Volume: 08 Number: 03 Category: 02 Page: 45-53 ISSN No: 1303-6319 Cilt: 08 Sayı: 03 Kategori: 02 Sayfa: 45-53

BASİT DOĞRUSAL REGRESYONDA SAĞLAM VE THEIL KESTİRİCİLERİNİN KARŞILAŞTIRILMASI

Onur TOKA* **Meral ÇETİN**** **Serpil AKTAŞ ALTUNAY*****

ÖZET

Basit doğrusal regresyon çözümlemesinde varsayımlar sağlandığında elde edilen en küçük kareler(EKK) kestirim değerleri, Gauss-Markov teoremine göre parametrelerinin doğrusal, yansız ve en küçük varyanslı kestiricileridir. Ancak, varsayımlar sağlanmadığında elde edilen kestiriciler, olması gereken özelliklerini kaybederler. Bu durumda regresyon çözümlemesinde alternatif olarak bazı parametrik olmayan ya da sağlam yöntemler kullanılır. Parametrik olmayan regresyon yöntemleri, hata dağılımına ilişkin normallik varsayımı gerektirmezler. Theil kestirimi, örneklemde aykırı değerler bulunduğunda sağlam sonuçlar vermektedir (Nevitt, J., Tam, H. P., 1998). Sağlam regresyon ise yine regresyon çözümlemesinde gerekli olan varsayımlar sağlanmadığında, olağan en küçük kareler yöntemine alternatif olarak önerilen bir yöntemdir. Bu çalışmada, basit doğrusal regresyonda aykırı değer varlığında sağlam regresyon yöntemler olan ortanca en küçük kareler(OEKK), en küçük kesilmiş kareler(EKKK), Huber'in M kestiricisi ve parametrik olmayan Theil regresyonu yöntemini karşılaştırmak için bir benzetim çalışması yapılmıştır. Elde edilen sonuçlar göreli etkinliklerine göre karşılaştırılmış ve yorumlanmıştır. Benzetim sonuçlarına göre sağlam yöntemden elde edilen kestiricilerin, Theil yönteminden elde edilen kestiricilere göre daha etkin sonuçlar verdiği görülmüştür.

Anahtar Kelimeler: Aykırı değer, Basit doğrusal regresyon, Robust regresyon, Theil yöntemi.

1. GİRİŞ

Birçok bilim insanı kendi çalışmalarında daha sağlam sonuçları bulduğunu gösterebilmek adına uygulamalı istatistiği sıklıkla kullanır. Eğer iki değişken arasında bir bağ olduğunu, bu değişkenlerden birinin diğerini açıklamada etkili olduğunu düşünüyorsa, en çok kullandıkları uygulamalı istatistik alanı ise regresyondur. Regresyon, bağımlı değişkenin, bir bağımsız değişken tarafından açıklanmasını oluşturan bir model olarak gösterildiğinde basit doğrusal regresyon olarak tanımlanmaktadır. Basit doğrusal regresyon çözümlemesinde varsayımlar sağlandığında elde edilen tahmin değerleri, Gauss-Markov teoremine göre parametrelerinin doğrusal, yansız ve en küçük varyanslı kestiricileridir. Ancak, varsayımlar sağlanmadığında elde edilen kestirimler sağlaması gereken özellikleri kaybederler. Bu durumlarda parametrik olmayan ve sağlam (robust) yöntemler kullanılabilmektedir. Sağlam regresyon yöntemlerinden ortanca en küçük kareler (OEKK), en küçük kesilmiş kareler (EKKK), Huber'in M kestiricisi kullanılmıştır. OEKK, Rousseeuw ve Leroy (1987) tarafından açıklanan hem x hem de y'lerdeki aykırı değerlere karşı sağlam bir kestiricidir. EKKK'da yine Rousseeuw ve Leroy (1987) tarafından artıkların karelerini belli bir fonksiyon altında minimize ederek aykırı değerlere karşı sağlam tahminler elde etmeye çalışan kestiricidir.

*Arş. Gör., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: onur.toka@hacettepe.edu.tr
**Doç. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: meral@hacettepe.edu.tr
***Doç. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: spxl@hacettepe.edu.tr

Huber (1973), M kestiricisini, artıkların bir fonksiyonunu minimize ederek belli bir denklem sistemiyle özellikle y yönündeki aykırı değerlere dayanıklı olan sağlam bir yöntem olarak tanımlanmıştır. Theil yöntemi ise sağlam kestiricilere alternatif olarak sunulmuştur. Basit doğrusal regresyonun eğim olarak da nitelendirilen β_1 katsayısı kestirimi için Sievers (1978) ve Scholz (1978), Randles ve Wolfes (1979), Dietz (1987), Nevitt ve Tam (1998), çeşitli eğim formülleriyle aykırı değerlere dayanıklı ağırlıklandırmalar geliştirmişlerdir. Bu çalışmada, basit doğrusal regresyonda aykırı değer varlığında sağlam regresyon yöntemleri ve parametrik olmayan Theil regresyonu yöntemini karşılaştırmak için bir benzetim çalışması yapılmıştır.

2. SAĞLAM REGRESYON

Basit doğrusal regresyon denklemi aşağıdaki gibi tanımlanır:

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad (1)$$

Burada y bağımlı değişkeni; x, bağımsız değişkeni tanımlarken e_i ise hata terimidir. β_0 ve β_1 değerleri, sırasıyla kesim noktasını ve eğimi vermektedir. Bu durumda eşitlik (1)'de verilen modelde yer alan katsayıların kestirimi için klasik regresyon çözümlemesine alternatif olarak bazı parametrik olmayan ya da sağlam(robust) yöntemler kullanılır. Sağlam regresyon yöntemi, hata terimi normal dağılmayan ve/veya aykırı değerler modeli etkilediği zamanlarda EKK yöntemine alternatif olarak kullanılmaktadır (Candan, 1995).

2.1 Ortanca En Küçük Kareler Kestirimi (Least Median Squares Estimation)

Ortanca en küçük kareler(OEKK) kestirimi, aykırı değerlerin ortaya çıkartılması için kullanılan sağlam bir yöntemdir. Rousseeuw (1984) tarafından önerilen bu yöntem, artık karelerinin toplamı yerine artık karelerinin ortancasının en küçüklenmesi düşüncesine dayanır. Bu çözümlemede,

$$e_i = y_i - x_i \beta$$

olmak üzere OEKK'yı,

$$\text{Minimize}_{\hat{\theta}}\, \text{ortanca}_i\, e_i^2 \qquad (2)$$

olarak verilmiştir. Bu yöntem hem x hem de y yönündeki aykırı değerlere karşı sağlamdır. Ancak OEKK, eşitlik (2)'de de görüldüğü gibi ortanca değerini en küçük yapmaya çalışırken kalan (n–1) gözlemi dikkate almaz. Bundan dolayı örneklem büyüklüğü arttıkça regresyon katsayılarının kestiriminde OEKK yöntemi, EKK'dan daha etkisiz olmaya başlar. Etkisizliği gidermek amacıyla Rousseeuw ve Leroy (1987)'de OEKK'in yerine en küçük kesilmiş kareler (EKKK) yönteminin kullanılabileceğini belirtmişlerdir (Ryan, 1997).

2.2 En Küçük Kesilmiş Kareler Kestirimi (Least Trimmed Squares Estimation)

Rousseeuw ve Leroy (1987) tarafından tanımlanan bu yöntem, artıkların kareleri $e^2_{(1)} \leq e^2_{(2)} \leq \cdots \leq e^2_{(n)}$ şeklinde en küçükten en büyüğe sıralanmak üzere eşitlik (3)'teki gibi verilir:

$$\text{Minimize}_{\hat{\theta}} \sum_{i=1}^{h} e^2_{(i)} \tag{3}$$

EKKK, yüksek bozulma noktasına sahip tahmin edicidir. Rousseeuw ve Leroy (1987), eşitlik (3)'teki h değerinin aşağıdaki gibi alınmasını önermişlerdir:

$$h = \lfloor n(1-\alpha) \rfloor + 1 \tag{4}$$

Eşitlik (4)'teki $\lfloor . \rfloor$ ifadesi, tam sayı kısmını ifade etmektedir.

Literatürde h değeri için örneklemin gözlem sayısına ve parametre sayısına bağlı olacak şekilde eşitlikler verilmiştir. h değerinin doğru tespit edilmesi önemlidir. Çünkü α kesilme oranı kullanılarak elde edilen h değeri ile veride bulunabilecek aykırı değerlerin doğru miktarı dışarıda bırakılırsa uygun bir kestirici elde edilmiş olur.

2.3 Huber'in M- Kestiricisi

Doğrusal regresyon için M kestiriciler, artık karelerin toplamından çok, artıkların bir fonksiyonunu en küçük yapan kestiricilerdir. M-kestiricileri, y yönündeki aykırı değerlere karşı sağlamdır. Huber (1973), regresyon M kestiricilerini,

$$\text{Minimize} \sum_{i=1}^{n} \rho(e_i) \tag{5}$$

olarak tanımlar.

Burada, $e_i = y_i - x_i\beta$ 'dır. Eşitlik (5)'i en küçükleyebilmek için β_j'ye göre ρ'nun birinci kısmi türevleri $(\psi = \rho')$ sıfıra eşitlenirse; denklem sistemi,

$$\sum_{i=1}^{n} x_{ij}\psi(e_i) = \sum_{i=1}^{n} x_{ij}\psi(y_i - x_i\beta) = 0 \tag{6}$$

biçiminde elde edilir. Genel olarak ρ türev fonksiyonu doğrusal değildir ve iteratif yöntemlerle çözülür. Huber'in M- Kestiricisi, y yönündeki aykırı değerlere karşı sağlam bir yöntemdir.

3. THEIL REGRESYONU

Theil yöntemi (Theil, 1950) parametrik olmayan yöntemlerden birisidir. Theil kestiricileri aykırı değerlerin bulunduğu durumda veri kümesi için uygulanan sağlam yöntemlere alternatif olarak sunulmuştur. Hataların herhangi bir dağılımı için parametrik olmayan yöntemler iyi sonuç vermelerine ve sağlam regresyon kestiricilerinin benzer sonuçlarına götürmelerine rağmen birçok hesaplama işlemi yapılmaktadır (Mutan, 2004).

Basit doğrusal regresyon modelinde β_0 ve β_1 parametrelerine karşılık gelen kestirimler Theil yöntemi ile elde edilebilir. Bu yöntemde, herhangi bir $i.$ ve $j.$ iki nokta arasındaki doğrunun eğimi için aşağıda verilen eşitlik tanımlanır:

$$S_{ij} = \frac{(Y_j - Y_i)}{(x_j - x_i)} \quad i < j \text{ ve } x_i \neq x_j \tag{7}$$

Eşitlik (7)'deki S_{ij} değerleri kullanılarak, basit doğrusal regresyon modelinde β_0 ve β_1 katsayılarının kestirimleri, sırasıyla eşitliklerdeki gibi bulunur:

$$\begin{aligned} \hat{\beta}_{1T} &= \text{or}\tan\text{ca}(S_{ij}) \\ \hat{\beta}_{0T} &= \text{or}\tan\text{ca}(Y) - \text{or}\tan\text{ca}(X) \end{aligned} \tag{8}$$

Dolayısıyla bu yöntemde olası tüm ikili eğimlerin ortancalarının alınmasıyla tüm örneklem noktalarından geçen doğrunun eğimi için sağlam bir kestirim elde edilir ki, bu kestirim örneklemdeki aykırı değerlere karşı dayanıklı bir kestirimdir (Nevitt ve Tam, 1998).

Theil kestiricisinin eşitliğini oluşturulan S_{ij} eğim değeri için zaman içerisinde farklı yaklaşımlarda olmuştur. Eğim için verilen farklı formüllerde aşağıdaki gibidir:

- β_1'in en küçük kareler kestiricisi β_{1SEKK} ile gösterilirse, eşitlik (9)'da ağırlık $w_{ij} = (x_j - x_i)^2$ olmak üzere S_{ij}'nin ağırlıklandırılmış ortalamadır.

$$\beta_{1SEKK} = \frac{\sum_{i<j} w_{ij} S_{ij}}{\sum_{i<j} w_{ij}} \tag{9}$$

- Randles ve Wolfe (1979), S_{ij}'nin ağırlıklandırılmamış ortalamalarını β_1'in kestiricisi olarak önermişlerdir:

$$\hat{\beta}_{1SO} = \frac{\sum_{i<j} S_{ij}}{N} \tag{10}$$

- Dietz (1987), Theil'in vermiş olduğu kestiriciyi her bir S_{ij} için bütün $i < j$ ve $x_i \neq x_j$ değerlerinde olasılıkları atanarak oluşturulan olasılık dağılımının ortancası olarak alındığını belirtmiştir. w_{ij} ağırlığının değiştirilmesiyle elde edilmiştir. S_{ij} için $w_{ij} \big/ \sum_{i<j} w_{ij}$ olasılıkları oluşturularak elde edilen olasılık dağılımının ortancalarından elde edilir. Ağırlıklar aşağıdaki gibi iki şekilde ele alınır:

- Sievers (1978) ve Scholz (1978)'un önerdikleri kestirim için ilk ağırlık $w_{ij} = (j-i)$ 'dir.
- Sievers (1978) ve Scholz (1978)'un önerdikleri kestirim için ikinci ağırlık $w_{ij} = (x_j - x_i)$ 'dir.

Theil, birçok kestirim yöntemi ile karşılaştırılmış ve kavram olarak örneklem verilerindeki aykırı değerlere dayanıklı eğim kestiricisi olarak yorumlanmıştır (Nevitt ve Tam, 1998).

4. BENZETİM ÇALIŞMASI

Bu bölümde çalışmanın amacına uygun olarak bir benzetim çalışması "S-Plus" paket programı kullanılarak yapılmıştır.

Benzetim çalışmasında basit doğrusal regresyon için, Theil kestiricilerinin, EKK, OEKK, EKKK ve Huber'in M-kestiricisinden daha tutarlı bir kestirici olup olmadığı incelenmiştir. Bu amaçla aşağıda verilen hata dağılımlarından, çeşitli bozulum oranlarında veri türetilerek 1000 tekrar üzerinden kestiricilerin parametre kestirimleri ve hata kareler ortalamaları (HKO) elde edilmiştir. Theil kestiricisinin verilerdeki bozulmalarda EKK'dan daha etkili sonuç verip vermediği ve sağlam yöntemlerden elde edilen kestiriciler ile Theil kestiricisinin sonuçları arasındaki farklılıklar HKO'larına göre karşılaştırılmıştır.

Benzetim çalışmasında hataların dağılımı çeşitli biçimlerde ve çeşitli bozulumlarla alınarak yukarıda belirtilen Theil yöntemi ve robust yöntemler uygulanmıştır.

Hataların dağılımı n=7, 10, 30, 50 gözlemleri için standart normal dağılımdan türetilmiştir. Bozulmalar ise n=10, 30, 50 durumları için sırasıyla %10 ve %30 olarak alınmıştır.

Normallik varsayımını bozmak için çeşitli bozulma oranları kullanılarak bozulmuş (contaminated) dağılımlar elde edilmiştir. Verilerde kullanılan bozulma oranları ve dağılımlar Tablo 1'de verilmiştir.

Tablo 1. Dağılımların açıklaması ve numaralandırılması

Dağılım No	Türetilen Dağılım
1	$\varepsilon \sim N(0,1)$
2	$\varepsilon \sim$Lognormal(0,1)
3	n=10 için %10 bozulma, N(0, 1)+N(0, 5)
4	n=10 için %10 bozulma, N(0, 1)+N(0, 9)
5	n=10 için %30 bozulma, N(0, 1)+N(0, 5)
6	n=10 için %30 bozulma, N(0, 1)+N(0, 7)
7	n=10 için %30 bozulma, N(0, 1)+N(0, 9)
8	n=30 için %10 bozulma, N(0, 1)+N(0, 5)
9	n=30 için %10 bozulma, N(0, 1)+N(0, 7)
10	n=30 için %10 bozulma, N(0, 1)+N(0, 9)
11	n=30 için %30 bozulma, N(0, 1)+N(0, 5)
12	n=30 için %30 bozulma, N(0, 1)+N(0, 7)
13	n=30 için %30 bozulma, N(0, 1)+N(0, 9)
14	n=50 için %10 bozulma, N(0, 1)+N(0, 5)
15	n=50 için %10 bozulma, N(0, 1)+N(0, 7)
16	n=50 için %10 bozulma, N(0, 1)+N(0, 9)
17	n=50 için %30 bozulma, N(0, 1)+N(0, 5)
18	n=50 için %30 bozulma, N(0, 1)+N(0, 7)
19	n=50 için %30 bozulma, N(0, 1)+N(0, 9)

Çalışmanın amacına göre yapılan benzetim çalışmasında elde edilen sonuçlar Tablo 2’de verilmiştir.

Tablo 2. Kestiricilere göre hata kareler ortalaması (1000 tekrar)

	Hata Dağılımı	EKK	OEKK	EKKK	M(Huber)	Theil
n=7	1	0.760639	1.907258	1.406432	0.804242	1.200008
	2	6.373845	3.289422	4.391503	4.044770	4.142326
n=10	1	0.525224	1.618579	0.797478	0.540453	0.766247
	2	6.101038	2.158424	2.778860	3.364054	4.195681
	3	1.608523	1.853754	0.790540	0.786184	1.187160
	4	3.951855	1.696885	0.713423	0.821764	1.218585
	5	2.073792	2.137039	2.150130	1.633030	2.281333
	6	3.661883	2.207651	3.568635	2.651068	3.494634
	7	5.507690	1.927702	5.420628	3.751212	4.166321
n=30	1	0.148816	0.662648	0.233411	0.156341	0.455342
	2	3.682325	0.728911	1.515712	1.854906	3.359937
	8	0.420624	0.665532	0.214987	0.208488	0.544231
	9	0.667655	0.634291	0.187397	0.207318	0.566440
	10	1.008647	0.631276	0.190876	0.223598	0.661829
	11	0.586829	0.744540	0.602068	0.430104	1.125093
	12	0.955765	0.721627	0.903869	0.546220	1.607480
	13	1.513571	0.670040	1.500407	0.734984	2.141729
n=50	1	0.083079	0.443300	0.142553	0.086583	0.440086
	2	3.011150	0.556451	1.402420	1.728431	3.352240
	14	0.245730	0.458338	0.125796	0.118967	0.476246
	15	0.384389	0.446537	0.118535	0.123855	0.457891
	16	0.555831	0.453147	0.111312	0.124860	0.486005
	17	0.303736	0.475230	0.332670	0.216577	0.657527
	18	0.545853	0.445948	0.606537	0.301645	0.828066
	19	0.870570	0.408874	0.940922	0.381872	1.259474

Tablo 2 incelendiğinde aşağıdaki sonuçlar elde edilmiştir. Buna göre;

n=7 durumu için, hatalar normal dağıldığında beklenildiği gibi, EKK yönteminden elde edilen HKO değerleri en küçüktür. Huber'in M kestiricisi, EKK'dan sonra diğer kestiricilere göre daha küçük HKO'na sahiptir. Dağılım lognormal olduğunda ise en küçük HKO'lı kestirici OEKK'dır. Huber'in M kestiricisi de diğer kestiricilere göre küçük HKO'ya sahiptir. Theil kestiricisi, lognormal dağılım durumunda EKK'den daha küçük HKO'sına sahip olduğu görülmüştür.

n=10 olduğunda, yine beklenildiği gibi normal dağıldığında en iyi kestirici EKK'dır ve Huber M kestiricileri de iyi kestiricilerdir. Hatalar lognormal dağıldığında OEKK en küçük HKO'na sahiptir. Lognormal dağılımda tüm sağlam kestiriciler Theil kestiricisinden iyi sonuç verirken Theil kestiricisi EKK'dan daha iyi sonuç vermiştir. Veriler %10 bozulduğunda EKKK ve Huber-M sağlam kestiricileri en iyi kestiricilerdir. Theil kestiricisi %10 bozulmada EKK'dan daha iyi sonuç vermiştir. %30 bozulma durumunda ise en iyi kestirici OEKK kestiricisidir. Theil kestiricisi sağlam kestiricilerden daha iyi sonucu hiç verememiştir ama EKK'dan daha düşük HKO'na sahip olarak EKK'dan daha iyi kestirici olmuştur.

n=30 olduğunda, sonuçlar %30 bozulmanın olduğu durumda biraz değişmektedir. Bozulmanın olmadığı durumda EKK yine en iyi kestirici olarak görülmektedir. Lognormal dağılımda OEKK kestiricisi iyi sonuç vermiştir. Theil kestiricisi, lognormal dağılımlarda sağlam kestiricilerden daha iyi sonuç verememesine rağmen EKK'dan daha iyi bir kestirici olmuştur. %10 bozulma durumunda EKKK ve Huber'in M kestiricisi iyi sonuçlar vermiş, Theil kestiricisi EKK'dan daha iyi olmaktan öteye geçememiştir. %30 bozulma durumunda ise sağlam kestiriciler arasında Huber'in M kestiricisi ve OEKK iyi sonuçlar vermiştir. Theil kestiricisi yine bozulma durumlarında sağlam kestiricilerden daha iyi sonuç vermediği gibi %30 bozulum durumunda EKK bile Theil kestiricisinden daha küçük HKO'sı vermiştir.

n=50 olduğunda, sonuçların hemen hemen n=30 durumuyla aynı olduğu görülmektedir. EKK'ler bozulum olmadığında yine en iyi kestiricidir. Lognormal dağılımda OEKK en iyi kestiricidir. Theil kestiricisi, veri sayısı arttıkça n=7, 10, 30 durumundakilerden farklı olarak EKK'dan daha yüksek HKO'na sahiptir. %10 bozulmada EKKK ve Huber'in M kestiricisi, %30 bozulmada ise Huber'in M kestiricisi en iyi kestiricidir. Theil kestiricisi yine EKK'dan bile daha büyük HKO'sı vererek çözümlemelerde etkili kestiriciler elde edememiştir.

5. SONUÇ VE TARTIŞMA

Literatürde çalışmalara bakıldığında en çok EKK ile Theil kestiricisinin karşılaştırılması yapılmış ve Theil kestiricisinin varsayım bozulumlarında EKK'ya göre daha sağlam olduğu sonuçlarına varılmıştır.
Bu çalışmada ise Theil kestiricisi EKK'ya göre 7, 10, 30 gözlemli verilerde ve %10'luk bozulumlarda daha iyi sonuç vermesine rağmen n=50 olduğunda %10 ve %30 bozulumlarda EKK Theil'den daha iyi sonuç vermiştir. Theil kestiricisi, parametrik olmayan bir yöntem olduğundan %10 ve %30'lu bozulumu olan belli bir parametrik sistemde sağlam kestiriciler kadar başarılı olmadığı görülmüştür.

EKK kestiricisinin, dağılımda herhangi bir bozulma olmadığında en iyi kestirici olduğu, ayrıca OEKK kestiricisinin lognormal dağılımlarda gözlem sayısı fark etmeksizin en önemli kestirici olduğu görülmüştür. Theil kestiricisi bu benzetim sonuçlarına göre genel olarak başarısız olmuştur. Diğer dağılım varsayımlarına göre olan davranışı da ayrıca incelenebilir.

6. KAYNAKLAR

Candan, M., 1995. Doğrusal Regresyon Çözümlemesinde Sağlam Kestiriciler. H. Ü. Fen Bilimleri Enstitüsü, Ankara (Türkçe).

Dietz, E. J., 1987. On Estimating a Slope and Intercept in a Nonparametric Statistics Course. North Carolina State University, USA (İngilizce).

Huber, P. J., 1973. Robust Regression: Asymptotics, Conjectures and Monte Carlo. Ann. Stat., 1, 199-821.

Mutan, O., 2004. Comparision of Regression Techniques via Monte Carlo Simulation. Graduate School of Natural and Applied Sciences of METU, Ankara (İngilizce).

Nevitt, J., Tam, H. P., 1998. A Comparison of Robust and Nonparametric Estimators Under the Simple Linear Regression Model. Multiple Linear Regression Viewpoints, 25, 54-69.

Randles, R. H., Wolfes, D. A., 1979. Introduction to the Theory of Nonparametric Statistics. Wiley & Son, New York.

Rousseeuw, P. J., 1984. Least Median of Squares Regression. J. Am. Stat. Assoc., 79, 971-880.

Rousseeuw, P. J., Leroy, A. M., 1987. Robust Regression and Outlier Detection. Wiley & Son, New York.

Ryan, T. P., 1997. Modern Regression Methods. John Wiley&Sons, Inc, New York.

Scholz, F. W., 1978. Weighted median regression estimates. The Annals of Statistics, 6(3): 603–609.

Sievers, G. L., 1978. Weighted rank statistics for simple linear regression. Journal of the American Statistical Association, 73(363): 628–631.

Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis. Ned. Akad. Wetensch. Proc. Ser. A, 53.

COMPARISON OF ROBUST AND THEIL ESTIMATORS IN SIMPLE LINEAR REGRESSION

ABSTRACT

According to Gauss-Markov Theorem, ordinary least squares (OLS) estimates are best linear unbiased estimators when the assumptions are satisfied in simple linear regression. However, these estimates in regression models are not resistant to deviations from classical assumptions. In this situation, some nonparametric or robust methods are used as alternatives in regression analysis. Nonparametric statistical methods require no normality assumption about the error distribution. Theil's estimator gives robust results when outliers are present in the sample (Nevitt, J., Tam, H. P., 1998). Robust methods are also alternative methods to ordinary least square estimates when the necessary assumptions in regression analysis are not satisfied. In this study, the robust methods of median least squares (MLS), least squares, Huber's M estimator and nonparametric Theil's regression method are compared by performing a simulation study for simple linear regression model when data consist of outliers. The performances of estimators are compared and interpreted with respect to their efficiency. According to simulation results, estimators obtained from robust methods provide more efficient results than the ones obtained from Theil's method.

Keywords: Outliers, Robust regression, Simple linear regression, Theil's method.