



Dayanıklı (Robust) Regresyon: Karşılaştırmalı Simülasyon Çalışması

Yasin Büyükkör^{1*}, Ali Kemal Şehirlioğlu²

¹ Karamanoğlu Mehmetbey Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Karaman, Türkiye (ORCID: 0000-0002-1006-0539)

² Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, İzmir, Türkiye (ORCID: 0000-0001-5190-6740)

(İlk Geliş Tarihi 22 Ocak 2020 ve Kabul Tarihi 22 Şubat 2020)

(DOI: 10.31590/ejosat.678710)

ATIF/REFERENCE: Büyükkör, Y. & Şehirlioğlu, A. K. (2020). Dayanıklı (Robust) Regresyon: Karşılaştırmalı Simülasyon Çalışması. *Avrupa Bilim ve Teknoloji Dergisi*, (18), 188-195.

Öz

Günümüzde birçok araştırmacı regresyon analizinde hata teriminin dağılışının Gausyan (Normal) olduğunu varsaymakta ve regresyon parametrelerinin tahminini yaparken için En Küçük Kareler (EKK) yöntemini kullanmaktadır. Ancak uygulamada normal dağılış varsayımı kabul edilse bile artıklar genellikle normal dağılıştan farklı bir dağılış göstermektedirler. Özellikle veri setinde bulunan sapan gözlemler veya sapan gözlem olduğundan şüphelenilen gözlemler, verilerin normallik varsayımını bozmakta ve EKK yöntemi ile yapılan parametre tahminleri hatalı (sapmalı) olmaktadır. Araştırmacılar böyle durumların üstesinden gelebilmek için son yıllarda sıklıkla kullanılan dayanıklı (robust) yöntemleri kullanmaktadırlar. Bu yöntemlerin arasında en çok kullanılan M- tahminciler (En Yüksek Olabilirlik tipi) gelmektedir. M- tahminleme yöntemi, En Çok Olabilirlik (MLE) yönteminin geliştirilmiş bir versiyonudur ve EKK yöntemi de bir M- tahminci olarak bilinmektedir. M- tahminleme yöntemi, eldeki veri setine uygun bir amaç fonksiyonunu minimize ederek parametre tahminlerini iteratif olarak elde etmektedir. Bu çalışmada farklı senaryolar ele alınarak EKK yöntemi, Huber M- tahminleme yöntemi ve Tukey Bisquare M- tahminleme yöntemi karşılaştırılmıştır. Ayrıca bu yöntemlerin amaç, etki ve ağırlık fonksiyonları incelenmiştir. Regresyon parametreleri tahminlenirken İteratif Olarak Tekrar Ağırlıklandırılan En Küçük Kareler (IRWLS) yöntemi kullanılmıştır. IRWLS yönteminde bir başlangıç çözümü uygun bir tahminleme yöntemiyle seçilir (Örn: EKK) ve M- tahminleme yöntemlerinin ağırlık fonksiyonları kullanılarak Ağırlıklı EKK yöntemiyle iteratif olarak parametre tahminleri elde edilir. Elde edilen parametre tahminleri Ortalama Karesel Hata (MSE), Sapma ve R² kriterleri açısından karşılaştırılmıştır. Eğer veri seti normal ise en kullanışlı yöntem EKK iken veri setinde kirlenme (contaminated) veya sapan gözlem olduğunda EKK yönteminin etkinliğini kaybettiği görülmüştür. Özellikle açıklanan değişken Y yönünde sapan gözlem olduğunda Huber ve Tukey M- tahminleme yöntemleri EKK'ya göre daha iyi sonuçlar vermektedir.

Anahtar Kelimeler: Dayanıklı (Robust) Regresyon, M- Tahminciler, Sapan Gözlem, İteratif Olarak Tekrar Ağırlıklandırılmış En Küçük Kareler

Robust Regression: A Comparative Simulation Study

Abstract

Today, many researchers assume that the distribution of the error term is Gaussian (Normal) in regression analysis and uses the Ordinary Least Squares (OLS) method to estimate the regression parameters. However, in practice, even if the distribution of errors is assumed to be normal, residuals are not generally normally distributed. Especially, if the data contains outliers or there are observations which

* Sorumlu Yazar: Karamanoğlu Mehmetbey Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Karaman, Türkiye, ORCID: 0000-0002-1006-0539, yasinbuyukkor@kmu.edu.tr

suspected to be outlier, the assumption of normality is violated and parameter estimates which made using the OLS, are biased. Researchers use robust methods to overcome when such problems occur. Among these methods, M-estimators (Maximum Likelihood Type) are the most used. The M- estimation method is a generalized version of the Maximum Likelihood (MLE) estimation method, and the OLS method is also known as an M-estimator. In the M-estimation method, it minimizes a objective function suitable for the data set and obtains parameter estimates iteratively. In this study, the OLS method, Huber M- estimation method and Tukey Bisquare M- estimation method were compared using different scenarios. In addition, the Objective, Influence and Weight functions of these methods were examined. Iteratively Re-Weighted Least Squares (IRWLS) method is used for parameter estimation. When using IRWLS method, an initial solution is selected by an appropriate estimation method (eg. OLS) and iterative solution is obtained by using the Weighted OLS method with the weight functions of the M-estimation methods. By comparing the obtained parameter estimates, Bias, Mean Squared Error (MSE) and R^2 criterias are used. If the data set is normal, the most useful method is OLS, whereas the OLS method has lost its efficiency when there is contaminated distribution or outliers in the data set. Huber and Tukey M- estimation methods give better results than OLS, especially when there is outlier in the Y direction.

Keywords: Robust Regression, M- Estimators, Outliers, Iteratively Re-Weighted Least Squares

1. Giriş

Birçok regresyon çalışmasında araştırmacılar genellikle ellerindeki veri setinin Gausyan (normal) dağılışa sahip olduğunu varsayarak parametre tahmini yapmaktadırlar. Ancak gözlemlerde gözlemlerin ortalamasından çok uzakta (sapan) değerler olduğunda veri setinin dağılışı normal dağılıştan farklılaşır. Örneğin Aksaraylı ve Pala (2018) çalışmalarında buna değinmiştir. Böyle durumlarda geleneksel yöntemlerle analiz yapmak tahmin sonuçlarını hatalı olarak verecektir (Hampel vd, 1986).

Veri setinde bulunan anormal durumlar genellikle;

- Ölçüm
- Kayıt ve
- Verileri bir yerden başka bir yere aktarırken ortaya çıkan hatalardan kaynaklanmaktadır.

Ancak bazı durumlarda veri setine ait doğal bir gözlem veya gözlemler bile sapan (outlier) gözlem olarak algılanabilir. Araştırmacı bu gözlemlerin analizini doğru yapmazsa, bu gözlemleri veri setinden dışlayabilir. Gözlemlerin özel durumlar haricinde dışlanması veya analizden tamamen çıkarılması parametre tahminleri ve modelin yorumlanması açısından hatalı sonuçlar ortaya çıkaracaktır. Özellikle regresyon analizinde sapan gözlem veya sapan gözlemlerin bulunması durumunda analiz yapılırken EKK yöntemini kullanmak sonuçların sapmalı çıkmasına neden olacaktır. Sapan gözlemlerin varlığı durumunda araştırmacılar bu gözlemlerden daha az etkilenen dayanıklı (robust) tahminleme yöntemlerine başvurmuşlardır. En sık kullanılan robust tahminleme yöntemi Huber (1964) tarafından geliştirilen M- Tahminleme (M- Estimators) yöntemidir. M- Tahminleme yöntemi En Çok Olabilirlik (MLE) yönteminin geliştirilmiş ve sapan gözlemlere dayanıklı versiyonudur (Stuart, 2011; Rousseeuw and Leroy, 1987; Andersen 2008).

Hata teriminin Olasılık Yoğunluk Fonksiyonu (OYF) $f(\varepsilon_i)$ olarak alınırsa β parametre vektörünün En Çok Olabilirlik (MLE) fonksiyonu;

$$\prod f(\varepsilon_i) = \prod f(y_i - x_i^T \beta) \quad (1)$$

yazılabilir. Eğer hataların dağılışı normal ise;

$$\sum \varepsilon_i^2 = \sum (y_i - x_i^T \beta)^2 \quad (2)$$

fonksiyonu minimize edilir. Ancak hataların dağılışı normal dağılıştan farklı olduğunda minimize edilecek fonksiyon uygun bir amaç fonksiyonuyla değiştirilerek parametre tahminleri yapılır. Eşitlik 2’de verilen amaç fonksiyonu aynı zamanda EKK yöntemi olarak bilinmektedir.

Çalışmada 2. bölüm genel olarak sıkça kullanılan M- Tahmincileri tanıtılacak ve son bölümde EKK, Huber ve Tukey M- Tahmincileri farklı simülasyonlar kullanılarak karşılaştırma yapılacaktır.

2. M- Tahminciler

EKK yöntemi, hata teriminin dağılışının normal olduğu varsayımıyla olabilirlik fonksiyonunun (Hata Kareler Toplamının) minimize edilmesiyle elde edilir. M- Tahminciler aynı fikirle hataların dağılışının normal olmadığı durumlarda (çarpık, kirlenmiş, basık, uzun kuyruklu vb.) farklı bir fonksiyon kullanarak MLE tahmini yapmaktadırlar. M- Tahminciler;

$$\sum \rho(\varepsilon_i) = \sum \rho(y_i - x_i^T \beta) \quad (3)$$

olarak gösterilerbilir. Burada $\rho(\varepsilon_i)$ amaç fonksiyonu (minimize edilecek fonksiyon) olarak adlandırılır ve sürekli ve türevlenebilir bir fonksiyondur. Minimize edilecek fonksiyon;

$$\min \sum \rho(r_i) = \min \sum \rho\left(\frac{\varepsilon_i}{s}\right) \quad (4)$$

ile gösterilir. Burada s standart sapmanın tahminidir ve genellikle medyana dayalı bir tahmindir. En çok kullanılan standart sapma tahmini;

$$s = MAD/0.6745 = \frac{\text{medyan}|\varepsilon_i - \text{medyan}(\varepsilon_i)|}{0.6745} \quad (5)$$

yazılır (Draper and Smith 2014: 572). Standart sapma tahmininin paydasında bulunan 0.6745 katsayısı ise veriler normal dağılışa sahip olduğunda σ ile MAD'ın eşit olmasını sağlamaktadır (Hogg, 1979). Eşitlik 4 parametrelerine göre türevlenirse;

$$\sum x_{ij} \psi\left(\frac{y_i - x_i^T \beta}{s}\right) = 0 \quad (6)$$

elde edilir. Burada $\psi(r_i) = \partial \rho(r) / \partial r$ 'dir ve Etki (Influence) Fonksiyonu olarak bilinir. Beaton ve Tukey (1974)' e göre ağırlık fonksiyonu;

$$w(r) = \frac{\psi(r)}{r} \quad (7)$$

dir. Genellikle dayanıklı (robust) regresyon analizinde Huber ve Tukey M- tahminciler yaygın olarak kullanılmaktadır. Hesaplama kolaylığı ve matematiksel olarak nispeten daha anlaşılır olması bu iki tahminleme yöntemini diğer yöntemlere göre popüler hale getirmiştir.

2.1. Huber M- Tahminciler

Huber (1964) tarafından geliştirilen M- Tahminleme yöntemi için amaç fonksiyonu;

$$\rho(r) = \begin{cases} \frac{1}{2} r^2 & , |r| < k \\ k|r| - \frac{1}{2} k^2 & , |r| \geq k \end{cases} \quad (8)$$

olarak bilinmektedir. Bu fonksiyonda bulunan k sabiti dönüm noktası (tuning constant) olarak adlandırılır ve değeri 1.345 olarak kullanılır. Burada k 'nın amacı eğer veriler gerçekten normal dağılışa sahipse kullanılan fonksiyonun etkinliğinin (efficiency) yaklaşık %95 olmasını sağlamaktır. Huber M- Tahmincilerin en önemli özelliği, fonksiyon $(-k, k)$ aralığında normal gibi davranırken diğer yerlerde Laplace (Double Exponential) dağılışı gibi davranmaktadır.

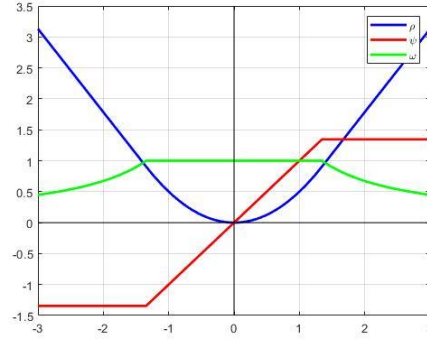
Huber M- tahminci için Etki fonksiyonu ve Ağırlık fonksiyonu;

$$\psi(r) = \begin{cases} r & , |r| < k \\ k \text{sign}(r) & , |r| \geq k \end{cases} \quad (9)$$

$$w(r) = \begin{cases} 1 & , |r| < k \\ \frac{k}{|r|} & , |r| \geq k \end{cases} \quad (10)$$

ile gösterilir. Amaç, Etki ve Ağırlık fonksiyonlarına ait grafik Şekil 1'de verilmiştir

Şekil 1. Huber Amaç, Etki ve Ağırlık Fonksiyonu



2.2. Tukey Biquare (Biweight) M- Tahminçiler

Beaton ve Tukey (1974) tarafından geliştirilen Biquare (Biweight) amaç fonksiyonu;

$$\rho(r) = \begin{cases} \frac{k^2}{6} \left\{ 1 - \left[1 - \left(\frac{r}{k} \right)^2 \right]^3 \right\} & , |r| < k \\ \frac{k^2}{6} & , |r| \geq k \end{cases} \quad (11)$$

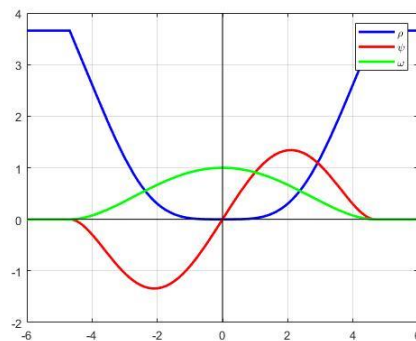
yazılabilir. Huber M- tahminçide olduğu gibi k dönüm noktası olarak bilinir ve değeri 4.685 olarak belirlenmiştir. Tukey Bisquare Etki ve Ağırlık fonksiyonu;

$$\psi(r) = \begin{cases} r \left[1 - \left(\frac{r}{k} \right)^2 \right]^2 & , |r| < k \\ 0 & , |r| \geq k \end{cases} \quad (12)$$

$$w(r) = \begin{cases} \left[1 - \left(\frac{r}{k} \right)^2 \right]^2 & , |r| < k \\ 0 & , |r| \geq k \end{cases} \quad (13)$$

ile gösterilir. Tukey Bisquare M- tahminci, dağılımın kuyruklarında bulunan gözlemlere 0 ağırlığını vermesi ve böylece bu gözlemleri analiz yapılırken etkisiz hale getirmesidir. Amaç, Etki ve Ağırlık fonksiyonlarına ait grafik Şekil 2’de verilmiştir.

Şekil 2. Tukey Bisquare Amaç, Etki ve Ağırlık Fonksiyonu



3. Simülasyon Çalışması

Simülasyon çalışmasında hatalar farklı özelliklere sahip dağılımlar kullanılarak türetilmiş ve tahminleme yöntemlerinin performansları değerlendirilmiştir. Karşılaştırma kriterleri olarak Sapma, Ortalama Karesel Hata (OKH) ve açıklama yüzdesi R^2 kullanılmıştır. Her bir simülasyon örnek hacimleri 30 ve 100 olacak şekilde 1000 tekrar üzerinden gerçekleştirilmiştir.

Uygulamada açıklayıcı değişken X değerleri $(-5,5)$ aralığında Uniform dağılımdan türetilmiştir. Gerçek regresyon parametreleri $\beta_0 = \beta_1 = 1$ olarak belirlenmiş ve böylece regresyon modeli;

$$Y = 1 + X + \varepsilon \quad (14)$$

Simülasyon çalışmasında kullanılan senaryolar;

- **Senaryo 1:** Hataların dağılışı standart normal dağılışı iken $\varepsilon \sim N(0,1)$
- **Senaryo 2:** Hataların dağılışı 5 serbestlik dereceli Student t dağılışı iken $\varepsilon \sim t(5)$
- **Senaryo 3:** Hataların dağılışının %80'i standart normal ve %20'si ortalaması 0 varyansı 10 olan normal dağılışı iken $\varepsilon \sim 0.8N(0,1) + 0.2N(0,10)$
- **Senaryo 4:** Hataların dağılışının %60'ı standart normal ve %40'ı ortalaması 0 varyansı 10 olan normal dağılışı iken $\varepsilon \sim 0.6N(0,1) + 0.4N(0,10)$
- **Senaryo 5:** Y yönünde bir adet sapan gözlem (standart normal dağılışa sahip gözlemlere 10 değeri eklenmiştir.)
- **Senaryo 6:** Y yönünde %20 oranında sapan gözlem (standart normal dağılışa sahip gözlemlere %20 oranında 10 değeri eklenmiştir.)

M- Tahmincilerle yapılan parametre tahminleri için İteratif Olarak Tekrar Ağırlıklandırılan En Küçük Kareler (IRWLS) yöntemi kullanılmıştır. Bu yöntemin algoritması;

1. Öncelikle EKK ile regresyon parametrelerinin tahminleri elde edilir.

$$b = (X'X)^{-1} X'Y$$

2. Elde edilen parametre tahminleri kullanılarak artıklar elde edilir.

$$e = Y - Xb$$

3. Artıklar kullanılarak standart sapmanın tahmini s (MAD) hesaplanır.

$$s = \frac{MAD}{0.6745} = \frac{\text{medyan}|e_i - \text{medyan}(e_i)|}{0.6745}$$

4. Elde edilen artıklar ve standart sapma kullanılarak standardize artıklar elde edilir.

$$r = e/s$$

5. Huber veya Tukey ağırlık fonksiyonları kullanılarak her bir artık için ağırlıklar hesaplanır.

6. Ağırlıklar kullanılarak parametre tahminleri Ağırlıklı EKK ile yapılır.

$$b = (X'WX)^{-1} X'WY$$

7. Yakınsama gerçekleşene kadar 2-7 arası tekrar edilir. Yakınsama kriteri;

$$\frac{\|b^i - b^{i-1}\|}{\|b^i\|} < 10^{-6}$$

Sapma ve OKH kriterlerini hesaplamak için;

$$OKH = \frac{1}{1000} \sum_{i=1}^{1000} (b - \beta)^T (b - \beta)$$

$$Sapma = \left| \frac{\sum_{i=1}^n b}{1000} - \beta \right|$$

eşitlikleri kullanılmıştır.

Tablo 1. Senaryo 1 için sonuçlar

Örnek Hacmi	Tahminleme Yöntemi	b ₀	b ₁	Sapma b ₀	Sapma b ₁	OKH (MSE)	R ²
n=30	EKK	0,910	1,054	<u>0,085</u>	<u>0,045</u>	<u>0,015</u>	0,892
	Huber	0,852	1,065	0,150	0,063	0,026	<u>0,908</u>
	Tukey	0,849	1,061	0,148	0,060	0,026	<u>0,908</u>
n=100	EKK	0,915	1,108	<u>0,080</u>	<u>0,035</u>	<u>0,015</u>	0,903
	Huber	0,878	1,083	0,120	0,085	0,022	<u>0,924</u>
	Tukey	0,869	1,080	0,128	0,082	0,023	<u>0,927</u>

Hataların standart normal dağılım gösterdiği durumda EKK yöntemi daha düşük Sapma ve MSE değerine sahipken daha düşük R² değerine sahiptir. Ancak bu değer diğer tahminleme yöntemleriyle karşılaştırıldığında önemsiz kalmaktadır. Örnek hacmi büyüdükçe sapma ve MSE daha azalmış buna karşılık açıklama oranı artmıştır.

Tablo 2. Senaryo 2 için sonuçlar

Örnek Hacmi	Tahminleme Yöntemi	b ₀	b ₁	Sapma b ₀	Sapma b ₁	OKH (MSE)	R ²
n=30	EKK	0,502	0,962	0,498	0,038	0,522	0,780
	Huber	0,549	0,978	0,453	0,023	0,206	0,835
	Tukey	0,572	0,994	<u>0,431</u>	<u>0,009</u>	<u>0,186</u>	<u>0,841</u>
n=100	EKK	0,870	1,047	0,130	0,047	0,453	0,856
	Huber	0,839	1,039	<u>0,159</u>	<u>0,040</u>	<u>0,027</u>	0,889
	Tukey	0,829	1,045	0,169	0,045	0,031	<u>0,901</u>

Hatalar daha uzun kuyruklu olan t(5) dağılımı gösterdiğinde EKK yönteminin etkinliği azalmakta ve dayanıklı yöntemlerin etkinlikleri artmaktadır. Örnek hacmi 30 olduğunda Tukey daha iyi sonuçlar verirken örnek hacmi 100 olduğunda Huber tahminleme yöntemi daha düşük sapma ve MSE değerleri vermiştir. Açıklama yüzdelere bakıldığında aralarında çok az bir fark olduğu görülmektedir.

Tablo 3. Senaryo 3 için sonuçlar

Örnek Hacmi	Tahminleme Yöntemi	b ₀	b ₁	Sapma b ₀	Sapma b ₁	OKH (MSE)	R ²
n=30	EKK	0,702	0,663	0,301	0,428	0,865	0,163
	Huber	0,866	0,983	0,149	0,029	0,025	0,745
	Tukey	0,931	1,026	0,098	0,014	0,014	0,935
n=100	EKK	1,587	1,116	0,210	0,100	0,651	0,361
	Huber	0,997	1,082	0,008	0,083	0,007	0,761
	Tukey	0,900	1,058	0,085	0,060	0,012	0,943

Artıkların %80'i standart normal dağılış gösterip %20'si ortalaması 0 varyansı 10 olan normal dağılıştan geldiğinde (contaminated normal) küçük örnek hacminde Tukey daha iyi performans gösterirken, büyük örnek hacminde Huber ve Tukey benzer davranmışlardır. Her iki örnek hacmi için Tukey daha yüksek açıklama yüzdesine sahiptir. Bunun muhtemel nedeni ise Tukey artıkların değeri 4.695 değerinin aştığında ağırlıklara 0 değerine vermektedir. EKK ise dağılış bozulmuş (contaminated) olduğunda daha büyük sapma ve MSE değerleri vermektedir.

Tablo 4. Senaryo 4 için sonuçlar

Örnek Hacmi	Tahminleme Yöntemi	b ₀	b ₁	Sapma b ₀	Sapma b ₁	OKH (MSE)	R ²
n=30	EKK	0,332	1,481	0,803	0,324	0,765	0,369
	Huber	0,330	1,176	0,711	0,190	0,549	0,495
	Tukey	0,698	0,994	0,545	0,037	0,416	0,803
n=100	EKK	0,825	1,134	0,175	0,134	0,935	0,225
	Huber	0,991	1,091	0,013	0,092	0,009	0,568
	Tukey	1,018	1,066	0,012	0,068	0,005	0,896

Artıkların dağılışı artık ayırt edilemez hale geldiğinde (%60'ı standart normal dağılıırken %40'ı ortalaması 0 varyansı 10 olan normal dağılıştan gelmektedir) EKK her iki durumda da düşük açıklama yüzdesine sahiptir. Dağılış büyük oranda bozulmaya uğradığı için veriler artık daha kararlı hale gelmektedir. Tukey her iki durumda da daha iyi sonuçlar göstermektedir.

Tablo 5. Senaryo 5 için sonuçlar

Örnek Hacmi	Tahminleme Yöntemi	b ₀	b ₁	Sapma b ₀	Sapma b ₁	OKH (MSE)	R ²
n=30	EKK	1,855	1,420	0,996	0,422	0,853	0,492
	Huber	0,962	1,011	0,018	0,019	0,003	0,858
	Tukey	0,879	0,988	0,101	0,008	0,012	0,945
n=100	EKK	1,298	1,200	0,341	0,213	0,547	0,662
	Huber	1,053	1,054	0,058	0,056	0,007	0,914
	Tukey	1,033	1,048	0,039	0,051	0,004	0,940

5. senaryoda hatalar normal dağılıştan türetilmişken üretilen sayılardan sadece biri deđiştirilerek 1 adet diđer verilerden çok yüksek (gross error, 30) bir veri eklenmiştir. Bu durumda tahminleme yöntemlerinden EKK büyük sapma ve MSE deđerleri üretmiştir. Küçük örnek hacminde MSE açısından Huber daha iyiyken, açıklama yüzdesi açısından Tukey daha iyidir. Büyük örnek hacminde ise sapma, MSE ve açıklama yüzdesi açısından Tukey daha iyi sonuçlar göstermektedir.

Tablo 6. Senaryo 6 için sonuçlar

Örnek Hacmi	Tahminleme Yöntemi	b ₀	b ₁	Sapma b ₀	Sapma b ₁	OKH (MSE)	R ²
n=30	EKK	2,646	1,020	1,659	0,655	0,951	0,430
	Huber	1,549	1,093	0,585	0,100	0,357	0,736
	Tukey	1,198	1,079	0,240	0,091	0,075	0,901
n=100	EKK	2,653	0,997	1,285	0,923	0,866	0,383
	Huber	1,494	1,126	0,531	0,128	0,304	0,736
	Tukey	1,161	1,122	0,193	0,129	0,059	0,913

6. senaryoda hatalar normal dağılıştan türetilmiş ancak üretilen sayıların %20'si (örnek hacmi 30 için 6 ve 100 için 20 adet) diđer verilerden yüksek ve sabit 10 deđeri ile deđiştirilmiştir. EKK yöntemi kötü sonuçlar verirken her iki örnek büyüklüğünde Tukey M- en iyi sonuçları vermiştir.

4. Sonuç

Çalışmada farklı tahminleme yöntemleri karşılaştırılmıştır. Karşılaştırılan tahminleme yöntemleri için;

- Artıkların dağılışı tahminleme yöntemi seçerken çok önemlidir.
- Eğer artıklar normal dağılışı gösteriyorsa en kolay ve en hızlı yapılan EKK yöntemi tercih edilmelidir.
- Eğer artıkların asıl dağılışının normal olduđu biliniyor ve verilerde sapan gözlem(ler) varsa Huber veya Tukey tarafından geliştirilen M- tahminleme yöntemlerine başvurulabilir.
- Bu yöntemler işlem süresi bakımından EKK'dan daha fazla sürmektedir.

Eđer artıklar normal dağılmıyorsa, bu yöntemler genellikle düşük etkinliğe sahiptir.

Kaynakça

- Aksaraylı, M., & Pala, O. (2018). A polynomial goal programming model for portfolio optimization based on entropy and higher moments. *Expert Systems with Applications*, 94, 185-192.
- Andersen, R. (2008). *Modern methods for robust regression* (No. 152). Sage.
- Beaton, A. E., ve Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16(2), 147-185.
- Draper, N. R., ve Smith, H. (2014). *Applied regression analysis* (Vol. 326). John Wiley & Sons.
- Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) *Robust statistics. The approach based on influence functions*. Wiley, New York.
- Hogg, R. V. (1979). Statistical robustness: One view of its use in applications today. *The American Statistician*, 33(3), 108-115.
- Huber, Peter J. Robust Estimation of a Location Parameter. *Ann. Math. Statist.* 35 (1964), no. 1, 73--101. doi:10.1214/aoms/1177703732.
- Rousseeuw, P. J., ve Leroy, A. M. (1987). *Robust regression and outlier detection* (Vol. 1). New York: Wiley.
- Stuart, C. (2011). *Robust regression*. Department of Mathematical Sciences, Durham University, 169.