



**EN KÜÇÜK KARELER REGRESYONUNA ALTERNATİF BİR YÖNTEM:  
KANTİL REGRESYON**

**Utku Kubilay ÇINAR\***

**ÖZ**

Kantil Regresyon yöntemi Koenker ve Bassett tarafından önerilmiş bir regresyon modelidir. Kantil Regresyon yöntemi, dayanıklı (robust) bir regresyon şeklidir. Basit regresyon modelleri ya da En Küçük Kareler yöntemi uç değerlere karşı esnek değildir. Kantil Regresyon yöntemi, diğer regresyon yöntemlerine göre uç değerlere karşı daha esnektir bu sebeple Kantil Regresyon yöntemi çok çeşitli uygulama alanlarında kullanılmaktadır. En Küçük Kareler yönteminde hataların varyansı konusunda bazı varsayımlar vardır. En Küçük Kareler (EKK) yöntemi, sapan değerlere (uç değerlere) karşı esnek olmadığından, yapılan tahmini etkilememektedir. Kantil Regresyon yöntemi, En Küçük Kareler yönteminin bu özelliğinden dolayı alternatif olarak geliştirilmiştir. Daha esnek bir yaklaşım olduğundan bazı varsayımlara gerek duymaz. Kantil Regresyon, doğrusal modelde bağımlı değişkenin dağılımının koşullu miktarlarını tahmin etmenin bir yoludur. Kantil Regresyonlar, veri kümelerinin koşullu dağılımındaki değişiklikleri görselleştirmek için kullanışlıdır. Kantil Regresyon yöntemi, özellikle uç değerlerin olduğu durumlarda kullanışlı bir yöntemdir. Bu çalışmanın amacı, Kantil Regresyon yönteminin etkililiğini tekrar test etmektir. Bu amaçla, R programı üzerinden yapay olarak veri üretilmiştir. Veri setinde uç değerler içeren bağımlı ve bağımsız değişkenler bulunmaktadır. Normal dağılım sergileyen veri seti üretildikten sonra uç değerler eklenip, toplam veri sayısı 400'de sınırlandırıldı. Yapay veri üretilmesinin sebebi ise teorik çalışmalarda, yapay veri üretmenin teoriyi daha iyi açıklayacağını düşünülmektedir. Normal dağılıma uygun veride, uç değerler eklenerek EKK yöntemi ve Kantil Regresyon yöntemi kıyaslanmıştır. Yapılan simülasyon çalışması sonunda uç değerler içeren veride, Kök Ortalama Hata Kare kriteri değerine göre Kantil Regresyon yönteminin performansı EKK yöntemine göre daha başarılı olduğu bulunmuştur.

**Anahtar Kelimeler:** Kantil Regresyon, en küçük kareler yöntemi, uç değerler

**A ALTERNATIVE TO ORDINARY LEAST SQUARES REGRESSION: QUANTILE  
REGRESSION**

**ABSTRACT**

Quantile Regression is a regression method which suggested by Koenker and Bassett. Quantile regression is a form of robust (outlier resistant) regression. Basic regression models or Least Square methods are not flexible to outlier values. Basic regression methods or estimators are affected by extreme values and Quantile Regression is less sensitive to extreme values than the other regression model. Therefore, Quantile Regression uses in a range of application settings. Least Squares methods has some assumptions about variance of errors. Quantile Regression approach is advanced like an alternative to the Least Squares method used for regression analysis assumptions. Quantile regression is a way to estimate the conditional quantiles of a dependent variables distribution in the linear model. Quantile Regressions are effective for visualizing changes in the conditional distribution of data sets. Quantile Regression

---

\*Yıldız Teknik Üniversitesi, Sosyal Bilimler Enstitüsü, Y.L.Öğrencisi,  
utkukubilaycinar@gmail.com, Orcid ID: 0000-0003-1450-427X

approach useful in such cases where especially the outlier values. The purpose of this study is to retest the effectiveness of the Quantile Regression method. For this purpose, synthetic data generated with the R program. There are dependent and independent parameters in dataset which contains extreme values. After synthetic dataset which exhibits normal distribution data has been generated, number of data was limited to 400. The reason why synthetic data is generated that it is thought that generating synthetic data will explain the theory better in theoretical studies. In the data exhibiting normal distribution, after adding the extreme values, Least Squares method and Quantile Regression method were compared. In the end of the simulation practice, according to RMSE criteria value, it is found that the performance of the Quantile Regression model is better than Least Squares model in the case of extreme values.

**Keywords:** Quantile Regression, least squares method, outlier values

### 1. GİRİŞ

Regresyon analizi, sık kullanılan istatistiksel yöntemlerin başında gelir. Regresyon analizinde çıkarsama amaçlandığı durumlarda, genellikle En Küçük Kareler (EKK) yöntemi kullanılır. EKK yöntemi mevcut varsayımlarından dolayı uç değerlere karşı duyarlıdır. Uç değerlerin (aykırı değer) varlığında, EKK'nın özelliklerinden olan etkinlik ve etkililik özelliklerini yitirebilir. Regresyon analizlerinde farklı modellerden biri ise Kantil Regresyon modelleridir. Kantil Regresyon, uç değerlere karşı daha dirençlidir. Kantil Regresyon yönteminin sapan değerlere karşı dirençli olmasının sebebi ise, bağımlı değişkenin dağılımını çeyreklere ya da parçalara bölebilmesindedir. "Kantil Regresyon yöntemi, bağımlı değişkenin dağılımına ait kovaryans etkilerini incelemeye olanak sağlamaktadır" (Çağlayan ve Arıkan, 2011).

"Deneysel çalışmalarda bağımlı değişkenin sadece ortalamasının değil, diğer kantil değerlerinin tahmininde de büyük bir öneme sahip olduğu düşünülürse, büyük maddi imkanlarla yürütülen mühendislik çalışmalarının değerlendirilmesinde Kantil Regresyonun kullanılması oldukça faydalı olacaktır" (Yavuz ve Aşık, 2017).

EKK yöntemi ile kurulan modeller, aşırı değerlere karşı hassas olduğundan ve mühendislik çalışmalarında gözlemlerin büyük maddi kaynaklarla edinildiğinden, Kantil Regresyon modellerinin bu alanlarda kullanılması yararlı olacaktır. Uç değerler kurulan modelin hatasını yükseltecektir. Kantil Regresyon ile bu uç değerlerden kaçınılabılır.

Kantil Regresyon yönteminde fonksiyonel durumun yani fonksiyonel ilişkilerin veri setinin, dağılımı belirlediği bir yaklaşım olması sebebiyle, aykırı değerlere karşı esnektir. Bu sebeple, yanlış fonksiyonel ilişkilerin oluşturduğu hatayı önler. Araştırmacı uç değerlerin durumuna göre modelini değiştirebilir (0.05; 0.25; 0.50; 0.75; 0.95 gibi kantiller için farklı modeller kurularak sonuç modeli oluşturulabilir). Kurulan her model için değişkenlerin anlamlılığı incelendikten sonra başarı ölçütüne göre sonuç modeli oluşturulur. Bu esneklik araştırmacıya uç değerlerden kaçınma imkânı sağlar. Farklı kantil değerleri ile kurulan modellerde, kantil değeri değişeceğinden modelin fonksiyonel ilişkisi de değişmiş olur. En küçük mutlak sapma tahmin edicileri, "Kantil Regresyonda kantil değerinin  $\frac{1}{2}$  olduğunda yani  $t=0.5$  olduğu durumunda, En Küçük Mutlak Sapma (LAD) analizi elde edilir" (Koenker ve Basett 1978).

Literatürde geleneksel regresyon modellerindeki bağımlı değişkenin, koşullu beklenen değeri açıklamayı amaçlarken, alternatif olarak kullanılan Kantil Regresyon ise koşullu kantillerdeki değişimi hedefler ve açıklar. Bağımlı değişkeni tahminleyecek bağımsız değişkenleri işlevsel olarak modeller.

Kantil Regresyon, klasik regresyon analizinin varsayımlarından olan hata terimlerine ilişkin normal dağılıma uygunluk varsayımını ihmal eden robust (dayanıklı) bir regresyon tekniği olarak ortaya çıkmıştır. “Her bir kantil değeri için hatayı minimum yapan parametre tahminlerini belirlemeye çalışır. Herhangi bir varsayıma sahip olmadığı için esnek bir yöntemdir” (Türkış, 2015). “Kantil Regresyon yöntemleri, bağımlı değişkenin koşullu dağılımının her noktası üzerindeki etkilerini dikkate almakta; dolayısıyla dağılımın asimetrik kuyruk özelliğini de analize dahil etmektedir” (Abadie, Angrist ve Imbens, 2002).

Kantil Regresyon modellerinde kantillere göre regresyon katsayıları hesaplanır. Kantil değerleri, modelin fonksiyonel yapısını değiştirmektedir. Uç değerlerin bulunduğu durumda, homojen olmayan veride daha iyi tahmin değerleri üretmektedir. Kantil Regresyon modellerinde, normal dağılım tablosundaki değerler üzerinden hesaplamalar yapılır. Normallik varsayımı sağlanamadığı durumlarda robust yöntemler kullanılır.

“Doğrusal Regresyonun ve EKK’nın varsayımları sebebiyle araştırmacılar çalışmalarında, uç değerlerin olduğu durumda Kantil Regresyon modelleri kullanmaları faydalıdır. Kantil Regresyon yönteminin kullanılmasının bazı avantajları vardır. Bunlar;

- Yanıtın her kantil değerinde, tahminlerin ve yanıtların ilişkisi tam olarak gösterilebilir.
- Klasik regresyon yöntemine göre varsayımları esnektir.
- Kantil Regresyon tahmin edicileri, asimptotik özelliklere sahiptir” (Wang ve Cook, 2017).

Chen ve Wei’den (2005), Yavuz ve Aşık’ın (2017) aktardığına göre, “Kantil Regresyon, özellikle koşullu kantillerin değişkenlik gösterdiği durumlarda kullanışlıdır. Kantillere bağlı olarak regresyon katsayıları belirlenir. Kantil Regresyon, ilk olarak regresyondaki klasik varsayımlardan hata terimlerinin normal dağılması varsayımını ihmal eden sağlam bir regresyon yöntemi olarak ortaya çıkmıştır”.

“Klasik EKK yöntemi, koşullu ortalamaya bağlı iken; Kantil Regresyon yöntemi ise koşullu kantil fonksiyonuna bağlıdır” (Koenker, 2005). Klasik regresyon yöntemlerinin aksine, hataların dağılımı ve hataların varyanslarının homojenliği varsayımı olmadığından daha esnektir. Medyan Regresyon, belirli kantiller ( $t=0.5$  kantil değeri) için genişletilmiş Kantil Regresyon durumudur. “Bu regresyon modelleri uç değerlere ve eğikliğe EKK yöntemlerinden daha az hassastır” (Saçaklı, 2005). Bu sebeplerle uç değerlerin varlığı durumunda, kurulan modelin kantil değerleri değiştirilerek uç değerlerden kaçınılır.

“F dağılım fonksiyonuna sahip Y bağımlı değişkenin, rastgele örnekleme; t x k boyutlu tasarım matrisi; b, tahmin edilecek katsayı vektörü ve  $e_t = y_i - x_i\beta$  hata değeri; t, regresyon kantil değeridir. Bunlara göre kurulan denklem;

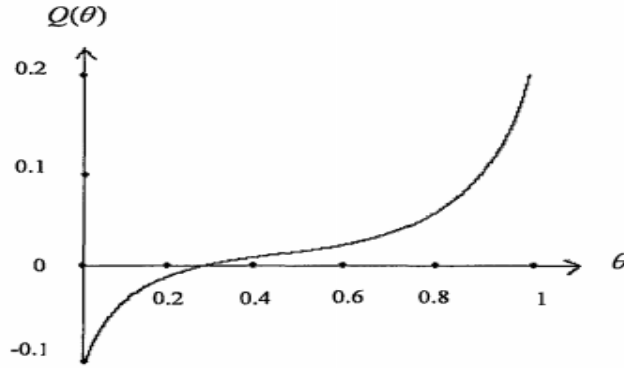
$$(0 \leq t \leq 1);$$

$$\min_{\beta \in \mathbb{R}^k} \left[ \sum_{t \in \{t; y_i \geq x_i b\}} \tau |y_i - x_i b| + \sum_{t \in \{t; y_i < x_i b\}} (1-\tau) |y_i - x_i b| \right] \quad (1.1)$$

denklemini minimum yapacak şekilde elde edilir” (Koenker ve Bassett, 1978).

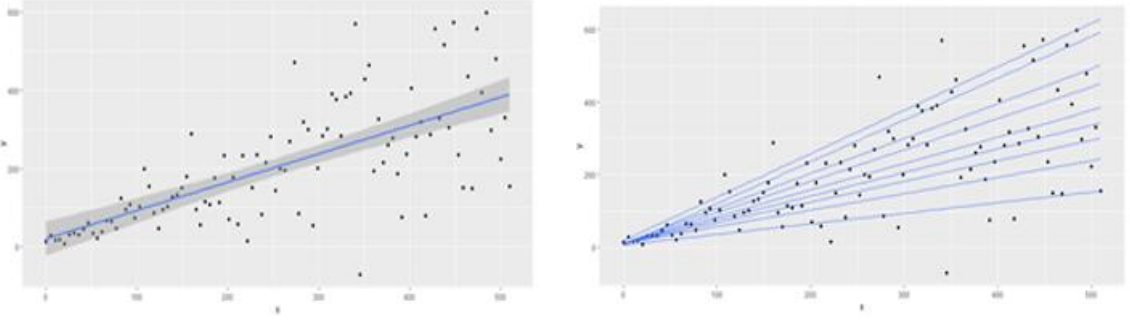
“Serinin mod değerinin olasılığı  $p\text{-mod} \geq 0.5$  ise, dağılım sola çarpıktır ve  $q(\theta)$  kantil yoğunluk fonksiyonu  $q(\theta) \leq q(1-\theta)$  durumunu sağlar,  $0 \leq \theta \leq 0,5$ 'tir. Kantil fonksiyonu da  $Q(\theta) \leq Q(1-\theta) \leq 2\theta(0,5)$  durumunu sağlar ve ortalama  $\leq$  medyan  $\leq$  mod sıralaması sağlanır” (Saçaklı, 2005).

Şekil 1.1. Kantil fonksiyonu



Kantil fonksiyonları, tüm olasılıklar için kantil değeri 0 ile 1 arasında olur. Medyan olarak ifade edilen değer, Kantil Regresyonda  $Q(0,5)$  yani  $t=0.5$  değeridir. Kantil değerlerinin kullanılabilmesi için normal dağılım tablosundan faydalanılır. “Hesaplamalar için normal dağılım tablosundan yararlanılır, örneğin 1.96 gibi bir değer 0.975 olasılığını alır.  $N(Q)$  standart normal dağılım için kantil fonksiyonuysa  $N(0.975)$  değeri 1.96’ ya eşittir” (Saçaklı, 2005). Şekil1.2. üzerinden Kantil Regresyon ile doğrusal regresyon kıyaslanmıştır.

Şekil 1.2. EKK model doğrusu ve Kantil model doğruları



Görsellerde görüldüğü üzere, uç değerler içeren veri setlerinde Kantil Regresyon uygulanarak regresyon doğrusunu, hatası en düşük olacak şekilde araştırmacı tarafından ayarlanabilir. Kantil model doğrusu şeklinde, yukarıdan aşağıya doğru sırasıyla 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 ve 0.9 kantil değerleri alınarak model doğrusu kurulur.

Kantil Regresyon uygulayan araştırmacı, uç değerlerin dağılımına göre kantil değerini belirler. Regresyon modelinin hatasını en düşük yapacak kantil değerini sonuç modelinde kullanır. Her bir kantil değeri için ayrı ayrı model oluşturulur. Araştırmacı, oluşturulan bu modeller arasından minimum hatalı olan Kantil Regresyon modelini seçer.

Değişen varyans (heteroscedasticity) sorunu, oluşturulan modelde önemli bir değişkenin modelde bulunmadığı durumlarda oluşabilir. Bununla beraber uç değerlerin bulunduğu durumlarda da değişen varyans problemi oluşabilir. Değişen varyans problemi ile t ve F testleri anlamsızlaşır. Tahminciler ise en düşük varyans değerleri olma özelliğini kaybeder. Böylelikle varyanslar büyür ve hatalar artar. Özellikle ekonomik verilerde (makro-mikro göstergeler, ekonometrik veriler, finansal tablolar, zaman serileri) uç değerlerin olması durumunda, sabit varyans ve hata terimlerinin normal dağılması gibi varsayımlar sağlanamayabilir. Bu durumda ise EKK yöntemi kullanılamayacaktır. Bu sebeple Kantil Regresyon yöntemi literatürde önerilmektedir.

“Modelde değişen varyanslılık ve/veya otokorelasyon olması halinde parametre tahminleri EKK yöntemi yapılırsa yansız, tutarlı ancak etkin olmayan parametre tahminleri elde edilir ki bu da elde edilen parametre tahminlerinin en iyi doğrusal yansız tahmin edici olma özelliğini ortadan kaldırır” (Ocak, 2010).

Bu sebeple Kantil Regresyon yöntemi, uç değerlere karşı esnek olduğundan değişen varyans probleminden ya da hata terimlerinin normal dağıldığı varsayımlarından etkilenmemektedir.

### 1.1 Kantil Regresyonun Özellikleri

Kantil Regresyon yöntemlerinin özelliklerinden dolayı bazı durumlarda kullanılması önerilmektedir. Bu özelliklerin bazıları;

- Kantiller, bağımlı değişkendeki sapan değere karşı daha esnektir (Robust özelliği).
- “Hata terimi normal dağılmadığında, Kantil Regresyon tahmincileri EKK tahmincilerinden çok daha etkin olabilir” (Ocak, 2010).

- Değişen varyansın belirlenmesine olanak sağlar.
- “Farklı kantillerde farklı sonuçların ortaya çıkması, bağımlı değişkenin koşullu dağılımının farklı noktalarındaki açıklayıcı değişkenlerdeki değişikliklere farklı tepki vermesi olarak yorumlanabilir” (Ocak, 2010).
- “Kantil Regresyon analizinde, kantil değerinin 0.5 olması durumunda En Küçük Mutlak Sapma (LAD) regresyon elde edilmektedir” (Altındağ, 2010).
- “Kantil Regresyon farklı kantil değerleri için  $y$ 'nin  $x$ ' e göre koşullu dağılımının tümü hakkında bilgi vermektedir” (Kurtoğlu, 2011).
- “EKK regresyon modelinde hata terimlerinin varyansları homojen varsayılırken, Kantil Regresyon modelinde ise hata terimlerinin değişkenliğine izin verilmekte ve varyans yapısına ilişkin herhangi bir varsayım bulunmamaktadır” (Yavuz, Aşık, 2017).
- “Kantil Regresyon amaç fonksiyonu için tahmin edilen katsayı vektörü bağımlı değişkendeki aşırı değerlere duyarlı değildir ve yerleşimin robust bir ölçüsüdür” (Saçaklı, 2005).

## 2. LİTERATÜR İNCELEMESİ

Literatür incelendiğinde EKK yöntemi tahmin ediciler ile Kantil Regresyon tahmin ediciler arasındaki kıyaslamalar dikkat çekmiştir. Taranan bilimsel çalışmalarda, veri setinde uç değerler (sapan değerler/outliers) bulunması halinde, Kantil Regresyonun daha başarılı sonuçlar verdiği belirtilmiştir. Kantil Regresyon analizi konusunda ilk araştırma Koenker ve Basett (1978) tarafından yapılmıştır. Kantil Regresyonda amaç, tahmin edicilerin mutlak hatalar toplamlarını minimum etmektir.

Keming Yu, Zudi Lu, Julian Stander (2003), Kantil Regresyon yönteminin EKK yönteminden daha detaylı bir istatistiksel model olduğunu belirtmiştir. Kantil Regresyon modeli, geleneksel regresyondan daha eksiksiz bir model sunmaktadır.

İrem Saçaklı (2005), değişkenler arasındaki ilişkilerin daha iyi açıklanabilmesi için alternatif regresyon modelleri kıyaslanmıştır. OECD ülkelerinin büyüme rakamları üzerine yapılan çalışmada, bazı ülkelerin uç değer olduğu belirtilmiştir. Uç değer içeren veriye farklı regresyon modelleri uygulanmıştır. Klasik tahmin modelleri, dağılımın normal olmasına dayandığı belirtilmiş ve robust yöntemler, klasik yöntemlere göre daha başarılı olduğu vurgulanmıştır. Uygulamada karşılaştırılan modeller arasında robust ve Kantil Regresyon modelleri daha iyi sonuç verdiği belirtilmiştir. Uygulama sonucunda büyüme hızının farklı olduğu ülkelerde, büyüme hızlarına etki eden değişkenlerinin farklı olduğu bulunmuştur.

Can Türkü (2015), sağkalım analizlerinde Kantil Regresyon ile Parametrik Regresyon yöntemlerinin performansları kıyaslanmıştır. Meme kanseri verisi üzerine yapılan simülasyon çalışmasında, heterojen yapıda üretilmiş veri setleri ve gerçek veri setleri üzerinde hem Parametrik Regresyon modellerinin hem de Kantil Regresyon modelinin sağkalım analizindeki performansları karşılaştırılmıştır. Çalışma sonucunda, farklı örneklem büyüklüklerinde Kantil Regresyon modeli hatalarının minimum olduğu belirtilmiştir. Homojen ve heterojen yapıdaki meme kanseri veri setinde, Kantil

Regresyon yönteminin performansının Parametrik Regresyon yöntemlerinden daha iyi olduğu belirtilmiştir.

Arzu Altın Yavuz, Ebru Gündoğan Aşık (2017), mühendislik alanında yapılan çalışmada, geleneksel regresyon analizlerine alternatif olarak uygulanan ve farklı bir yaklaşım olan Kantil Regresyon yöntemi kullanılmıştır. Çalışma verisi olarak beton kırma deneyine ilişkin değerlere uygulanmıştır. Kurulan farklı modellerin bağımlı değişkeni açıklama oranı EKK yöntemi ile kurulan modelden az olmadığı belirtilmiştir. Ayrıca varsayımsal bozulmalara karşı da dayanıklı olduğundan çıkarsama amaçlı kullanılabileceği vurgulanmıştır.

Atilla Ahmet Uğur, Demet Özocaklı (2018), gıda açığı bulunan ülkeler arasında gıda güvencesizliğini farklı düzeyleri ile bağımsız değişkenler arasındaki ilişkiyi, Kantil Regresyon yöntemi ile incelenmiştir. Ayrıca oluşturulan Kantil Regresyon yöntemi ile gaussian sabit etkisi panel tahmincileri ile de karşılaştırılmıştır. Çalışmada, gıda güvencesizliğinin kantilleri ve kantil tahminleri ile gaussian sabit etki panel tahmincileri arasında önemli farkların gözlemlendiği belirtilmiştir. Gelişmiş su kaynaklarına erişimin yüzdesinde ise kantiller arasında daha az farklılaşma görüldüğü gösterilmiştir.

İlkay Altındağ (2010), çalışmanın ilk uygulaması Konya ilinin güneşlenme süresi üzerine yapılmıştır. İkinci uygulama verisi ise apartmanın yaşı, hırsız sayısı ve gelir değişkenlerine ilişkin dört bağımsız değişkenle Kantil Regresyon uygulaması yapılmıştır. Konya ilinin güneşlenme süresi üzerine yapılan çalışmada Kantil Regresyon yöntemi ile EKK yöntemi kıyaslanmıştır. Ortalama hata kare ölçütlerine göre kıyaslandığında kantil değeri 0.45 seçildiğinde EKK yönteminden daha düşük hata verdiği belirtilmiştir. Çalışmanın ikinci uygulamasında, ortalama hata kare kriterine göre kıyaslandığında kantil değeri 0.5 olarak belirlendiğinde en düşük hata değeri gözlemlendiği belirtilmiştir. Kantil Regresyon analizi sonuçlarında, 0.5 kantil değerinde, ortalama hata kare ve ortalama mutlak hata değerleri en düşük değerleri gözlemlenmiştir.

Literatürde kabul görmüş deneysel çalışmaların büyük bir bölümü, basit regresyon analizlerinde EKK yönteminin kullanıldığı görülmüştür. Kantil Regresyon yönteminde, sadece bağımlı değişkenin ortalama değerleri değil aynı zamanda araştırmacıya, verinin dağılımı hakkında da bilgi verir.

### 3. YÖNTEM VE VERİ

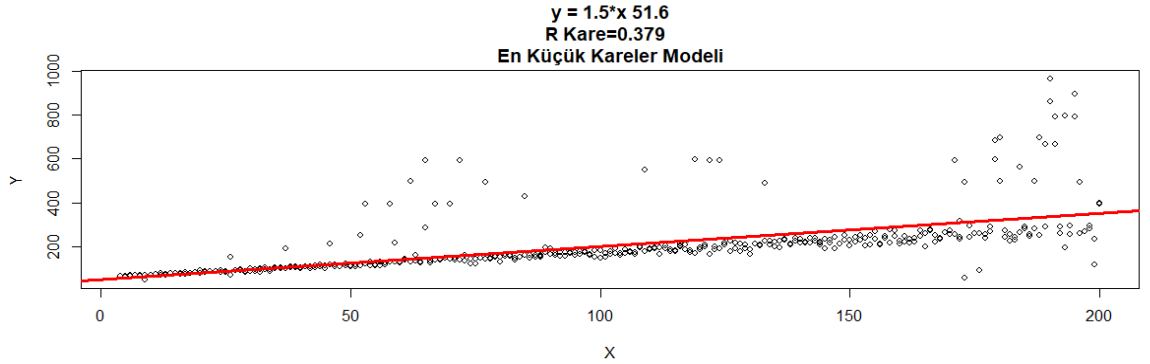
R programı üzerinden yapay olarak üretilen veride, uç değerler içeren bir bağımlı ve bir bağımsız değişken bulunmaktadır. Normal dağılım sergileyen veri seti üretildikten sonra uç değerler eklenip, toplam veri sayısı 400'de sınırlandırıldı. Kıyaslama yapılırken başarılı modeli seçme kriterleri olarak; Kök Ortalama Hata Kare, Ortalama Mutlak Hata ve Ortalama Mutlak Yüzde Hata değerleri kullanıldı. Analizler R Studio 3.5.0 programı üzerinden gerçekleştirildi. Bu çalışmada, uç değerlerin varlığında EKK yöntemi ile Kantil Regresyon yöntemi kıyaslandı. Bu çalışmanın amacı, EKK yönteminin kullanıldığı modellerde uç değerlerden dolayı gerekli varsayımlar sağlanamadığında, kullanılabilecek Kantil Regresyon yöntemini tanıtmak, uygulama çalışmasında uç değerler içeren ve yapay olarak üretilen veri setlerinde Kantil Regresyonu açıklamaktır.

#### 4. UYGULAMA

Uç değerler içeren veri setinde Kantil Regresyon yöntemi ile EKK yönteminin kıyaslanması için öncelikle EKK yönteminin uygulanacağı regresyon modeli oluşturuldu. Kantil Regresyon yönteminde, modelin hataları hakkında bir varsayım bulunmadığından uç değerlere karşı esnek, böylelikle uç değerlerin olduğu verilerde ya da değişen varyans (heteroscedasticity) barındıran veri setlerinde bu iki model kıyaslanmıştır.

Öncelikle EKK yöntemi kullanılarak regresyon analizi uygulandı. EKK yönteminin varsayımları uç değerlerden önemli ölçüde etkilenmektedir. Uç değerler eklenmeden önce yapılan modelinde, gerekli tüm varsayımlar sağlanmasıyla beraber değişen varyans problemi de mevcut değildi. EKK yöntemi ile kurulan modelde regresyon katsayıları %95 güven aralığında istatistiksel olarak anlamlı olduğu görüldü. Uç değerler eklenerek oluşturulan EKK yöntemi modelinde, denklemin çıkarsama ve durum belirleme amaçlı kullanılabilmesi için önce de değinildiği gibi hataların normal dağılım sergilemesi ve aykırı değer içermemesi gerekmektedir.

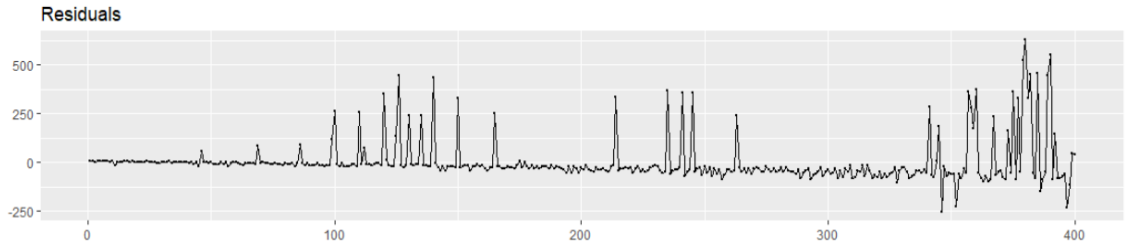
**Şekil 4. 1.** Uç değerlerin bulunduğu verinin regresyon doğrusu ve modeli



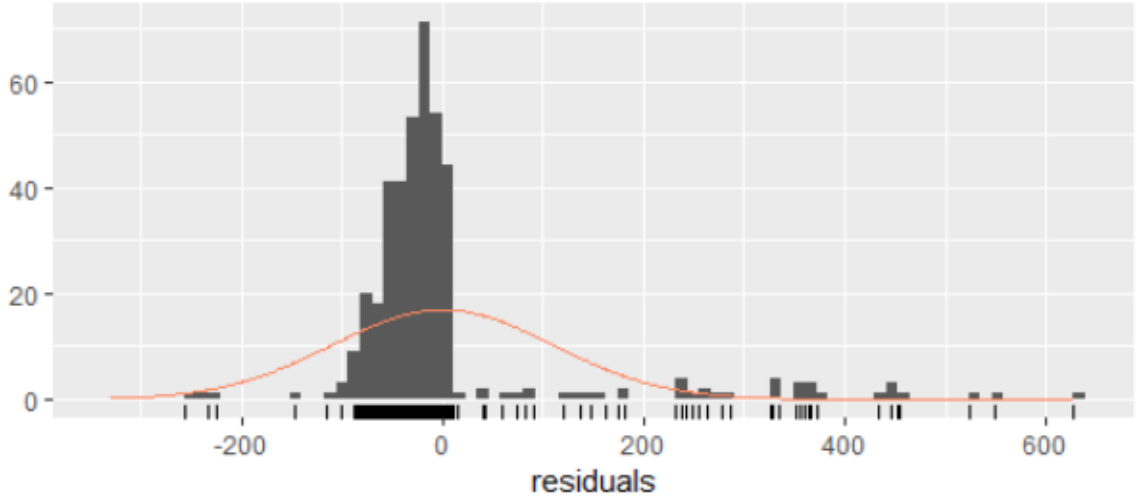
EKK yöntemi ile incelenen modelin çıkarsama ve durum belirleme amaçlı kullanılması, hata terimlerinin normal dağılıma sahip olmadığından ve aykırı değer içerdiğinden, değerlendirilen EKK yöntemine göre model oluşturmak uygun değildir. Bu model yardımıyla yapılacak çıkarsamalar gerçeği yansıtmayacaktır. Bu uygulamanın amacı, uç değerlerin varlığında bozulan EKK yöntemlerinin, alternatif yöntem olan Kantil Regresyon yöntemi ile regresyon modeli oluşturmaktır.

Veri setine uç değerler eklenmesiyle beraber değişen varyans problemi oluşmuştur.

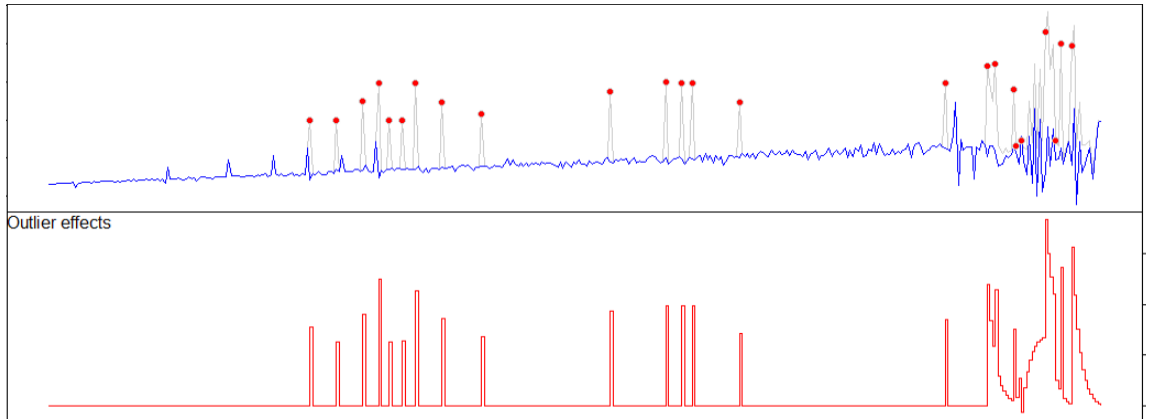
**Şekil 4.2.** EKK yöntemi ile kurulan modelin kalıntı dağılımı





**Şekil 4.3.** EKK yöntemi ile kurulan modelin kalıntılarının histogramı

EKK yönteminin uygulandığı modelin kalıntı dağılımı Şekil 4.2 ve Şekil 4.3 incelendiğinde, uç değerler modelin anlamlılığını ve uygulanabilirliğini değiştirmiştir. Uç değerler eklenmeden önce oluşturulan modelde dağılım normallik gösterirken uç değerlerden sonra verinin normal dağılımı bozulmuştur. Değişen varyans problemi incelenirken Breusch-Pagan testi uygulanmıştır.

**Şekil 4.4.** Veri setinde bulunan uç değerler ve uç değer etkisi

Veri setinde, uç değer etkisi gösteren veriler Şekil 4.4' teki gibidir. Değişen varyans, hata terimi varyansının tüm gözlemler için eşit olmaması durumudur. Değişen varyans problemi, önemli bir parametre modelin dışında kaldığı durumlarda, uç değerlerin olduğu durumlarda ya da model kurma hatasının olduğu durumlarda oluşur. “Değişen varyans problemini gidermek için değişken dönüşümü yapılmalı, modele önemli bir parametre eklenmeli veya ağırlıklı EKK yöntemi kullanılmalıdır” (Bager ve Odah, 2017). Değişen varyans problemi, Breusch-Pagan testi ile test edilmiştir.

**Tablo 4.1.** Breusch-Pagan testi çıktısı

Breusch-Pagan test	Breusch-Pagan test Değeri	p Değeri
BP	25.383	0.001

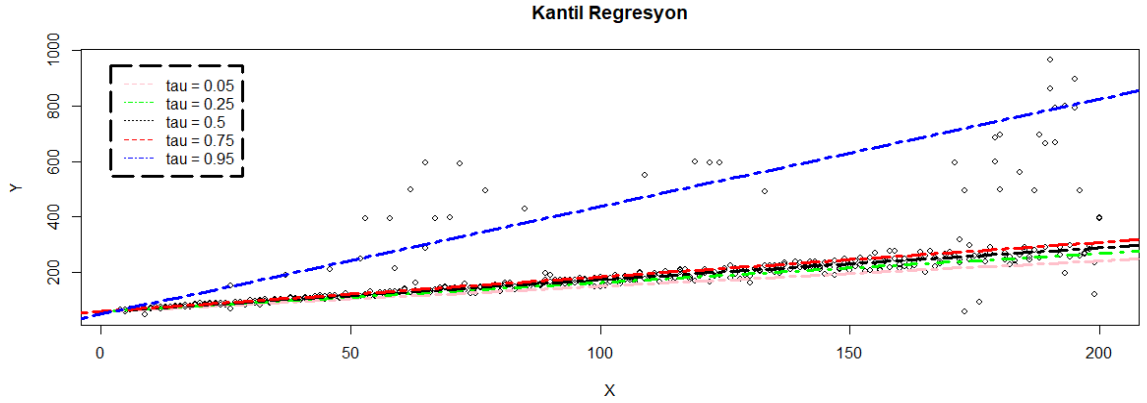
Ho: Değişen varyans yoktur (sabit varyanslı)

H1: Değişen varyans mevcuttur

Tablo 4.1' de p değeri incelendiğinde, %95 güven aralığında yokluk hipotezi reddedilir (alternatif hipotezi reddedecek kadar yeterli kanıt yoktur) ve veride değişen varyans (heteroscedasticity) mevcut olduğu görülür. Uç değerler, veri setinde bağımlı ve bağımsız değişkenin dağılımını ve varyansını değiştirdiği yorumu yapılır. Breusch-Pagan testi sonucunda sabit varyanslı olmadığı sınınanmıştır.

Kantil Regresyon yönteminin kullanılabilir olmasının sebeplerinden biri uç değerlere karşı esnek olmasıdır ve uç değerlerden etkilenecek varsayımları bulunmamasıdır. Değişen varyans problemi, EKK yönteminde önemli bir sorundur çünkü hata terimi varyansının tüm gözlemler için eşit olması gerekmektedir. Kantil Regresyon yöntemi, değişen varyansın olduğu durumlarda, EKK yöntemine alternatif olarak geliştirilmiş ve literatürde bu durumlarda kullanılması önerilmektedir.

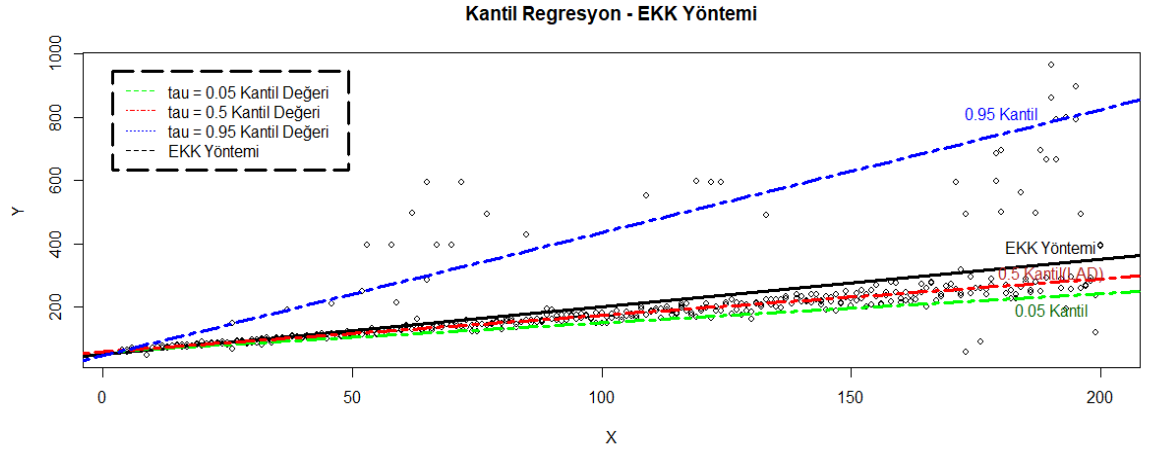
**Şekil 4.5.** Kantil değerlerine göre değişiklik gösteren regresyon doğruları



Şekil 4.5' te görüldüğü üzere araştırmacı, uç değerlerin dağılımına göre kantil değerini ayarlayabilir ve modelini bu regresyon doğrusunu veren model üzerine kurabilir. Uç değerlerin varlığında EKK yöntemi varsayımlarından olan sabit varyans özelliğini yitirebilir fakat kantil Regresyon yönteminde böyle bir varsayım olmadığından uç değerlerden etkilenmez ve uç değerlerin konumlarına göre en düşük hatalı model tercih edilir.

Uç değerlerin dağılımına göre EKK yöntemi ile kurulan regresyon doğrusu sapma gösterir ve uç değerlerin olduğu tarafa yakınsar. Kantil Regresyonda uç değerlerin dağılımı ve konumuna göre dengelenerek model kurulur. Bu sebeple uç değerlere karşı esnektir ve araştırmacıya kolaylık sağlar.

Şekil 4.6. EKK yöntemi ve kantil regresyona göre kurulan regresyon doğruları



Şekil 4.6’ da Y değerindeki uç değerlerin dağılımı, EKK modeli regresyon doğrusunu yukarı doğru maskeleymiş ve hatayı arttırmıştır. Kantil Regresyon yöntemi ile oluşturulan modeller, kantil değerine göre (uç değerlerin dağılımına/konumuna göre) değiştirilebildiğinden hatası en düşük olan anlamlı model seçilebilir. Uç değerler, Kantil Regresyon doğrusu maskeleyemez çünkü araştırmacı kantil değerini sapan değerlerin dağılımına göre belirler, bu sebeple uç değerlerden kaçınılır.

Şekil 4.6’ da uç değerlerin, regresyon eğrisini değiştirdiği görülür. Kantil Regresyon doğruları incelendiğinde, 0.95 kantil değerine sahip olan modelin regresyon doğrusu, uç değerlerden geçtiği görülür. EKK yöntemi ile kurulan modelin regresyon doğrusu, 0.5 kantil değerine sahip Kantil Regresyon doğrusuna göre uç değerlere daha yakındır. Bu durumda model hatasının yüksek olması beklenir.

Farklı kantil değerlerine sahip Kantil Regresyon modelleri, EKK yöntemi ile kıyaslanmıştır. Yeni gözlem değeri modele eklendiğinde, oluşan test hataları tablodaki gibidir;

Tablo 4.2. EKK ve Kantil Regresyon model hataları

	RMSE	MAE	MAPE
Model (EKK)	124.06	66.09	22.45
Model_1 (Kantil 0.05)	126.47	56.28	33.15
Model_2 (Kantil 0.1)	123.32	51.06	28.66
Model_3 (Kantil 0.9)	183.29	155.00	42.65
Model_4 (Kantil 0.9)	115.17	42.68	21.19

0.5)			
Model_5 (Kantil 0.95)	295.52	249.28	52.22

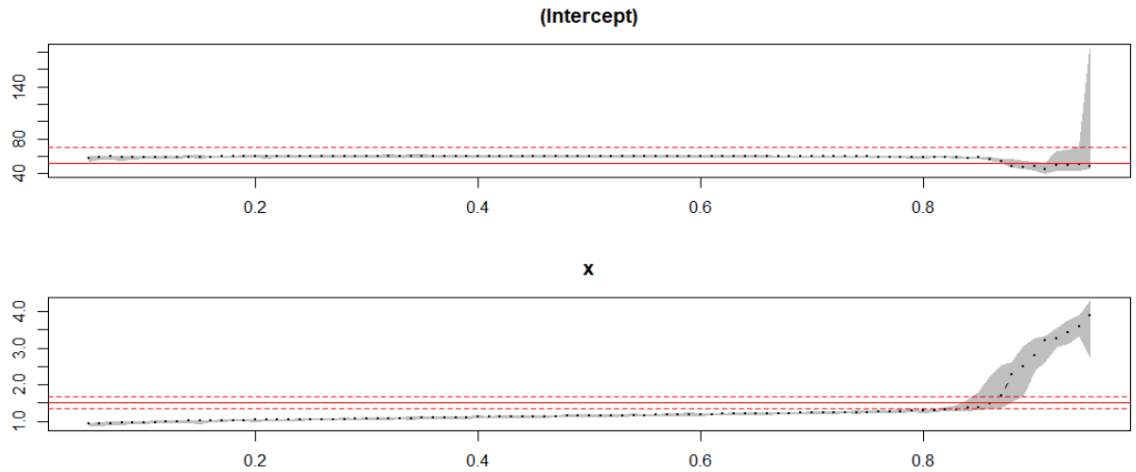
Tablo 4.2' e göre en düşük hatayı veren model, 0.5 kantil değeri (Q(0.5)) ile Model\_4 olmuştur. Kantil değerinin 0.5 olmasıyla En Küçük Mutlak Sapma (LAD) regresyon analizi elde edildi. Kantil değeri 0.5 olan Model\_4'ün sonuçları;

**Tablo 4.3.** Model\_4 katsayı tablosu

Model 4	Kat Sayı	Standart Hata	t Değeri	p değeri
Sabit	60.28	1.15	52.23	0 .001
X	1.14	0.009	115.57	0 .001

Tablo 4.3' te p değerleri incelendiğinde model sabitinin ve X parametresinin %95 güven aralığında istatistiksel olarak anlamlı olduğu görülür.

**Şekil 4.7.** Model\_4'ün sabit ve bağımlı değişkeninin katsayı ve güven bandları tahmin grafiği



Şekil 4.7' de görülen her siyah nokta, x ekseninde belirtilen kantilin eğim katsayısıdır. Kesikli ve düz çizgiler %95 güven aralığında alabileceği değerleri gösterir ve bu bölgenin değerleri EKK yöntemi kullanılarak belirlenir. Eğer tüm noktalar, güven aralığı bölgesinin içinde olmuş olsaydı, doğrusal regresyon eğiminin bir bağımlı ve bir bağımsız değişkenin arasındaki ilişkiyi tanımlamak için yeterli olduğu yorumu yapılabilirdi.

## 5. SONUÇ

Regresyon analizinde EKK yönteminin bazı varsayımları vardır. Bu varsayımlar, uç değerlerden büyük oranda etkilenir. Bu etki sonucunda varsayımlar bozulabilir ve bu durumda da EKK yöntemi uygulanamaz hale gelebilir. Kantil Regresyon yöntemi, uç değerlere karşı esnektir ve araştırmacıya özgür bir alan verir. Bu sebeple Kantil Regresyon yöntemi, uç değerlerin veride bulunması durumunda EKK yöntemine alternatif olarak geliştirilmiştir.

Literatürde geleneksel regresyon modellerindeki bağımlı değişkenin, koşullu beklenen değeri açıklamayı amaçlarken, alternatif olarak kullanılan Kantil Regresyon yönteminde ise koşullu kantillerdeki değişimi hedefler ve açıklar. Kantil Regresyon yönteminde, sadece bağımlı değişkenin ortalama değerleri değil aynı zamanda araştırmacıya, verinin dağılımı hakkında da bilgi verir.

Kantil Regresyon yönteminde, uç değerlerin dağılımına göre kantil değerleri ayarlanabildiğinden uç değerlerden/sapan değerlerden, kantil değerleri değiştirilerek bu uç değerlerden arındırılabilir. Literatürde, uç değerlerin varlığında Kantil Regresyonun, EKK yöntemine göre daha başarılı sonuçlar verdiği belirtilmiştir. Bu çalışmada da uç değerlerin bulunduğu veride, 0.5 kantil değerine (Q(0.5)) sahip Kantil Regresyon modeli, EKK yöntemine göre oluşturulmuş modelden daha başarılı sonuç verdiği görülmüştür. EKK modelinde Kök Ortalama Hata Kare değeri 124.06 iken 0.5 kantil değerine sahip Kantil Regresyon modelinde ise Kök Ortalama Hata Kare değeri 115.17 olmuştur. Normal dağılım sergileyen veri seti R programında yapay olarak üretildikten sonra uç değerler eklenip, toplam veri sayısı 400 de sınırlandırılmıştır. Bu veri setinde Kantil Regresyon modeli daha başarılı olduğu görülmüştür.

Bu çalışmada, EKK yöntemine alternatif olarak geliştirilen Kantil Regresyon yöntemi ile EKK yönteminin kıyaslanması amaçlanmıştır. Yapılan uygulama sonucunda sabit varyans varsayımını bozabilecek düzeyde uç değerler içeren veri setinde, Kantil Regresyon yönteminin EKK yöntemine göre daha başarılı sonuçlar verdiği gözlenmiştir. Bu uygulamadan ulaşılan sonuç ile uç değer içeren verilerde yapılacak analizlerde ve çalışmalarda araştırmacılara yol gösterici bir çalışma olabilir.

## KAYNAKLAR

ABADIE, A., Angrist, J. & Imbens, G., (2002), Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. **Econometrica**, 70(1): 91-117.

ABDULLAHI, I., (2011), Analysis of Quantile Regression as Alternative to Ordinary Least Squares Regression, Unpublished Master Thesis, Nijerya: Ahmadu Bello University.

ALTINDAĞ, İ., (2010), Kantil Regresyon ve Bir Uygulama, Yayınlanmamış Yüksek Lisans Tezi, Konya: Selçuk Üniversitesi.

- BAGER, A., S., M., Odah, M., H. & Mohammed, B., K., (2017), Using Approach Quantile Regression to Determine the Factors Affecting Measuring Capacity in Iraq, **American Review of Mathematics and Statistics**, 5(1): 35-44
- CHEN, C. & Wei, Y., (2005), Computational Issues for Quantile Regression, **The Indian Journal Of Statistics**, 67(2): 399-417.
- COOK, D., Wang, W. & Wang, E., (2017), Package: quonkar, Quantile Regression Outlier Diagnostics with K Left Out Analysis, <https://cran.r-project.org/web/packages/quokar/quokar.pdf>, Erişim Tarihi: 22 Eylül 2018.
- ÇAĞLAYAN, E. ve Arikan, E., (2011), Determinants of House Prices in Istanbul: A Quantile Regression Approach, **Qual Quant**, 45(2): 305-317
- FIRPO, S., Fortin, N. & Lemieux, T., (2009), Unconditional Quantile Regressions, **Econometrica**, 77(3): 953-973
- KESKİN, B., (2012), Sağlam Bir Çıkarsama Yöntemi: Kantil Regresyon, Yayınlanmamış Yüksek Lisans Tezi, Muğla: Muğla Sıtkı Koçman Üniversitesi
- KOENKER, R. & Bassett, G., (1978), Regression Quantiles, **Econometrica**, 46(1): 33-50
- KOENKER, R. & Bassett, G., (1982), Robust Tests of Heteroscedasticity based on Regression Quantiles, **Econometrica**, 50(1): 43:61
- KOENKER, R. & Hallock K., F., (2001), Quantile Regression, **Journal of Economic Perspectives**, 15(4): 143-156
- KOENKER, R., (2005), **Quantile Regression**, London: Cambridge University Press.
- KURTOĞLU, F., (2011), Kantil Regresyon: Teorisi ve Uygulamaları, Yayınlanmamış Yüksek Lisans Tezi, Adana: Çukurova Üniversitesi
- OCAK., B., (2010), Dik Ağırlıklandırmanın M-Regresyon Üzerine Etkisi, Yayınlanmamış Yüksek Lisans Tezi, Konya: Selçuk Üniversitesi.
- ÖZEL, H., A. ve Sezgin, F., (2013), Ticari Serbestleşme-Ekonomik Büyüme İlişkisinin Bootstrap Kantil Regresyon Yardımıyla Analizi, **İstanbul Üniversitesi İktisat Fakültesi Mecmuası**, 62(2): 283:303.
- SAÇAKLI, İ., (2005), Kantil Regresyon ve Alternatif Regresyon Modelleri ile Karşılaştırması, Yayınlanmamış Yüksek Lisans Tezi, İstanbul: Marmara Üniversitesi.
- STANDER, J., Yu, K. & Lu, Z., (2003), Quantile Regression: Applications and Current Research Areas, **The Royal Statistical Society**, 52(3): 331-350
- TÜRKİŞ, C., (2015), Sağkalım Analizinde Kantil Regresyon ve Parametrik Regresyon Modellerinin Performanslarının Karşılaştırması, Yayınlanmamış Yüksek Lisans Tezi, Aydın: Adnan Menderes Üniversitesi.
- UĞUR, A., A. ve Özocaklı, D., (2018), Gıda Güvencesizliğinin Bazı Belirleyicileri (Kantil Regresyon Yöntemi ve Sabit Etki Panel Yönteminin Karşılaştırılması), **Sosyoekonomi**, 26(35): 195-205.

YAVUZ, A., A. ve Aşık, E., G., (2017), Kantil Regresyon, **Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi**, 9(2), 138:146.

YILDIRIM, N., (2010), En Küçük Kareler, Ridge Regresyon ve Robust Regresyon Yöntemlerinde Analiz Sonuçlarına Aykırı Değerlerin Etkilerinin Belirlenmesi, Yayımlanmamış Yüksek Lisans Tezi, Adana: Çukurova Üniversitesi.

YORULMAZ, Ö., (2009), Dayanıklı Regresyon Yöntemi ve Çeşitli Sosyal Veriler Üzerine Aykırı Gözlemlerin Teşhisi, **Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi**, 12(21): 76-88.