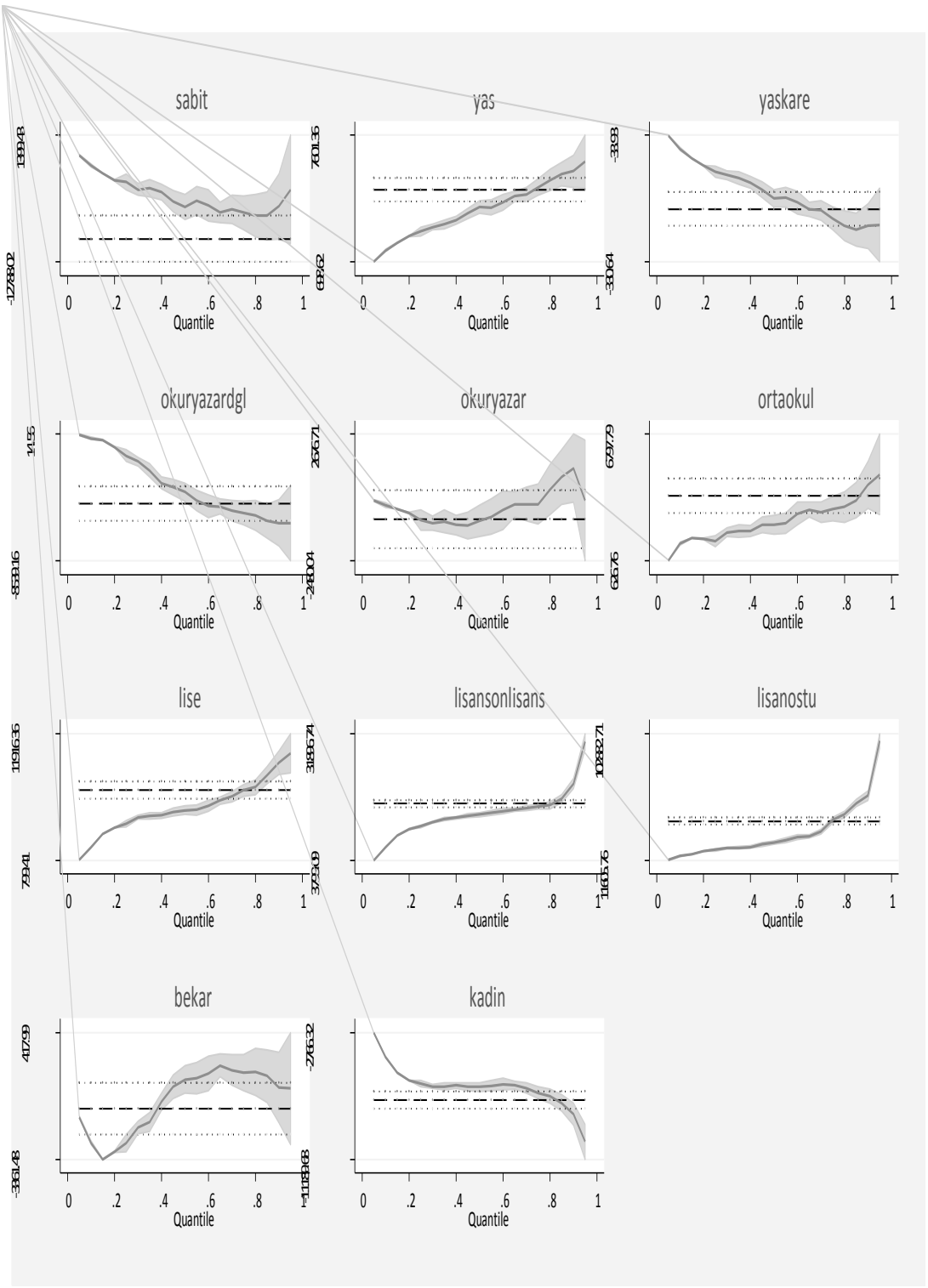
Tablo 9: 2002 Yılı Bağımsız Değişkenlerin Katsayı Grafikleri



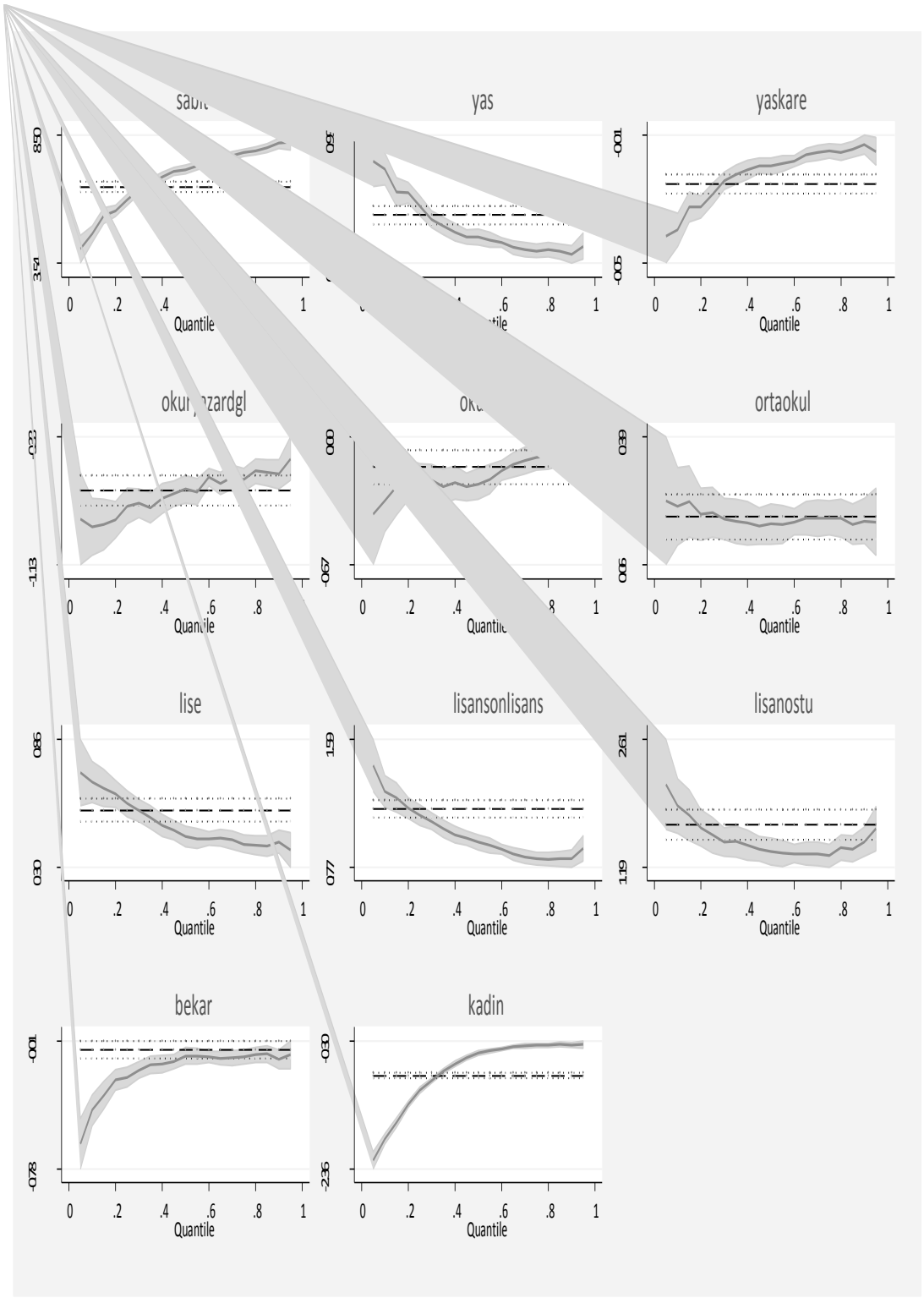
Tablo 10: 2002 Yılı Bağımsız Değişkenlerin Katsayı Grafikleri (Logaritmik)



Tablo 11: 2002 Yılı Bağımsız Değişkenlerin Katsayı Grafikleri



Tablo 12: 2002 Yılı Bağımsız Değişkenlerin Katsayı Grafikleri (Logaritmik)



3.8. KANTİL REGRESYON MODELLEMESİ

Kantil regresyon modellerinin en önemli avantajlarından biriside veri setini belli kantillere (19 kantile kadar) ayırmasıdır. Bu çalışmada veri setini temsil etmek amacıyla kantil değerleri sırasıyla 0.s10, 0.25 0.50, 0.75, 0.90 olarak ele alınmıştır. Ayrıca bu kantiller 2002 ve 2010 yılları itibari ile karşılaştırılmış ve bağımlı değişken olan hanehalkı bireylerinin fert gelirleri bir de logaritmik dönüşüm yapılarak da modeller yeniden tahminlenmiştir. Modeller STATA 12 programında sqreg komutuyla elde edilmiştir. Bu kod tahminleme yöntemi olarak doğrusal programlama ya da bootsrap yöntemi yerine simülasyon yöntemi ile elde edilmiştir. Tüm modelerde tahmin edilen değişkenler %5 anlam düzeyinde anlamlıdır.

Model yorumlaması kolay olması açısından 2002 ve 2010 yıllarının katsayı yorumları yapılmıştır. Tablo 14, 2002 yılı incelendiğinde yaşın gelir üzerindeki etkisi pozitif iken deneyimin etkisi negatif olmaktadır. Bekarların evlilere göre kadınlarında erkelere göre geliri daha az tahminlenmiştir. Eğitim durumu ise incelendiğinde ilkokulu bitirenlerin diğer eğitim düzeyleriyle kıyaslandığında okuryazar ve diplomasız okur yazarlara göre daha fazla olmasına karşın diğer eğitim düzeylerine göre daha düşüktür. Özellikle eğitim düzeyi artıkça gelirin yüzdesel olarak artığı görülmektedir.

Kantilere (p=0.10, p=0.25, p=0.50, p=0.75, p=0.90) bakıldığında katsayıların yönlerinde herhangi bir değişkenlik görülmezken, katsayı değerlerinde farkılıklar olmuştur. Örneğin daha alt kuyruklarda yani kantilin 0.10 olduğu düşük gelirli bireylerde eğitimin ilkokul mezunlarıyla kıyaslandığında düzeyi artıkça bireyin gelir yüzdesinin de artığı görülmektedir.Alt kuyruklu kantiller üst kuyruklu kantiller ile kıyaslandığında eğitimin gelir üzerinde belirginliği azalmaktadır. Bu da daha düşük gelirli bireylerin eğitimin gelirlerinde daha büyük bir etki yaptığını göstermektedir.

2010 veri setine göre benzer yorumları yapmak mümkündür. 2002 yılına göre katsayılar daha büyük oranda tahminlenmiştir.

Tablo 13: Kantiller Arası Karşılaştırma: 2002 Yılı

	Kantil 0.10		Kantil 0.25		Kantil 0.50		Kantil 0.75		Kantil 0.90	
	Katsayı	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S.Hata**
Bağımlı değişken: gelir										
Yaş	81.36986	18.94415	147.8289	10.3387	207.2639	13.38956	321.7226	16.87176	438.5316	43.88341
yaşkare	-.8219178	.194963	-1.381579	.1093994	-1.842425	.1570913	-2.80544	.2217157	-3.411058	.5839811
Eğitim										
Okur Yazar Değil	-106.8493	34.4741	774.079	75.52775	-1754.368	94.47154	-2405.05	108.7806	-3010.301	272.6818
Okur Yazar	-83.83562	36.51178	-369.5	63.54101	-673.1972	111.9601	-787.4588	144.8532	-550.7228	416.1781
İlk Okul	333.6986	60.51592	400.9474	53.7656	534.9552	49.31946	818.8027	71.04573	1099.652	176.1321
Ortaokul ve Dengi	480.137	80.48688	1178.579	47.35922	1514.385	90.9401	2390.478	121.6384	3040.426	243.1291
Lise ve Dengi	2247.945	120.4382	3395.632	141.9612	4224.805	231.2485	5223.734	299.3426	8846.724	884.7376
Lisans ve Önlisans	5495.479	315.8678	5914.079	603.7537	8822.435	2090.55	15844.96	3826.425	26013.19	5639.004
Medeni Durum										
Bekar	-982.1918	125.5974	-745.1579	75.11315	-339.9007	69.097	-358.6451	108.189	-461.7812	279.5091
Cinsiyet										
Kadın	-1407.945	37.38081	-1833.553	49.39489	-1880.827	58.96255	-2212.815	77.30831	-2610.071	217.5585
Sabit	-497.2603	439.4071	-1269.079	227.5427	-1527.482	268.9304	-2310.583	345.9174	-2738.633	868.2806

Temel sınıflar: Her iki modelde 2010 yılına göre ilk okul mezunu evli erkek (sadece çalışan bireyler modele dahil edilmiştir).

Tüm katsayılar %1, %5 ve % 10 anlamlılık düzeyinde anlamlıdır.

Tablo 14: Kantiller Arası Karşılaştırma : 2002 Yılı

	Kantil 0.10		**Kantil 0.25**		**Kantil 0.50**		**Kantil 0.75**		**Kantil 0.90**	
	Katsayı	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S.Hata**
Bağımlı değişken: log(gelir)										
Yaş	.1404589	.0104968	.1020687	.0051014	.0917736	.0042314	.096176	.0047777	.096534	.0052266
yaşkare	-.0013722	.0001217	-.0009604	.0000602	-.0008451	.0000492	-.0008773	.0000552	-.0008459	.0000562
Eğitim										
Okur Yazar Değil	-.7577195	.0947427	-.7409753	.0721019	-.7093859	.0564508	-.6558846	.0390885	-.6991879	.0353443
Okur Yazar	-.4361686	.1120763	-.4176688	.0773272	-.3859778	.0604319	-.3557804	.0496987	-.4385542	.0491759
İlk Okul	.23763	.0638829	.1945733	.0227594	.1461504	.0213775	.1294096	.0244798	.1825182	.0450957
Ortaokul ve Dengi	.5812373	.0380992	.4952592	.0252141	.4757754	.02352	.434376	.0293445	.4161867	.023988
Lise ve Dengi	1.091479	.0540253	.9445638	.030126	.8551431	.02279	.7875983	.0308446	.8265345	.057214
Lisans ve Önlisans	1.372539	.1814768	1.498836	.1606474	1.383392	.0772721	1.320719	.0914828	1.178058	.0848037
Medeni Durum										
Bekar	-.3762342	.0819193	-.2491028	.0400588	.0331023	-4.77	-.1263902	.0310218	-.0884846	.0542526
Cinsiyet										
Kadın	-1.436255	.1095858	-.9250854	.0697867	.0349274	-13.64	-.3538817	.0279623	-.3363276	.03181
Sabit	4.044256	.2230394	5.361776	.1035281	.0888799	67.52	6.336868	.1031553	6.691306	.1223414

Temel sınıflar: Her iki modelde 2010 yılına göre ilk okul mezunu evli erkek (sadece çalışan bireyler modele dahil edilmiştir).

Tüm katsayılar %1, %5 ve % 10 anlamlılık düzeyinde anlamlıdır.

Tablo 15: Kantiller Arası Karşılaştırma : 2010 Yılı

	Kantil 0.10		Kantil 0.25		Kantil 0.50		Kantil 0.75		Kantil 0.90	
	Katsayı	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S. Hata**	**Katsayı**	**S.Hata**
Bağımlı değişken: gelir										
Yaş	1328	204.4637	2355.385	204.8841	3678.066	167.5077	4771.081	283.1725	5638	560.9885
yaşkare	-66.4	10.43185	-119.7692	12.35972	-181.9702	10.16562	-228.5541	18.39767	-246.8	34.88421
Eğitim										
Okur Yazar Değil	-265.6	58.02962	-1420	138.0879	-3882.214	297.7261	-5297.595	312.003	-5969	653.876
Okur Yazar	-199.2	56.04796	-803.3847	146.6188	-837	305.8931	-162.1622	456.5772	1300.4	784.4941
İlk Okul	1477.6	227.7773	1596.615	253.85	2393.643	252.8218	3172	372.2807	4334.6	779.1178
Ortaokul ve Dengi	1952	253.7023	4002.769	184.2414	5193.369	224.5399	6928.311	255.1914	9352.8	608.1304
Lise ve Dengi	6614.4	396.6297	11307.85	261.4212	13908.64	250.0208	15666.26	286.1926	20771	1453.327
Lisans ve Önlisans	14534.4	1074.753	19076.54	1063.271	24333.64	2450.192	40466.11	7120.804	58230	17236.81
Medeni Durum										
Bekar	-2869.6	299.2165	-2881.231	252.9089	-978.0356	207.8559	-778.4459	331.3213	-1225.4	501.483
Cinsiyet										
Kadın	-4380	119.2459	-6176.539	165.8243	-6392	176.9596	-6770.96	230.9374	-8190	407.5356
Sabit	-1994.4	932.6439	-3820.615	807.3508	-6612.072	693.9512	-7286.852	1035.735	-6577.8	2202.665

Temel sınıflar: Her iki modelde 2010 yılına göre ilk okul mezunu evli erkek (sadece çalışan bireyler modele dahil edilmiştir).

Tüm katsayılar %1, %5 ve % 10 anlamlılık düzeyinde anlamlıdır.

Tablo 16: Kantiller Arası Karşılaştırma: 2010 Yılı

	Kantil 0.10		**Kantil 0.25**		**Kantil 0.50**		**Kantil 0.75**		**Kantil 0.90**	
	Katsayı	**S.Hata**	**Katsayı**	**S.Hata**	**Katsayı**	**S.Hata**	**Katsayı**	**S.Hata**	**Katsayı**	**S.Hata**
Bağımlı değişken: log(gelir)										
Yaş	.816879	.0545496	.6242647	.0410874	.4618984	.0215366	.3888052	.0157606	.3705947	.0269328
Yaskare	-.0419486	.0032556	-.0314472	.0024041	-.023274	.0012682	-.0187845	.0009551	-.0169546	.0016801
Eğitim										
Okur Yazar Olmayanlar	-.8896363	.1740865	-.7605461	.072976	-.6561642	.0487733	-.5975996	.0620131	-.5625839	.0312801
Okur yazar	-.3324634	.1418755	-.2432658	.065926	-.2459203	.0482669	-.1062222	.0397503	-.0862128	.0528667
Ortaokul	.2123098	.0745777	.1978341	.0276067	.1679204	.0221122	.1832476	.0208541	.1754449	.0264874
Lise (mesleki ve teknik lise)	.6743271	.0509131	.5808258	.0219993	.4378968	.0266545	.4041331	.0239555	.4133416	.0323252
Yüksek Öğretim	1.255517	.0338139	1.10248	.0227219	.929938	.021149	.8280029	.0256923	.8271767	.0389891
Lisans Üstü	1.868166	.0647278	1.546946	.0487866	1.36527	.0666775	1.320746	.0783348	1.473003	.2088647
Bekar	-.4307433	.0914145	-.2317358	.0481305	-.1037696	.0297057	-.1101508	.0161433	-.1241708	.0200394
Kadın	-1.873931	.0875302	-1.094745	.0442714	-.4984355	.0326045	-.371106	.0208422	-.3656326	.0282445
Sabit	4.667427	.2298591	5.984491	.1785955	7.171453	.0924615	.0621807	.0621807	8.186789	.1007521

Temel sınıflar: Her iki modelde 2010 yılına göre ilk okul mezunu evli erkek (sadece çalışan bireyler modele dahil edilmiştir).

Tüm katsayılar %1, %5 ve % 10 anlamlılık düzeyinde anlamlıdır.

SONUÇ

Doğrusal regresyon modellerinde temel amaç hataların toplamının minimizasyonunu sağlamaktır. En çok tercih edilen yöntem olarak En Küçük Kareler (EKK) Yöntemi kullanılmaktadır. Veri setini yapısına bağlı olarak aşırı değer ve sapan gözlemlerin olması veya hataların varsayımına uyulmaması durumun EKK yöntemi ile elde edilen parametrelere güvenilemeyecek ve aynı zamanda istatistikî testler için güvenirlik sorunu yaşanacaktır. Buna bağlı olarak EKK yöntemi yerine alternatif regresyon yöntemlerine başvurmak gerekecektir.

Alternatif yöntem olarak bahsedilen robust yöntemler veri setinde aşırı gözlem olması durumunda kullanılması avantajlı olacaktır. En Küçük Mutlak Sapma (LAD) hataların karelerinin toplamı yerine hataların mutlak toplamını minimize etmektedir. Kantil Regresyon (QR) modelinin en belirgin özelliği diğer modellerin aksine ortalamaya dayalı tahmin yapmak yerine veri setini 19 kantile kadar bölerek daha geniş ve kapsayıcı tahminler yapabilmektedir. Bu yöntemin diğer bir avantajı da normalliğin sağlanmaması durumunda da etkin tahminciler elde edebilmesidir. Normal dağılımın sağlandığı durumlarda EKK ile 0.5'inci kantil benzer sonuçlar verecektir.

2002 ve 2010 yılları kıyaslandığında ortamla gelirde 2002 de yıllık ortalama gelir 4600 lira iken 2010 yıllında ortalama fert geliri 12450 düzeyine çıkmıştır. Bu da kriz sonrası Türkiye'deki istikrarlı ekonomik politikaların incelenen hanelerin gelirini yaklaşık artırdığını göstermektedir. On yıllık süre içerisinde ortlama yaşın 35'ten 38'e çıkmıştır. Eğitimdeki artış daha çok yüksek eğitim düzeyinde olmuştur.

QR ve EKK sonuçları incelendiğinde Hanehalkı bireylerinin gelirini yas ve aldıkları eğitim pozitif etkilerken, deneyim olarak açıklanan yaşın karesi değikeni ile bireyin kadın ve bekar olması ise gelirlerin negatif yönde etkilemiştir.

2002 ve 2010 yıları kıyaslamasında dikkat çekici bir unsurda 2002 yıllında ilkokulu bitirenlerle diğerleri arasındaki gelir farkı iki katından fazla olmasına rağmen bu 2010 yılında 0.5 katına düşmüştür.

Kantil regresyon modellerini alternatif regresyon modellerinden ayıran önemli bir özellikte bağımsız değişkenlere ait katsayı grafikleridir. 2002 yılına ait yaş değişkenin grafiği incelendiğinde gelirin 0.5 ile 0.85'inci kantiler üzerene etki

pozitif iken bundan sonra gelirin etkisi negatif olmaya başlamıştır. Kantiler itibariyle bakıldığında katsayıların yönlerinde herhangi bir değişkenlik görülmezken, katsayı değerlerinde farkılıklar olmuştur. Örneğin daha alt kuyruklarda yani kantilin 0.10 olduğu düşük gelirli bireylerde eğitimin ilkokul mezunlarıyla kıyaslandığında eğitim düzeyi artıkça bireyin gelir yüzdesinin de artığı görülmektedir.Alt kuyruklu kantiller üst kuyruklu kantiller ile kıyaslandığında eğitimin gelir üzerinkde belirginliği azalmaktadır. Bu da daha düşük gelirli bireylerin eğitimin gelirlerinde daha büyük bir etki yaptığını göstermektedir.

Sonuç olarak 2002 ve 2010 yılları kıyaslandığında ortalama gelirde bir artış olduğu ve eğitim düzeyinin daha düşük gelir gruplarında daha fazla etkin olduğu görülmektedir.

KAYNAKÇA

Akdeniz, Fikri. **Olasılık ve İstatistik**, Baki Kitapevi, Adana, 2002.

Altındağ, İrem. **Kantil Regresyon ve Bir Uygulama**. (Yayınlanmış Yüksek Lisans Tezi) Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya, 2010.

Andrew, R. Gray ve Stephen G. MacDonell. "A Comparison of Alternatives to regression Analysis Model Building Techniques to Develop Predictive Equations for Software Metrics"**, The Information and Software Technology**, Cilt:39, 2005, ss. 425-437.

Bidabad, Bijan. "L_1 Norm Computational Algorithms", http://www.bidabad.com/doc/l1-article6.pdf (18.12.2012).

Birkes, David. ve Yadolah, Dodge. **Alternative Methods of Regression**, John Wiley & Sons,Inc., Canada, 1993.

Bloomfield, Peter ve William L. Steiger. **Least Absolute Deviations Theory, Applications, and Algorithms**, Boston, 1983.

Briand, Lionel C., R. Basili Victor ve M. Thomas William. "A Pattern Recognition Approach for Software Engineering Data Analysis"**, İEEE Transactions on Software Engineering**, Cilt:18, 1992, ss. 931-932.

Cade, Brian S. ve Jon D. Richards. "Permutation Test For Least Absolute Deviation Regression", **International Biometric Society**, Cilt52, 1996, ss.886-902.

Davidson, Russel ve James Gordon. Mackinnon. **Econometric Theory and Methods**, Oxford University Press, 2004.

Draper, Norman, R. ve Harry Smith. **Applied Regression Analysis**, John & Sons, Canada, 1998.

Faraway, J.Julian. **Extending the Linear Model With R: Generalized Linear, Mixed Effects and Nonparametric Regression Models**.: Chapman & Hall., Boca Raton CRC, 2005.

Faraway, J. Julian. **Linear Models with R**, Chapman&Hall / Crc, USA, 2005.

Ferguson, Thomas Shelburne. **Matematical Statistics: A Desicion Theoretic Approach**, Academic Press, New York, 1967.

Gilchrist, Warren. **Statistical Modelling with Kantile Functions**, Cherman & Hall/Crc Press, London, 2000.

Gürler Kiren, Özlem ve Şenay Üçdoğruk. "Türkiye'de Cinsiyete Göre Gelir Farklılığının Ayrıştırma Yöntemiyle Uygulanması"**, Journal of Yasar University**, Cilt:12, ss.571-589.

Haifeng, Chen ve Peter Meer. "Robust Resgression With Projection Based M-Estimators", **Proceedings of the Ninth IEEE international Conference on Computer**, 2003, s.1.

Hao, Lingxin ve Daniel Q. Naiman. **Quantile Regression**, Sage Publications, 2007.

Hawley, V. Robert ve Neal, C. Gallagher. On Edgeworth's Method for Minimum Absolute Error Linear Regression, **İEEE Transaction on signal processing**, Cilt:42, 1994, ss. 2045-2054.

Johnston, Jack ve John, Dinardo. **Econometric Methods**, Mcgraw Hill, Madison, 1997.

Karagöz, Murat. **İstatistik Yöntemleri**, Ekin Yayın Dağıtım, Bursa, 2009.

Koenker, Roger. **Quantile Regression**, Cambridge University Press, New York, 2005.

Koenker, Roger ve Jose, A.F. Machado. "Goodness of Fit and Related inference Processes for Quantile Regression", **Journal of the American Statistical Association,** Cilt:94, 1999, ss. 1296-1310

Koenker, Roger. **Quantile Regression**, Cambirdge University Pres, New York, 2005.

Kuenker, Roger ve Vasco, D'Orey. **Computing Regression Quantiles**, Applied Statistics, Cilt.46, 1987, ss. 383-393.

Latif, Öztürk. **Doğrusal Regresyonda Sağlam kestirim Yöntemleri ve Karşılaştırılmaları**, (Yayınlamış Doktora Tezi), Mimar Sinan Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 2003.

Moshe, Buchinsky. **The Theory and Practice of Kantile Regression. Pablished Doctoral Dissertation**. Cambridge, Massachusetts: Graduate Faculty of Harvar University. 1991.

Parzen, Emanuel. "Nonparametric Statistical Data Modeling"**, Journal of the Ametican Statistical Association**, Cilt:74, 1979, ss. 105-121.

Parzen, Emanuel. "Kantile Probability and Statistical Data Modeling" **Statistical Science**, Cilt:19, 2004, ss. 652-662.

Ramanathan, Ramu. **Introductory Econometrics With Applications**, The Dryden Press, San Diego, 1998.

Ronchett, Evezio. **Statistical Data Analysis Based on The L Norm and Realted Methods,** Bounded Influence in Regression, North Holland, 1991.

Rousseeuw Peter J.. “Least Median of Squares Regression”, **Journal of The American Statistical Association**, Cilt:79, 1984, ss.871-880.

Saraçoğlu, Bedriye ve Ferhan Çevik. **Matematiksel İstatistik Olasılık ve Önemli Dağılımlar**, Gazi Büro Kitabevi, Ankara, 1995.

Sheskin, David J.. **Handbook of Parametric and Nonparametric Statistical Procedeures,** CRC Pres Company, Washington D.C., 2004.

Steven, P. Ellis. “ Instability of least Squares, Least Absolute Deviation and Least Median of Squares Linear Regression”, **Statistical Science**, Cilt.13, 1998, ss. 337-344.

Şanlı, Kamile. “Yöneylem Araştırması Derneği”, **YA/EM Doktora Öğrencileri Kolokyumu**, http://www.yad.org.tr/oturum2.pdf, (16.07.2011).