Süleyman Demirel Üniversitesi
İktisadi ve İdari Bilimler Fakültesi
Y.2005, C.10, S.1 s.263-279.

AŞIRI DEĞER İÇEREN VERİ KÜMELERİNDE HATA TERİMLERİNİN BİNOM DAĞILIMA UYDUĞU DURUMDA EKK VE ROBUST LTS REGRESYON TAHMİNCİLERİNİN SİMULASYON ÇALIŞMASI İLE KARŞILAŞTIRILMASI

THE COMPARISON OF OLS METHOD WITH ROBUST LTS REGRESSION ESTIMATORS SIMULERTION IN CASE OF BINOMIAL DISTRIBUTION OF ERROR TERMS IN DATA SETS INCLUDING OUTLIERS

Dr.Latif ÖZTÜRK
Dr. Hakan TÜRKAY*

ÖZET

Doğrusal regresyon modellerinin çözümü için kullanılabilecek bir çok alternatif yöntem vardır. Bu yöntemler arasında en çok tanınan, kullanılan ve uygulamada neredeyse standart bir mekanizma halinde gelen yöntem EKK(En Küçük Kareler)'dır. Yöntemin dayandığı, aralarında hata terimlerinin normal dağılıma uyduğu varsayımı da olan önemli bir çok varsayım söz konusudur. Bu varsayımların sağlandığı durumda şüphesiz EKK tahmincisi en iyidir. Ancak, uygulamada sıklıkla karşılaşılan durum varsayımların bazılarının sağlanamaması halidir ve bu durumda EKK istenen özelliklere uygun parametre tahminleri yapamaz. Bu sorun karşısında (varsayımlardan sapma halinde) varsayımlardan sapmaya duyarsız olan robust regresyon yöntemleri geliştirilmiştir. Bu çalışmada, hata terimlerinin Binom dağıldığı durumda robust regresyon yöntemlerinden biri olan LTS (Least Trimmed Squares) tahmincisinin EKK ile karşılaştırılması simulasyon çalışması yardımıyla yapılmıştır. Karşılaştırma, veri kümesinin aşırı değerler (outliers) içerdiği ve içermediği durumlarda gerçekleştirilmiştir. Yöntemlerden elde edilen parametre tahminleri başlangıçta atanan parametre değerlerinden gösterdikleri sapma miktarlarına göre iki yöntemin performansı değerlendirilmiştir.

ABSTRACT

There are many alternative methods which can be used for the solution of linear regression models. Among these methods, LS is the best known, the most commonly used method, and it is assumed as the standard mechanism in applications. There are many assumptions, one of which is that

* İnönü Üniversitesi İ.İ.B.F. Ekonometri Bölümü

"error terms are distributed normal", on which the method is based. When the assumptions are proved, LS is the best method to be used. However, in practice, all of the assumptions mostly can not be proved, and in such a situation, LS can not estimate the parameters with required properties. Towards this problem (deviations from assumptions) robust regression methods which are insensitive to deviations from assumptions, are developed. In this study, LTS (Least Trimmed Regression), one of the robust regression methods, is compared with LS by using simulation approach, within the case that error terms are distributed binomial. This comparison is held within the situations where data include outliers and vice-versa. The performance of the methods are evaluated according to the bias between the parameter estimate which are obtained by using the regression methods and appointed as initial parameter values.

Robust Regresyon, LTS, Aşırı Değer, Simulasyon
Robust Regression, LTS, Outliers, Simulation

1. GİRİŞ

Regresyon analizinde öncelikli amaç, örneklem regresyon fonksiyonunu temel alarak anakütle regresyon fonksiyonunu olabildiğince doğruya yakın tahmin etmektir. Örneklem regresyon fonksiyonunu bulmak için kullanılan çeşitli yöntemler arasında en yaygın tanınan ve kullanılan yöntem, Carl Friedrich Gauss tarafından bulunan Sıradan En Küçük Kareler (EKK) yöntemidir. Yöntemin ana fikri kalıntı karelerinin toplamını minimum yapmak ($\underset{\hat{\theta}}{Min} \sum_{i=1}^{n} r_i^2$) ve bu yolla en iyi tahmin değerlerine ulaşmaktır. Bu minimizasyon işlemi sonucunda bulunan tahminciler en küçük kareler tahmincileri diye adlandırılmakta ve klasik istatistiğin çok önemli bir temel taşını oluşturmaktadır. Bu tahminciler, klasik doğrusal regresyon modelinin varsayımları veri iken sapmasız doğrusal tahminciler içerisinde en küçük varyanslı olanlardır; yani en iyileridir. Bu ifade, Gauss-Markov teoremi diye adlandırılır.

EKK yönteminin bu ölçüde yaygın kullanılmasının nedeni, kavramın kolaylıkla anlaşılabilmesi ve uygulanabilmesidir. Yöntemin bulunduğu zamanlarda bilgisayarların olmaması ve EKK tahmincilerinin veriler üzerinde matris cebri uygulamalarıyla kolayca hesaplanabilir olması, yalnızca EKK tahmincisini uygulanabilir bir yaklaşım yapmıştır. Gauss, hata teriminin normal (Gaussyen) dağıldığı durumda EKK'nın en iyi olduğunu gösterdikten sonra, Gaussyen varsayımlar ve EKK yöntemi uygulamada standart bir mekanizma haline getirilmiştir. Ancak, verilerin çoğu zaman klasik varsayımları tamamen sağlamadığı uzun yıllardır bilinen bir gerçektir. (örneğin, hata teriminin dağılımı her zaman normal değildir.) Buna bağlı olarak, bu sorunu aşabilmek ve varsayımların sağlanmadığı durumlarda da regresyon doğrusunu gerçeğe en yakın biçimde tahmin etmek amacıyla robust (dayanıklı) regresyon tahminciler önerilmiştir.

Regresyon analizi uygulamalarında, hata terimi sadece normal dağılıma uymamakla kalmayıp, bazı durumlarda kesikli dağılımlara uygun bir dağılımda göstermektedir. Örneğin, gölge bağımlı değişkenli doğrusal olasılık modellerinde hata terimi binom dağılıma uymaktadır. Bundan dolayı, bu çalışmada hata teriminin Binom dağılıma uyduğu durumda EKK ve yüksek bozulma sınırına sahip dayanıklı tahmincilerden biri olan LTS (Least Trimmed Square : En küçük Kırpılmış Kareler) tahmincilerinin EKK yöntemiyle karşılaştırılması yapılarak, yöntemlerin üstünlük ve zaafları ortaya çıkarılmaya çalışılmıştır.

2. EKK YÖNTEMİNİN AŞIRI DEĞERLERE KARŞI DUYARLILIĞI

Bir veri kümesi dört tip gözlem değeri içerebilir : Düzgün (regular) gözlemler, dikey aşırı değerler, iyi kaldıraç noktaları ve kötü kaldıraç noktaları (yüksek etkili gözlemler).[1]

Aşırı değerler, tipik olmayan ve verinin geri kalan kısmının karakteristik dağılımına uygun görüntü vermeyen gözlem değerleridir. Bu değerler, temel olgunun (değişkenin) özelliklerini yansıtıyor olabileceği gibi, regresyon analizi uygulamalarında sıklıkla karşılaşılan bir durum olan bazı gözlem değerlerinin ölçme, kaydetme, kopyalama veya aktarma hataları nedeniyle yanlış bir değer olarak alınmasından da kaynaklanabilir. Sonuç olarak; "Regresyon aşırı değerleri, verilerin çoğunluğu tarafından oluşturulan doğrusal kalıba uymayan gözlemlerdir. Bu tip bir gözlem olağanüstü bir Y_i veya olağanüstü bir $(X_{i1},,.....,X_{ip})$ veya da her ikisi birlikte olabilir."[2]

Bağımlı değişken Y_i'nin gözlem değerlerinden (X_k, Y_k) gibi bir nokta Y_i değerlerinin çoğunluğundan oldukça uzak bir yerde yer almakta ise, bu nokta Y yönünde aşırı değer olarak adlandırılır. Söz konusu nokta, EKK doğrusu üzerinde oldukça büyük bir etkiye sahip olacak ve doğruyu tamamen değiştirebilecektir. Bu tip dikey aşırı değerler büyük pozitif veya büyük negatif kalıntı (r_i) değerlerine sahiptirler. Ancak, bu durumda elde edilen kalıntı değeri (r_i) şüpheli bir büyüklüktür. Bu tip aşırı değerler kalıntıların listesinden veya kalıntıların grafiğinin çizimi yardımıyla tespit edilebilir.

Ancak, büyük hataların sadece bağımlı değişkende ortaya çıkması için herhangi bir neden yoktur, bu durum mutlak anlamda, açıklayıcı değişkenlerin birinde aşırı değerlerin bulunmasına çok benzerdir. Genelde örneklemde ki X_i değerlerinin büyük çoğunluğundan çok uzakta yer alan

---

[1] Peter J. Rousseeuw, Katrien Van Driessen, "Computing LTS Regression for Large Data Sets" (Düzenlenmiş), *Technical Report*, University of Antwerp,1999,s.2.

[2] Peter J. Rousseeuw, Stefan Van Aelst, "Positive-Breakdown Robust Methods in Computer Vision" **Computing Science and Statistics**, Vol:31,1999, s. 451.

X_k gibi bir değer olduğunda (X_k, Y_k) değerine X yönünde aşırı değer denir. Veri kümesinin merkezine uzak olan bu nokta, regresyon doğrusunu büyük bir güçle çektiğinden mekanikteki kaldıraç kavramına benzer olarak kaldıraç noktası (leverage point) olarak adlandırılmaktadır. (X_k, Y_k) noktası verilerin çoğunluğu tarafından oluşturulan örüntüye uygunsa yani çoğunluk veri tarafından belirlenen regresyon doğrusuna yakın bir yerde ise iyi kaldıraç noktası olarak adlandırılır. [3] Çoğu zaman regresyon aşırı değerlerinin EKK kalıntılarının incelenmesiyle tespit edilebileceği düşünülmektedir. Ne yazık ki, aşırı değerlerin kaldıraç noktası olduğu durumlarda bu doğru değildir. Kaldıraç noktaları, çok büyük kalıntı değerlerine sahip olacağından $\sum_{i=1}^{n} r_i^2$ değerine büyük bir ilave katkıda bulunur. Bu durumda, gerçek regresyon doğrusu EKK yaklaşımıyla seçilemez, çünkü EKK doğrusu söz konusu noktaya doğru kayacağından büyük kalıntı değeri gerçekte olduğundan küçük, diğer noktaların kalıntıları ise olması gerekenden büyük görünecektir. Eğer "en büyük EKK kalıntılarına sahip olan noktaların atılması" şeklinde bir ilke uygulanırsa, iyi veri noktaların atılması söz konusu olabilir. Regresyon aşırı değerleri (ister X, isterse Y yönünde olsun) EKK yöntemi için ciddi bir tehdit oluşturmaktadırlar.[4]

Sonuç olarak, EKK gibi klasik parametre tahmin yöntemlerinin anlaşılması ve hesaplanması kolay olmasına rağmen, verinin içerdiği az sayıda aşırı değer bile söz konusu yöntemler tarafından tahmin edilen doğruyu keyfi bir büyüklükte etkilemekte ve parametre tahmin değerleri üzerinde önemli bir rol oynamaktadırlar. Bundan dolayı, dayanıklı tahminci olarak adlandırılan ve bahsedilen sorundan fazlaca etkilenmeyen tahmincilere ilgi hızla artmaktadır.

3. ROBUST (DAYANIKLI) REGRESYON TAHMİNCİLERİNİN KULLANILMA NEDENLERİ

Doğrusal regresyon analizi uygulamalarının bir çoğunda, serilerde aşırı değerlerin olması veya hata teriminin normal dağılımdan uzaklaşması gibi sorunlar ortaya çıkmaktadır. Söz konusu sorunların varlığı durumunda, daha önce belirtilen varsayımlardan küçük sapmalara bile aşırı duyarlılık göstermesi özelliğinden dolayı, EKK tahmincileri istenen özelliklere uygun tahminler verememektedir. Bu nedenle, EKK tahmincilerinde ortaya çıkan sorunların üstesinden gelebilecek ve aşırı değerlere karşı duyarsız olan dayanıklı regresyon yöntemlerine ihtiyaç duyulmuştur. Bu yöntemler, hata terimi normal dağılmadığında ve aşırı değerlerin varlığı durumunda klasik

[3] Peter J. Rousseeuw, Annick M. Leroy, **Robust Regression and Outlier Detection**, New York, John Wiley, 1987, s.6.

[4] Peter J. Rousseeuw, Annick M. Leroy , **a.g.e.**, s.7-8.

yöntemlere göre daha uygun sonuçlar vermektedir. Daha önce bahsedildiği gibi, verinin aşırı değerler içermesi durumunda, öncelikle verinin çeşitli dışlama kuralları kullanılarak aşırı değerlerden temizlenmesi, daha sonra kalan veri üzerinde klasik tahmin ve test yöntemlerinin uygulanmasının yeterli olacağı düşünülebilir. Ancak, EKK yöntemiyle yapılan tahminde regresyon doğrusu aşırı değerlere doğru çekileceğinden bu değerlere ait kalıntılar oldukça küçük olabilecektir. Bu durum, söz konusu aşırı değerleri maskeleyebilecektir. EKK yönteminin veri kümesi içinde aşırı değerlerin varlığı halinde kullanılamamasındaki diğer nedenler aşağıdaki gibi sıralanabilir: [5]

1. Yukarıdaki iki aşamayı (aşırı değerlerin teşhisi ve parametrelerin tahmini aşamalarını) birbirinden ayırabilmek nadiren mümkündür. Örneğin, çok değişkenli regresyon problemlerinde, parametreler için dayanıklı tahminlere başvurmadan aşırı değerlerin (doğru bir şekilde) tespit edilmesi zordur.

2. Orijinal gözlem değerleri kütlesi, arasına büyük hatalar karıştırılmış normal gözlemlerden oluşmuş olsa bile, temizlenmiş veri normal olmayacaktır ve istatistiksel hatalar meydana gelecektir (yanlış dışlamalar veya alıkoymalar nedeniyle). Sonuç, gerçekte normal olmayan bir dağılımdan türetildiği durumdakinden daha da kötü olacaktır. Dolayısıyla, klasik normal teorinin temizlenmiş örneklemlere uygulanabilirliği söz konusu değildir. Bu durumdaki, iki aşamalı yöntemin işleyişi, doğru bir dayanıklı yöntemin işleyişinden çok daha zor olabilir.

3. En iyi dışlama yöntemlerinin performanslarının bile, iyi bir dayanıklı yöntemin performansına yeterli düzeyde ulaşamadığı da deneysel olarak gösterilebilen bir gerçektir. Dayanıklı yöntemler gözle görülür biçimde çok daha iyidir, çünkü bunlar değerlerin tamamen kabulü ile tamamen dışlanması arasında sorunsuz bir geçişi sağlayabilirler.

Ayrıca, aşırı değerlerin teşhisi, "ne kadar uç noktada" olduğunu ölçmek için bir ölçüye ihtiyaç duymaktadır. Bu ölçü, veri için bazı modellerden ve bu modelden ayrılığın bazı ölçülerinden ortaya çıkmaktadır. Çoklu aşırı değerler, söz konusu ölçünün kendisinin bir aşırı değer tarafından bozulmuş olabileceği tehdidini taşımaktadır. Aşırı değerler ölçüsünün kirlenmesi, aşırı değeri tespit edicisini bozmaktadır ve tabi ki aynı zamanda bu aşırı değeri tespit etme usulü üzerine kurulan herhangi bir tahminciyi de bozmaktadır. [6]

---

[5] Peter J. Huber, **Robust Statistics**, New York, John Wiley and Sons, 1981, s.4-5.

[6] N. Billor, A.S. Hadi, P.F. Vellman, "BACON : Blocked Adaptive Computational Efficient Outlier Nominators", **Computational Statistics and Data Analysis**, Vol.34, s.280.

4. ROBUST REGRESYON TAHMİNCİLERİNİN AMAÇLARI

Robust (dayanıklı) regresyon tahmincileri, bir çok amaca yönelik olarak son yıllarda giderek artan bir hızda istatistik ve ekonometri de regresyon modellerinin parametrelerinin tahmin edilmesi için kullanılmaktadır. Bu yöntemlere başvurulmasındaki amaçlar aşağıdaki gibi özetlenebilir:[7]

1. Veri kümesine en iyi uyumu gösteren yapıyı tespit etmek,

2. Eğer gerekli görülürse, uzaktaki veri noktalarını (aşırı değerleri) veya daha ileri analizler için yapının temelindeki sapmaları teşhis etmek,

3. Yüksek düzeyde etkili veri noktalarını teşhis etmek ve bunlar üzerine dikkat çekmek,

4. Kuşkulanılmayan ardışık bağımlılık veya daha da genel olarak varsayılan korelasyon yapısındaki sapmalar ile ilgilenmektir.

Gözlem değerlerinin küçük bir kısmında büyük sapmaların bulunması küçük sapmalar olarak kabul edilebilirler; dayanıklı yöntemlerin öncelikli hedefi, bazı klasik yöntemlerin aşırı duyarlılığı da göz önüne alındığında, büyük hatalara karşı koruma sağlamasıdır. Literatürde dayanıklı yöntemler için açıkça veya ima yolu ile çok sayıda başka amaçlara da (örneğin, yüksek asimptotik görece etkinlik veya yüksek mutlak etkinlik gibi) yer verilmektedir. Ne var ki, bu amaçlar ikinci derecede önemlidir ve bunlara yukarıda bahsedilen amaçlardan daha fazla önem verilmemelidir.[8]

5. BOZULMA SINIRI (BREAKDOWN POİNT)

Aşırı değerlere karşı nicel dayanıklılığın bir ölçüsü bozulma sınırıdır. "İlk olarak Hodges tarafından 1967'de kullanılan bozulma sınırı kavramı tek boyutlu konum tahmincileri ile sınırlıydı. Daha sonra Hampel tarafından 1968 ve 1971'de yapılan çalışmalarla genelleştirilmiştir."[9] Hampel 1971'de çok daha genel bir tanımlama yapmıştır, bu tanımlamada asimptotik olduğundan ve dolayısıyla matematiksel bir ifade olduğundan çok fazla yaygınlaşamadı. Ancak, 1983'de Donoho ve Huber tarafından yaygın bir şekilde tanınır hale getirilmişlerdir. Bozulma sınırının Donoho ve Huber tarafından tanımlanan basit bir sonlu örneklem versiyonu şöyledir:[10]

---

[7] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, Werner A. Stahel, **Robust Statistics**, New York, John Willey & Sons, 1986, s.11.

[8] Peter J. Huber, **a.g.e.**, s.5-6.

[9] Yijun Zuo, "Some Qantitative Relationships Between Two Types of Finite Sample Breakdown Point", **Statistics and Probability Letters**, vol.51, s.369-375, 2001.

[10] Peter J. Rousseeuw, Annick M. Leroy , **a.g.e.**, s.9.

Z, n tane gözlemden oluşan bir örneklem ve T bir regresyon tahmincisi varsayılırsa;

$$Z = \{(X_{11},.....X_{1p},Y_1).....(X_{21},.....X_{np},Y_n)\} \quad (1)$$

Z örneğine T yöntemi uygulanarak $T(Z) = \hat{\theta}$ şeklinde katsayılar vektörü bulunabilir. Burada, orijinal veri değerlerinden herhangi m tanesinin yerine keyfi bir değer yazılarak elde edilen kirletilmiş bir Z' düşünelim. Böyle bir kirletme ile neden olunabilen sapma bias(m;T,Z) ile gösterilirse;

$$\text{Sapma}(m,T,Z) = \sup_{Z'} \|T(Z') - T(Z)\| \quad (2)$$

Eğer sapma (m;T,Z) sonsuz ise, m tane aşırı değerin T üzerinde büyük bir etkiye sebep olduğu, hatta tahmincinin bozulduğu söylenebilir. Bu nedenle, Z sonlu örneğinde T tahmincisinin bozulma sınırı,

$$\varepsilon_n^*(T,Z) = \min\left\{\frac{m}{n}, sapma\,(m;T,Z)\ \text{sonsuz}\right\} \quad (3)$$

şeklinde tanımlanır. Bozulma değeri veya sınırı bir anlamda en kötü durum senaryosudur. "Kabaca ifade edilecek olursa, bozulma sınırı tahmincilerin başa çıkabilecekleri kötü aşırı değerlerin oranının sınırını verir." [11] Diğer bir deyişle; bozulma sınırı T tahmincisinin T(Z) den oldukça uzak keyfi değerler almaya başlamasına neden olabilecek en düşük kirlenme oranıdır. [12] Yani, bir tahmincinin bozulma sınırı, tahmincinin keyfi olarak büyük sapma değerleri almasına yol açabilen aşırı değerlerin oranı olarak tanımlanır. [13] EKK'nın bozulma sınırı, örneklemdeki tek bir aşırı değerden bile etkilendiği için 1/n'dir, dolayısıyla asimptotik durumda n sonsuza giderken sıfıra yaklaşır.

Bir parametre tahmin yönteminin çöktüğü nokta, yani bozulma sınırı, sonuçta elde edilecek tahmin değerini anlamsızlaştıracak kadar kirletebilen veya yönlendirebilen örneklemdeki noktaların en küçük kısmıdır. [14] Sonuç olarak, bozulma sınırı bir tahmincinin anlamsız sonuçlar vermeden kullanılabileceği, gerçekte "kötü" kirlenmenin ne kadar olabileceği konusunda kaba bir fikir verir. Bozulma sınırı, tahmincinin kararlılığının global bir ölçüsüdür. Bundan dolayı, T tahmincisi tamamen işe yaramaz hale gelmeden önce, kirlenmiş modelin varsayılan modelden ne kadar

---

[11] Peter J. Huber, **a.g.e.**, s.13.
[12] Peter J. Rousseeuw, Annick M. Leroy, **a.g.e.**, s.10.
[13] J. Matoušek, D.M. Mount, N.S. Natenyahu, "Efficient Randomized Algorithms For The Repeated Median Line Estimator**", Algorithmica,** 20, 1998, s.136.
[14] Feifang Hu, Jianhua Hu, "A Note On Breakdown Theory For Bootstrap Methods", **Statistics and Probability Letters**, 50, 2000, s.49-53.

uzaklaşabileceğini gösterir. Aynı zamanda, bozulma sınırı, modele bağlı olmaması anlamında da global bir ölçüdür. [15]

6. ROBUST REGRESYON TAHMİNCİLERİNİN ÖZELLİKLERİ

Ele alınan herhangi bir istatistiksel yöntem aşağıdaki özelliklere sahip olmalıdır: [16]

1. Yöntem, varsayılan modelde makul bir ölçüde iyi (optimal veya optimale yakın) etkinliğe sahip olmalıdır.

2. Model varsayımlarından küçük sapmaların yöntemin performansını sadece düşük bir düzeyde zayıflatması anlamında dayanıklı olmalıdır. Yani, sonraki (bir tahminin asimptotik varyansı veya testin gücü ve seviyesi ile tanımlanan) modelde hesaplanan nominal değere yakın olmalıdır.

3. Modelden biraz daha büyük sapmalar, büyük bir yıkıma neden olmamalıdır.

Yukarıda sayılan özellikler ile daha önce belirtilen dayanıklı yöntemlerin kullanım amaçları da dikkate alındığında, dayanıklı bir yöntem, veri kümesinin çoğunluğuna uygun yapıyı tanımlama ve aşırı değerleri teşhis etme özelliklerine de sahip olmalıdır. Sonuç olarak, dayanıklı bir tahmincinin sahip olması gereken özellikler aşağıdaki gibi özetlenebilir: [17]

1. Bu tahminciler, küçük kirlenmelere küçük tepkiler göstermelidir.

2. Büyük kirlenmelerin varlığında bile güvenilir olmalıdır.

3. Herhangi bir miktardaki kirlenmenin etkisi üzerinde bir sınıra sahip olmalıdır.

4. Yuvarlamaya ve gruplamaya düzgün olarak tepki göstermelidir.

5. Veriden açıkça görülen aşırı değerleri ayıklayabilmelidir.

6. Olabildiğince küçük bir varyansa sahip olmalıdır.

---

[15] Robert G. Staudte, Simon J. Sheather, **Robust Estimation and Testing**, New York, John Wiley , 1990, s.56.

[16] Peter J. Huber, **a.g.e.**, s.5.

[17] Frank R. Hampel, "The Influence Curve and Its Role in Robust Esmation", **Journal of the American Statistical Assocation**, 69, 1974, s.383-393.

7. EN KÜÇÜK KIRPILMIŞ KARELER (LTS :LEAST TRİMMED SQUARES)

Yüksek bozulma sınırına sahip tahminciler hem X hem de Y yönündeki aşırı değerlerin varlığı durumunda (çok sayıda olsa bile) güvenilir parametre tahminleri elde etme sorununa çözüm getirmektedirler. Diğer dayanıklı regresyon tahmincileri (M, L ve R tipi tahminciler) Y yönündeki aşırı değerlere karşı direnç gösterseler bile, X yönündeki aşırı değerler (kaldıraç noktaları), bu tahmincileri baskı altına alabilmekte ve yönlendirebilmektedir. Yüksek bozulma sınırı ile ifade edilen husus, ele alınan tahmincinin bozulma sınırının asimptotik olarak %50 sınırına ulaşmasıdır. Bu oran bozulma sınırı için mümkün olan en iyi sonuçtur. Bundan dolayı, %50 bozulma sınırına sahip tahminciler, yüksek bozulma sınırlı tahminciler (HBE: High Breakdown Estimators) olarak adlandırılırlar. Bu tahminci grubunun öncülüğünü Rousseeuw tarafından 1984 de öne sürülen LMS (Least Medyan Squares : En küçük Medyan Kareler) ve LTS (Least Trimmed Squares : En küçük Kırpılmış Kareler) tahmincileri yapmıştır. Daha sonra, S tahminciler, MM tahminciler ve LTA (Least Trimmed Absolute Deviations : En küçük Kırpılmış Mutlak Sapmalar) tahmincileri gibi yüksek bozulma sınırına sahip diğer tahminciler önerilmiştir.

LTS tahmincisi, EKK'ye benzer bir minimizasyon işlemi ile elde edilir.

$$\min_{\hat{\theta}} \sum_{i=1}^{h} \left(r^2\right)_{i:n} \tag{4}$$

Burada, $\left(r^2\right)_{1:n} \le \left(r^2\right)_{2:n} \le ..... \le \left(r^2\right)_{n:n}$ küçükten büyüğe doğru sıralanmış kalıntı kareleridir ve h toplama dahil edilen kalıntı karelerinin sayısıdır. Bu formülün EKK ile tek farkı en büyük kareli kalıntıların toplama işlemi dışında bırakılmasıdır. Bu nedenle, LTS, (n-h) sayıda en büyük kalıntıya sahip gözlemin aşırı değer olarak belirlenip, hesaplamanın dışında bırakılmasıyla bir dayanıklı tahmin elde etme yöntemidir. Bu, en küçük EKK amaç fonksiyonlu h birimden oluşan alt küme bulmayla eş değerdir. LTS tahmini, bu h noktaya uyan EKK tahminidir. "Görülmektedir ki, LTS tahmin sürecinde iyi noktaların büyük çoğunluğu kullanılması gerektiğinden, doğru bir tahmin elde etmek için h sayısı veri kümesindeki iyi noktaların sayısına olabildiğince yakın olmalıdır. Bu gerçekleştirildiğinde, LTS mümkün en iyi tahmini verir." [18] Uygulamada, h'nin değerinin belirlenmesi oldukça zordur ve subjektiftir. Eğer varsayılandan daha fazla veri kirliliği ile karşı karşıya kalınırsa LTS dayanıklılık zaafı gösterir. Bunun aksine, eğer kirlilik varsayılan düzeyden daha az ise, parametre tahminleri doğruluk bakımından sorun çıkarır ve tüm iyi noktalar hesaba katılmadığından dolayı etkinlik zaafı

[18] Hakan Türkay, "Doğrusal Regresyon Modellerinin Robust (Dayanıklı) Yöntemlerle Tahmini ve Karşılaştırmalı Uygulamaları", Yayınlanmamış Doktora Tezi, İstanbul Üniversitesi, 2004, s.89.

gösterirler. "h değerine ve aşırı değerli veri yapısına bağlı olarak LTS oldukça etkin olabilir. Aslında, aşırı değerli noktaların tamamı kırpılırsa, hesaplama açısından EKK'ya eşittir." [19]

Daha önce yapılan simulasyon çalışmalarında, LTS oldukça iyi sonuçlar vermektedir. Örneğin, LTS, LMS ve EKK'nın karşılaştırıldığı bir simulasyon çalışmasında, normal dağılan hata terimlerinde aşırı değerlerin varlığı durumunda LTS'nin diğerlerine göre daha sapmasız parametre tahminleri ürettiği gösterilmiştir. [20] LTS yönteminin amaç fonksiyonu düzgün bir fonksiyondur ve buna bağlı olarak yerel etkilere karşı daha az duyarlıdır. LTS, asimptotik olarak normal dağılıma uyar ve yakınsama oranı $n^{-1/3}$ 'tür. Bu avantajlarına rağmen, LTS'nin hesaplanması oldukça zordur. Bunun dışında, LTS amaç fonksiyonunun matematiksel optimizasyonlara elverişli olmaması ve buna bağlı olarak tahminlerin hesaplanmasının zorluğu önemli sorunlar olarak görülmektedir. Ayrıca, LTS amaç fonksiyonu, sıkı bir dışlama usulü üzerine kurulmuştur. Yani, eldeki bir veri noktası tahmin sürecinde ya tamamen hesaba dahil edilir, yada tamamen sürecin dışında bırakılır. Bu durum, eğer şüpheli bölgelerde veri noktaları var ise, iyi bir yol değildir.

7. 1. LTS Tahmincisinin Hesaplanması

LTS tahmincisinin kesin hesaplamalarında, öncelikle örneklemden h boyutlu alt örneklerin tümü çekilir. Bu, C_n^h kadar alt kümenin elde edilmesi anlamına gelir. Sonra, bu h boyutlu C_n^h tane alt kümenin her biri için regresyon doğrusu uydurulur. Buradan, amaç fonksiyonunda belirtildiği gibi, her bir örnek için hesaplanan kalıntıların en küçük h tanesinin toplamı bulunur. Bu toplamlar arasından minimumu seçilerek LTS çözümü olarak alınır. Burada karşılaşılan en büyük sorun, örneklemden çekilecek alt kümelerin sayısının devasa boyutlara ulaşabilmesidir. Sadece küçük örneklemlerde bu yolla LTS çözümünün elde edilebileceği açıktır. Örneklem boyutu büyüdükçe işlem yoğunluğu ve harcanan zaman çok artacak ve belli bir aşamadan sonra LTS çözümüne ulaşmak imkansız hale gelecektir.

LTS tahmincisi, söz konusu imkansızlıklar nedeniyle kesin algoritma yerine genellikle yaklaşık çözümler veren algoritma ile elde edilir. Progress programında kullanılan yaklaşık algoritma şöyle işlemektedir : Parametre sayısı p boyutlu alt örneklemler çekilerek bunların her biri için bir regresyon doğrusu uydurulur ve bu doğrular yardımıyla bulanan kalıntıların kareleri küçükten büyüğe doğru sıralanarak en küçük h tanesinin toplamı alınır. Bu şekilde elde edilen toplam değerlerinden minimumu veren regresyon doğrusu LTS tahmini olarak belirlenir.

---

[19] Randall E.Schumacker, Michael P. Monahan, Robert E. Mount, "A Comparison of OLS to LTS and MM Robust Regression in S-PLUS", **Southwest Educational Research Association 25th Annual Meeting**, Austin, Texas,2002,s.3.

[20] Latif Öztürk, "Doğrusal Regresyonda Sağlam Kestirim Yöntemleri ve Karşılaştırılmaları" Yayınlanmamış Doktora Tezi, Mimar Sinan Üniversitesi, İstanbul, 2003, s.118.

7.2. LTS Tahmincisinin Özellikleri

LTS tahmincisinin özellikleri aşağıdaki gibi özetlenebilir: [21]

1. LTS tahmincisi için her zaman bir çözüm vardır.

2. LTS tahmincisi, regresyon eş değişim (equivariant), ölçek eş değişim ve afine eş değişim özelliklerine sahiptir.

$$\sum_{i=}^{h}\left[\left(Y_i+\mathbf{x_i}\mathbf{V}\right)-\mathbf{x_i}\left(\boldsymbol{\theta}+\mathbf{V}\right)\right]^2_{i:n}=\sum_{i=}^{h}\left[\left(Y_i-\mathbf{x_i}\boldsymbol{\theta}\right)^2\right]_{i:n} \quad (6)$$

$$\sum_{i=1}^{h}\left[cY_i+\mathbf{x_i}\left(c\theta\right)\right]^2_{i:n}=c^2\sum_{i=1}^{h}\left[\left(Y_i-\mathbf{x_i}\boldsymbol{\theta}\right)^2\right]_{i:n} \quad (7)$$

$$\sum_{i=1}^{h}\left[Y_i-\left(\mathbf{x}_iA\right)\left(\theta A^{-1}\right)\right]^2_{i:n}=\sum_{i=1}^{h}\left[\left(Y_i-\mathbf{x_i}\boldsymbol{\theta}\right)^2\right]_{i:n} \quad (8)$$

Eğer p>1, $h=\left[n/2\right]+\left[(p+1)/2\right]$ ve gözlemler genel durumda ise,(yani, gözlemlerin p tanesi θ için tek bir çözüm veriyorsa) LTS yönteminin bozulma sınırı;

$$\varepsilon^*=\frac{\left[(n-p)/2\right]+1}{n} \quad (9)$$

şeklinde tanımlanır. "Daha büyük h değerleri için, LTS'nin bozulma sınırı, $\varepsilon^* \cong (n-h)/n$ 'dir. Örneğin, h ≅ 0,75n seçildiğinde yöntemin bozulma sınırı $\varepsilon^* \cong 0{,}25$ olacaktır." [22] LTS yöntemi, LMS'de olduğu gibi verilen şartlar altında %50 bozulma sınırına ulaşmaktadır. Genel olarak, h kırpma oranı, α 'ya bağlanır. Örneğin, $h=\left[n(1-\alpha)\right]+1$ olduğunda bozulma sınırına $\varepsilon^*=\alpha$ 'dır.

3. Eğer p>1 ve gözlemler genel durumda ve $y_i=X_i\theta$ 'yı sağlayan gözlemlerin $(n+p-1)/2$ den kesinlikle daha çok olan bazı θ 'lar var ise , o zaman LTS çözümü θ 'ya eşit olur.

4. LTS tahmincisi, normal dağılım durumunda da aynı asimptotik etkinliğe sahiptir. M tahmincileri gibi,

---

[21] Janne Heikkilä , **Robust Regression**, çevrimiçi: http://www.ee.oulu.fi/~jth/robust.pdf, 05.01.2003.

[22] Mia Hubert, Peter J. Rousseeuw, Stefan Van Aelst, "Robustness", ***Encyclopedia of Actuarial Sciences***, Edited by B. Sundt , J. Teugels, John Wiley, New York, 2003, s.6.

$$\psi(t) = \begin{cases} t, & |t| < \Phi^{-1}(1-\alpha/2) \quad ise \\ 0, & \text{diger durumda} \end{cases} \tag{10}$$

şeklinde tanımlanır. Bu tanım, Huber-tipi sıçrayan (skipped) ortalama ile aynıdır.

6. LTS yönteminin asimptotik etkinliği, eğer α = 0,5 (ε^* = % 50) ise sadece %7,2 olarak bulunur. Ancak, h değeri n değerine yaklaştıkça, LTS'nin etkinliği EKK'nın etkinliğine yaklaşacaktır.

8. UYGULAMA

Uygulamada, hata terimlerinin binom dağılıma uyduğu durumda EKK ve robust LTS regresyon tahmincilerinin simulasyon yöntemiyle karşılaştırılması yapılacaktır. Burada amaç; EKK ve robust LTS yöntemlerinden elde edilen parametre tahminlerinin simulasyon sürecinde başlangıç değeri olarak verilen parametre değerlerinden ne kadar sapma gösterdiklerinin ölçülmesidir. Bir başka ifadeyle amaç, Binom dağılımlı hata terimleri durumunda söz konusu regresyon yöntemlerinin hangisinin daha az sapmalı tahminler ürettiğinin gösterilmesidir. Bu amaca yönelik olarak, simulasyon sürecinde kullanılacak olan veriler hipotetik olarak türetilmiştir. Elde edilen veriler ve bu verilerin türetilme şekli aşağıda gösterilmiştir:

1. Düzgün dağılımdan 0 ile 1 aralığında iki tane (r_1 ve r_2) rassal sayının elde edilmesi

2. Elde edilen bu rastsal sayılar Z olarak tanımladığımız aşağıdaki eşitlikte yerine koyularak bir N(0,1) dağılmış değişken değeri elde edilir.

$$Z = \sqrt{(-2\ln(r_1)} \quad \text{Cos}(2\Pi r_2) \tag{11}$$

3. N(0,1) dağılmış söz konusu Z değeri; $U = \mu + \sigma\ Z$ eşitliğinde yerine konularak U rassal değişkeni elde edilebilir.[23]

Bu açıklamalar ışığında elde edilen N(10, 6) dağılmış veriler aşağıdaki gibidir.

[23] An Introduction to Random Number Generation and Simulation, http://dmawww.epfl.ch/benarous/Pmmi/interactive/rng7.htm, 20.04.2004.

Tablo 1: Hipotetik olarak elde edilen yaklaşık olarak N(10,6) dağılmış veriler.

12,16774	11,68551	8,67988	10,70425	7,054214	13,56055	13,16472
13,93235	7,90377	10,30552	10,14960	9,86481	15,85245	11,27039
7,95474	10,92810	9,52312	5,24065	11,01156	5,81505	8,91277
10,47066	3,98880	13,67944	10,46852	7,994942	9,34246	11,91694
12,54917	9,75686	9,84980	9,04269	11,05708	6,54892	11,62940

Min: 3,9888 Ortalama: 10,113642 Medyan: 10,305520
Std, Sapma: 2,575998

8.1. Binom Dağılıma Uyan Hata Teriminin Elde Edilmesi

X, p parametresi ile bir Bernoulli rassal değişkeni olduğunda; X'in dağılım fonksiyonu,

$$F(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \le x < 1 \\ 1, & 1 \le x \end{cases} \qquad (12)$$

şeklinde verilmektedir. Dolayısıyla,

$$Fx^{-1}(u) = \begin{cases} 1, & 1-p < u \\ 0, & dd \end{cases} \qquad (13)$$

şeklinde alınır ve u'nun yerine 1-u koyulursa, p parametresi ile Bernoulli rassal değişkeninin U(0,1) rassal değişkeninden eğer u<p ise 1 ve diğer durumlarda 0 alınarak elde edilir. Bernoulli rassal değişkenini elde edildikten sonra bu rassal değişken kullanılarak Binom rassal değişkeni aşağıdaki gibi elde edilir:[24]

1. Başlangıç değeri olarak X=0 alınır.
2. N tane döngü yapılarak yukarıda elde ettiğimiz B(1,p) rassal değişkeni X değişkenine eklenerek (X=X+B(1,p)) Binom rassal değişkeni elde edilir.

8.2. Simulasyon Uygulamasının Algoritması

1. Anakütle regresyon parametre değerleri simulasyonun başlangıcında $\beta_1 = 1$ ve $\beta_2 = 2$ olarak kabul edilmiştir.
2. Örneklem büyüklüğü n=35 seçilmiş ve her gözlem için normal dağılımdan elde edilen X~$[\mathrm{N}(10,6)]$ değerleri sabitlenmiştir.
3. Hata terimleri (u_i değerleri) rassal olarak binom dağılımından elde edilmiştir. p=0.1, p=0.5 ve p=0.9 için 100 ve 200 döngülü simulasyon uygulaması yapılmıştır.
4. Anakütle katsayı değerleri ve hata terimleri bilindiğine göre,

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (14)$$

[24] An Introduction to Random Number Generation and Simulation , http://dmawww.epfl.ch/benarous/Pmmi/interactive/rng6.htm , 20.04.2004.

eşitliğinin kullanımıyla Y_i değerleri bulunmaktadır.

5. Bulunan bu Y_i değerleri daha önce sabitlenmiş X değerlerine dayandırılarak, tahmin değerleri $\hat{\beta}_1$ ve $\hat{\beta}_2$ elde edilir.
6. Bu tahminciler her defasında aynı $\beta_1 = 1, \beta_2 = 2$, X değerlerinin ve farklı u_i değerlerinin kullanımıyla çok sayıda deneme yapılarak çok sayıda $\hat{\beta}_1$ ve $\hat{\beta}_2$ değeri elde edilir.
7. Çok sayıda elde edilen bu $\hat{\beta}_1$ ve $\hat{\beta}_2$ değerlerinin toplanıp deneme sayısına bölünmesiyle $\overline{\hat{\beta}_1}$ ve $\overline{\hat{\beta}_2}$ değerleri elde edilir. Bu $\overline{\hat{\beta}_1}$ ve $\overline{\hat{\beta}_2}$ değerleri birinci adımda verdiğimiz $\beta_1 = 1, \beta_2 = 2$ değerlerine eşitse tahmin edicinin "sapmasız" olduğunu göstermektedir.
8. Tablolarda verilen sapma miktarları, tahmin değerleriyle $\beta_1 = 1, \beta_2 = 2$ değerleri arasındaki farkın mutlak değeridir. ($|\overline{\hat{\beta}_i} - \beta_i|$)

Bu uygulama, normal dağılmış veriler ile bu verilere bir ve iki aşırı değer (leverage point, X_6=11.68551 yerine $X_6=X_6$+25.00 ve X_{11}=8.679885 yerine $X_{11}=X_{11}$+(-20)) eklendiği durumdaki şekliyle tekrar ele alınarak sonuçları karşılaştırılmıştır. Orijinal veriler ve aşırı değer bulunan veriler karşısında EKK ve LTS regresyon tahmincilerinden elde edilen parametre tahminlerinin sapma miktarları karşılaştırılmıştır. LTS regresyon tahmincisinde h değeri α=0,25 alınarak,

$$h = \left[n(1-\alpha)\right] + 1 \qquad (15)$$

eşitliğinden elde edilmiştir.

Tablo 2**:** 100 Deneme Sonucunda EKK ve LTS Yöntemlerinden Elde Edilen Parametre Tahminlerindeki Sapma Miktarları B(N,p)

	EKK Sapma Miktarı		**LTS Sapma Miktarı**		**Hata Terimlerinin Dağılımı B(N,p)**
	Sabit (β_1=1)	**Eğim (β_2=2)**	**Sabit (β_1=1)**	**Eğim (β_2=2)**	
Orijinal Değerler İçin	3,65956	0,017248	3,28937	0,025903	B (35, 0.1)
	17,35927	0,010299	17,38000	0,001261	B (35, 0.5)
	31,54559	0,004457	31,36990	0,037611	B (35, 0.9)
Tek Aşırı Değer Varlığında	15,42959	1,272161	3,15577	0,026130	B (35, 0.1)
	29,42862	1,273484	17,46000	0,000630	B (35, 0.5)
	43,40972	1,271358	31,63000	0,010965	B (35, 0.9)
İki Aşırı Değer Varlığında	17,82749	1,475443	3,51336	0,000843	B (35, 0.1)
	31,97494	1,480539	17,62350	0,005900	B (35, 0.5)
	45,91887	1,478116	31,5800	0,013247	B (35, 0.9)

Tablo 2'de görüldüğü gibi 100 denemelik simulasyon sürecinde, orijinal veri kümesinde (herhangi bir aykırı değerin bulunmadığı durumda) EKK ve LTS yöntemleri benzer sonuçlar vermiştir. Söz konusu iki yöntemden elde edilen sabit parametre tahminleri başlangıç değerinden büyük sapmalar gözlenirken eğim parametresi tahminlerinde başlangıç değerine oldukça yakın, küçük sapmalar görülmektedir. Eğim parametresi tahminlerinde p=0.1 ve 0.9 için EKK yöntemi daha küçük sapma gösterirken p=0.5 için LTS yöntemi daha küçük sapma göstermiştir. Ancak, tek veya iki aşırı değerin bulunduğu durumlarda, EKK yönteminden elde edilen sonuçlar başlangıç değerinden oldukça uzaktır, yani EKK yönteminden elde edilen hem sabit parametre hem de eğim parametresi tahminleri büyük sapmalar göstermektedir. Buna karşılık, LTS yöntemiyle elde edilen tahminler sabit parametre için EKK yöntemine benzerlik gösterirken eğim parametresi tahmininde orijinal veri kümesinde olduğu gibi küçük sapmalar göstermektedir. Bu durum LTS yönteminin Binom dağılımlı hata terimlerinde de aykırı değerlerden fazlaca etkilenmediğinin ve EKK yönteminin ise tek bir aşırı değerin bile etkisinde kalarak daha fazla sapmalı sonuçlar verdiğinin bir göstergesidir.

Tablo 3: 200 Deneme Sonucunda EKK ve LTS Yöntemlerinden Elde Edilen Parametre Tahminlerindeki Sapma Miktarları B(N,p)

	EKK Sapma Miktarı		**LTS Sapma Miktarı**		**Hata Terimlerinin Dağılımı B(N,p)**
	Sabit (β_1=1)	**Eğim (β_2=2)**	**Sabit (β_1=1)**	**Eğim (β_2=2)**	
Orijinal Değerler İçin	3,60578	0,012348	3,60438	0,009423	B (35, 0.1)
	17,34241	0,010109	17,89660	0,029253	B (35, 0.5)
	31,42134	0,007487	31,48000	0,000000	B (35, 0.9)
Tek Aykırı Değer Varlığında	15,49504	1,280112	3,38500	0,004794	B (35, 0.1)
	29,43881	1,281681	17,84290	0,024552	B (35, 0.5)
	43,33189	1,263859	31,64890	0,016021	B (35, 0.9)
İki Aykırı Değer Varlığında	17,79300	1,469606	3,73168	0,022984	B (35, 0.1)
	31,70959	1,460223	17,89500	0,049518	B (35, 0.5)
	45,77887	1,465779	31,53000	0,001462	B (35, 0.9)

Simulasyonda deneme sayısının 200 çıkarıldığı durumda EKK ve LTS regresyon tahmincilerinden elde edilen parametre tahminlerine ilişkin Tablo 3'de gösterilen sonuçlar incelendiğinde 100 denemeli duruma benzerlik gösterdiği gözlemlenmektedir. Yani, aykırı değer içermeyen orijinal veri kümesinde EKK ve LTS tahminleri birbirine yakınken tek veya iki aykırı değer içerilen durumda eğim parametresi için LTS yönteminin EKK yöntemine göre daha sapmasız tahminler verdiği görülmektedir.

9. SONUÇ

Hata terimlerinin binom dağılıma uyduğu durumda, EKK ve LTS regresyon tahmincilerinin karşılaştırıldığı simulasyon çalışması sonucunda; orijinal veri kümesinde EKK ve LTS yöntemlerinin benzer sonuçlar verdiği, aşırı değerlerle verinin kirletildiği durumda ise LTS tahmincisinin bariz bir şekilde daha az sapmalı parametre tahminleri ürettiği görülmektedir. Buna göre; hata terimlerinin binom dağılıma uyduğu durumda, eğer veri kümesi aşırı değerler içeriyorsa, regresyon parametrelerinin LTS yöntemiyle tahmini daha sağlıklı sonuçlar verecektir. Bu durum dikkate alınarak; çeşitli teşhis araçları kullanılarak bir veya daha fazla aşırı değer tespit edilirse, doğrudan LTS yöntemiyle regresyon parametrelerinin tahmin edilmesi daha uygun bir yaklaşım olacaktır. EKK yönteminden elde edilen kalıntı değerlerinin aşırı değerlerin tespiti konusunda yetersizliği de göz önüne alınırsa; hem aşırı değerlerin tespiti, hem de parametre tahminlerinin sapmasızlığı için LTS yönteminin gerekliliği ortaya çıkmaktadır.

KAYNAKÇA

1. BİLLOR, N.; HADİ, S.A.; VELLEMAN, P.F. (2000), "BACON : Blocked Adaptive Computational Efficient Outlier Nominators", **Computational Statistics and Data Analysis**, Vol.34, s.279-298.

2. HAMPEL, F. R. (1974), "The Influence Curve and Its Role in Robust Esmation", Journal of the American Statistical Assocation, 69.

3. HAMPEL, F.R.; RONCHETTI E.M.; ROUSSEEUW P.J.; STAHEL W.A.(1986), **Robust Statistics**, New York: John Wiley.

4. HEİKKİLÄ,J.(2003),Robust Regression, Erişim: http://www.ee.oulu.fi/~jth/robust.pdf, Erişim Tarihi: 05.01.2003.

5. HU, F.; HU, J. (2000), "A Note On Breakdown Theory For Bootstrap Methods", Statistics and Probability Letters, 50, s.49-53.

6. HUBER, Peter J. : **Robust Statistics**, New York , John Wiley, 1981.

7. HUBERT, M.; ROUSSEEUW, P.J.; VAN AELST, S.(2003), "Robustness", Encyclopedia of Actuarial Sciences, Edited by B. SUNDT, J. TEUGELS, New York: John Wiley.

8. MATOUŠEK, J.; MOUNT, D.M.; NATENYAHU, N.S.(1998), "Efficient Randomized Algorithms For The Repeated Median Line Estimator", Algorithmica, 20.

9. ÖZTÜRK, L.(2003), "Doğrusal Regresyonda Sağlam Kestirim Yöntemleri ve Karşılaştırılmaları", Yayınlanmamış Doktora Tezi, Mimar Sinan Ün.

10. ROUSSEEUW, P.J.; LEROY, A.M.(1987), **Robust Regression and Outlier Detection**, New York: John Wiley.

11. ROUSSEEUW, P.J.; VAN AELST, S. (1999), "Positive-Breakdown Robust Methods in Computer Vision" Computing Science and Statistics, Vol:31, s. 451-460.

12. ROUSSEEUW, P.J.; VAN DRİESSEN, K.(1999), "Computing LTS Regression for Large Data Sets" (Düzenlenmiş), Technical Report, University of Antwerp.

13. SCHUMACKER, R.E.; MONAHAN, M.P.; MOUNT R.E.(2002), "A Comparison of OLS to LTS and MM Robust Regression in S-PLUS", **Southwest Educational Research Association 25th Annual Meeting**, Austin, Texas,2002.

14. STAUDTE, R.G.; SHEATHER S.J.(1990), **Robust Estimation and Testing**, New York : John Wiley.

15. TÜRKAY, H. (2004), "Doğrusal Regresyon Modellerinin Robust (Dayanıklı) Yöntemlerle Tahmini ve Karşılaştırmalı Uygulamaları", Yayınlanmamış Doktora Tezi, İstanbul Üniversitesi.

16. ZUO,Y.(2001), "Some Qantitative Relationships Between Two Types of Finite Sample Breakdown Point", Statistics and Probability Letters, vol.51, s.369-375.

17. _____An Introduction to Random Number Generation and Simulation, Erişim:http://dmawww.epfl.ch/benarous/Pmmi/interactive/rng6.htm , Erişim Tarihi : 20.04.2004.

18. _____An Introduction to Random Number Generation and Simulation, Erişim:http://dmawww.epfl.ch/benarous/Pmmi/interactive/rng7.htm , Erişim Tarihi : 20.04.2004.