

BÖLÜM I

1. GİRİŞ

Ekonometrik modeller çeşitli ekonomik, matematiksel ve istatistiksel sınırlamalar altındadır. Bu nedenle geçerli sonuçlara ulaşmak için bir çok testin yapılması gerekmektedir. Kurulmuş bir (ekonomik) modelin verilerinde ardışık-bağımlılık, değişen-varyans veya UDler mevcut ise sıklıkla kullanılmakta olan EKK¹ ya da verilerin tamamını dikkate alan herhangi başka bir tahmin edici kullanıldığında bu yöntemler verilerin içerdiği genel bilgiyi yansıtmamaktadır. Bu çalışma modellerde yer alan ekonomik veya matematiksel sınırlamalardan daha çok istatistiksel sınırlamalar ve özellikle de UDler² ile ilgilenmektedir.

UD varlığında EKK tahminlerinin çökebilmesi büyük bir tehlike yaratmaktadır, çünkü EKK tahminleri çok yaygın olarak kullanılmakta ve uzun zamandan beridir en kaliteli verilerin bile UD içerdiği, buna ilaveten kaliteli iktisadi veri olmaya çok az sayıda aday veri olduğu bilinmektedir (Orhan vd., 2001). Rousseeuw ve Leroy (1987), Rousseeuw ve Wagner (1994), Knez ve Ready (1997) çalışmalarında UDlerin çeşitli

¹ En Küçük Kareler

² Uç değerler, ing outlier

olumsuz etkileri sergilenmiş ve böyle durumlarda UDlerin tespiti için SR³ yöntemleri önerilmiştir.

Bir veri kümesinde UDler, verilerin yarısından daha az miktarda olmalarına rağmen, verilerin çoğunun vermek istediği bilgiye engel olan ve sonuçlar üzerinde yanıltıcı etkiler yaratan veriler olarak adlandırılabilir (Rousseeuw ve Zomeren, 1990). Takip eden bölümlerde de gösterildiği gibi, bu durum veri kümesinde bir tane dahi UD bulunması durumunda bile geçerlidir ve bu sebeple güvenilir tahminler elde edilememektedir.

Halbuki Türk iktisadi verilerinde UD bulunma olasılığı, ekonomik olarak daha istikrarlı bir yapıya sahip gelişmiş ülkelerden daha fazladır. Bunun sebebi ise, tanımdan da anlaşılacağı gibi, veri kümesine normal zamanlarda etkili olmayan ve çok nadiren görülen çeşitli sosyo-ekonomik değişkenlerin/değişimlerin Türkiye’de daha fazla/sık etkili olması ve böylece UD oluşumuna sebebiyet verebilmesidir.

Literatürde veri kümesinde sadece bir tane UD bulunması durumunda bunları tespit etmek için güvenilir ve kolay uygulanabilir teknikler mevcuttur. Klasik UD tespit yöntemleri için Hadi ve Simonoff (1993) çalışmasında geniş bir literatür bulunabilir, fakat birden çok UD mevcut olması durumunda bu yöntemler yetersiz kalmaktadır. Birbirine yakın UDler bazen birbirlerini maskeleyebilmekte⁴ ve tespit edilebilmelerini güçleştirmektedir. Hatta bu UDler SR tekniklerinde güvenilir verilerin bile UD olarak

³ Sağlam Regresyon, ing. Robust Regression

⁴ ing. Masking

görünmesine sebep olabilmektedirler⁵. Takip eden üçüncü bölüm bu çalışma için türetilmiş verilerle bu duruma çok güzel örnekler içermektedir. Bütün bunlara rağmen SR tekniklerine duyulan bu ilgisizlik şu nedenlere bağlı olabilir (Orhan vd., 2001):

- Büyük veri setlerinde SR tekniklerine gerek olmadığı, yeterli veri ile doğru tahminlerin elde edilebileceği inancı⁶.
- Görsel yollarla UDlerin tespit edilebileceği, EKK artıklarına bakarak beklenmedik artıkların UD olarak adlandırılabilceği inancı.
- Sağlam analiz sonuçlarının yorumlarının alışılmadık olması.
- Gerçek veri setlerine uygulanan sağlam analiz sonuçlarının başarısının farkında olunmaması.

Veri kümesinde birden çok UD mevcut olması durumuna önlem olarak SR literatüründe UDLere karşı ÇDT⁷ geliştirilmiştir. ÇDT ediciler yüksek sayıda ve kötü yerleştirilmiş UDler için güvenilir parametre tahminleri yapma problemi ile ilgilidir. Farklı kriterler bulunmasına rağmen genel prensip verileri bir çok parçaya bölüp verilerin çoğunu, kullanılan kritere göre kapsayan bir büyük altküme ve geri kalanını dışlayan altkümeler oluşturmaktır (Hawkins ve Olive, 1999). Bu sayede UDlerden etkilenmeyen ve verilerin çoğundaki bilgiyi yansıtan tahminler üretilebilmektedir.

Bu çalışmanın amaçlarından biri Türk ekonomik verilerinde UD tespit etmek için mümkün olan en verimli UD tespit yöntemlerini kullanmaktır. Bu amaç için güncel UD

⁵ Yükleme, ing. Swamping

⁶ Kısaca büyük örneklem özellikleri geçerliliğini kaybetmekte

⁷ Çökmeye Dayanıklı Tahmin Ediciler, ing. Robust Breakdown Estimators

tespitinde ne tür felsefe/bakış açılarının halen geçerli olduğu araştırılmış ve yeni olarak literatüre katılmış olan bir kısım tahmin ediciler ile SR literatürünün yönü belirlenmiştir. Son yıllarda bilgisayar teknolojisindeki gelişmelere paralel olarak çok çeşitli UD tespit yöntemleri geliştirilmiş ve halen geliştirilmektedir. Sürekli bir gelişme halinde olan bir konu olması sebebi ile bu konu ile ilgilenen araştırmacılar, son yenilikleri takip etmekte güçlük çekmektedirler. Bu nedenle bu çalışma bir başka amaç olarak araştırmacılara yol göstermek için literatürde bu konu ile ilgili güncel bulguları bir araya getirmeyi amaçlamaktadır. Bu çalışmada, neden mevcut regresyon yöntemlerinin geçerli ve güvenilir tahminler vermeye yetmediği; SR literatüründe kesin kabul görmüş, evrensel bir yöntemin mevcut olup olmadığı; UDlerin ekonomi verileri üzerinde anlamlı bir etkisi olup olmadığı sorularına cevap aranmaya çalışılmaktadır.

Bu çalışma dört bölümden oluşmaktadır. Birinci bölümde giriş kısmı sunulmuştur. Takip eden ikinci bölümde, UD tanımı, önemi ve etkileri verilmiş; buna ek olarak güncel UD tespit yöntemleri literatür taraması yardımıyla sunulmuştur. Üçüncü bölümde kesit veriler kullanılarak Türk ekonomi verilerinde UD varlığı araştırılmış ve bilimsel çıkarsama üzerindeki etkileri sorgulanmıştır. Son bölüm ise, bulgular, tartışma ve yorum kısmına ayrılmıştır.

1.1. Literatür Taraması

Literatür taramalarına göre, uygulamalı ekonometride SR teknikleri nadiren kullanılmaktadır (Orhan vd., 2001). Bu durum UD etkilerinin ve öneminin göz ardı edildiğini göstermektedir. UDlerin iktisadi veriler üzerindeki etkilerini incelemek için SSCI ve online veri tabanları başta olmak üzere yapılmış Internet taraması sonucunda, dünya literatüründe Orhan ve diğerleri (2001) ve Türkiye’de Erlat (2005), Erlat (2003), Küçükkocaoğlu ve Kiracı (2003), Atuk ve Ural (2002) dışında bu durumu dikkate alan fazla sayıda iktisadi çalışmaya rastlanmamıştır.

Orhan ve diğerleri (2001) çalışmasında, UDlerin tespitini ve etkilerini gösteren 3 uygulama yapılmıştır. İlk uygulama, 1960-1985 yılları arasında tüm dünya çapında 61 ülkeyi kapsayan ulusal büyüme ile ilgili bir regresyon analizini içermektedir. Modelde GSYH⁸ bağımlı değişken, işgücü büyümesi, relatif GSYH, malzeme yatırımı, malzeme dışı yatırımlar bağımsız değişkenleri oluşturmaktadır. Modelden beklenti, malzeme yatırımı ile bağımlı değişken arasında ilişki olmasıdır. Standartlaştırılmış EKKK kalıntıları yardımıyla UD tespiti yapılmıştır. UD olarak tespit edilen ülke veri setinden çıkarıldıktan sonra tekrar regresyon yapıldığında sonuçlar beklendiği gibi olmuştur. EKK yöntemi uygulandığında malzeme dışı yatırımlar anlamsız bulunurken, SR teknikleri uygulandığında malzeme dışı yatırımlar anlamlı bulunmaktadır. R² ve F değerleri artmaktadır. İkinci uygulamada, OECD ülkelerine SOLOW büyüme modeli uygulanmıştır. Veri seti 22 ülkeden oluşmaktadır. Standartlaştırılmış EKKK

⁸ Gayri Safi Yurtiçi Hasıla, ing. Gross Domestic Product, GDP

kalıntılarının mutlak değeri 2,5 değerinden büyük olan veriler kötü kaldıraç noktaları olarak tespit edilmiş ve veri setinden çıkarılmışlardır. Anakitlenin yıllık büyümesini gösteren N değişkeninin logaritması, veri setinde UDler mevcut iken sınırdan da olsa anlamlı bulunurken, veri setinden UDler çıkarıldıktan sonra anlamsız olarak tespit edilmiştir. Yapılan üçüncü uygulamada ise, hisse senedi getirileri modellenmiştir. Yine standartlaştırılmış EKKK kalıntı değerlerinden yararlanılarak UD tespitinde bulunulmuştur. UD olarak tespit edilen 2 yıla ait veriler veri setinden silinmiş ve tekrar analiz edilmiştir. Daha önceden anlamsız olarak bulunan enflasyon değişkeni anlamlı olarak bulunmuştur. Modelin uyumu artmıştır. Sonuç olarak UD olarak tespit edilen veriler gözlemlerden çıkarılınca sonuçlar teori ile uyumlu olmaktadır. Çalışma sonucunda, SR tekniklerini uygulamada katlanılan zorluklar yanında yararlarının da çok fazla olduğu desteklenmektedir.

Erlat (2005) çalışması, enflasyon oranlarını dikkate alan bir zaman serileri çalışmasıdır. Erlat (2003) çalışması ise reel döviz kuru ile ilgili bir çalışmadır. Her iki çalışmada da zaman serileri analizinden önce UD varlığına bakılmış ve etkileri dikkate alınmıştır. Erlat (2003) çalışmasında uygulanan birim kök testinde, tespit edilen UDler kukla değişken olarak alınıp etkisi ortadan kaldırılmaya çalışılmıştır. Sonuçlara göre, birim kök testi uygulanırken UD varlığı göz önüne alınmalıdır.

Atuk ve Ural (2002) çalışması mevsimselliği giderme yöntemlerini karşılaştırmaktadır. Bu yöntemler UDlere karşı duyarlıdır ve her çeşit UDi tespit edebilmektedirler. Fakat yöntemler aynı verideki UDleri farklı çeşit olarak tespit

etmektedirler. Çalışma genel anlamda mevsimsellik ile ilgilenmektedir ve UD tespitine değinmektedir.

Küçükkocaoğlu ve Kiracı (2003) çalışması Finansal Varlıkları Fiyatlama Modeli'nin (CAPM) beta katsayısını hesaplamak için kullanılan EKK yöntemi ile SR yöntemleri sonuçlarını karşılaştırmaktadır. Çalışma sonucunda SR teknikleri beta katsayısını hesaplarken çok daha başarılı olmuştur. SR yöntemlerinden EKOK yöntemi kullanılmıştır. Elde edilen örnek sonuçlar karşılaştırıldığında, EKOK yöntemine göre elde edilen beta değeri birden büyükken, EKK yöntemine göre elde edilen beta değeri birden küçüktür. Özetle EKOK yöntemi ile bulunan beta katsayısı hisse senedinin riskini beklentilere uygun olarak yansıtmaktadır.

BÖLÜM II

2. UÇ DEĞERLER, KULLANIM ALANLARI VE TESPİT YÖNTEMLERİ

Bu bölümde UDlerin bilimsel sorgulama üzerindeki olumsuz etkileri, bu çalışma için türetilmiş örneklerle gösterilmiştir. Daha sonra zaman süreci içinde geliştirilmiş alternatif UD tespit ve çalışma prensipleri özetlenmiştir. Son olarak da güncel UD tespit yöntemlerinin nasıl bir bakış açısı izlediğine dair bir literatür taraması yapılmış ve yöntemler gruplandırılarak sunulmuştur. Yöntemlerdeki gelişmeler ile SR literatürünün gittiği yön belirlenmeye çalışılmıştır. İzleyen bölümde kısa bir UD ve EKK tanımına ve örneklerle UDlerin etkisine yer verilecektir. Daha sonraki altbaşlıkta güncel UD tespit yöntemlerinin temelinde yatan, ilk ortaya çıkmış ve halen kullanılmakta olan tahmin edicilerin özellikleri açıklanacaktır. Üçüncü altbaşlıkta güncel UD tespit yöntemlerine değinilecek, son olarakta çalışmada kullanılan yöntemler açıklanacaktır.

2.1 Uç Değerler ve EKK

EKK yöntemi, 1800'lü yıllarda bulunuşundan beri, veriden rahatça türetilmesi, kolay anlaşılır ve hızlı uygulanabilir olması gibi avantajlara sahip olması sebebiyle sıkça tercih edilen bir yöntemdir. Algoritmanın temeli, (1) nolu denklemde gösterildiği gibi, hata kareleri toplamının minimizasyonunu veren tahminlere dayanmaktadır (Rousseeuw ve Leroy, 1987).

$$\text{Minimize}_{\hat{\theta}} \sum_1^n (y_i - \hat{y}_i)^2 \quad (1)$$

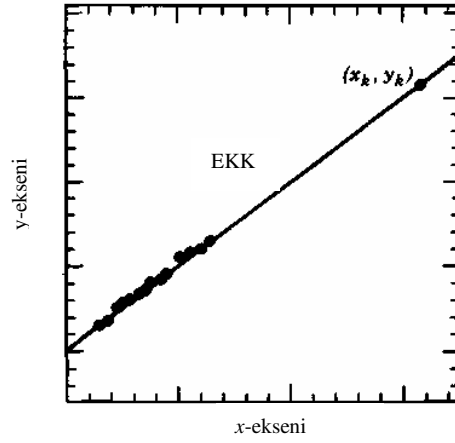
EKK yöntemi her tahmin edilen değerin (\hat{y}_i), gözlenen değerden (y_i) farkının karesini dikkate aldığından her UDin tahmin üzerine yanıltıcı etkisi çok fazla olacaktır. Bunun sebebi ise UDlerin diğer verilerin çoğundan daha yüksek artık değerlerine sahip olmasıdır. Halbuki, (1) denkleminde EKKler, artıkların karelerinin toplamını indirmeye çalışmakta ve yüksek kalıntı değerlerinin etkileri kareleri oranında olmaktadır. EKKler UDlerin yüksek kalıntı değerlerini düşürmekte, fakat iyi verilerin kalıntılarını artırmakta ve böylece kalıntı kareleri toplamı daha az olmaktadır. Böylece EKKler yüksek kalıntı değerlerine izin vermemektedir.

Rousseeuw ve van Zomeren (1990) çalışmasına göre, bir veri setinde çok çeşitli UDler olabilir. Verinin büyük çoğunluğunun vermek istediği bilgiye engel olan fakat ilk bileşeni UD olmayan, sadece y_i ⁹ değeri UD olan bir (x_i, y_i) noktası dikey UD, sadece x_i ¹⁰ noktası UD olan (x_i, y_i) noktası kaldıraç noktası olarak adlandırılır. Verinin büyük bir çoğunluğunun vermek istediği bilgiyi izleyen (x_i, y_i) kaldıraç noktasına, iyi kaldıraç noktası; diğer kaldıraç noktalarına ise kötü kaldıraç noktası adı verilir. Özetle, bir veri setinde dört çeşit veri bulunabilir. Bunlar, UD içermeyen veriler, dikey UDler, iyi kaldıraç noktaları, kötü kaldıraç noktalarıdır. Çoğu veri setleri bu dört çeşit verinin hepsini birden içermez. İyi kaldıraç noktalarına örnek şekil 2.1.'de (Rousseeuw ve Leroy, 1987: 6) verilebilir. Kötü kaldıraç noktalarının EKK tahminleri üzerindeki etkisi

⁹ (x_i, y_i) noktasını oluşturan ikinci bileşen

¹⁰ (x_i, y_i) noktasını oluşturan birinci bileşen

çok fazladır (Rousseeuw ve Leroy, 1987:6). Çünkü normalden farklı seyir gösteren x noktaları EKK doğrusunu yana doğru eğmekte, kalıntıları çok artırmakta ve böylece kalıntı kareleri toplamı çok yükselmektedir. Kötü kaldıraç noktalarına örnek olarak, şekil 2.2. ve şekil 2.3 verilebilir



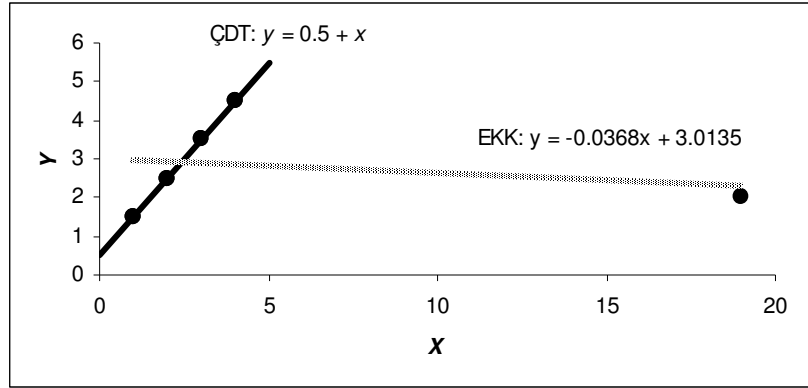
Şekil 2.1. İyi kaldıraç noktası

2.1.1. Örneklerle Uç Değerlerin Etkisi

Bu kısımda amaç UDlerin bilimsel sorgulamayı nasıl olumsuz yönde etkilediği ve etkili değişkenleri nasıl etkisiz gösterdiğini sunmaktır. Aşağıdaki tablo 2.1’de yer alan veriler ve şekil 2.2. bu çalışma için üretilmiş olup, bu amaç için uygundur.

Tablo 2.1. Veri

Y	X
1.5	1
2.5	2
3.5	3
4.5	4
2	19

**Şekil 2.2.** Uç-değerlerin EKK tahmin edicisi üzerindeki etkisi

Şekil 2.2. de de görülebileceği gibi bir tane veri EKK tahmininin ya da verilerin tamamı kullanılarak yapılabilecek herhangi bir tahminin çökmesine ve verilerin çoğunun yapmış olduğu tahminle ilgisi olmayan bir tahminin ortaya çıkmasına sebep olmuştur. Takip eden tablo 2.2. ise EKK regresyonunun sonuçlarını özetlemektedir.

Tablo 2.2. Tablo 2.1.'de yer alan verilerin regresyon sonuçları

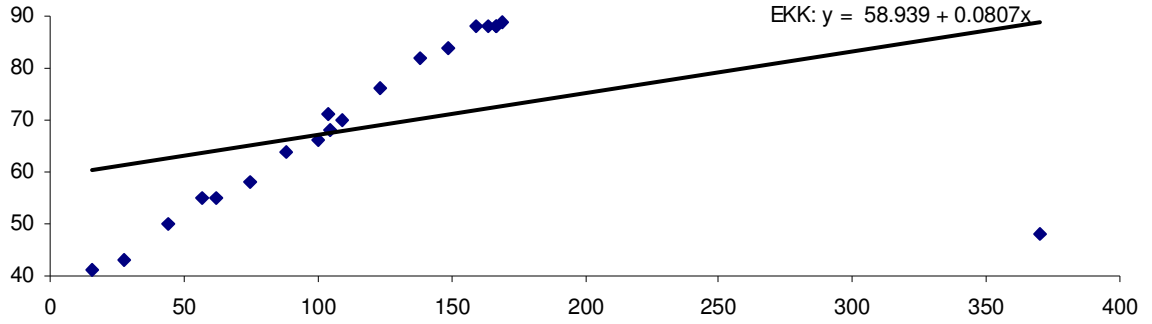
	<i>katsayılar</i>	<i>standart hata</i>	<i>T stat</i>	<i>p-değeri</i>	<i>regresyon istatistikleri</i>	
Kesişim	3.0134	0.80203	3.7572	0.0329	Çoklu R	0.228108
X	-0.0368	0.09069	-0.4057	0.7121	R²	0.052033
					Ayarlı R²	-0.263955
					Standart Hata	1.353785
					Gözlem	5

Sonuçlar incelendiğinde x değişkeninin anlamlı bir değişken olduğunu verilerin desteklemediği ve %5 güven aralığında sadece kesişimin anlamlı olduğu görülebilmektedir. Regresyondan x değişkeni atıldıktan sonra sadece kesişimin anlamlı olduğunu gösteren aşağıdaki sonuçlar elde edilmektedir. UDlerin varlığı anlamlı bir değişkeni anlamsız gibi gösterebilmektedir.

Tablo 2.3. X değişkeni atılarak tablo 2.1.'de yer alan veriler için regresyon sonuçları

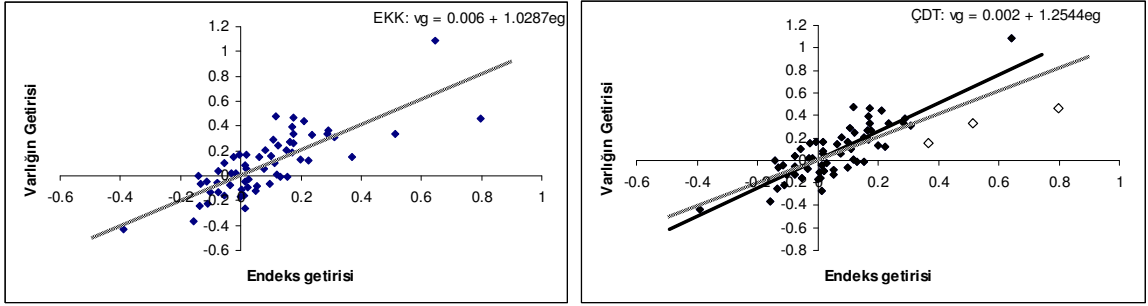
	<i>katsayılar</i>	<i>standart hata</i>	<i>T stat</i>	<i>p-değeri</i>	<i>regresyon istatistikleri</i>	
Kesişim	2.8	0.53851	5.19946	0.00651	Standart Hata	1.2041598
					Gözlem	5

EKK yönteminin en cazip özelliklerinden birisi büyük örneklem özelliklerine sahip olmasıdır. Bu bağlamda verilerde yer alan az sayıda UDin etkisinin veri boyutu arttıkça ortadan kalkacağı inancı mevcuttur, fakat bu tehlikeli bir önyargıdır. Takip eden şekil 2.3.'deki veriler Daniel ve Wood'dan (1971) alınmıştır. Bu şekil çok sayıda veri mevcut iken bir tane UDin bile sonuçları değiştirebildiğini göstermektedir. Rousseeuw ve Leroy (1987:69) çalışmasında örnekleme yer alan UD sayısı %0 civarında olduğunda, yani yok denebilecek kadar az sayıda UD bulunduğu bile, EKK tahminlerinin çöktüğünü simülasyonla göstermektedirler. Knez ve Ready çalışmalarında (1997) EKK tahminlerinin, büyük örneklerde dahil, UDLere karşı duyarlılığını göstermektedirler. Bu sebeple EKK için büyük örneklem özellikleri UD varlığında geçerliliğini kaybetmektedir.



Şekil 2.3. Tek bir UDin EKK tahmini üzerindeki etkisi

Önemli bir başka tehlike ikinci boyutta UDlerin görsel olarak her zaman tespit edilebileceğinin düşünülmesidir. Küçükkocaoğlu ve Kiracı (2003) çalışmasından alınan şekil 2.4. bunun her zaman mümkün olmayacağını göstermektedir, çünkü veri sayısı artıkça görsel olarak verilerin yoğunluğu ayırt edilememektedir. Bununla birlikte yine aynı şekil EKK kalıntılarına bakarak UD tespit edilemeyeceğini göstermektedir. Soldaki şekilde EKK tahmininin verilerin çoğunun yönünü uygun bir şekilde yansıttığı düşünülebilir ve hatta en üst sağda yer alan veri UD olarak değerlendirilebilir. Ama ÇDT ile UDler arandığında sağdaki şekilde yer alan içi boş üç noktanın UD olduğu tespit edilir. Görsel olarak UD tespiti ile daha kapsamlı bilgi için Rousseeuw ve Van Zomeren (1990) incelenebilir ve konu ile ilgili gerçek hayattan bir çok örnek Rousseeuw ve Leroy (1987) çalışmasında bulunabilir.



Şekil 2.4. UDLerin kendilerini maskeleyebilmeleri ve iyi verileri UD gibi göstermelerine örnek

Yukarıda yer alan örnekler UD tespiti konusunda karşılaşılan zorluklar ve sonuçlar üzerindeki yıkıcı etkileri sergilemek açısından önemlidir. Üçüncü bölümde Türk verilerinde böyle durumların mevcudiyeti sorgulanmakta ve varsa etkileri sergilenmektedir.

2.1.2. Zaman Serilerinde Uç Değerler

Zaman serisi ve öngörü yöntemleri mühendislik, ekonomi, fizik bilimleri gibi pek çok alanda kullanılmaktadır. Bu kısımda, bu yaygın kullanımını da göz önünde bulundurarak, zaman serilerinde UD varlığı ve çeşitlerine değinilecektir.

Zaman serisi analizinde kullanılan paket programlar Box ve Jenkins (1976) yöntemine dayanmaktadır. Bu yöntemde ne yazıkki UDLere karşı dayanıklı değildir (Rousseeuw ve Leroy, 1987). Sağlam zaman serisi analizleri için geniş literatür taraması Martin ve Yohai (1984) çalışmasında bulunabilir.

$$x_t - \alpha_1 x_{t-1} - \dots - \alpha_p x_{t-p} = e_t + \beta_1 e_{t-1} + \dots + \beta_q e_{t-q} \quad (2)$$

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + e_t \quad (3)$$

3. denklemde de görüldüğü gibi, x_t , gecikmiş x_{t-1}, \dots, x_{t-p} bağımsız değişkenlerince¹¹ açıklanan bir bağımlı değişkendir. e_t ortalaması 0, varyansı σ^2 olan bir hata terimidir. 2. denklemdeki eşitliğin sol tarafı **AB(p)**¹² olarak adlandırılır (Rousseeuw ve Leroy, 1987). p derecesi x değişkeninin kaçınıcı dereceden geçmiş değerine bağlı olduğunu gösterir. Eşitliğin sağ tarafı, **HO(q)**¹³ kısmıdır. Hata terimlerinin doğrusal birleşiminden oluşmaktadır. Model ABHO(p,q) modeli olarak gösterilmektedir.

Örnek AB(1) modelinde:

$$y_t = \gamma + \alpha_1 y_{t-1} + e_t \quad (4)$$

t dönemindeki y değeri kendisinin bir önceki dönemdeki değerine bağlı olan bağımlı değişken olarak tanımlanmaktadır. *Birinci Dereceden Ardışık Bağlanımlı* bir modeldir. AB(1) modelinde y_t değişkeninin durağan¹⁴ olabilmesi için $|\alpha_1| < 1$ olmalıdır.

¹¹ ing. lag value

¹² Ardışık Bağlanım, ing. Autoregressive, AR(p)

¹³ Hareketli Ortalamalar, ing. Moving Average, MA(q)

¹⁴ Geleceğin geçmiş gibi olacağı varsayımı

Örnek HO(2) modelinde:

$$y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} \quad (5)$$

t dönemindeki y değeri hata terimlerinin doğrusal bir fonksiyonuna bağlıdır. *İkinci Dereceden Hareketli Ortalamalar Modelidir.*

y kuşkusuz hem AB, hem HO özellikleri taşıyabilir ve ABHO modeli olarak adlandırılır. Genel olarak bir $ABHO(p,q)$ sürecinde p tane ardışık bağlanım, q tane hareketli ortalama terimi bulunur. $ABBHO^{15}$ modeli yaygın adıyla Box-Jenkins modelidir (Gujarati, 1999). $ABBHO$ modeli durağan olmayan yani bütünleşik olan zaman serisi modellerinde kullanılır. Bütünleşik zaman serisinin ortalaması, varyansı, ortak varyansı zamanla değişir. Böyle modellerde seri durağan olana kadar farkları alınır. Yani bir seriyi durağan yapmak için d kez fark alınır da buna $ABHO(p,q)$ modeli uygulanırsa, başlangıçtaki zaman serisine $ABBHO(p,d,q)$ modeli denir. Burada d seriyi durağan duruma getirmek için kaç kez farkının alınması gerektiğini gösterir. $ABBHO(p,0,q)$ $ABHO(p,q)$ modeli anlamına gelir. Aynı zamanda $ABHO(0,q)$ modeli de $HO(q)$ anlamına gelir (Gujarati, 1999). Gujarati'ye göre (1999) durağanlık varsayımının yapılmasının nedeni şöyle açıklanabilir:

“B-J [Box-Jenkins]'in amacı, örneklem verilerini türettiği düşünülebilecek bir istatistik modelini belirlemek ve tahmin etmektir. Tahmin edilen bu model kestirim için kullanılacaksa, modelin özellikleri zaman içinde, özellikle de gelecek dönemlerde değişmemelidir. Öyleyse durağan veri gereksiniminin basit nedeni, bu verilerden çıkarsanan herhangi bir

¹⁵ Ardışık Bağlanımlı Bütünleşik Hareketli Ortalama, ing. Autoregressive Integrated Moving Average, ARIMA

modelin de durağan ya da kararlı olabilmesi, dolayısıyla da kestirim için geçerli bir temel sağlayabilmesi gerekliliğidir”¹⁶.

Şekil 2.5. (a) da Rousseeuw ve Leroy’dan (1987) alınan UD olmayan bir zaman serileri veri seti örneği verilmiştir. Böyle verilerde kesişim ve eğim parametrelerini doğru bir şekilde tahmin etmek zor değildir. y_t ’nin t zamanına ve $t-1$ gecikmiş değerine karşı serpilme diyagramı örneği verilmiştir. Fox (1972) ve Martin (1981)’e göre zaman serilerinde 2 çeşit UD bulunmaktadır. Bunlardan ilki *Yenilikçi*¹⁷ UDlerdir. Örnek şekil 2.5. (b)’de verilmiştir. Burada t_0 zamanına ait 1 tane UD içeren veri seti gösterilmektedir. A noktası dikey yönde bir UD’dir.

İkinci çeşit UDLere ise, *Katkısal*¹⁸ UDler adı verilmektedir. Y_t ’nin kendi değerleri bozulup kirlendiğinde oluşmaktadır. Böylece gözlemler ABHO modeline uymamaktadır. Hata terimlerinin çoğunun sıfır olmasının yanında sıfıra eşit olmama olasılığı sıfır olmadığından¹⁹ $E(u_i)=0$ varsayımı yapılamamaktadır.²⁰ Hata terimlerinin bağımsız ve belli bir dağılıma sahip olan rassal değişkenler olduğu varsayımı ayrılmış UDleri²¹ oluşturur. Şekil 2.5.(c)’de ayrılmış katkısal UDLere örnek gösterilmiştir. Şekil 2.5.(d) katkısal UDLere örnek olarak verilmiştir. Uygulamada daha sık olarak katkısal UDler ile karşılaşmaktadır. EKK yöntemi böyle durumlarda sapmalı tahminler vermektedir. Eğer

¹⁶ Michael Pokorny, *An Introduction to Econometrics*, Basil Blackwell, New York, 1987: 343

¹⁷ ing. innovation

¹⁸ ing. additive

¹⁹ $P(v_t \neq 0) = \varepsilon$

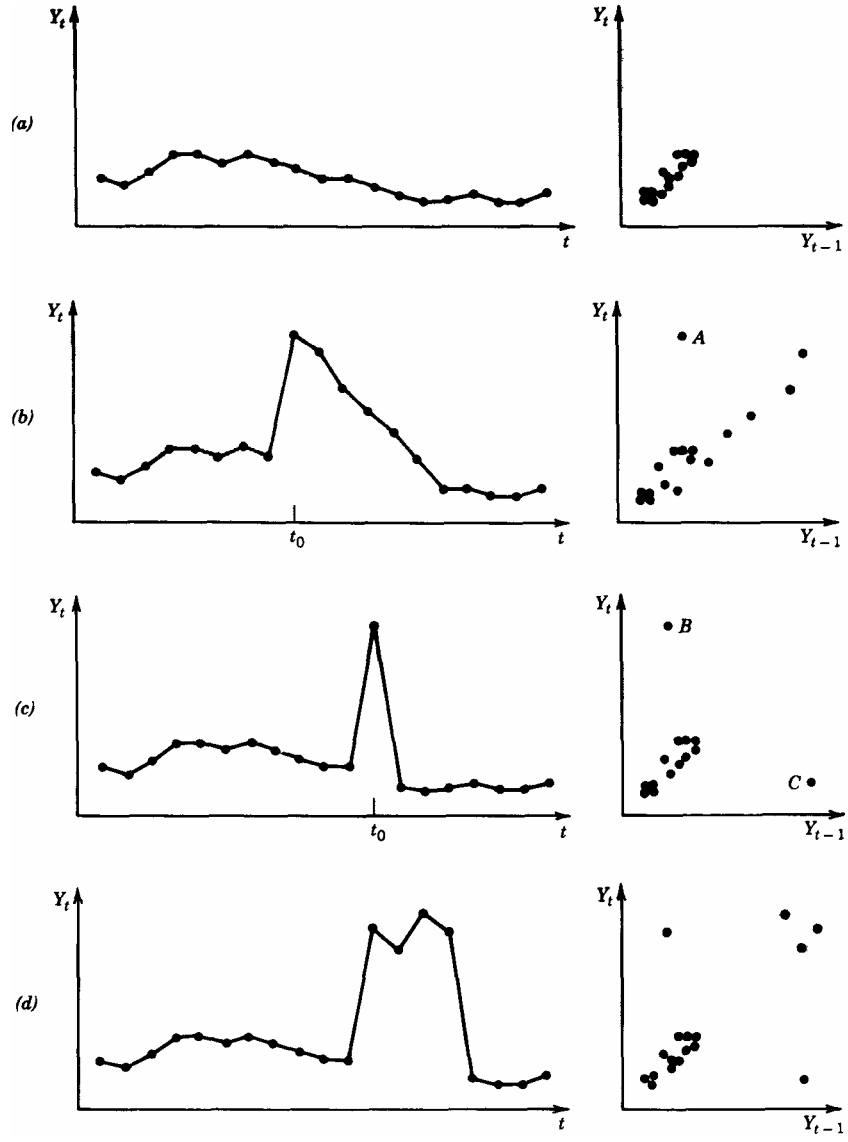
²⁰ u_i : Hata Terimi

²¹ ing. isolated outliers

ABHO modeli sadece yenilikçi UD içeriyorsa M ve GM tahmin edicileri kullanılabilir (Rousseuw ve Leroy, 1987).

Zaman serilerinde 4 durum ile karşılaşmaktadır (Barnett ve Lewis, 1994):

- Bütün UDler katkısız UD olabilir.
- Bütün UDler yenilikçi UD olabilir.
- Bütün UDler aynı çeşit olabilir, ama çeşidi bilinmeyebilir.
- Her iki UD çeşidinde mevcut olabilir.



Şekil 2.5. $Y_t - Y_{t-1}$ zaman serileri serpilme diyagramı²²

²² Rousseeuw ve Leroy, 1987: 276

2.2. İlk UD Tespit Yöntemleri

Bu kısımda UD tespit yöntemlerine kuşbakışı bir bakış sağlamak için yöntemler ilgili alt başlıklar ile gruplanarak sunulmuştur.

2.2.1. Doğrudan Yöntemler

Doğrusal model için doğrudan yöntemler veri miktarı ile oynayarak UD'leri tespit prensibine dayanır. Bunlardan geri tarama²³ yöntemlerinde ilk önce bütün veri kümesi ele alınır ve kullanılan yöntemdeki kritere göre UD olarak tespit edilen veriler elenir ve UD'siz bir küme elde edilmeye çalışılır. Bir başka yöntem ileri tarama²⁴ yöntemidir. Bu yöntemde UD'siz bir veri altkümesi tespit edilir ve bu altküme ile uyumlu veriler altkümeye katılarak altküme büyütülür. Bu süreç bazı veriler altkümeye katılmak için uygun kritere ulaşamayınca kadar devam eder ve bu kalan veriler UDler olur. Geri tarama yöntemleri tercih edilmemektedir (Atkinson ve Riani, 1997). Bunun en basit açıklaması daha önceki bölümde anıldığı gibi çok sayıda UD bulunması durumunda bunların kendilerini maskeleyebilecekleri ve UD olmayan verileri de UD gibi göstermesine sebep olacağı sebebi sayılabilir. İleri tarama yöntemlerinde öne çıkanlar şunlardır:

Hadi ve Simonoff (1993) yönteminde, EKK'den elde edilen düzeltilmiş

²³ ing. backward search

²⁴ ing. forward search

kalıntılardan²⁵ mutlak değeri en küçük olanlardan, modeldeki parametre sayısından bir fazlası adedinde, temel bir altküme ele alınır. Verilerin yarısı kadarını içerecek şekilde EKK'e göre düzenlenmiş kalıntıların en küçük değerlerini kullanarak bir alt küme oluşturulur. Bu verilere ilaveten t dağılımına göre tespit edilmiş bir kritik değere kadar olanlar dışındakiler UD olarak tespit edilir. Hadi ve Simonoff (1997) bu algoritmaya bir geliştirme önermişlerdir.

Pena ve Yohai (1995) etki matrisi²⁶ algoritması, etki matrisini kullanarak birim vektörlerinde çökme noktalarını araştırır. Etki matrisi EKK kalıntılarının köşegen matrisi ve varyansın temsilcisi bir matris ile işleme tabi tutulduğu bir yöntemdir. Eğer bu matris içerikleri 2,5 değerini geçerse o değerler UD olarak tanımlanır. Fakat prosedür nadiren de olsa, UD olmayan verileri UD olarak tespit etmektedir.

Swallow ve Kianifard (1996) yineleyerek kalıntı ileri taraması yönteminde ilk defa çoklu UD tespiti için UDlerden etkilenmeyecek standart sapma kullanmaları bir yeniliktir. Algoritma EKK artıklarını hesaplayarak standart sapmaları bulur, fakat kaldıraç noktalarına oldukça duyarlıdır.

²⁵ Düzeltilmiş kalıntılar $a_i = e_i / \sqrt{1 - p_{ii}}$ şeklinde tanımlanmaktadır. Formülde e_i hata terimini, p_{ii} ise P ($P = X(X^T X)^{-1} X^T$) matrisinin i . köşegen elemanını göstermektedir (X , $n \times k$ boyutlu k açıklayıcı değişkenlerinden oluşan matris).

²⁶ ing. influence matrix

2.2.2. Kümeleme Yöntemleri ile UD Tespiti

UD tespiti için verilerin birbirlerine yakınlıklarını ölçüp kümeler oluşturarak verilerin çoğunun oluşturduğu kümeden uzak ve azınlıkta olan verileri UD olarak tespit mantığına dayanmaktadır. Bunu başarmak için ilk ortaya atılan yöntemlerden birisi MU²⁷ olmuştur. Klasik fakat dayanıklı olmayan bir yöntemdir. Bütün veriler kullanılarak verilerin merkezi noktası ve sapması tespit edilir, fakat birden çok UD bu istatistikleri olumsuz etkilemekte ve UDler kendilerini maskeleyebilmektedir. Hesaplaması kolay olmasına rağmen amaç fonksiyonunda verilerin tamamını dikkate alan bir mesafe ölçütü ortaya koyduğu için UD'leri tespit edememektedir (Rousseeuw ve Leroy, 1987; Rousseeuw ve Zomeren, 1990). Çok Değişkenli Kırpılma²⁸ ve En Yüksek Olabilirlik Kırpılma²⁹ gibi yöntemlerin hepsi MU'nı temel aldıkları için sağlamlıkları tartışılabilir. Bu yöneme alternatif takip eden yöntemler geliştirilmiştir.

X artıkları (Martens ve Naes, 1989) gibi klasik tekniklerde her veri için orijinal verilerin bağımsız değişkenleri için bir toplam sapma değeri ve tahminler için toplam bir sapma değeri hesaplanır ve karşılaştırılır. Anormal değerler UD olarak tespit edilir.

EKOD (Rousseeuw, 1984) ilk ve halen yaygınlıkla kullanılmakta olan bir SR kümeleme yöntemidir. Bu yöntemde amaç verilerin yarısına yakın bir altküme ile bu verilerin varyans-kovaryans matrisini hesaplayıp bu matrisi en küçük yapacak temel veri altkümesini tespit etmeye çalışmaktır. Daha sonra bu temel altküme etrafında kıkare

²⁷ Mahalanobis Uzaklıkları, ing. Mahalanobis Distance

²⁸ ing. Multivariate Trimming, MVT

²⁹ ing. Maximum Likelihood Trimming, MLT

dağılımına göre kritik mesafe tespit edilir ve bu alanın dışında kalan veriler UD olarak tespit edilir. EKOD çökmeye karşı dayanıklı en yüksek değere sahiptir. (Rousseeuw ve Leroy, 1987; Lopuhaa ve Rousseeuw, 1991). İlerleyen kısımlarda EKOD yönteminin gelişmiş versiyonu olan HEKOD³⁰ yönteminin altında daha detaylı açıklaması bulunabilir. Birden çok altküme ile hesaplama yapmak gerektiğinden veri sayısı çok olduğunda kesin sonuca ulaşmak imkansız olmakta ve özel durumlar dışında sadece birden çok rasgele seçilmiş altküme için yaklaşık tahminler hesaplanmaktadır (Hardin ve Rocke 2004). Kikare dağılımına göre kritik değer bulmak yerine Hardin ve Rocke (1999) F dağılımı yönteminin ve Hardin ve Rocke (2004) çoklu kümeler için F dağılımı yönteminin daha verimli olduğunu göstermektedirler. Hardin ve Rocke (1999) F dağılımı yönteminde UDler F dağılımının yüzdelik dilimlerindeki kümelenmelere bakarak ve UD bütün kümelerden uzak olmasına göre tespit edilebilmektedir. Bu yöntemleri kullanırken veri kümesinde yer alan grupların bilindiği, eliptik ve sabit olduğu varsayılmaktadır ki UD tespiti için büyük bir kısıtlama olmaktadır.

EKHE³¹ (Rousseeuw, 1984) tahmin yönteminde amaç verilerin yarısına yakınına içeren bir elips tespit edip bu elipsten belirli sınır dışında kalanları UD olarak değerlendirmektir.

Sebert vd. (1998) Kümeleme³² Algoritması çalışmasında, EKK'den elde edilen standartlaştırılmış (ilgili değer ve standart sapma oranı) tahmin değerleri ve

³⁰ Hızlı En Küçük Varyans-Kovaryans Determinantı, ing. FAST-MCD

³¹ En Küçük Hacimli Elipsoid, ing. Minimum Volume Elipsoid, MVE

³² ing. Clustering

standartlaştırılmış artık değerleri için Euclidean matrisi ile kümeleme algoritmasını kullanır. Algoritmanın önemli noktası, tek en büyük kümeyi bulmak ya da verilerin çoğunluğunu UD dışı olarak sınıflandırmasıdır.

2.2.3. Görsel Yöntemler

İki parametreye kadar doğrusal model ile çalışılması durumunda görsel olarak UDler uygun çizimle tespit edilebilmektedir. KAMda³³ (Preparate ve Shamos, 1985) verilerin her koordinatı için merkez nokta tespit edilir ve bu merkezden en yüksek uzaklıklar hesaplanır. Bu uzak noktalar bütün verileri içerecek biçimde birleştirilir ve içbükey bir alan oluşturulur. Bu alan dışında kalan tahmin noktaları UD olarak tespit edilir.

Bazı grafik yöntemler UDleri tespit etmek için çok boyutlu regresyonda mümkün olduğu kadar az bilgi kaybı ile boyut düşürmeyi ve grafik yardımı ile UDleri tespit etmeyi amaçlamaktadır. Bunlar DOVT³⁴ (Cook ve Weisberg, 1991), DGR³⁵ (Li, 1991), Temel Hessian Yönler³⁶ (Cook, 1998; Li, 1992) ve Grafik Regresyon (Cook, 1998) yöntemleridir. Amaç farklı yöntemlerle merkezi alt uzayı iki ya da üç boyutlu uzaylara indirgeyip görsel olarak tahmin ederek UDleri açığa çıkarmaktır. Bu yöntemlerle ilgili literatür taraması Cook (1998) çalışmasında bulunabilir. DOVT yöntemi DGR

³³ Konveks Ayıklama Metodu, ing. Convex Hull Method

³⁴ Dilimli Ortalama Varyans Tahmini, ing. Sliced Average Variance Estimation, SAVE

³⁵ Dilimli Geri Regresyon, ing. Sliced Inverse Regression, SIR

³⁶ ing. Principal Hessian Directions

yöntemine göre daha kapsamlı olarak bilgi içermesine karşın artan değişken sayısı ile hesap yükü artacaktır. DOVT ve DGR birbirlerini tamamlayacak bir şekilde kullanılmaları daha uygun olmaktadır (Cook ve Critchley, 2000).

Belirsiz tahminlerin kullanımı³⁷ için Martens tarafından bir formül geliştirilmiş ve Hoy vd. (1998) bu formülü iletmişlerdir. Bu formül sayesinde her veri için bir belirsizlik değeri hesaplanabilmekte ve yüksek belirsizlik değerleri UD olarak değerlendirilebilmektedir. Pierna vd. (2002) bu değer UD tespitinde kullanılabildiğini göstermektedirler. Fakat Küçükkocaoğlu ve Kiracı (2003) çalışmasında yer alan, önceki bölümde verilen, şekil 2.4.'te veri yoğunluğu arttıkça sınıra yakın UD'lerin görsel olarak tespit edilemediği görülmektedir.

2.2.4. Dolaylı Yöntemler

SR literatüründe bu yöntemler UDleri tespit etmek için doğrudan UD için müdahale etmeden tahmin elde edildikten sonra UD tespit etme mantığına dayanmaktadır. Bu yöntemlerden ilki M tahmin edicileridir (Huber, 1973). Burada amaç kalıntıların kareleri toplamı yerine yine amaç fonksiyonunda bütün verileri dikkate alarak kalıntıları farklı bir şekilde (fonksiyon olarak) değerlendirmektir ve bu sebeple çökmeye dayanıklı değildir (Rousseeuw ve Leroy, 1987). M tahmin edicilerin en eskilerinden birisi olan EKMS³⁸ yönteminde kalıntıların mutlak değerleri toplamı

³⁷ ing. uncertainty estimates

³⁸ En Küçük Mutlak Sapma, ing. Least Absolute Deviation, LAD

indirgenmektedir. Bütün verileri dikkate alması sebebi ile UD'lerden etkilenmekte ve aynı EKK etkilerine maruz kalmaktadır (Rousseeuw, 1984).

Huber (1973, 1981) M tahmin edicilerini EKK yöntemine alternatif olarak tanımlamıştır. Bu yöntem tekrarlanan regresyon ve bu tekrarda sürekli verilerin ağırlığının değiştirilmesi esasına dayanmaktadır. Kalıntıların mutlak değerleri belirli bir eşik değerinin altında olunca tam ağırlıkla regresyona katılır, fakat eşik değerini geçince daha düşük ağırlıkla değerlendirilir. Veri kümesinde bir tane UD bulunması durumunda normallik çizimine bakarak bir tane UD tespit edilebilmekte, fakat birden çok UD bulunması durumunda UD'lerin birbirlerini maskeleymesi sebebi ile bu yöntem başarısız kalabilmektedir. Huber M tahmin edicisi UD'lere dayanıklıdır, fakat kaldıraç verilere karşı dayanıklı değildir. Bunun üzerine GM³⁹ tahmin edicisi Maronna vd. (1979) tarafından bulunmuştur. Amaç verilerin etkisini düşürmek için onlara değişken ağırlık tanımadır. GM tahmin edicisi SR tekniği olmasına rağmen, çökmeye çok dayanıklı değildir, çünkü çökmeye dayanıklılık noktası üst sınırı parametre sayısının tersi kadardır. Tukey Biweight Regresyonu (Mosteller ve Tukey, 1977) yönteminde farklı bir fonksiyon kullanılmıştır ve yöntem Huber M'den daha yaygın kullanılmakta fakat Huber M birden çok UD'ler tespit durumunda daha başarılı olabilmektedir (Hund vd., 2002).

ÇDTlerden, S tahmin edicisi Rousseeuw ve Yohai (1985) tarafından tanıtılmıştır. Amacı kalıntıların bir fonksiyonu olan sapma fonksiyonunu indirgemektir. Bu bağlamda türetilmiş olan Tau tahmin edicisinde Yohai ve Zamar (1988) standartlaştırılmış kalıntı

³⁹ Genelleştirilmiş M

değerleri ile varyansın birlikte formülünü indirgemeye çalışmaktadırlar.

EKKK⁴⁰ (Rousseeuw, 1984) tahmin edicisi ilerleyen kısımlarda detaylı bir şekilde açıklanacaktır.

EKHSK⁴¹ (Rousseeuw, 1984) tahmin edicisinde amaç öyle bir tahmin bulmak ki bu tahmin sonucu oluşan kalıntıların kareleri büyükten küçüğe sıralandığında hedef alınan belirli bir sıradaki kalıntının karesi başka tahminlerin aynı sıradaki sıralanmış kalıntı karesi değerinden daha düşük olsun. Eğer medyan kalıntı değeri indirgenmeye çalışılırsa buna özel olarak EKOK⁴² denilmektedir (Hampel, 1975; Rousseeuw, 1984). Bu tahmin edicilerin en çok eleştirilen özelliği asimptotik olarak etkin olmamasıdır. Bu yöntemin farklılaşmış versiyonunu UESK⁴³ ile Lee vd. (1998) geliştirmişlerdir.

MU tahmin (Yohai, 1985) edicisi çökmeye dayanıklı ve yüksek etkinliğe sahip üç aşamadan oluşan bir tahmin edicidir. İlk aşama S tahmin edicisini kullanır ve kalıntıları elde edilir. İkinci aşamada, kalıntıları kullanarak M tahmin edicisi hesaplanır. Son aşama olarak, çok yüksek kalıntılara 0 ağırlık veren bir fonksiyon ile M tahmin edicisi hesaplanır. Bunun gibi türetilmiş tahmin ediciler sayesinde SR literatürü çok zenginleşebilmektedir.

Projeksiyon İzleme⁴⁴ (Li ve Chen, 1985) Metodu temel bileşenler analizinden

⁴⁰ En Küçük Kırpılmış Kareler, ing. Least Trimmed Sum of Squares, LTS

⁴¹ En Küçük Hedeflenmiş Sıralı Kalıntı Karesi, ing. Least Quantile of Squares, LQS

⁴² En Küçük Ortanca Kalıntı Karesi, ing. Least Median of Squares, LMS

⁴³ Uyumlu En Küçük Sıralı Kareler, ing. Adaptive Least kth Order Least Squares, ALKS

⁴⁴ ing. Projection Pursuit

daha sağlam Huber M tahminini kullanarak ve medyan değerlerle çalışarak daha sağlam tahminler elde etmektedir.

Coakley ve Hettmansperger (1993) tahmin edicisi EKKK metodunu kullanır. Sonuçları da belirli ağırlıklarla tekrar hesaplar. Ağırlıklar EKHE ile elde edilir. Böylece dayanıklılık elde edilir.

EKKMK⁴⁵ (Bassett, 1991) tahmin edicide amaç öyle bir tahmin bulmak ki bu tahmin sonucu oluşan kalıntıların mutlak değerleri büyükten küçüğe sıralandığında hedef alınan belirli bir sıradaki kalıntıdan küçük bütün kalıntılar toplanmakta ve bu toplam başka tahminlerin sonucu oluşan toplamdan daha düşük yapılmaya çalışılmaktadır. Büyük veri setleri için uygulanması ve hesaplanması kolay, istatistiksel açıdan etkin bir tahmin edicidir. Veri setlerinde kayıp değer olduğunda onların tahmininde de işe yaramaktadır.

EKÇF⁴⁶ (Croux vd., 1994) tahmin edicide amaç kalıntı çiftlerinin farklarının mutlak değerlerini sıralayıp en küçük çeyreği minimize etmektir. Fakat EKÇF modelde kesişim parametresini içermemektedir.

EKKF⁴⁷ (Stromberg vd. 1995) tahmin edicisinde amaç kalıntı çiftlerinin farklarının karelerinin en küçük çeyreğinin toplamını minimize etmektir.

⁴⁵ En Küçük Kırılmış Mutlak Kareler, ing. Least Trimmed sum of absolute deviations, LTA

⁴⁶ En Küçük Çeyrek Farkları, ing. Least Quartile Difference, LQD

⁴⁷ En Küçük Kırılmış Farklar, ing. Least Trimmed Differences, LTD

Farklı bir OO⁴⁸ sınaması testini Wang vd. (1997) UD tespiti için uygulamışlardır. Bu yöntem sayesinde farklı dağılımlardan gelmiş veriler içinden UD'leri tespit edebilmişlerdir. Wang vd. (1997) çalışmasında yapılmış simülasyonlarda testin gücü yüksek çıkmıştır.

TBA⁴⁹ yönteminde veri kümesindeki varyasyonun çoğunu mümkün olduğu kadar çok değişken için veren temel bileşenler aranmakta ve veri miktarını indirgemekte kullanılmaktadır. Zhang vd. (1999) ve Lalor ve Zhang (2001) tarafından UD tespitinde kullanılmıştır.

2.2.5. Karşılaştırmalar

Tablo 2.4. EKOK, EKKK, EKÇF ve EKKF yöntemlerini karşılaştırmaktadır. Bu yöntemler yaygın olarak kullanılması ve programlarının herkese açık olması sebebi ile tercih edilmiştir. EKÇF ve EKKF tahmin edicileri verideki küçük kaymalara dayanıklı olmamalarına rağmen yüksek etkinlikleri bu tahmin edicileri EKOK ve EKKK'e göre tercih edilir yapmaktadır. Buna ilaveten EKKF ve EKÇF normal veriler için EKÇF ve EKKK'e karşı üstünlüğü vardır. Hettmansperger ve Sheather (1992) ve Sheather vd. (1997) verilerde ufak kaymalarda EKÇF (EKOK) tahmininin dikkate değer bir biçimde oynadığını belirtmektedirler.

⁴⁸ Olabilirlik Oranı, ing. Likelihood Ratio, LR

⁴⁹ Temel Bileşenler Analizi, ing. Principal Component Analysis, PCA

Tablo 2.4. EKOK, EKKK, EKÇF ve EKKF Karşılaştırması

	Gauss etkinliği	Çökme noktası (asimptotik olarak tahmini)
EKOK	%0 ^a	%50 ^c
EKKK	%7 ^b - %8 ^a	%50 ^c
EKÇF	%67 ^a	%50 ^a
EKKF	%66 ^a	%50 ^a

Açıklama: ^a Stromberg vd. (2000), ^b Agullo (2001), ^c Rousseeuw ve Leroy (1987)

Hawkins ve Olive (1999) çalışmalarında EKKMK, EKOK ve EKKK metotlarını karşılaştırmış ve bu yöntemlerin daha kolay uygulanması için algoritmalar geliştirmişlerdir. EKKK yöntemindeki etkinlikten biraz fedakarlık yapılarak EKKMK için kesin sonucun daha kolay elde edilme durumunun mümkün olduğunu göstermişlerdir. Bu çalışmanın gözlemlerine göre veri kümesi büyük olduğunda EKKMK yönteminin EKOK ve EKHSK yöntemlerine karşı çok cazip bir seçenek olduğu görülmektedir. Buna ilaveten EKKMK yönteminin etkinliğinin EKKK kadar olduğunu belirtmekte ve kayıp veriler olması durumunda bile çalıştığını göstermektedir. EKOK ve EKKK yöntemleri gibi EKKMK, MU, S ve Tau tahmin edicileri için ilk tahmin olarak kullanılabilir ve hesaplanması daha kolaydır.

Yukarıda anılan EKKMK, EKHSK ve EKKK tahmin edicileri yapmış oldukları tahminlerde UDleri tespit etmede verilerin büyük bir kısmını dikkate alarak tahmin yapmakta ve bu verilerin davranışı dışında bir davranış gösteren verileri UD olarak tespit etmekte ve çoğu zaman başarılı olmaktadır. Fakat veri kümesinde birden çok UD bulunması ve bu verilerin diğer verilerin büyük bir kısmı ile doğrusal uyumluluk

göstermesi durumunda sağlam tahmin edicilerin tam oturma⁵⁰ özelliği UDleri tespit edememektedir ve hatta veri sıkışması problemi ortaya çıkmaktadır. Stromberg (1993) bu istikrarsızlığa çare olarak EKHSK için verilerin sadece yarısının eğilimine göre değil de bütün olası veri kapsamları için tahmin yapıldığında bu problemin olmayacağını söylemektedir. Bir başka problem tahmin edici aranırken verilerin büyük kısmı dikkate alındığından geri kalan veriler diğer verilerle uyumlu olması durumunda dahi UD olarak değerlendirilmektedir. Bu sebeple uyumlu verilerin dahil edilmesi konusunda daha sonra bazı sorunlar ortaya çıkmaktadır.

Stromberg vd. (2000) çalışmalarında EKKF anlatılmaktadır. Hesaplanması zor olmasına rağmen, uygun bir çözüm algoritması kullanılarak elde edileceğinden ve diğer ÇDTe göre daha güvenilir olduğundan söz etmektedirler.

Agullo (2001) çalışmasında, EKKK tahmin edicisi için iki algoritma geliştirmiştir. İlk algoritma tahmini sonuç üretir, fakat ikinci algoritma kesin sonucu verir ve ayrıntılı dallama-budaklama tekniğine dayalıdır. Gerçek ve simulasyon verilerine uygulanması sonucunda göstermiştir ki, literatürdeki tekniklere nazaran daha az hesaplama maliyeti mümkündür.

Yukarıdaki algoritmaların hepsinde veri kümesinin belirli miktarda bir altkümüsi kullanılarak tahmin yapılmakta ve tahmin ile kalıntılar hesaplanmaktadır. Daha sonra bu kalıntılar kullanılarak amaç fonksiyonunun değeri hesaplanmakta ve alt kümeler içinde amaç fonksiyonunu en iyi bir biçimde sağlayan en iyi alt küme tespit edilmektedir. Eğer

⁵⁰ ing. exact fit

veri sayısı az ise oluşturulabilecek bütün alt küme kombinasyonları için bu işlemler tekrarlanmakta, fakat veri sayısı çok olması durumunda bu işlemlerin hepsini tamamlamak imkansızlaşmaktadır. Böyle durumlarda çok defalar rasgele veri alt kümeleri alınmakta ve iyi bir tahmin aranmaktadır.

2.3. Güncel UD Tespit Yöntemleri

Güncel UD tespit yöntemleri daha önce bahsi geçen kısıtlayıcı varsayımlarda azalma sağlamak ve daha geniş bir veri yelpazesinde daha performanslı çalışmaktadırlar. 2000 yılından önceki yöntemler için Melouna ve Militký (2001) iyi bir kaynaktır. Bu sebeple 2000 yılından sonra gelişmeler ne yönde olmuş aşağıda anılmaktadır.

Arslan ve Billor (2000) çalışmasında M tahmin edicisine dayalı LM⁵¹ tahmin edicisi anlatılmıştır. Sonuçlara göre, Liu tahmin edicisi bağımlı değişkendeki büyük değişimlerden etkilenmektedir. Fakat LM yöntemi UD'lerin etkisini azaltmaktadır.

ÇRA⁵² yönteminde her değişken bağımlı değişken gibi düşünülüp sırayla diğer bütün değişkenlerle regresyona tabi tutulur. Her veri için değişkenlerin değeri ile regresyon sonucu çıkan tahmin arasındaki farkların mutlak değerleri toplanır. Bu toplam ne kadar yüksekse UD olma olasılığı o kadar yüksek olacağı düşünülür. Lalor ve Zhang (2001) tarafından UD tespitinde kullanılmıştır.

⁵¹ Liu tipi M

⁵² Çoklu Regresyon Analizi, ing. Multiple Regression Analysis, MRA

Jiang, Tseng ve Su (2001) çalışmasında iki kademeli bir kümeleme algoritması UD tespitinde kullanılmaktadır. Aynı kümede olan veriler aynı özelliklere sahip olur, yani aynı küme içinde hepsi UD olabilir ya da hepsi UD olmayabilir. Birinci aşamada veri belirli miktarda altkümelere ayrılır ve bu kümelerin sayısı sabit kalacak bir şekilde bir bağlantı bulununcaya kadar iyileştirme yapılır. İkinci kademede kümelerin merkezi budak olacak şekilde ağaçlar oluşturulur ve iki yeni alt ağaç oluşacak şekilde en uzun dal bölünür. Az sayıda dalı olan ağaçlar UD içerdiği varsayılır ve kırılır.

Wisnowski vd. (2001) çalışmasındaki yöntemler daha düşük boyutlarda, az UD yüzdesine sahip verilerde, düşük kaldıraç veri uzaklıkları, yüksek UD kalıntı uzaklıkları, yüksek sayıda çoklu nokta kümelerinin olduğu verilerde daha iyi sonuçlar vermektedir. Kalıntı uzaklıkları yüksekse Hadi ve Simonof (1993) versiyonu tavsiye edilmektedir. Simpson ve Montgomery (1998) tahmin edicileri ile EKKK-EKOD (Rousseeuw ve Zomeren 1990) metodu SR yöntemleri içinde en iyi sonuçları vermiştir.

Kwon vd. (2001) iki farklı HU^{53} kullanmışlardır. Piramit yapısını kullanan HU eşleştirme algoritması hem gerçek hem de simulasyon verilerine uygulanmıştır. Bu yöntemlerden ilki, M-HU, M tahmin edicisine dayalıdır, ikincisi, EKKK-HU, EKKK metoduna dayalıdır. Bu yöntemler hesaplama zamanını kısaltmaktadır ve etkinliği çeşitli uygulamalarla test edilmiştir.

Rio, Riu ve Rius (2001) çalışmasında BKK⁵⁴ yöntemini kullanılarak UDLeri

⁵³ Hausdorff Uzaklığı, ing. Hausdorff Distance

⁵⁴ Bivariate Küçük Kareler, ing. Bivariate Least Squares, BLS

tespit etmek için Cook (1977) çalışmasının mantığını temel alarak grafiksel kriter geliştirilmiştir. BLS parametre hesaplamaları yaparken y ve x eksenlerdeki sapmaları dikkate alarak tahminde bulunmaktadır.

Hund vd. (2002) çalışmalarında iki teknik açıklanmıştır. Bunlar EKK artıklarının yarı normal çizim grafiği ve etkilerin normal olasılık çizim grafiği yöntemleridir. UD tespitinde daha duyarlı olan Huber M ve Tukey Biweight regresyonunun EKK ile karşılaştırmalı uygulamaları verilmiştir. Bu sayede iki grafiksel teknik UDleri tespit etmeye çalışmaktadır.

Arslan vd. (2002) makalelerinde, kısıtlanmış M tahmin edicisi⁵⁵ ve S tahmin edicilerini açıklamışlardır. Kısıtlanmış M tahmin edicileri (Mendes ve Tyler, 1995) çökmeye dayanıklılığı ve yüksek asimptotik etkinliğe sahip olmasıyla SR tahmin edicilerine alternatif olarak ortaya çıkmıştır. Kısıtlanmış M tahmininde kullanılan algoritma S tahmin edicileri içinde düzenlenebilir. Kısıtlanmış M ve S tahmin edicilerinin EKK ile karşılaştırılması yapılmıştır.

Pierna vd. (2002) çalışmalarında, OF⁵⁶ (Rimbaud vd. 1999) diğer teknikler ile karşılaştırmak için kullanılmıştır. Sonuç olarak, diğer metotlar ile karşılaştırıldığında belirsiz tahminlerin kullanımı ve KAM UDlerin tespiti için daha pratik yol olduğu bulunmuş, fakat OFin da eklenmesiyle UD olmayanlar tespit edilebilmektedir. KAM,

⁵⁵ ing. Constrained M Estimators

⁵⁶ Olası Fonksiyonlar, ing. Potential Functions

UD tespiti için görsel olarak iyi bir metot; BM⁵⁷ ise, klasik metotlara benzemektedir. Her iki metot da pratik birer alternatiftir. Karşılaştırmalar sonucunda ek olarak bir bilgi içeren tek metot OFMdur.⁵⁸ SR metotların dezavantajı, her yeni bir örnek eklendiğinde tüm modelin elden geçirilmesi gerekliliğidir. Belirsiz tahminlerin kullanımı, Konveks Ayıklama Metodu ve Olası Fonksiyonlar metodu birlikte kullanıldığında UD ve iyi verilerin tespiti pratik bir yoldur.

Hubert vd. (2005) çalışmalarında sağlam bileşenler analizi için yeni bir yöntem tanıtmıştır. Klasik TBA metodu verinin kovaryans matrisine dayalıydı ve UD'lere karşı oldukça duyarlıydı. Geçmişte iki sağlam yöntem geliştirilmiştir. İlki, birim vektörlere dayanan düşük boyutlu verilerle sınırlanan, EKOD ve S tahmin edicileri gibi yöntemler, ikincisi, çok boyutlu verilerle çalışan, Projeksiyon İzlemeye dayalı yöntemlerdir. STBA⁵⁹ yöntemi her iki yöntemi de kapsayan yeni bir yöntemdir. STBA metodu, daha dayanıklı tahminler yapmakta ve hesaplanması daha hızlıdır. Bu tekniklerin yüksek boyutlu olmaları önemlidir.

Wang ve Suter (2003) çalışmalarında çökmeye dayanıklı tahmin edicilerde görülen veri sıkışması etkisini ortadan kaldırmak için EKSF⁶⁰ yöntemini önermektedirler ve örneklerle EKOK ve EKKK'den daha performanslı olduğunu göstermektedirler. Bu yöntem EKKK ile birlikte simetrik mesafe ölçütünü içerecek şekilde tasarlanmıştır, fakat bu simetrik hesaplamalar daha fazla zaman gerektirmektedir.

⁵⁷ Belirsizlik Metodu

⁵⁸ Olası Fonksiyonlar Metodu

⁵⁹ ing. Sağlam Temel Bileşenler Analizi, ROBPCA

⁶⁰ En Küçük Simetri Farkı, ing. Least Trimmed Symmetry Distance, LTSD

SEKKMK, SEKHSK, SEKKK (Olive ve Hawkins, 2003) tahminlerin türetiliş amacı, daha önce anılan klasik EKKMK, EKHSK ve EKKK tahmin ediciler üzerinde ilaveler yapılarak daha dirençli olan SEKKMK, SEKHSK, SEKKK tahminlerinin elde edilmesidir. Amaç kapsanan veri miktarına bağlı olarak sağlam parametre tahmini yapmak ve bir ayarlama parametresi belirleyip hedef alınan kalıntı değerinden büyük kalıntı değerini UD olarak tespit etmektir. Bu sayede klasik SR parametrelerinin UD olarak nitelediği verilerin bir kısmı bir parametreye göre temel veri miktarı olarak tespit edilmektedir. Bu yöntem literatürde detaylı bir şekilde incelenmemiş olup UDleri hangi oranda tespit edebiliyor, maskeleye ve veri sıkışmasına karşı ne oranda başarılı incelemek gerekmektedir.

Wu ve Chow (2004) Kohonen tarafından önerilen KOH⁶¹ yöntemi bir sinirsel ağ uygulamasıdır.

Dahl ve Naes (2004) çalışmasında veri içinde birbirinden farklı gruplar olduğu düşünülürse Procrustes Mesafesi ile Hiyerarşik Kümeleme önermektedir. Bu sayede verideki farklı yapılar veya UDler tespit edilebilmektedir. Verilerin Procrustes Mesafesi ve Hiyerarşik Kümeleme Analizi, birbirine yakın verileri birararaya getirme prensibine dayanmaktadır.

Multihalver (Fernholz vd., 2004) metodunda veriler iki eşit parçaya bölünüp her parça için ve parçaların birbirlerine etkileri (farkları) için istatistikler oluşturulmaktadır. Önerilen algoritma mümkün olduğu kadar çok ve farklı yarıya bölünmüş veri

⁶¹ Kendinden Organize Harita, ing. Self Organizing Map, SOM

altkümeleri için tekrarlanır ve belirli etki sınırının üzerindeki değerler UD olarak tespit edilir.

Zhao vd. (2004) çalışmalarında RBF-PLS, Prescott testi ve Çok-Katmanlı İleri Ağları sinir ağı UD tespiti için önerilmektedir. Amaçları doğrusal olmayan veya model yapısı bilinmeyen durumlarda başarılı UD yöntemi olarak ortaya çıkarmaktır.

DDclust ve DDclass (Jörnsten, 2004) yönteminde bir verinin bulunduğu kümenin derinliği ile komşu kümenin derinlik farkı ve ortalama mesafelerin kümelere göre normalize edilmiş farkların ağırlıklı ortalaması hesaplanmaktadır. Buna ilaveten bu çalışma bir verinin bulunduğu kümeyi temsil etme özelliğini ölçen veri derinliği için bir kategori önermektedir.

Liu, Shah ve Jiang (2004) Eşanlı Veri Filtre-Temizleyicisi yönteminde özellikle vurgu yapılacak nokta veriler toplandıkça eşanlı olarak UDleri tespit etmeye çalışması ve Kalman Filtresi ile birlikte “temiz” UDleri de tespit etmesidir. Diğer metotlara avantaj olarak bu metodun özelliklerinden biri, model ile ilgili ön bilgiye gereksiniminin olmamasıdır, çünkü genelde varsayım UDlerin verilerin büyük çoğunluğunun istatistiksel dağılımını takip etmeyen gözlemler olması ve UD tespit yöntemlerinin çoğu verilerin özdeş ve bağımsız dağılım özelliğine sahip olduğu varsayımdır. Bu yöntemin bir başka üstünlüğü otokorelasyonlu verilere uygulanabilir olması ve eşanlı olarak UD bulup yerine uygun tahmin koymasıdır. Bu yöntemin çalışma prensibi Martin ve Thompson (1982) çalışmasında önerilen ve belirli bir zaman aralığı için veri tahmininin

bir geri zaman aralığı için sağlamlaştırılmış katsayılarla hesaplanması üzerinedir.

Hardin ve Rocke (2004) metodu F dağılımını kullanarak kümeler içinde UD hesaplaması yapmaktadır. Ki-kare ve F testi sonuçlarına göre bir takım sınır değerleri tespit edilmekte, ve bu değerleri aşan veriler UD olarak adlandırılmaktadır.

Filzmoser vd. (2004) çalışmasındaki çoklu UD tespit yönteminde yeni bir metot olarak normal dağılımdan gelen UDLer ile farklı bir dağılımdan gelen (uç) değerleri tespit edebilebilmesidir. Buna ilaveten görsel olarak UDLeri tespit etmek için bir metot sunulmaktadır. UDLeri tespit etmek için temel olarak verilerin UDLerden etkilenmeyen merkezi ve uzaklıklarını tespit için EKOD ve sağlam parametre tahminleri için EKKK kullanmışlardır.

Tao, Wu ve Wang (2004) geliştirdikleri metot ile doğrusal bir modelde bu modelden sabit uzak mesafe dahilindeki noktalara pozitif değer veren geri kalanlara ise negatif değer veren bir fonksiyon tanımlayabilmektedirler. Bu durum başarılı bir istatistiksel öğrenme algoritması olan DVM⁶² tanımlanmasında yardımcı olacaktır. Tax ve Duin (1999) çalışmasındaki Destek Vektör Veri Tanımlaması⁶³ yöntemi gibi burada da amaç bütün verileri içeren en küçük hacmi bulmaktır.

Choulakian (2005) çalışmasında TBA yöntemini bir adım daha ilerleterek L1-norm TBA yaklaşımını oluşturmuştur. Her iki yöntem de çökmeye dayanıklı değildir, fakat bu çalışmada gerekli hesaplamaları yapmak için üç tane algoritma tanıtmakta ve

⁶² Destek Vektör Makinesi, ing. Support Vector Machine

⁶³ ing. Support Vector Data Description

UD tespiti için sınır değerler tespit etmektedir.

2.4. Bu Çalışmada Kullanılan Yöntemler

Bu çalışmada kesit verilerde yer alan UDLeri tespit etmek için iki metot birarada kullanılmıştır. Bu metotlardan ilki Rousseeuw ve Van Driessen (1998) tarafından geliştirilen olan HEKKK⁶⁴ algoritması ile ikincisi Rousseeuw ve Van Driessen (1999) tarafından geliştirilmiş olan HEKOD⁶⁵ algoritmalarıdır. Bu programlar çok boyutlu veya yüksek sayıda veri ile çalışmayı mümkün kılmaktadır ve Internet'ten (<http://www.agoras.ua.ac.be/>) ayrı ayrı tedarik edilebilmektedirler. Bu iki algoritma iki farklı ÇDT üretmektedir ve birarada kullanıldığında sonuçlar üzerinde olumsuz etki yaratan mümkün olduğu kadar az sayıda UDLeri tespit edilebilmektedirler. İşleyiş prensipleri takip eden altbaşlıklarda açıklanmıştır.

2.4.1. Hızlı En Küçük Kırpılmış Kareler

EKKK yöntemi Rousseeuw tarafından (1983, 1984: 876) yüzde 50`ye yakın, $(n-h)/n$, çökmeye dayanıklılık oranı ile sağlam bir tahmin edici olarak literatüre sunulmuştur. EKKK algoritmasının geliştirilmiş algoritması olan HEKKK algoritması uygulanmaktadır. Bunun sebebi ise EKKK yönteminde yüksek boyutlu verilerde yüksek

⁶⁴ Hızlı En Küçük Kırpılmış Kareler, ing. Fast Least Trimmed Squares, FAST-LTS

⁶⁵ Hızlı En Küçük Ortak Varyans Determinantı, ing. Fast-Minimum Covariance Determinant, FAST-MCD

hesaplama zamanları ve hatta sonuca ulaşamayacak kadar çok hesaplama yapma sıkıntısıyla karşılaşılmasıdır.

EKKK tahmin edicisinde amaç (5) nolu denklemde gösterildiği üzere h ($h < n$) tane artık⁶⁶ (r) kareleri toplamını minimize etmektir. Belirli bir tahmin için önce artık kareleri hesaplanır ve daha sonra büyükten küçüğe $(r^2)_{1:n} \leq \dots \leq (r^2)_{n:n}$ şeklinde sıralanır. EKKK yöntemi EKK yöntemine çok benzemektedir, tek farkı n tane gözlem yerine, h tane en küçük gözlem amaç fonksiyonunda dikkate alınır ve büyük artıklar toplam dışında bırakılır. Böylece büyük artıkların tahminleri saptırması engellenmiş olur. h , $n/2$ 'ye yaklaştığında tahmin edici en sağlam sonucu vermektedir.

$$\text{Min}_{\hat{\theta}} \sum_{i=1}^h r_{i:n}^2 \quad (6)$$

$$r^2 = (y_i - \hat{y}_i)^2 \quad (7)$$

θ : Tahmin Edici

y_i : Gözlenen Değer

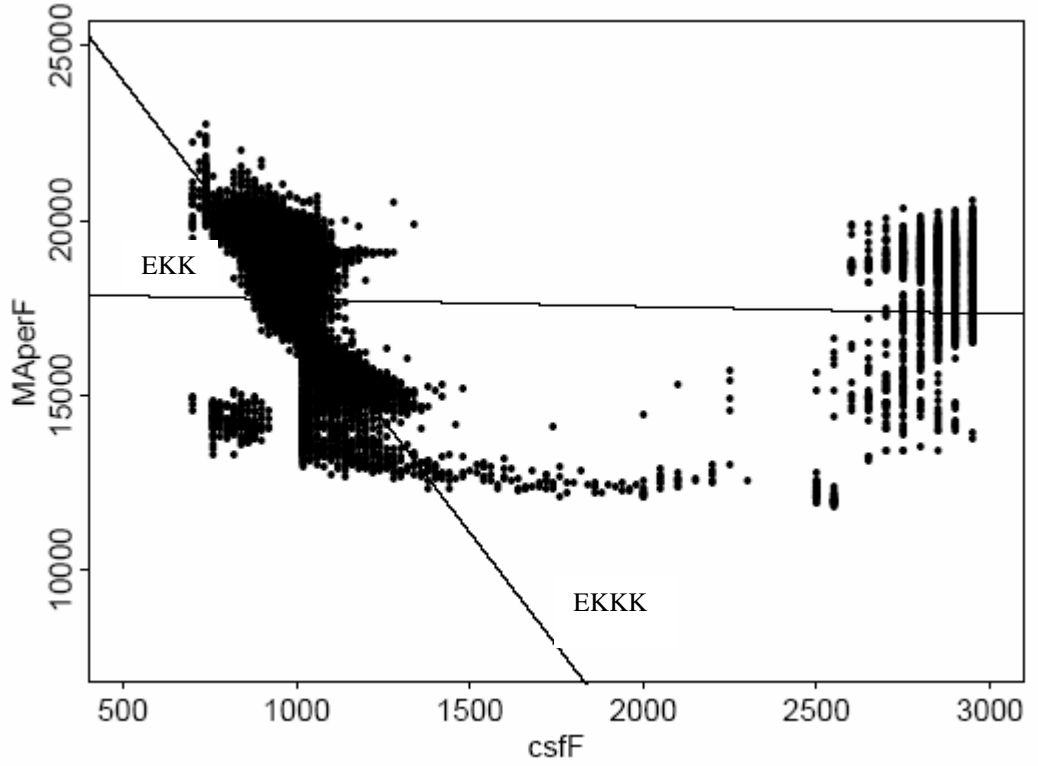
\hat{y}_i : Tahmin Edilen Değer

EKKK yönteminin bazı üstünlükleri, amaç fonksiyonunun başka sağlam regresyon yöntemlerine oranla daha problemsiz olması ve istatistiksel açıdan daha etkin olmasıdır, çünkü EKKK tahmin edicisi asimptotik olarak normaldir (Hössjer, 1994).

⁶⁶ ing. residuals

Geliştirilen HEKKK yöntemi bu konuda meşhur başka yöntemlerden daha hızlıdır ve çok büyük örneklem kümeleri ile çalışabilmektedir. Buna ilaveten küçük ve az boyutlu verilerde kesin EKKK tahmin edicisini ve büyük verilerde hızlandırıcı kısaltmalar ile yaklaşık bir tahmin bulmaktadır. Bu durum algoritmayı daha kullanışlı yapmaktadır. Her iki EKKK algoritmasında amaç, EKK yönteminde olduğu gibi bütün verileri dikkate alarak bir indirgeme yapmak olmayıp, bunun yerine verilerin %50 ya da %75 miktarına yakın bir veri altkümesi miktarının EKK tahminini amaç fonksiyonu olarak belirleyip bu değeri indirmeye çalışmaktır. Amaç fonksiyonunda hangi veri altkümesinin en küçük değeri vereceği ancak deneme yanılma yöntemi ile bulunabilmektedir. Bu durum yüksek hesaplama zamanları ve bilgisayar programlama bilgisi gerektirmektedir ki sağlam regresyonun şimdiye kadar popüler olmamasının önünde en büyük engel olarak bu gerçek durmaktadır.

Takip eden şekil 2.6. HEKKK yönteminin başarısını göstermektedir. Şekilde $n=56,744$ örneklem için HEKKK ve EKK tahminleri karşılaştırılmış ve sonuçlar 3 dakikada elde edilmiştir (Rousseeuw ve Driessen, 1998:4). EKK tahmininin UDlerden olumsuz yönde etkilendiği ve yanlış tahmin doğrusu verdiği görülebilir.



Şekil 2.6. EKKK ve EKK doğrusu karşılaştırması

2.4.2. Hızlı En Küçük Ortak-Varyans Determinantı

Diğer algoritma olan HEKOD ise EKOD diye bilinen en küçük ortak-varyans determinantı algoritmasının gelişmiş bir sürümünü uygulamaktadır. Bu geliştirilmiş algoritma küçük ve az boyutlu verilerde kesin EKOD tahmin edicisini bulmakta veya büyük miktarda verilerde hızlandırıcı kısaltmalar ile yaklaşık bir tahmin bulmakta ve bu durum algoritmayı daha kullanışlı yapmaktadır. Her iki EKOD algoritmasında da amaç verilerin %50 ya da bu civarda bir miktarda veri altkümüsi oluşturmak ve bunun EKOD

tahminlerini hesaplamaktır. Bu mümkün olduğu kadar çok farklı veri altkümelerine defalarca yapılır ve her defasında hesaplanmış amaç fonksiyon değerleri içinde en küçük değeri veren altküme dikkate alınır. Bu veri altkümesi son bir kez daha işleme tabi tutulur ve bu verilerin merkezinden uzaklığa göre mesafeler hesaplanır ve belirli bir mesafeden sonrakiler kaldıraç nokta(lar)⁶⁷ olarak tespit edilirler. Etkili noktalar EKK tahminlerini hem çok olumlu hem de çok olumsuz etkileyen verilerdir. HEKKK'de olduğu gibi amaç fonksiyonunda hangi (%50 ya da %75) veri kümesinin en küçük değeri vereceği ancak deneme yanılma yöntemi ile bulunabilmektedir ve bu durum yüksek hesaplama zamanları gerektirmektedir.

EKOD yöntemi Rousseeuw (1983,1984) tarafından geliştirilen oldukça sağlam bir tahmin yöntemidir. Yöntemin amacı, n gözlem içinden varyans-kovaryans matrisi determinantı en küçük olan h tanesini belirlemektir. Tavsiye edilen $h = \frac{n + p + 1}{2}$

olmasına rağmen $\frac{n + p + 1}{2} \leq h \leq n$ olan herhangi bir tamsayı h olarak alınabilir.

EKOD, EKHE⁶⁸ ile aynı çökme noktasına sahiptir. Her iki yöntemde kümeleme analizi mantığına dayanmaktadır. Eğer α ($0 < \alpha < \frac{1}{2}$) arasında iken, UD oranının α olduğu biliniyorsa, EKOD veya EKHE yöntemleri ile çalışılmaktadır. EKOD yöntemi EKHE yöntemine tercih edilmektedir, bunun için bir çok sebep sayılabilir. EKOD yöntemi EKHE yönteminden istatistiksel olarak daha etkin tahminler vermektedir, çünkü

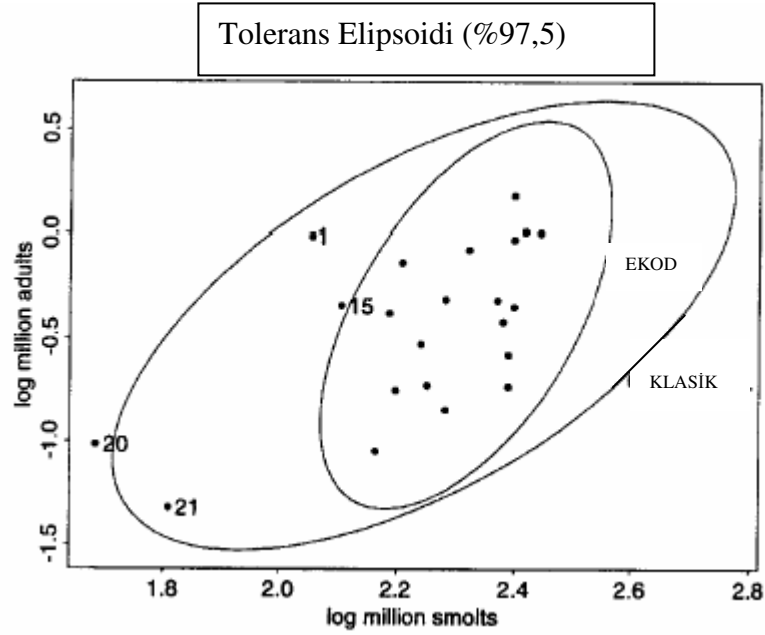
⁶⁷ ing. leverage point

⁶⁸ En Küçük Hacimli Elipsoid, ing. Minimum Volume Elipsoid, MVE

asimptotik olarak normaldir (Butler, Davies ve Juhn, 1993). EKOD ile elde edilen sağlam uzaklıklar EKHE'ye göre daha kesindir ve çoklu UDleri daha iyi tespit edebilir. Bu avantajlarına rağmen EKOD daha az kullanılmaktadır, bunda hesaplama zorluğunun olması başlıca sebep olarak sayılabilir.

Verilerin boyutu ve büyüklüğü EKOD hesaplamalarında sıkıntılar yarattığı için, HEKOD yöntemi geliştirilmiştir. $n > 1000$ büyüklüğündeki veri setlerinin bilinen EKOD yöntemi ile uygulaması yok iken, $n = 50,000$ büyüklüğündeki veri setleri HEKOD yöntemi ile 15 saniyede uygulanabilmektedir. $n = 677$, $p^{69} = 9$ değişkenli ve $n = 137,256$, $p = 27$ değişkenli veri setlerine uygulanmış örnekler Rousseeuw ve Driessen (1999) çalışmasında bulunabilir.

⁶⁹ kesişim parametresi de dahil toplam katsayı miktarı



Şekil 2.7. EKOD tolerans elipsoidi⁷⁰

Bu iki algoritmayı bir arada kullanmaktan amaç birbirlerinin eksik taraflarını tamamlayıp zararlı UDlerin tespit edilmesidir. HEKKK algoritması UDlerin hepsini bulabilmektedir, fakat yüksek standart hata oranına sahip verilerde başarısız kalmaktadır. Yüksek standart hatalı verilerde sadece dikkate alınan düşük standart hatalı veri altkümesi için UD tespit edilmesi, verinin tümünün standart hatası alt kümenin standart hatası ile uyumlu olmadığı durumlarda, gereğinden fazla veri UD olarak tespit edilmesine sebep olmaktadır. Bu durumu HEKOD algoritması çözmektedir çünkü birbirine yakın (cluster) verileri bulmakta ve uzak olanları etkili nokta olarak tespit etmektedir. İki algoritma birleştirildiğinde hem UD hem etkili nokta olan veri altkümesi zararlı UD olmakta ve gerekli tedbirler alınmaktadır.

⁷⁰ Rousseeuw ve Driessen (1999: 220)

BÖLÜM III

3. TÜRK EKONOMİK VERİLERİNDE UÇ-DEĞERLERİN VARLIĞI VE BİLİMSEL ÇIKARSAMA ÜZERİNDEKİ ETKİLERİ

3.1. Türk Modellerinde Uç Değerler ve Etkileri

İkinci Dünya Savaşından sonra Türkiye’de çok partili hayata geçilmesiyle birlikte bu sistemin olumlu yanlarıyla birlikte olumsuz etkileri de kendisini göstermiştir. Bu olumsuzluklardan biri de ekonomide devletin büyük ağırlıkla yer alması sebebi ile ekonomik hayatın bundan etkilenmesidir. Devlette devamlılığın olmaması, politik istikrarsızlık, farklı ideolojik programlar, popülist politikalar ve diğer sebepler ekonomide normal zamanlardan farklı hızlı inişler ve çıkışlara, değişimlere ya da şoklara sebep olmuştur. Bu hızlı değişimlerin düzensiz aralıklarla ya da ekonominin öngöremeyeceği zamanlarda olması bütün ekonomik değişkenlerin bu değişimlere her zamanki gibi tepki vermemesi fikrini akla getirmektedir. Bu fikri temel alan bu bölümde Türk ekonomik verilerinde bu tür değişimlerin UD oluşturup oluşturmadığı ve bu uçdeğerlerin EKK sonuçları ile bilimsel sorgulamayı ne oranda etkilediği incelenmektedir.

Bu fikri sınamak için SSCI⁷¹ taranmış fakat Türk verileri ile EKK yöntemini kullanan çalışma bulunamamıştır. Daha sonra Internet'te yer alan bu özelliğe sahip çalışmalar taranmış ve verileri türetilen ilk iki çalışma örnek olarak sunulmaktadır.

3.2. Para Talebi ile Kayıtdışı Ekonomiyi Ölçmek⁷²

Kayıtdışı faaliyetlerin nakit para ile sürdürüldüğü varsayılacak olursa, kayıtdışı ekonomi Tanzi (1983) çalışmasında varsayıldığı gibi para talebine bakılarak tahmin edilebilir. Bu amaçla takip eden tabloda yer alan değişkenler, beklenen işaretler ve model tanımlanabilir. Tablo 3.2. regresyon sonuçlarını göstermektedir.

Tablo 3.1. Para talebi ile kayıtdışı ekonomiyi ölçmek modelinde değişkenler, beklenen işaretler ve veri kaynakları

Bağımlı değişken		Veri Kaynağı
<i>DRP</i> : Dolaşımdaki reel para		Merkez Bankası
Bağımsız değişkenler	Beklenen işaret	
<i>RGSMH</i> : Reel GSMH	Artı	Merkez Bankası
<i>VRG</i> : Vergi gelirlerinin konsolide bütçeye oranı	Artı	Gelirler ⁷³
<i>FZ</i> : Tasarruf mevduatı <u>nominal</u> faiz oranları	Eksi	Merkez Bankası
<i>L</i> : Logaritma		
<i>D</i> : Birinci fark	Yıllar:	1970-1998
Model		
$DLDRP = \beta_0 + \beta_1 DLRGSMH + \beta_2 VRG + \beta_3 DFZ$		

⁷¹ ing. Social Science Citation Index

⁷² Çalışmalar ilham amacı ile kullanılmıştır ve çalışmaları eleştirme amacı güdülmektedir. Çalışmaların yayınlanmış olmaması ve sadece modellerinden ilham alınması sebebi ile referans verilmemiştir.

⁷³ <http://www.gelirler.gov.tr/> (Son erişim: Ekim 2004)

Tablo 3.2. Para talebi ile kayıtdışı ekonomiyi ölçmek modelinde regresyon sonuçları

	<i>katsayılar</i>	<i>standart hata</i>	<i>t stat</i>	<i>p-değeri</i>	<i>regresyon istatistikleri</i>	
Kesişim	-0.2584	0.1119	-2.3078	0.0295642	Çoklu R	0.6519
<i>RGSMH</i>	0.7135	0.4241	1.6823	0.1049568	R ²	0.4250
<i>VRG</i>	0.3089	0.1353	2.2834	0.031171	Ayarlı R ²	0.3560
<i>FZ</i>	-0.0018	0.0005	-3.2205	0.0035332	Standart Hata	0.0403
					Gözlem	29
					F	6.1606

Tablo 3.2. incelendiğinde katsayı işaretleri beklendiği gibi çıkmasına rağmen değişkenlerden anlamlı olması gereken *RGSMH* anlamsız fakat anlamsız olması beklenen *FZ* anlamlı çıkmaktadır. İnsanların tasarruf davranışı düşünülecek olursa paranın alım gücünü arttıracak olan etmen reel faizlerdir, bu sebeple nominal faizlerden çok reel faizlerin para talebinde etkili olması beklenmektedir. Reel faizlerin sabit olduğu ülkelerde nominal faizlerin *FZ* anlamlı çıkması teorik olarak mümkündür, fakat reel faizlerin uzun süre negatif (1980 öncesi) gittiği ve daha sonraki yıllarda dalgalandığı Türkiye’de, faizleri modelde anlamsız olması beklentisini yaratmaktadır.

Veriler UD için incelendiğinde üç tane veri geri kalanlardan farklı davranış göstermektedir; bunlar 1987, 1988 ve 1998 yıllarının verileridir. Türkiye’nin bu yıllarda yaşamış olduğu özel durumlar olup olmadığı incelendiğinde takip eden gerçekler tespit edilebilmektedir: 1988’de seçimlerden sonra kamu açıklarındaki artış ve mali piyasalarda yaşanan dalgalanmalar sonucu faizler yükseldi; döviz rezervinde azalma oldu ve 4 Şubat tarihinde bir dizi ekonomik kararlar alındı (Ayaz ve Asma, 2004). Buna

ilaveten 1988 yılından itibaren ekonominin döviz açığı büyüdü, resmi döviz kuru (TCMB tarafından belirlenen kur) ile serbest döviz kuru arasında açıklık genişledi. Başta ihracatçılar olmak üzere tüm kesimler devalüasyon beklentisi içine girdiler. Bir yandan devalüasyon beklentisi öte yandan hızlanan enflasyon sonucu, reel faiz düzeyi % 2-%2.5 oranında geriledi. Devalüasyon beklentisi ve reel faiz oranındaki düşme para ikâmesi olgusunu hızlandırdı. Tasarrufçular TL den kaçarak dövize ve altına yöneldiler. Reel sektörlerde yatırıma yönelmeyen likidite fazlası mali piyasalarda ciddi dengesizlikler yarattı. Ülke 1988-1989 stagflasyonuna sürüklendi. Türkiye 1988 ve 1989 yıllarında stagflasyon yaşamış ve bu stagflasyonu 4 Şubat 1989'de alınan mali önlemlerle aşmaya çalışmıştır (Şahin, 2002). 1998 yılında Rusya'daki mali çöküşler sebebiyle Türk ekonomisi olumsuz etkilendi, 6 milyar dolar dolayında yabancı finansal yatırım Türkiye'yi terk etti, ekonomi darboğaza sürüklendi (Ayaz ve Asma, 2004).

Bu durumlar ekonomik değişkenler arasındaki ilişkiyi normal zamanlardakinden farklı olacak şekilde bir müddet değiştirebilecek olaylardır ve bu durumlar UD oluşumuna sebep olmuşlardır. UD çıkarıldıktan sonra tablo 3.3. ile ifade edilmiş beklenen sonuçlar elde edilmektedir. Tablo 3.2. ve 3.3. karşılaştırıldığında katsayıların ve buna bağlı tahminlerin büyük ölçüde farklı olduğu ortaya çıkmaktadır.

Tablo 3.3. UDler çıkarıldıktan sonra para talebi ile kayıtdışı ekonomiyi ölçmek modelinde regresyon sonuçları

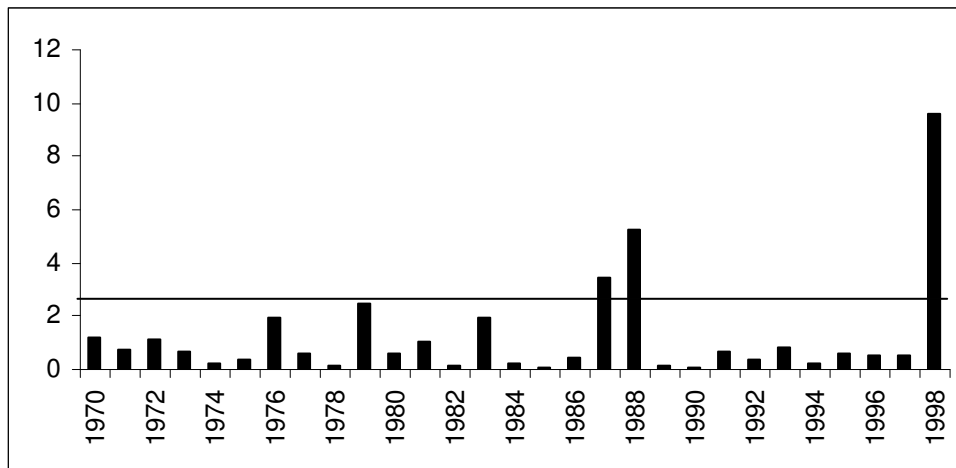
	<i>katsayılar</i>	<i>standart hata</i>	<i>t stat</i>	<i>p-değeri</i>	<i>regresyon istatistikleri</i>	
Kesişim	-0.3956	0.1050	-3.7670	0.00106	Çoklu R	0.7445
RGSMH	2.0298	0.5426	3.7407	0.00113	R ²	0.5543
VRG	0.4225	0.1235	3.4198	0.00242	Ayarlı R ²	0.4936
FZ	0.0028	0.0013	2.0148	0.05630	Standart Hata	0.0341
					Gözlem	26
					F	9.1233

Burada akla gelebilecek bir başka soru Türk Ekonomisinde yer alan başka çalkantılar neden UD olarak kendini göstermemiştir? Bunun için SMK⁷⁴ değerlerine bakarak cevap verilebilir.

Kalıntı değerleri şekil 3.1.'de incelendiğinde başka yıllarda da para talebinde dikkate değer değişimlerin olduğu görülmektedir. Genelde kriter 2.5 değerinin üzerindeki SMK değerlerini UD olarak kabul etme yönündedir ve bunun sonucu olarak önceki paragraflarda yer alan veriler UD olarak tespit edilmiştir. Diğer büyük kalıntı değerleri incelendiğinde öngörülebileceği gibi ağır bunalım yılı olan 1979'da (Şahin, 2002) yüksek bir SMK değeri mevcuttur. Bu yıl da verilerden çıkarılıp regresyon yapılacak olursa katsayılar ve anlamlılıklarda büyük değişme olmazken standart sapma ve R² terimleri iyileşmektedir.

⁷⁴ Standartlaştırılmış Mutlak Kalıntı

Veriler JBN⁷⁵ testine tabi tutulduğunda hem verilerde UD mevcutken, hem de verilerden UDler çıkarıldıktan sonra kalıntıların normallik varsayımının red edilemediği ortaya çıkmıştır. Kısacası verilerde normallikten sapmalara bakarak UDlerin mevcudiyeti tespit edilememekte ve EKK kalıntıları UDler ile birlikte bile normallik göstermektedir. Bu örnekten çıkan başka bir özellik çoklu-doğrusallık etkisi konusundadır. UDler ile birlikte kısmi korelasyon katsayıları ve yan regresyona bakıldığında değişkenler arasında herhangi bir çoklu-doğrusal bağlantı görülmemektedir. UDler çıkarıldıktan sonra ise *RGSMH* ve *FZ* arasında çoklu-doğrusal bağlantı olduğu ortaya çıkmaktadır. Buradan çıkartılabilecek en önemli çıkarımlardan biri UDlerin verilerin kendi aralarındaki ilişkileri bozmaları ve dolaylı olarak da bunu tahminlere ve kalıntılara yansıtılmalarıdır. Kalıntıların değişkenlere bağlı grafikleri çizildiğinde veya White değişen-varyans testi uygulandığında her iki durumda da değişen-varyans bulunamamıştır.



⁷⁵ Jarque-Berra Normallik

Şekil 3.1. Para talebi ile kayıtdışı ekonomiyi ölçmek modelinde ÇDT ediciler ile oluşan SMK değerleri

3.4. Döviz Kuru Modeli

Bu bölümde Internet'te (<http://www.usfca.edu/economics/veitch/econ621.htm>) yer alan ve döviz kurunun davranışını açıklamak için kurulmuş bir modelin alıntısıdır. Takip eden tablo 3.4. modeldeki değişkenleri ve tablo 3.5. regresyon sonuçlarını göstermektedir.

Sonuçlar incelendiğinde 3 tane veri UD olarak ortaya çıkmaktadır ve bu değerler çıkarıldıktan sonra tablo 3.6.'da yer alan sonuçlar elde edilmektedir. Bu veriler Ocak-Şubat, Mart-Nisan 1994 ve Kasım-Aralık 2000 yıllarıdır. Bu UDlerin tespit edilmesi çok isabetlidir, çünkü aşağıdaki Şahin (2002) çalışmasında belirtildiği gibi o dönem döviz açısından işlerin her zamanki gibi gitmediği bir dönemdir.

“...uluslararası derecelendirme kuruluşları Türkiye'yi yatırım yapılabilir ülke derecesinden spekülâtif ülke derecesine düşürdüler. Türkiye'nin kredi notunun düştüğünün medyada ilan edilmesi, finansman piyasasındaki kargaşayı artırdı. Döviz hücüm başladı, döviz fiyatları çılgın bir tempoda yükseldi. TCMB, döviz fiyatlarını tutmak ve ekonomideki fazla likiditeyi emmek için 1994'ün ilk aylarında piyasaya 3.5 milyar dolar döviz sattı. Fakat bu ortamda TCMB rezervlerinin tümü de satılsa döviz kurlarını tutmanın mümkün olmayacağı anlaşıldı ve uygulamadan vazgeçildi. 5 Nisan 1994 istikrar Programı açıklanıncaya kadar döviz kurları süratle yükseldi yani TL serbest piyasada devalüe oldu.... (Şahin, 2002)”

“Üst üste ortaya çıkan yolsuzluklar, bazı bankaların içinin boşaltılarak TMSF'na devredilmesi, kamu bankalarının görev zararları sonucu nakit sıkıntısına düşmesi ve bankacılık sistemi içinde yükümlülüklerini yerine getirememesi sonucu Kasım 2000’de ülkeden para çıkışı yaşandı. Borsa düştü, faiz ve döviz kurları yükselmeye başladı, fakat bu kriz Şubat 2001’e kadar ertelenebildi. 19 Şubat 2001’de ise pamuk ipliğine bağlı ekonomik "istikrar" koptu ve Türkiye Cumhuriyet tarihinin en ağır ekonomik krizlerinden birisine sürüklendi. Üstelik bu kriz bir istikrar programı uygulamada iken, geliyorum diye diye geldi. (Şahin, 2002)”

Tablo 3.4. Döviz kuru modelinde değişkenler, beklenen işaretler ve veri kaynakları

Bağımlı değişken		Veri Kaynağı
<i>DK</i> : Döviz Kuru		IFS
Bağımsız değişkenler	Beklenen işaret	
<i>UR</i> : Uluslararası rezervler	Eksi	IFS
<i>TÜFE</i> : Tüketici fiyatları endeksi	Artı	IFS
<i>PT</i> : Para tabanı	Artı	IFS
<i>PA</i> : Para arzı	Artı	IFS
<i>FO</i> : Türk faiz oranları	Eksi	IFS
<i>EÜ</i> : Endüstriyel üretim	Eksi	IFS
<i>NI</i> : Net İhracat	Eksi	IFS
<i>F</i> : Birinci fark		
<i>B</i> : Yüzde değişim	Aylar:	1992:1-2000:11
Model		
$BDK = \beta_0 + \beta_1 BUR + \beta_2 BTÜFE + \beta_3 BPT + \beta_4 BPA + \beta_5 FFO + \beta_6 BEÜ + \beta_7 FNI$		

Tablo 3.5. Döviz kuru modelinde regresyon sonuçları

	<i>katsayılar</i>	<i>standart hata</i>	<i>T stat</i>	<i>p-değeri</i>	<i>regresyon istatistikleri</i>	
<i>Kesişim</i>	-3.364	0.755	-4.453	$2.24 \cdot 10^{-05}$	Çoklu R	0.81
<i>BUR</i>	-0.138	0.0445	-3.112	0.00243	R ²	0.656
<i>BTÜFE</i>	1.421	0.133	10.642	$4.9 \cdot 10^{-18}$	Ayarlı R ²	0.632
<i>BPT</i>	0.374	0.109	3.415	0.000928	Standart Hata	3.479
<i>BPA</i>	-0.065	0.0483	-1.345	0.181	Gözlem	106

<i>FFO</i>	0.0053	0.0126	0.419	0.675	<i>F</i>	26.797
<i>BEÜ</i>	-0.114	0.0461	-2.479	0.0148		
<i>FNİ</i>	-0.00275	0.000969	-2.836	0.00554		

İki regresyon sonucu karşılaştırıldığında teorik olarak faiz oranlarının döviz kurunda etkili olması gerekirken anlamlı çıkmamaktadır. UDler çıkarıldıktan sonra beklenen sonuçlar elde edilmektedir. İlk regresyonda beklenen sonuçlara ulaşılmaması için sebep çoklu-doğrusal-bağlantı olabilir mi diye yan regresyon ile kontrol edildiğinde faiz oranı için çoklu-doğrusal-bağlantı bulunamamıştır. Normallik testi yapıldığında UDli veri kümesinde normallik hipotezi red edilmiş, fakat UDler çıkarıldıktan sonra bu hipotez red edilememiştir. Son ilginç sonuç, UDler ile birlikte değişen-varyans mevcut iken UDler çıkarıldıktan sonra değişen-varyansın kalktığı görülmektedir. UD veren bu üç ay dışında 2.5 SMH değerinden büyük kalıntılara bakıldığında hiçbir değer bulunamamıştır.

Tablo 3.6. UDler çıkarıldıktan sonra döviz kuru modelinde regresyon sonuçları

	<i>katsayılar</i>	<i>standart hata</i>	<i>T stat</i>	<i>p-değeri</i>	<i>regresyon istatistikleri</i>	
<i>Kesişim</i>	2.0856	0.516	4.034	0.00011	Çoklu R	0.725
<i>BUR</i>	-0.0695	0.0230	-3.011	0.00334	R ²	0.527
<i>BTÜFE</i>	0.393	0.0932	4.216	5.67·10 ⁻⁵	Ayarlı R ²	0.492
<i>BPT</i>	0.148	0.0579	2.562	0.0119	Standart Hata	1.748
<i>BPA</i>	-0.0554	0.0245	-2.256	0.0263	Gözlem	103
<i>FFO</i>	0.0348	0.00711	4.901	3.92·10 ⁻⁶	<i>F</i>	15.12
<i>BEÜ</i>	-0.061	0.0236	-2.584	0.0112		
<i>FNİ</i>	-0.00147	0.000494	-2.973	0.00373		

BÖLÜM IV

4. BULGULAR, TARTIŞMA ve YORUM

Araştırmada, Türkiye gibi gelişmekte olan ülkelerin ekonomisinde risk ve krizlerle karşılaşılma olasılığının diğer ülkelere göre daha fazla olduğu görüşü göz önünde bulundurularak bu kriz, şok ve başka etkenlerin UD oluşumuna sebebiyet verebileceği ve bu nedenle UDlerin tahminleri saptırıcı etkilerinin olacağı hipotezi kurulmuştur. Bu hipotezden yola çıkarak öncelikle teknolojik gelişmelere paralel olarak gelişen bir literatür olan *Güncel Uç Değer Tespit Yöntemleri* Literatürü taranmış ve tespit yöntemleri gruplandırılarak sunulmuştur. Eski ve genel yöntemlere ulaşılmış ve bu yöntemlerin gelişmiş versiyonları belirtilmiştir. Çoğu yöntemin avantajı yanında bazı dezavantajları vardır. Teknolojik gelişmelerle birlikte iyi yönde gelişmeler olacağı beklenmelidir.

UD tespiti doğrultusunda yapılan kısmi literatür taramasında yöntemlerin nihai ve her alanda kullanılacak bir UD tespit yöntemine doğru ıraksamadığı

görülmektedir. İstanbul Üniversitesi tarafından düzenlenen “7. Ulusal Ekonometri ve İstatistik Sempozyumu”ndaki bir tartışmada bu konu ile uğraşan araştırmacılar tarafından, Gentleman, J. F. ve Wilk, M. B. (1975a, 1975b) yönteminin evrensel bir yöntem olduğu fakat bilgisayar uygulamasının olmaması nedeniyle büyük örneklem üzerinde uygulanamadığı belirtilmiştir. Araştırma sonuçlarından elde edilen yöntemler tespit yaklaşımlarına göre gruplandırılmıştır. Gruplandırma ile tüm UD tespit yöntemlerine toplu bir bakış sağlanmış ve literatürde birbirinden oldukça farklı onlarca görüşe rastlanmıştır.

Son literatür yeni fikirleri içermekte eski yöntemlere ilaveler yapmakta, eski yöntemleri kullanarak yeni yöntemler geliştirmektedir. Eski yöntemlerin eksik yönleri hızla gelişen bilgisayar teknolojisi yardımıyla geliştirilmekte ve iyileştirilmektedir. Bilgisayar teknolojisinin gelişimi ile bu iyileştirme sürecidir. Klasik yöntemler olan EKKK, EKOK, EKOD hala kullanım alanı bulabilmekte ve SAS gibi paket programlarına girmektedirler. En büyük eksiklik ise bütün bu yöntemlerin karşılaştırmasını yapacak çalışmaların azlığıdır. Bütün bu sebeplerden sonra çıkan sonuç verilerde UD bulunması konusunun bilim adamları tarafından ciddiye alındığı ve ezbere yapılacak bir analizin başka sorunlara yol açabileceğidir. Fakat uzmanlık alanı olmasa da araştırmacıların UD varlığını ve etkilerini göz önüne aldığı çalışmaların sayısı yeterli değildir.

Üçüncü bölümde UD etkisi uygulamalar ile desteklenmiştir. Az miktarda olmalarına rağmen, verilerin çoğunun vermek istediği bilgiye engel olan, veriler

arasındaki ilişkiyi bozan ve sonuçlar üzerinde yanıltıcı etkiler yaratan UDlerin bu özellikleri örneklerle gösterilmiştir. Buna ilaveten UDlerin kendilerini maskelemeleri sonucu klasik tespit edilme yöntemleri ile tespit edilemeyeceğine yine örnekler ile değinilmiştir. Bu örneklerden bir kısmı bu çalışma için özel olarak türetilmiş verilerden ve son ikisi de Internet'te yapılan arama sonucu model aşamasındaki Türk verileri ilgili çalışmalardan oluşmaktadır. Türk verileri ilgili olarak tahmin ediciler, veriler ve sonuçlarla ilgili takip eden bulgular elde edilmiştir:

1. Her iki Türk modeli için türetilen verilerde UD tespit edilmiştir. Bu durum genelleştirilemez, fakat uzun zamandan beridir en kaliteli verilerin bile UD içerdiği, buna ilaveten kaliteli iktisadi veri olmaya çok az sayıda aday veri olduğu ve Türk iktisadi verilerinde UD bulunma olasılığının ekonomik olarak daha istikrarlı bir yapıya sahip gelişmiş ülkelerden daha büyük olasılık olduğu konusunda bir tespittir. Bir sonraki aşama daha çok sayıda model için bu çalışmayı tekrarlamak ve çok nadiren görülen çeşitli sosyo-ekonomik değişkenlerin/değişimlerin Türk verilerinde ne oranda olumsuz etkiye yol açtığını ortaya çıkarmaktır. Önemli bir başka nokta, bu çalışmada kullanılan örnekler özellikle seçenekler içinden seçilip sunulmamış olup verileri türetilen ilk iki örnek olması açısından önemlidir. Türk ekonomik verilerinin UD açısından sorgulanması gerekmektedir.
2. Ani ekonomik değişimler/şoklarla UD arasında bir bağlantı bulunmuş ve UD olarak tespit edilen zamanlarda ilgilenilen ekonomik değişkeni etkileyen anormal

bir davranışın varlığı tespit edilmiştir. Sağlam regresyonda ÇDT edicilerin kalıntılarından bu tür değişimler tespit edilebilmekte ve bu kalıntılar kendilerini yüksek SMK değeri olarak göstermektedir. Buradan, sağlam tahminlerin normal zamanlarda daha güvenilir tahminler yapacağını, şok zamanlarında ise ne olacağını ancak şokun niteliğini bilerek cevap verebileceğimiz sonucunu çıkarabiliriz.

3. UD mevcudiyetinde beklenen bir bulgu Türk verilerinde de EKK tahminlerinin çökmesidir. Buna bağlı olarak bu tahminlerden türetilebilecek herhangi değerlendirme kriterleri de çökmektedir. Kalıntılara, normallik testlerine, çoklu-doğrusal-bağlantı testlerine ya da değişen-varyans testlerine bakıldığında UD mevcudiyeti konusunda herhangi bir ipucu elde edilememektedir. Bu büyük bir tehlikedir çünkü bir tane bile UD EKK tahminlerini çöktürmektedir. Bu çalışmada sunulan örnekler ile verilerde UD bulunması durumunda EKK tahminlerinin tam ters veya güvenilir sonuçlara sebebiyet vereceği sergilenmiştir.
4. UDler bilimsel sonuç çıkarmaya darbe vurmaktadır, çünkü UD mevcudiyetinde sonuçlar bilimsel teoremleri desteklemeyebilmektedir. Buna karşı tedbir alınmaması durumunda hipotez aşamasında bulunan bir çok gözlem teorem veya kanun olarak kabul edilmeden elenebilir. Bu durum bilim yolunda bir duraksamaya sebep olabilecek büyük bir tehlikedir. Türk örneklerinde gösterildiği gibi buradan çıkan sonuç, bilimsel hipotezlerin ya da teoremlerin savunduğu fikir “sağlam” (regresyon sınavından geçmiş) veriler ile

sınanmadıkça kabul görmeleri imkansız olacaktır, çünkü sonuçlar yanıltıcı olmaktadır.

5. UD mevcudiyetinde ve yokluğunda verilerin yansıttığı bilgi EKK tahmini ile birlikte kullanıldığında çok farklı olabilmektedir. Normallik, çoklu-doğrusal-bağlantı veya değişen-varyans tespiti konusunda UDLer ile birlikte tam tersi sonuçlar çıkabilmektedir.

Bütün yukarıda anılan sorunlar EKK yerine bütün verileri dikkate alan başka bir tahmin edici de kullanılsaydı yine ortaya çıkacaktı, çünkü bu tür tahmin ediciler bütün kalıntıları belirli bir amaç fonksiyonuna göre indirgemektedirler. Bu durum UDLere yüksek kalıntı değerleri aktarılmasına engel olmakta ve böylece verilerin çoğunun yansıtmak istediği bilgi tahmin edilememektedir. SR teknikleri verilerin yarısına kadar kısmının yüksek kalıntı miktarları almasına izin verdiği için verilerin çoğundaki bilgi yansıtılmamaktadır.

UDler çeşitli sebeplerden ötürü oluşabilmektedir. Bu çalışma kapsamında Türkiye’de ekonomik şokların (buhran zamanlarının veya politik istikrarsızlıkların) UDLerle ilişkisi incelenmiştir. Buna ilaveten yanlış ölçümler, model hataları veya insani faktörler bu tür verilere sebebiyet verebilmektedir ve başka bilimsel çalışmalar ile Türk verilerinde incelenmesi gerekmektedir.

SONUÇ

Sonuç olarak UDler vardır ve bilimsel çalışmalarda etkileri göz önünde bulundurulmalıdır. Çalışma sonucuna göre UDler EKK tahminlerini çökertmektedir bu nedenle sağlam regresyon yöntemlerini kullanmak hem normal zamanlarda hemde şok zamanlarında daha güvenilir tahminler verecektir. Bu çalışmada UDler çıkarıldıktan sonra geri kalan veriler ile tekrar regresyon yapılmıştır. Fakat bu konu ile ilgili olarak Rousseeuw ve Zomeren (1990) sağlam regresyon kullanılarak amacın UDleri tespit etmek olduğunu ve tespit edildikten sonra bu konuda uzman kişinin bu değerlerle ne yapması gerektiğini tavsiye etmektedirler. Bu çalışma sonucu elde edilen tecrübe UDin veri kümesinden çıkarılması ya da modele uygun değişkenler katıldıktan sonra kullanılması yönündedir. Rousseeuw ve Leroy'a göre (1987) UDlerin modelden çıkarılmasının veri kaybı olarak düşünülmemesi gerekmektedir, çünkü çıkarılmadığı durumlarda yanlış tahminler elde edilebilmektedir.

KAYNAKLAR

- Agullo, J. 2001. "New Algorithms for Computing the Least Trimmed Squares Regression Estimator". *Computational Statistics & Data Analysis*, 36, 425-439.
- Arslan, O, Edlund, O., ve Ekblom, H. 2002. "Algorithms to Compute CM- and S-estimates for Regression". *Metrika*, 55, 37-51.
- Arslan, O. ve Billor, N. 2000. "Robust Liu Estimator for Regression Based on an M-Estimator". *Journal of Applied Statistics*, 7, 39-47.
- Atkinson, A.C., Riani, M. 1997. "Bivariate Boxplots, Multiple Outliers, Multivariate Transformations and Discriminant Analysis: the 1997 Hunter Lecture". *Environmetrics*, 8, 583-602.
- Atuk, O. ve Ural, B. P. 2002. "Seasonal Adjustment Methods: An Application to the Turkish Monetary Aggregates". *Central Bank Review* 1 (2002) 21-37.
- Ayaz, M.F. ve Asma, M.Z. 2004. Anketör : ZEKİ M. *Ekonomik Krizin Batmana Yansımaları*. Erişim: [http://www.batman.smmmo.org.tr/ekonomik_kriz.htm]. Erişim Tarihi: 13.11.2004.
- Bassett, G. W. 1991. "Equivariant, Monotonic, 50% Breakdown Estimators". *American Statistician*, 45, 135-137.
- Butler, R.W., Davies, P.L. ve Jhun, M. 1993. "Asymptotics for the Minimum Covariance Determinant Estimator". *The Annals of Statistics*, 21, 1385-1400.

- Choulakian, V. 2005. "L1-Norm Projection Pursuit Principal Component Analysis". *Computational Statistics & Data Analysis*, baskıda.
- Coakley, C.W., Hettmansperger, T.P. 1993. "A Bounded Influence, High Breakdown, Efficient Regression Estimator". *Journal of the American Statistical Association*, 88, 872–880.
- Cook, R. D. ve Critchley, F. 2000. "Identifying Regression Outliers and Mixtures Graphically". *Journal of the American Statistical Association*, 95, 781.
- Cook, R. D. ve Weisberg S. 1991. Discussion of a paper by K. C. Li. *Jornal of American Statistical Association*, 86, 328–32.
- Cook, R.D. 1977. "Detection of Influential Observations in Linear Regression". *Technometrics*, 19, 15-18.
- Cook, R.D. 1998. "Principal Hessian Directions Revisited". *Journal of the American Statistical Association*, 93(441), 84–100.
- Croux C., Rousseeuw P.J. ve Hössjer O. 1994. "Generalized S-estimators". *Journal of the American Statistical Association*, 89, 1271-1281.
- Dahl, T. ve Naes, T. 2004. "Outlier and Group Detection in Sensory Panels Using Hierarchical Cluster Analysis with the Procrustes Distance". *Food Quality and Preference*, 15, 195–208.
- Daniel, C. ve Wood, F. S. 1971. *Fitting Equations to Data*, New York: John Wiley & Sons.

- Davies, P. L. and Gather, U. 1993. "The Identification of Multiple Outliers (with discussion)". *Journal of Statistical Planning and Inference*, 122, 65-78.
- Erlat, Haluk 2005. "Long Memory in Turkish Inflation Rates". <http://www.sba.luc.edu/orgs/meea/volume3/HErlat/>.
- Erlat, Haluk 2003. "The Nature of Persistence in Turkish Real Exchange Rates". *Emerging Markets Finance and Trade*, 39, 2.
- Fernholz, L.T., Morgenthaler, S. ve Tukey, J.W. 2004. "An Outlier Nomination Method Based on the Multihalver". *Journal of Statistical Planning and Inference*, 122, 125– 139.
- Filzmoser, P., Garrett, R. G. ve Reimann C. 2004. "Multivariate Outlier Detection in Exploration Geochemistry". *Computers & Geosciences*, baskıda.
- Gentleman, J. F. ve Wilk, M. B. 1975a. "Detecting Outliers in a Two-Way Table I. Statistical Behavior of Residuals". *Technometrics*, 17: 1-14.
- Gentleman, J. F. ve Wilk, M. B. 1975b. "Detecting Outliers in a Two-Way Table II. Supplementing the Direct Analysis of Residuals". *Biometrics*, 31: 387-410.
- Gujarati, D. N. 1999. *Temel Ekonometri* (Çev. Ümit Şenesen ve Gülay Günlük Şenesen), Literatür Yayıncılık, İstanbul.
- Hadi, A.S., Simonoff, J.S. 1993. "Procedures for the Identification of Multiple Outliers in Linear Models". *Journal of the American Statistical Association*, 88, 1264–1272.

- Hadi, A.S., Simonoff, J.S. 1997. "A More Robust Outlier Identifier for Regression Data". *Bulletin of the International Statistical Institute*, 281–282.
- Hampel, F. 1975. "Beyond Location Parameters: Robust Concepts and Methods (with discussion)". *Bulletin 40th Session of the ISI*, XLVI (1), 375–391.
- Hampel, F. R. 1973. "Robust Estimation: A Condensed Partial Survey". *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 27, 87-104.
- Hardin, J. ve Rocke, D. 2002. "The Distribution of Robust Distances". <http://handel.cipic.ucdavis.edu/~dmrocke/Robdist5.pdf>, son erişim: 8 Mart 2005.
- Hardin, J. ve Rocke, D. M. 2004. "Outlier Detection in the Multiple Cluster Setting Using the Minimum Covariance Determinant Estimator". *Computational Statistics & Data Analysis*, 44, 625-638.
- Hawkins, D. M. ve Olive D. J. 1999. "Applications and Algorithms for Least Trimmed Sum of Absolute Deviations Regression". *Computational Statistics & Data Analysis*, 32, 119-134.
- Hettmansperger, T. P. ve Sheather, S. J. 1992. "A cautionary note on the method of least median squares". *American Statistician*, 46: 79-83.
- Hoy, M., Steen, K. ve Martens, H. 1998. "Review of Partial Least Squares Regression Prediction Error in Unscrambler". *Chemometrics and Intelligent Laboratory Systems*, 44, 123-133.
- Hössjer, O. 1994. "Rand-Based Estimates in the Linear Model with High Breakdown Point". *Journal of the American Statistical Association*, 89, 149-158.

- Huber, P. J. 1973. "Robust Regression: Asymptotics, Conjectures, and Monte Carlo". *The Annals of Statistics*, 1, 799-821.
- Huber, P. J. 1981. *Robust Statistics*, New York: John Wiley and Sons.
- Hubert, M., Rousseeuw, P.J. ve Branden, K.V. 2005. "ROBPCA: A New Approach to Robust Principal Component Analysis". *Technometrics*, 47, 64-79.
- Hund, E., Massart, D. L. ve Verbeke, J. S. 2002. "Robust Regression and Outlier Detection in the Evaluation of Robustness Tests with Different Experimental Designs". *Analytica Chimica Acta*, 463, 53-73.
- Jiang, M.F., Tseng, S.S. ve Su, C.M. 2001. "Two-phase Clustering Process for Outliers Detection". *Pattern Recognition Letters*. 22(6/7): 691-700.
- Jörnsten, R. 2004. "Clustering and Classification Based on the L1 Data Depth". *Journal of Multivariate Analysis*, 90, 67-89.
- Justel, A. ve Pena, D. 1996. "Gibbs Sampling Will Fail in Outlier Problems With Strong Masking". *Journal of Computational and Graphical Statistics*, 5, 176-189.
- Kiracı, A. 1996. *Robust Regression and Applications*. Yayınlanmamış Yüksek Lisans Tezi, Bilkent Üniversitesi Sosyal Bilimler Enstitüsü.
- Knez, P.J. ve Ready, M.J. 1997. "On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions". *Journal of Finance*, 52, 1335-1382.

- Küçükkocaoğlu, G. ve Kiracı, A. 2003. "Güçlü Beta Hesaplamaları". MÖDAV, *Muhasebe Bilim ve Dünya Dergisi*, 5, 4, 109-125.
- Kwon, O.K., Sim, D.G. ve Park, R.H. 2001. "Robust Hausdorff Distance Matching Algorithms Using Pyramidal Structures", *Pattern Recognition*, 34, 2005-2013.
- Lalor, G.C. ve Zhang, C. 2001. "Multivariate Outlier Detection and Remediation in Geochemical Databases", *The Science of the Total Environment*, 281, 99-109.
- Lee, K.M., Meer, P. ve Park R.H. 1998. "Robust Adaptive Segmentation of Range Images". *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(2), 200-205.
- Li G., Chen Z. 1985. "Projection-Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo". *Journal of the American Statistical Association*, 80(391), 759-766.
- Li, K.C. 1991. "Sliced Inverse Regression for dimension reduction (with discussion)". *Journal of the American Statistical Association*, 86, 316-342.
- Li, K.C. 1992. "On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma", *Journal of the American Statistical Association*, 87, 1025-1039.
- Liu, H., Shah, S. ve Jiang, W. 2004. "On-line Outlier Detection and Data Cleaning". *Computers Engineering*, 28, 1635-1647.
- Lopuhaa, H.P. ve Rousseeuw P.J. 2001. "Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices". *The Annals of Statistics*, 19, 229-248.

- Maronna, R. A., Bustos, O. H. ve Yohai, V. J. 1979. "Bias- and Efficiency-Robustness of General M-Estimators for Regression with Random Carriers". In *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblatt (Eds.). Lecture Notes in Mathematics 757. Springer, Berlin, 91-116.
- Martens, H. ve Naes, T. 1989. *Multivariate Calibration*, John Wiley & Sons Ltd., Chichester.
- Martin, R. D., ve Thomson, D. J. 1982. "Robust-Resistant Spectrum Estimation", *Proceeding of the IEEE*, 70, 1097–1115.
- Melouna, M., Militký, J. 2001. "Tutorial: Detection of Single Influential Points in OLS Regression Model Building", *Analytica Chimica Acta*, 439, 169–191.
- Mendes, B. ve Tyler, D.E. 1995. "Constrained M Estimates for Regression". In: *Robust Statistics; Data Analysis and Computer Intensive Methods*, Lecture Notes in Statistics 109. Springer, New York, 299-320.
- Millar A.M. ve Hamilton D.C. 1999. "Modern Outlier Detection Methods and their Effect on Subsequent Inference". *The Journal of Statistical Computation and Simulation*, 64, 2.
- Mosteller, F. ve Tukey, J.W., 1977. "Data Analysis and Regression". Addison-Wesley, Reading, MA, 351-365.
- Olive, D. J. ve Hawkins, D. M., 2003. "Robust Regression with High Coverage". *Statistics & Probability Letters*, 63, 259-266.

- Orhan, M., Zaman, A. ve Rousseeuw, P.J. 2001. "Econometric Applications of High-Breakdown Robust Regression Techniques". *Economics Letters*, 71, 1–8.
- Pena, D. ve Yohai, V. 1995. "The Detection of Influential Subsets in Linear Regression by Using an Influence Matrix". *Journal of the Royal Statistical Society Ser. B*, 57, 145-156.
- Pierna J. A. F., Wahl, F., Noord, O. E. ve Massart, D. L. 2002. "Methods for Outlier Detection in Prediction". *Chemometrics and Intelligent Laboratory Systems*, 63, 27-39.
- Preparate, F.P. ve Shamos, M. I. 1985. *Computational Geometry, An Introduction*, Springer-Verlag.
- Reinmann, C., Filzmoser, P., Garrett, R.G. 2005. "Background and Threshold: Critical Comparison of Methods of Determination". *Science of the Total Environment* (baskıda).
- Rimbaud, D.J., Bouveresse, E, Massart, D.L., de Noord, O.E. 1999. "Detection of Prediction Outliers and Inliers in Multivariate Calibration". *Analytica Chimica Acta*, 388, 283-301.
- Rio, F.J, Riu, J. ve Rius X. 2001 "Graphical Criterion for the Detection of Outliers in Linear Regression Taking into Account Errors in Both Axes". *Analytica Chimica Acta*, 446 (2001), 49–58.
- Rousseeuw, P. J. 2004. Antwerp Group on Robust & Applied Statistics, <http://agoras.ua.ac.be/> (10 Temmuz 2005, WEB).

- Rousseeuw, P. J. ve Leroy, A.M. 1987. *Robust Regression and Outlier Detection*, John Wiley, New York.
- Rousseeuw, P.J. ve Yohai, V. 1984. “Robust Regression by Means of S-Estimators”, in *Robust and Nonlinear Time Series Analysis*, (edited by J. Franke, W. Härdle and D. Martin), Lecture Notes in Statistics No. 26, Springer Verlag, Berlin, 256-272.
- Rousseeuw, P.J. 1984. “Least Median of Squares Regression”. *Journal of the American Statistical Association*, 79, 871-880.
- Rousseeuw, P.J. ve Driessen, K. 1998. *Computing LTS Regression for Large Data Sets*, Technical Report. Belgium: Universitaire Instelling Antwerpen (UIA).
- Rousseeuw, P.J. ve Driessen, K. 1999. “A Fast Algorithm for the Minimum Covariance Determinant Estimator”. *Technometrics*, 41, 212–223.
- Rousseeuw, P.J. ve Zomeren B.C. 1990. “Unmasking Outliers and Leverage Points”. *Journal of the American Statistical Association*, 85, 411, 633-639.
- Şahin, H. 2002. *Türkiye Ekonomisi*, Bursa: Ezgi Kitapevi Yayınları.
- Sebert, D.M., Montgomery, D.C., Rollier, D. 1998. “A Clustering Algorithm for Identifying Multiple Outliers in Linear Regression”. *Computational Statistics & Data Analysis*, 27, 461–484.
- Sheather, S. J., McKean, J. W. ve Hettmansperger, T. P. 1997. “Finite Sample Stability Properties of the Least Median of Squares Estimator”. *Journal of Statistical Computing and Simulation*, 58(4), 371–383.

- Simpson, J.R., Montgomery, D.C. 1998. "A Compound Estimator for Robust Regression". *Naval Research Logistics*, 45, 125–134.
- Stromberg, A. J., Hawkins, D. M. and Hössjer, O. 1995. "The Least Trimmed Differences Regression Estimator and Alternatives". Technical Report 1995:26, Dept. Mathematical Statistics, Lund Univ. and Lund Institute of Technology. Z .
- Stromberg, A. J., Hössjer, O. ve Hawkins, D. M. 2000. "The Least Trimmed Differences Regression Estimator and Alternatives". *Journal of the American Statistical Association*, 95, 451, 853.
- Stromberg, A.J. 1993. "Computing the Exact Value of the Least Median of Squares Estimate and Stability Diagnostics in Multiple Linear Regression". *SIAM Journal on Scientific Computation*, 14, 1289-1299.
- Swallow, W., Kianifard, F. 1996. "Using Robust Scale Estimates in Detecting Multiple Outliers in Linear Regression", *Biometrics*, 52, 545–556.
- Tanzi, V. 1983. "The Underground Economy in the US: Annual Estimates, 1930-80". *IMF Staff Papers*.
- Tao, Q., Wu, G. ve Wang, J., "A New Maximum Margin Algorithm for One-Class Problems and its Boosting Implementation". *Pattern Recognition*, baskıda.
- Tax, D. Ve Duin, R. P. W. 1999. "Support Vector Domain Description". *Pattern Recognition Letters*, 20, 1191-1199.

- Wang, H. ve Suter, D. 2004. "Using Symmetry in Robust Model Fitting". *Pattern Recognition Letters*, 24, 2953–2966.
- Wang, S., Woodward, W.A., Gray, H.L., Wiechecki, S. ve Sain, S.R. 1997. "A New Test for Outlier Detection From a Multivariate Mixture Distribution". *Journal of Computational and Graphical Statistics*, 6, 285-299.
- Wisnowski, J. W., Montgomery, D. C. Ve Simpson J. R. 2001. "A Comparative Analysis of Multiple Outlier Detection Procedures in the Linear Regression Model". *Computational Statistics & Data Analysis*, 36, 351-382.
- Wu, S. ve Chow T.W.S. 2004. "Clustering of the Self-Organizing Map Using a Clustering Validity Index Based on Inter-Cluster and Intra-Cluster Density". *Pattern Recognition*, 37, 175 – 188.
- Yohai, V. J. 1985. "High Breakdown-Point and High Efficiency Robust Estimates for Regression". *Technical Report No. 66*, Dept. Statistics, Univ. Washington, Seattle.
- Yohai, V.J. ve Zamar, R. 1988. "High Breakdown-Point Estimates of Regression by Means of the Minimization of an Efficient Scale". *Journal of the American Statistical Association*, 83, 406-413.
- Zhang, C.S., Wong, P.M. ve Selinus, O. 1999. "A Comparison of Outlier Detection Methods: Exemplified with an Environmental Geochemical Dataset". *The 6th International Conference on Neural Information Processing*, Perth, Australia, November 16-20, 1, 183-187.
- Zhao, W., Chen, D. ve Hu S. 2004. "Detection of Outlier and a Robust BP Algorithm Against Outlier". *Computers and Chemical Engineering*, 28, 1403–1408.