

38468

T.C
EGE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI
10.0600.0000.057

Yüksek Lisans Tezi

*KALİTE KONTROL AMAÇLI ZAMAN SERİLERİNDE
SAPAN DEĞER (OUTLIER) ANALİZİ*

Hazırlayan : Arş.Gör. Ahmet KAYA
Danışman : Prof. Dr. Onur BASKAN

Bu tez, aşağıda isimleri yazılı juri,

Prof. Dr. Onur BASKAN (Başkan)
Prof. Dr. Fikret İKİZ.....(Üye)
Prof. Dr. Cudi OKUR.....(Üye)

tarafından oy birliği ile kabul edilmiştir.

Bornova - İZMİR
1995

İÇİNDEKİLER	SAYFA
ABSTRAKT	1
ABSTRACT	1
ÖNSÖZ	2
BÖLÜM 1.	3
1. GİRİŞ	3
BÖLÜM 2.	9
2. KALİTE KONTROL	9
2.1 Giriş	9
2.2 Kalite'nin Gelişimi	9
2.3 Kalite	10
2.4 Kalite'nin Ölçüsü ve Değişkenliği	11
2.4.1 Genel Nedenler	11
2.4.2 Özel Nedenler	11
2.5 Kalite Kontrol	12
BÖLÜM 3.	14
3. ZAMAN SERİLERİ	14
3.1 Giriş	14
3.2 Tek Değişkenli Zaman Serileri Tahmin Modelleri	15
3.2.1 Trend Analizi Yöntemi	16
3.2.2 Hareketli Ortalamalar	17
3.2.3 Üssel Düzeltme Yöntemi	17
3.2.4 Uyarlayıcı Arındırma Yöntemi	18
3.2.5 Box-Jenkins(B.J) Tahmin Modelleri	19
3.2.5.1 Durağan Zaman Serisi($AR(p)$, $MA(q)$, $ARMA(p,q)$) Modeller	20
3.2.5.1.1 AR(Auto Regresif) Modeller	20
3.2.5.1.2 MA(Moving Average) Modeller	23
3.2.5.1.3 $ARMA$(Auto Regresif Moving Average) Modeller	25
3.2.5.2 Durağan Olmayan Zaman Serisi Modeller	25

İÇİNDEKİLER	SAYFA
3.2.5.2.1 ARIMA(Auto Regresif Integrated Moving Average) Modeller	26
3.2.5.3 Modelin Uygunluk Testi	27
3.2.5.4 Box-Jenkins Yönteminin Üstün ve Zayıf Yönleri	28
3.2.5.4.1 Üstün Yönleri	28
3.2.5.4.2 Zayıf Yönleri	29
BÖLÜM 4.	30
4. SAPAN DEĞERLER	30
4.1 Giriş	30
4.2 Basit Örneklemeler İçin Sapan Değerler	31
4.2.1 Thompson Yöntemi	31
4.2.2 Pearson ve Chandra Sekar Yöntemi	31
4.3 Zaman Serilerinde Sapan Değerler	33
4.3.1 Birinci Tip Sapan Değer(Additive Outlier)	36
4.3.2 İkinci Tip Sapan Değer(Innovation Outlier)	36
4.4 Sapan Değerlerin Tespiti	36
4.4.1 İteratif Yöntem	37
4.4.2 İteratif Yöntem'in Adımları	37
4.4.3 İteratif Yöntem'in Akış Şeması	39
4.4.4 İteratif Yöntem'in Performansı	40
4.5 Sapan Değerlerin Kalite Kontrol Yorumları	48
4.5.1 Birinci Tip Sapan Değerler	48
4.5.2 İkinci Tip Sapan Değerler	48
BÖLÜM 5.	50
5. UYGULAMA	50
5.1 Uygulama Sonuçları Tartışma ve Yorumlar	62
SUMMARY	64
KAYNAKLAR	65
EKLER	69
EK-1 İteratif Yöntem İçin Bilgisayar Programı	69

ABSTRAKT

Çok yaygın olarak kullanılan kalite kontrol yöntemlerinin etkinliğini artıracak yöntemler her zaman söz konusu olmuştur.

Bu çalışmanın kapsamı içerisinde istatistiksel kalite kontrolü için önemli sayılabilecek bir yöntem tanıtılmış ve uygulama olarak kimyasal bir üretim sürecinden iki saatte bir elde edilmiş 197 gözlem incelenmiştir. Bu uygulama işleminde eşit zaman aralıkları ile elde edilen gözlemlere uygun zaman serisi modeli uydurulmuş ve modelden büyük oranda sapan iki adet gözlem tesbit edilerek tipleri belirlenmiştir. Daha sonra sapan gözlemlerin etkisi giderilerek, gözlem değerleri için tekrar kontrol işlemi yapılmış ve sapan gözleme rastlanmadığı görülerek kontrol işlemine son verilmiştir. Bu arada sapan gözlem değerlerin, modelin parametre değerleri ve varyans üzerindeki etkileri ortaya konmuş ve kalite kontrol yorumları yapılmıştır.

ABSTRACT

A search for a more effective method to detect outliers for quality control methods which have been used for years in our country but whose effect has been discussed, has always been very popular.

In this study a method, which is important for statistical quality control, has been introduced, and 197 observations obtained from a chemical production process every two hours were investigated. In this application, a suitable time series model has been fitted to the observations obtained in equal time intervals, and as a result of two outlying observations and their types have been detected.

Later, the effect of these two mentioned outlying observations has been overcome, and the control procedure has been repeated since no outlying observations were found, the control procedure has been stopped.

Meanwhile, the effect of outlying observations on the parameters and the variance were investigated, and the quality control interpretation of the outliers were made.

ÖNSÖZ

Yirminci yüzyılın ikinci yarısında sanayi ve diğer sektörlerde kalite devriminin başlamasıyla birlikte yeni bir döneme girilmiş, işletmeler arası rekabet ön plana çıkmıştır. Kalite kontrol'ün tam anlamıyla kullanıma geçirilmesi firmaların temel amacı olan kar'ı maksimize eden unsurlardan biri olmuştur. Korumacılığın kaldırılması ve gümrük duvarlarının yıkılması ile birlikte yabancı sermaye'ye geniş imkanlar tanınmış, bunun sonucu olarak, güçlü ve dinamik kuruluşlar ulusal sınırların ötesine çok daha kolay bir şekilde erişmişlerdir. Bu yönü ile sanayi ve ticarete, genel anlamda ekonomik bir yayılmanın söz konusu olduğu söylenebilir. Günümüzde bütün kesimlerce tartışılan gümrük birliğinin sağlanması durumunda, Türkiye üretmiş olduğu mallar ile gelişmiş batılı ülkelerle rekabet etmek durumunda kalacak ve kalite kontrolü, kalitatif ve kantitatif bazda kullanıma geçirmek durumu söz konusu olacaktır. Ayrıca Türkiye' nin bu konuda daha nitelikli uzmanların yetişmesi için çalışmalara başlaması gerekmektedir.

İşte ülkemiz için bu kadar güncel ve bir o kadar da önemli olan kalite kontrol kavramı yine son derece güncel olan zaman serilerinde sapan değerlerle ilişkilendirilerek çalışılmıştır. Son yıllarda özellikle japon, çin ve koreli bilim adamlarının ilgi odağı haline gelen sapan değerler (Outliers) kavramı söz konusu ülkelerde kalite kontrol aracı olarak kullanılmaktadır. Ülkemiz sanayi ve bilimi için büyük önem taşıyacağı düşünülen bu konunun bundan sonra daha etkin bir şekilde çalışılacağı düşünülebilir.

Böyle önemli bir konuda çalışma yapma fırsatı veren, danışmanlığımı üstlenen, çalışmalarımı yakından takip ederek tezimin sonuçlanmasına katkıları olan sayın hocam Prof.Dr. Onur BASKAN'a saygılar sunuyorum. Ayrıca daima önerileri ve yardımları ile bana destek olan, bölümdeki bütün hocalarıma, yardımlarını esirgemeyen değerli arkadaşım Uzman Mehmet ESEN' e teşekkürlerimi sunuyorum.

Bu tez Ege Üniversitesi Araştırma Fonu Saymanlığı tarafından desteklenmiştir.

BÖLÜM 1

1. GİRİŞ

Belirli bir konu ile ilgili olarak verilerin uygun şekilde toplanması, açık ve anlaşılır bir tarzda ifade edilmesi büyük önem taşımakta, genel anlamda istatistiğin önemli bir aşamasını meydana getirmektedir. Ancak, özellikle son yıllarda genel anlamda istatistiğin diğer aşamaları, yani verilerin analizi ve istatistik karar teorisi, önem açısından ön plana geçmiştir. Özellikle sınırlı miktarda bilgilerden sonuç çıkarmak ve bu sonuca göre karar vermek gerektiği hallerde, istatistik yöntemlerin uygulanması, çok çeşitli alanlarda ilerleme imkanı yaratmıştır(Köksal, 1985).

Veri analizine dayalı istatistik metodlardan biri de istatistiksel kalite kontrolü'dür. Bu yöntemlerin 50-60 yıldan beri yaygın olarak kullanılmaya başlanmasıyla büyük yararlar sağlanmış ve benzer yöntemler, işletmelerin diğer faaliyetlerinde de başarı ile kullanılmaya başlanmıştır. Örneğin, bir fabrikanın bir gün içinde ürettiği mamüllerin hepsini teste tabi tutmak mümkün değil ise, kalite hakkında bir sonuca varabilmek için mamüller arasından sınırlı bir miktar seçerek karar verilebilir. Benzer şekilde, yeni bir mamülün piyasaya çıkarılıp çıkarılmaması ile ilgili bir karar vermek için muhtemel alıcılar arasından örneklemeler seçerek bulunan sonuçlarla, mamülün satılabilirliği ile ilgili bir sonuca varılabilir.

Esas itibariyle kalite kontrol, üretilen mamülün önceden belirlenen standartlara uygun olması ve piyasa şartlarında rekabet edebilmesi açısından da son derece önemlidir. Eğer üretim süreci kontrol altında değil ise, problemin kaynağı ve çözülebilirliği konusunda eldeki ölçütler yetersiz kalabilir.

Kalite kontrolde yapılan üretimi, cinsi ve niteliği gözönünde bulundurarak iki başlık altında toplamak gerekir. *Eğer yapılmış üretim, ölçülebilir, tartılabilir ve bir birimle ifade edilebilir özelliklerden bir veya birkaçına sahip ise, o taktirde üretimin sürekli olduğundan söz edilir.* Aksi durumda kesikli üretim süreci söz konusu olur. Bu çalışmanın kapsamı içerisinde yapılacak uygulama sürekli üretim süreci için bir örnek olacaktır.

Ancak zaman serilerinde kesikli ve sürekli kavramı, elde edilen gözlem değerlerinin eşit zaman aralıklarıyla elde edilip edilmemesi durumuna göre belirlenmektedir. Yani zaman serilerini oluşturan gözlem değerleri, zaman içinde eşit olmayan aralıklarla elde ediliyorsa,

meydana gelen seri, sürekli zaman serisidir. *Eğer gözlem değerleri belirli veya eşit zaman aralıklarıyla elde ediliyor iseler böyle seriler, kesikli zaman serileri olarak bilinirler.*

Uygulamada, genelde ekonomik olmakla birlikte bir çok amaca yönelik zaman serileri ile çalışılmaktadır. Mesela, ABD'de finansal politikaları tahminlemek amacıyla *Giorgio Canarella ve Stephen K.Polard* tarafından yapılan çalışma, *Akaike(1973), Schwarz(1978)* ve *Hannan(1980)* tarafından geliştirilen, ARIMA modellerin optimal özelliklerini yansıtan ölçüt çalışma esas alınarak yapılmıştır(*Pena, 1987*).

Buna benzer olarak, ARIMA ve ekonometrik modeller arası bağlantı kullanılarak, ARIMA modeller üzerine çalışmalar yapılmıştır. Bu çalışmalar yine, ABD' de *Nelson(1973), Granger ve Newbold(1977), Neftçi(1982), Harvey ve Tood(1983)*, tarafından yapılmıştır. Daha sonraları *Berkman(1980), Simpson(1980)* ve *Hafer(1980)* ARIMA modeller üzerine yapılmış yeni tanımlamalarla tahmin edilmiş parametre değerleri üzerine ayrıntılı çalışmalar yapmışlardır(*Pena, 1987*).

Ayrıca, yine ABD'de *Froyen(1974), Barro(1978), Niskanen(1978), Levy(1981), Sims(1980), Mishkin(1982), Illmakunnas ve Tsurumi(1985)* ve daha bir çok araştırmacı, ekonomik ve ekonometrik amaçlı çalışmalar yapmışlardır(*Pena,1987*).

Ekonomik amaçlı çalışmalar yanında, meteorolojik çalışmalarda, sağlık bilimlerinde, hava ve çevre kirliliği ile ilgili çalışmalarda ve sözkonusu kirliliklerin yol açtığı hastalıklar ile ilgili çalışmalarda artan oranda ikili zaman serileri (Binary Time Series) ile çalışılmaktadır. Bu amaçla ABD'de sağlık ve çevre koruma ajansı tarafından çalışmalar yapılmış ve yapılmaktadır(*Fitzmaurice ve Lipsitz, 1995*).

Zaman serileri ile ilgili yapılan çalışmalardan biri de kısmi otokorelasyon hata kareler toplamına dayalı zaman serisi modelleri için iyi uyum testleri ile ilgili çalışmalardır. Bu çalışmalardan biri de *Anna Clara Monti(1984)* tarafından yapılmıştır. Bu amaçla geliştirilmiş bir çok test bulunmaktadır. Sözkonusu testler, *Quenouille(1947,1949), Bartlett(1954), Box & Pierce(1970), Ljung & Box(1979), Ansley & Newbold(1979)* ve *Godfrey(1979)* tarafından geliştirilmiştir(*Monti, 1994*).

Ülkemizde ise, ekonomik eğilim ve tahminleri belirlemeye yönelik çalışmaların büyük bir kısmı zaman serilerinin kullanımı ile yapılmaktadır. Özellikle; milli gelir, üretim, fiyatlar,

istihdam gibi reel ekonominin temelini oluşturan kavramlar için zaman serisi analizleri çok kullanılmaktadır. Ayrıca Merkez Bankası kur tahminleri, hazine'nin bütçe ile ilgili tahminlemeleri, borsa'nın hisse senedi eğilimleri, Devlet İstatistik Enstitüsü tarafından yapılan enflasyon tahminleri ve daha bir çok tahmin zaman serisi analizleri ile yapılmaya başlanmıştır.

Zaman serileri değişik kullanım alanları ile birlikte son zamanlarda, artan oranda kalite kontrol' de üretim amaçlı süreçler için de kullanılmaya başlanmıştır.

Görüldüğü gibi, uygulamada üzerinde en çok çalışılan seriler, ekonomik eğilim ve sonuçları tahminlemeye yönelik, günlük, aylık veya yıllık bazda elde edilmiş kesikli zaman serileri olmaktadır.

Çalışmada incelenen sapan değer (Outlier) kavramı ise ilk kez 1838 yılında *Bessel* tarafından Alman astronomlar okulunda kullanılmaya başlanmıştır. Daha sonraları istatistik biliminin gelişmesi ve araştırma çalışmalarının artmasıyla doğrusal modeller, regresyon modelleri, çok değişkenli istatistik, deney düzenleme ve son olarak zaman serilerinde, sapan değer çalışmaları yapılmaya başlanmıştır(*Tatlıdil, 1981*).

Zaman serilerinde sapan değerler kavramı ise ilk defa *Fox(1972)* tarafından çalışılmıştır. *Fox*, otoregresif modellerde sapan değerleri ortaya çıkarmak amacıyla olabilirlik oran kriteri adını verdiği bir ölçüt geliştirmiş ve ortaya çıkan sapan değerleri 1 ve 2.tip sapan değerler biçiminde tanımlamıştır. Ayrıca *Fox*, güç fonksiyonları üzerine çalışmıştır(*Fox, 1972 ve Hawkins, 1981*).

Bundan sonra bir çok araştırmacı, *Fox'* un yaptığı çalışmaları daha da geliştirerek, bütün ARIMA modelleri kapsayacak şekilde, çoklu sapan değerleri taramaya yönelik yöntemler geliştirmişlerdir. Simülasyon çalışmaları, bu yöntemlerden en etkin olan yöntemin *Chang(1982), Chang ve Tiao(1983)*, tarafından geliştirilen iteratif yöntem adını verdikleri yöntem olduğunu ortaya koymuştur. *Hillmer(1983) Tsay(1986), Pena(1987), Abraham ve Yatawara(1988), Bruce ve Martin(1989)* iteratif yöntemi tartışmışlardır. Bununla birlikte *Abraham ve Yatawara(1988)* Lagrange çarpanlar metodu veya skor tabanlı sapan değer testleri üzerine çalışmalar yapmışlardır. *Pena(1987), Abraham ve Chuang(1989) ve Bruce ve*

Martin(1989) sapan deęer taramada, sapan olan gozlemin silinmesi esasına dayalı testler ile birlikte zaman serilerinde etkili gozlemler zerine alıřmalar yapmıřlardır(*Ljung, 1993*).

Dięer yandan *Harrison ve Stevens(1976)*, *Smith ve West(1983)*, *West, Harrigon ve Migon(1985)* ve *West(1986)* zaman serilerinde sapan deęerler elde etmek iin sıralı tarama(Sequential detecting) yntemleri zerinde alıřmıřlardır(*Tsay, 1988*).

Sıralı tarama yntemleri zerine alıřmalar doęrusal regresyon problemleri iin de alıřılmıřtır. Bu amala, *Mararsinghe(1985)*, *Kianiard ve Swallow(1990)* tarafından sıralı strateji testi adıyla bir yntem geliřtirilmiřtir(*Pena ve Yokai, 1995*).

Ayrıca, *Jones(1980)*, *Ljung(1982, 1989)*, *Harvey ve Pierce(1984)*, *Kohn ve Ansley(1986)*, *Wincek ve Reinsel(1986)* ve dięerleri, gozlem deęerlerinden bazılarının kayıp olması durumunda, sapan deęerlerin saptanmasına ynelik olarak olasılık oran kriterleri geliřtirmıřlerdir(*Ljung, 1993*).

Zaman serilerinde sapan deęerlerin etkisi zerine alıřmalar, ilk defa *Box & Tiao(1965)* tarafından alıřılmıřtır. *Box & Tiao* yaklařımı, *Chen ve Tiao(1986)* tarafından, *Chang ve Tiao(1983)*'n iteratif yntemi kullanılarak genelleřtirilmiřtir. Son zamanlarda ise *Tsay*, buna yeni eklemeler yapmıřtır. Birinci dereceden AR modellerde sapan deęer etkileri ise *Whichern, Miller ve Hsu(1976)* tarafından yapılmıřtır. Bu arařtırmacılar, varyans deęiřimlerinin parametre deęerleri zerindeki etkisini incelemiřlerdir(*Tsay, 1988*).

Zaman serilerinde sapan deęerlerin saptanmasına ynelik alıřmalardan biri de Robust Procedure adı verilen ve *Denby ve Martin(1979)* tarafından geliřtirilen yntemdir. Bu yntem *Martin ve Yokai(1985)* tarafından da alıřılmıřtır. Ancak *Chang ve Tiao(1983)* sz konusu yntemin 2. tip sapan deęerlerin teřhis edilmesinde etkin olmadıęını ortaya koymuřtur (*Tsay,1988*).

Bruce ve Martin(1989), 1. tip sapan deęerin saptanması ve parametre tahminlemesine ynelik alıřmalar yapmıřlardır. Arařtırmacılar, ayrıca sapan deęerlerin sadece bulunduęu pozisyonu deęil, yakın pozisyonları da etkiledięini ortaya koymuřlardır(*Tsay,1988*).

Chang(1988) ise, yanlış belirlenen sapan deęer tiplerinin, test yntemlerinde etkinlik kaybına neden olduęunu ortaya koymuřtur(*Muirhead, 1986*).

Sapan deęerleri tespit etmede kritik deęer olarak kullanılan C deęerlerini saptamak amacıyla Monte-Carlo simülasyonları kullanılmıştır. Ayrıca, *Berman(1964)* tarafından bulunan asimptotik sonuçların da kullanılabilir olduęu yine Berman tarafından ispat edilmiştir(*Ljung 1993*).

Sapan deęerler kavramı, bir çok arařtırmacı tarafından farklı biçimlerde tanımlanmıştır. Mesela sapan deęer, *Bross(1961)* tarafından, "*görünümü geri kalan gözlem deęerlerinin görünümüne uymayan bir veya bir kaç gözlem*" olarak tanımlanmaktadır. *Stefansky(1972)*, ise "*kestirim deęerlerinden çok farklı olan gözlem deęerleri*" şeklinde bir tanım vermiştir. Aynı tanım, "*örneklemedeki dięer gözlemlerden farklı olan ve kestirim sonuçlarını büyük ölçüde etkileyen az sayıda gözlem deęeri şeklinde de verilmiştir.*" Bu ve buna benzer bir çok tanım verilebilir. Ancak *bir gözlemin sapan olabilmesi deęer olarak büyük yada küçük olmasıyla ilişkilendirilemez.* Gerçek model denklemini elde edildiğinde büyük hata deęerine sahip olan gözlem veya gözlemlerin sapan olmaları olanaklıdır. Bu durumu gözönünde bulundurarak, "*gözlem ve kestirim deęerleri arasındaki farkın en büyük olduęu gözlem, sapan gözlem olabilir.*" denilebilir.

Sapan deęerler, model yanlışlığı, gerekli bazı dönüşümlerin yapılmamış olması, ölçüm, tartım , ya da kaydetme hataları gibi çeşitli nedenlerden kaynaklanabilmektedir (*Andrews ve Pregibon, 1977*).

Örneklemedeki bazı gözlemler dięer gözlemlerden farklı bir veya daha çok etken sonucunda sapan deęerler oluşturabilmektedir. Bu deęerler doğal rastgelelik sonucunda da ortaya çıkabilmektedir(*Tatlıdül, 1981*).

Zaman serilerinde sapan deęerleri ortaya çıkarmak amacıyla bir çok yöntem geliştirilmiştir. Simülasyon çalışmaları ve denemeler, bu yöntemlerden en etkin olanın Chang-Tiao ve Chen'in(1988) geliştirdikleri ve "iteratif yöntem" adını verdikleri yöntem olduğunu ortaya koymuştur. Söz konusu yöntem, dięer yöntemlere göre uygulanması kolay ve pratik olmasına rağmen yine de bazı karmaşık matematiksel işlemleri gerektirmektedir(*Chang, Tiao ve Chen, 1988*).

Kalite kontrolde, sürecin kontrol altında olup olmadığını belirlemek amacıyla, süreçten elde edilen gözlemlerin bir zaman serisi oluşturduęu düşünülerek serinin uyum gösterdięi

parametre tahminleri elde edilmekte, daha sonra iteratif yöntem ile seride ortaya çıkması muhtemel iki tip sapan değer belirlenmektedir. *Sapan değer tiplerine göre problem; giridi, çıktı ya da süreç problemi olarak sınıflandırılmaktadır.* Daha sonra sapan değerlerin etkileri giderilmekte ve seride sapan değer kalmayınca kadar işlem tekrar edilmektedir(Acar ve Booth,1987).

Bu çalışmanın amacı, Box-Jenkins tahmin modelleri olarak da bilinen stokastik zaman serisi modellerinin (AR, MA, ARMA, ARIMA) tanıtımını yapmak, zaman serilerinde ortaya çıkması muhtemel iki tip sapan değeri belirlemek, ortaya çıkan sapan değer tiplerine göre kalite kontrol yorumları ve önerileri kapsamaktadır. Çalışmada, zaman serileri ve kalite kontrol kavramlarına bağlı olarak, ikinci bölümde kalite ve kalite kontrol ile ilgili tanımlar, üçüncü bölümde zaman serisi modellerinin tanıtımı, dördüncü bölümde basit örnekler için sapan değerlerin saptanmasına yönelik test yöntemleri, bunun yanında zaman serilerinde karşılaşılabilecek muhtemel sapan değerler ve sapan değeri saptamak amacıyla *Chang-Tiao* ve *Chen*'in geliştirmiş oldukları iteratif yöntem tanıtılacaktır. İteratif yöntem sonucu ortaya çıkan sapan değerler, tiplerine göre sınıflandırılmakta ve bunların kalite kontrol amaçlı süreçler için ne anlam taşıdığı ifade edilmektedir. Beşinci bölümde ise uygulama sonuçları ve tartışma yer almaktadır.

BÖLÜM 2

2. KALİTE KONTROL

2.1 Giriş

Kalite devrimi 1950'de *Edwards Deming* isimli bir istatistikçinin Japonya'ya, Japon yöneticilerine kalite yönetim ilkelerini öğretmek için gitmesiyle başlamıştır. Japonya bu ilke ve yöntemleri öğrenmekle kalmamış, bunları özümsemiş, uygulamış ve daha da geliştirmiştir. Bir kaç yıl içerisinde, çürük mallar için ikaz etiketi olan Japon malları, üstün kalite ve değer in sembolü durumuna gelmiştir. Özellikle Amerika ve diğer batılı ülkeler, Japonya'nın pazarlarını küçültüğünü görünce, endüstrileri için kaçınılmaz olduğuna inandıkları kalite ve kontrol kavramını felsefesiyle uygulamaya sokmuşlardır. Örneğin Amerika'da kalite devrimi 1980'lerin başında otomobil ve elektronik sanayilerinin Japon ürünlerinin amansız rekabeti karşısında büyük kayıplar vermesi sonucunda gerçekleşmiştir. Günümüzde ise bütün dünya ülkeleri kalite kontrol kavramının gereğini ve önemini göz önünde bulundurarak, hızlı bir şekilde kullanmaya ve bu alanda uzmanlar yetiştirmeye başlamışlardır (*Kavrakoğlu, 1992*).

2.2 Kalite'nin Gelişimi

Kalite gelişimi, 1920'lerde İngiliz istatistikçi Sir *R.A.Fisher* tarafından, tarım alanlarında istatistiksel deney tasarımı kullanılarak, bulunmuş ve geliştirilmiştir.

Kısa sürede, Amerika' da tarım üretiminin geliştirilmesi için yoğun olarak uygulanmış ve Amerika' nın bu alanda dünyada lider konumuna gelmesine büyük katkı yapmıştır. Sonraları kimya ve ilaç endüstrisinde de kullanılmış olmasına rağmen, imalat sektöründeki son yıllara kadar kısıtlı kalmıştır. Amerika'nın imalat sektörünü bir bakıma yeniden keşfetmesi, 1980'den sonra Japon kalite uygulamalarının incelenmesi sonucunda başlamıştır.

Dünyada kalite alanındaki gelişmelerin durumuna bakıldığında Japonya'nın en önde olduğu görülebilir. Ancak Amerikada 1980' lerde başlayan kalite çalışmaları bütün hızı ile devam

ediyor. Amerika, Japonya ile birlikte Batı Avrupa ve Pasifik ülkeleri de kalite gelişiminde önemli mesafeler almaya başlamışlardır.

2.3 Kalite

Kalite latince "qualitas" demektir ve nasıl oluştuğu anlamına gelen "qualis" sözcüğünden gelmektedir. Sözcük anlamı hangi nesne için kullanılıyorsa onun gerçekte ne olduğunu belli etmek amacını taşır. Kalite sözcüğünün tanımında önemli derecede karışıklıklar vardır. Çünkü kalite zaman içinde değişen koşullara uygun olarak şekil değiştirir.

Bu nedenle kalite' nin tanımı da çeşitli değişimlere uğramıştır. Bu değişim, tanımların önce yalnız tüketiciye yönelik düşünülmesinden kaynaklanmış ancak daha sonra bu tek yönlülük kaldırılarak kalite kavramı genelleştirilmiştir. Yapılmış tanımlardan hareketle, genel bir tanım vermek gerekirse kalite, *bir ürün ya da hizmetin özellikleri itibariyle toplumun beklentilerini karşılayacak düzeyde olmasıdır* denilebilir. Örneğin, suyun kalitesi, içme suyu saf, mikrop ve zehirli maddelerden yoksunlukla; yıkanma suyu yumuşaklık kir ve yağların giderilmesi ile; kağıt yapımında, eriyik halindeki zerreciklerden ve renklendirici maddelerden yoksunluğu ile belirlenir. Görüldüğü gibi her durumda suyun kalitesi belirli bir kullanım için suyun uygunluğunu saptayan bir grup özellikleri taşır (Baskan, 1994).

Bunun yanında kalite' nin *tasarım kalitesinin, tasarıma uygunluk kalitesinin, kullanımda bekleneni verme kalitesinin* kontrolleri ile sağlanabileceği anlaşılmıştır.

Tasarım kalitesi denince akla, üretilen ürünün boyut, ağırlık, hacim, dayanıklılık gibi fiziksel yapı ile ilgili spesifikasyonların belirlenmesi gelir.

Tasarıma uygunluk kalitesi ise, üretilen bir ürünün tasarım aşamasında belirlenen tasarım kalitesi ile ilgili spesifikasyonlara uyma derecesidir. Tasarıma uygunluk kalitesinin kontrolü ham ve yardımcı maddelerin sağlanmasından, üretilen malın ambalajlanıp depolanmasına

kadar geçen tüm aşamaları kapsar ve istatistiksel yöntemlerin en yaygın biçimde uygulandığı alanlardır.

Kullanımda bekleneni verme kalitesi ise bir tüketici tarafından satın alındığında ne derece hizmet verdiğinin ölçüsüdür(Baskan,1994).

2.4 Kalitenin Ölçüsü ve Değişkenliği

Üretim işlemi çok sayıda karmaşık faktörlerin etkisi altında meydana gelir. Bunun için değişkenlik nedenlerinin tümünü saptamak ve saptananların etki derecelerini kesin bir şekilde belirtmek olanaksızdır. Öte yandan, bir üretim işlemini etkileyen tüm faktörlerin, etki dereceleri ile birlikte bilindiği varsayılsa bile bunlar zaman içinde değişkenlik gösterir. Değişkenlik nedenleri, çok ve etkileri farklı olmakla birlikte kalite açısından ikiye ayrılırlar.

2.4.1 Genel Nedenler

Üretimi etkileyen faktörlerin hepsinde sürekli var olan nedenlerdir. Bunlar: kalite programının niteliği, hammadde özelliklerinin belirlenmesi, işletme koşulları, makinaların durumları, işçilerin fabrikadaki sosyal durumları şeklinde sıralanabilir.

2.4.2 Özel Nedenler

Üretim faktörlerinde sadece bir veya bir kaçında zaman zaman araya giren aksaklıklardır. Bunlar, makinalardan birinin arızalanması, işçilerden birinin işi aksatması hammaddelerin istenilen nitelikte olmaması gibi aksaklıklardır(Baskan, 1994).

Kalitenin ölçüsünü veren boyut, konum, montaj, çalıştırma ve benzeri değişkenliğin gözlenebileceği genel nedenlerle ilgili konularda ölçme işlemleri çok önemlidir. Kalite kontrolde tasarım ve uygunluk kaliteleri ancak ölçme faaliyetleri sonunda önem kazanır. Üretilen ürünle standart değer arasında bir fark görüldüğünde bu farklılığın gerçek mi yoksa rastgele mi oluştuğu düşünülür. Böyle bir sorun karşısında, üretimde yeni ölçmeler yaparak aynı farklılığın görülüp görülmediği araştırılır.

Buna göre *kalite kontrol faaliyetleri içinde ölçme aletlerinin seçimi, genelleştirilmesi, kullanılması, bakımı ile ölçme yöntemlerinin uygulanmasından oluşan faaliyetler topluluğuna ölçme yöntemi denir*. Günümüz üretim koşullarında ölçme işlemleri, üretim öncesi ölçmeler, üretim sırasında ölçmeler, üretim sonrası ölçmeler olmak üzere üç grupta toplanır(Başkan, 1994).

2.5 Kalite Kontrol

Kalite kontrolü çeşitli biçimlerde tanımlama olanağı vardır. Sanayide kalite kontrolü, genellikle ürün kalitesinin istenen düzeyde tutulması, geliştirilmesi ve üretimde hatalı ürün verilmesinin önlenmesi ya da azaltılması anlamına gelir. Ürün kalitesi alıcı tarafından, aranan belirli koşullar olup ürünün esas kullanım durumu ve fiziksel özelliklerine ait garanti belgesi, dayanma süresi ve güvenilirlik, satış fiyatı gibi önlemlerin alınması gerekir. Bunların başta gelenleri uygun hammadde sağlanması, kaliteli makina, tesis ve cihaz kullanılması, makina ayarlarının uygun olması ve benzeridir.

Görüldüğü gibi tasarım, üretim ve kontrol fonksiyonları birbirinden çok farklı olmakla birlikte, ürünün standart uygunluğunu en az masrafla sağlayabilmek için aralarındaki düzenlemeyi sağlayacak etkili bir yöntem aracına gereksinim vardır ve araç, kalite kontrol'dür.

Kalite kontrol, üretim satın alma ve başka alanlarda kalitenin sağlanması, sürdürülmesi ve yükseltilmesi çalışmalarını planlama ve geliştirme yolu ile üretimin, tüketici açısından en ekonomik düzeyde ve yüksek kalitede yapılmasına olanak sağlar. Bu nedenle kalite kontrol'ün önemi artık tartışılmaz biçimde kabul edilmiştir.

O halde kısaca kalite kontrol, *bir işletmenin kalite hedeflerine ulaşması, kalite isteklerinin sağlanması için kullanılan uygulama yöntemleri ve faaliyetlerin tümü olarak tanımlanır*.

Kalite kontrol faaliyetlerinin gerçekleştirilmesi çoğu kez gözlemlerin, üretim sürecinden örnekleme yoluyla alınmasını gerekli kılmaktadır. Bu, hem zaman hem de maliyet açısından

gerekli, bazen kaçınılmazdır. *O halde örnekleme, üretim sürecinden elde edilen mamüllerin tümü üzerinde inceleme yapmak yerine, mamüller arasında belli teknikler kullanılarak, sürecin özelliklerini yansıtacak bir alt grubun alınıp incelenmesi işlemidir.*

Seri imalata dayanan büyük miktarlarda malların belli zaman aralıklarına göre, partiler halinde teslimi söz konusu olduğu zaman malların saptanan spesifikasyonu karşılayıp karşılamadığı test edilerek kabul veya red edilir. Olasılıklı olmayan örnekleme yöntemlerinin işletmelerde en çok kullanılan türlerinden biri olan bu sözkonusu yöntem kabul örneklemesi olarak bilinir (*Başkan,1994*).

Kabul örneklemesinin başlıca iki amacı vardır. Bunlar,

- 1) Kötü kaliteye karşı tüketiciyi korumak,
- 2) Üreticiyi iyi kaliteye özendirip kötü kalite için cezalandırmaktır.

Kabul örneklemesi tek bir partinin kalitesini belirlemek için kullanılan bir yöntem olmadığı gibi üretim işleminde kaliteyi kontrol altına almak için kullanılan bir yöntem de değildir.

Kabul örneklemesi, çok sayıda partilerin teslimine dayanan bir işlemde taraflarca saptanan ortalama kalitenin ekonomik ve güvenilir bir biçimde elde edilmesini sağlar.

Kabul örnekleme türlerinin üç çeşidi vardır. Bunlar,

- 1) Tek kabul örneklemesi,
- 2) Çift kabul örneklemesi,
- 3) Çoklu kabul örneklemesi.

Uygulandığı kalite niteliği bakımından ise kabul örnekleme türleri , sürekli değişkenler için ve süreksiz değişkenler için kabul örnekleme türleri olmak üzere ikiye ayrılmaktadır(*Başkan,1994*).

BÖLÜM 3

3. ZAMAN SERİLERİ

3.1 Giriş

Zaman serileri, verilerin oluş zamanları esas alınarak elde edilen serilerdir(Colton, 1968).

Bir başka deyişle, *zamanın fonksiyonu olarak elde edilmiş verilerin oluşturduğu seriler zaman serileri olarak bilinir.* Tahminler yapmak ve gelecek hakkında bilgi edinmenin en iyi yollarından birisi zaman serisi analizidir. Zaman serisi analizinin yapılabilmesi için, parametre tahminleri için yanlılığa sebep olan, yani ileri sürülen varsayımların göz önünde tutulmamasına yol açan *trend, konjonktürel dalgalanma, mevsimlik etki ve rasgele değişkenliklerin* kontrol edilmesi gerekmektedir(Çömlekçi, 1982).

Zaman serileri çeşitli amaçlar için analiz edilir. Bu amaçlar aşağıdaki gibi sıralanabilir.

Zaman serisi *unsurlarına ayırma*, zaman serileri arasındaki *ilişkiyi açıklama, kontrol* ve ileriye dönük *tahmin* amaçlarından oluşmaktadır.

Zaman serileri analizi' nin en önemli amacı bu serilerin ileriye dönük tahmin amacıyla analiz edilmesidir(Chatfield, 1980 s: 7-9).

Bu çalışmanın kapsamı içerisinde ileriye dönük tahmin amacı ile yapılan parametre kestirimleri kullanılarak, kontrol işlevi yapılmaktadır. Yani zaman serilerinin iki amacı kullanılarak bir sentez yapılmaktadır.

Zaman serilerinin ileri dönük tahmin yöntemleri iki grupta toplanabilir. Bunlar *Çok değişkenli* ve *Tek değişkenli zaman serileri* ile ilgili tahmin yöntemleridir.

Çok değişkenli zaman serileri iki veya daha fazla zaman serisi arasındaki sebep-sonuç ilişkisini tanımlayan ve daha sonra tahmin ve kontrol amacıyla kullanılan yöntemlerdir. Tahmin edilecek değişken ile bu değişkeni açıklayan diğer değişkenler arasında mantıksal ilişkiler varsa ve bu değişkenlerin zaman aralıklarıyla aldığı sayısal değerler mevcut ise bir ilişki modeli kurulur. Literatürde dönüşüm fonksiyon modelleri, dinamik regresyon modelleri ve çok değişkenli zaman serileri analizi olarak bilinen yöntemler buna örnek olarak verilebilir.

Çok değişkenli zaman serileriyle ilgili tahmin modelleri, tahmin sistemiyle ilgili herşeyin bilindiğini dikkate alır ve birbiriyle ilişkili olayların tahmin edilmesini sağlar. İlişki modellerine dayanarak yapılan tahminlerin hatası düşük olabilir. Ancak tahmin sistemiyle ilgili herşeyin bilinmesi çoğu zaman mümkün olmayabilir veya mümkün olsa bile analiz için uygun olmayabilir. Bunu ekonomide bir örnekle ifade etmek gerekirse, milli gelir değişkeni, çok sayıda ekonomik değişken için açıklayıcı değişken durumundadır. Ancak yıllık değerleri bilinen bir değişkenin aylık değerlerini istatistiksel yöntemlerle belirlemek mümkün olsa da, belirlenen değerler sağlıklı olabilir. Bu nedenle aylık verileri esas alan ekonometrik çalışmalarda kullanılan ilişki modellerinde sağlıklı olmayan verilerin kullanılması ile elde edilen bulguların güvenilirliği düşük olabilir. Bu nedenlerden dolayı zaman serilerinin ileriye dönük tahmininde, tek değişkenli zaman serileriyle ilgili tahmin yöntemleri yaygın bir şekilde kullanılmaktadır(Özmen, 1986).

3.2 Tek Değişkenli Zaman Serileri Tahmin Modelleri

Tek değişkenli zaman serileri ile ilgili tahmin yöntemleri, zamana bağlı tek bir değişkene ait verilerin mevcut olması durumunda kullanılan ve sadece ileriye dönük tahmin yapmaya imkan veren istatistiksel yöntemlerdir. Bu yöntemler zaman serilerinin bu günkü ve geçmiş dönem gözlem değerlerini kullanarak, serilerin gelecek dönem tahmin değerlerinin elde edilmesini sağlarlar. *Yöntemlerin dayandığı varsayımlar şunlardır.*

- 1) Bir zaman serisinde mevcut olan zaman serisi unsurlarının gelecek dönemde de aynı kalacağı kabul edilir. Bu varsayım nedeniyle geçmiş dönem gözlem değerlerine dayanarak gelecek dönem tahmini değerlerin elde edilmesi sağlanır.
- 2) Bu yöntemler, zaman serisini meydana getiren unsurları birbirlerinden ve tesadüfi unsurlardan ayırmak suretiyle serinin gelecekte alabileceği değeri tahmin etmeyi amaçlar.
- 3) Bu yöntemler, eşit zaman aralıklarıyla elde edilen gözlem değerlerinden meydana gelen kesikli zaman serilerine uygulanır(Özmen,1986).

Tek değişkenli zaman serileriyle ilgili tahmin yöntemleri, geliştirildikleri tarih sırasına göre tanıtılacaktır.

3.2.1 Trend Analizi Yöntemi

Geliştirilen yöntemler arasında en eski olanı trend analizi yöntemidir. Hesaplanması ve anlaşılması kolay olan bu yöntem, günümüzde orta ve uzun dönem tahmin amacıyla sık kullanılan sayısal tahmin yöntemlerinden biridir.

Trend analizinin esası, zamana bağlı herhangi bir olaya ait değerlerin serpmeye diyagramında göstermiş oldukları serpmeye uygun bir matematiksel fonksiyon belirlemek ve bu fonksiyonla ilgili olayın zamana göre nasıl bir eğilim gösterdiğini belirlemektir. Kısacası biri açıklayıcı, diğeri açıklanan değişken kabul edilen iki vasıf arası ilişki matematiksel olarak aşağıdaki gibi ifade edilir(*Chambers, Mullick ve Smith, 1971*).

$$X_t = f(t) \dots\dots\dots(3.1)$$

Trend analizi ile ilgili tahmin yönteminin kullanılabilmesi için enaz 7 yıla ait verilerin bulunması gerekir. Yöntemde sakınca olarak görülebilecek unsurlardan birincisi tahmin işleminde sadece iki değişkenin dikkate alınmasıdır. Oysa, ekonomik olayların meydana gelmesinde etken olan çok sayıda faktör bulunur. Bunun yanında ARIMA modeller için öngörülen model denkleminin yeterli olup olmadığına ilişkin test , trend analizi için sözkonusu değildir. Ancak tahmin hatalarının karesini minimize eden denklem yeterli kabul edilir(*Whellwright ve Makridakis, 1973 s:93*).

Diğer taraftan trend analizi ile elde edilen tahmin değerleri için kurulacak güven aralıkları, tahminin yapıldığı zaman aralığı ile trendin tahmininde kullanılan verilerin kapsadığı zaman arasındaki farka bağlı olduğundan, uzun dönem tahminleri için kurulacak güven aralıkları oldukça geniş olacaktır. Bu nedenle yapılan tahminlerin güvenilirliği azalır. Ayrıca trend analizi, mevsim unsurunu dikkate almaz(*Özmen, 1986*).

3.2.2 Hareketli Ortalamalar

Bu tahmin yönteminin esası, bir zaman serisindeki gözlem değerlerini belirli büyüklükteki kümeler halinde toplamak, her küme için aritmetik ortalama hesaplamak ve bu ortalamayı, ait olduğu kümenin en yeni terimini izleyen terimin tahmin değeri olarak kabul etmektir (*Whellwright ve Makridakis, 1973*).

$$X_{t+1} = \frac{1}{N} [X_t + X_{t-1} + \dots + X_{t-N+1}] \dots\dots\dots (3.2)$$

Hareketli ortalamalar tahmin yöntemi gözlem değerlerinin oluşumunda rasgeleliğin yüksek olduğu, buna karşılık birbirini izleyen gözlem değerleri arasındaki otokorelasyonun düşük olduğu zaman serisinde uygulanır. Kısa dönem tahmin amacıyla kullanılabilir olan bu yöntemin uygulanabilmesi için çok sayıda gözlem değerine gereksinim vardır.

Hareketli ortalama yöntemine yapılan en önemli eleştiri, bu yöntemin sadece hareketli ortalama dönemindeki tarihi verilere eşit ağırlık vermesi, eski dönemleri bütünüyle görmezden gelmesidir (*Şahin, 1982 s:163*).

Bu yöntemin kullanılmasıyla yapılacak olan uzun dönem tahmin değerlerinin doğruluk derecesi düşüktür. Ancak bir aylık ön dönem için yapılan tahminlerin güvenilirliği yüksek olabilir (*Chambers, Mullick ve Smith, 1971 s: 55*).

Hareketli ortalamalar tahmin yöntemi de trend analizi tahmin yöntemi gibi mevsimsel veriler için uygun bir yöntem değildir.

3.2.3 Üssel Düzeltme Yöntemi

Üssel düzeltme yönteminin kuramsal esasları ilk defa 1958 yılında *C.C. Holt* tarafından ortaya konmuştur. *Holt* tarafından geliştirilen üssel düzeltme yöntemi mevsim ve trend unsuru içermeyen basit formdaki zaman serileri için uygulanmıştır. *Brawn*, yöntemle tam açıklık kazandırmış ve uygulama alanına koymuştur. 1960 yılında *Winter* üssel düzeltme yöntemini mevsimsellik gösteren ekonomik zaman serileri haline getirmiştir. Bu yöntemlerin her biri

kendinden önce geliştirilmiş olan yöntemlerin dezavantajlarını avantaja dönüştürmeyi amaç edinmiştir. Bu nedenle üssel düzeltme yöntemlerinin bazıları diğerlerine nazaran daha çok yönlü, bazıları hesaplama açısından çok karmaşık, bazıları fazla bilgisayar zamanı alır (*Whellwright ve Makridakis, 1973 s:58-100*).

Üssel düzeltme tahmin yöntemleri temel özellik açısından hareketli ortalama tahmin yöntemine benzer. Fakat üssel düzeltme yöntemleri zaman serilerinin tüm gözlem değerlerini gözönünde bulundurdıkları ve seri değerlerine bu günkü dönemden uzaklıklara göre azalarak tartı verdikleri için hareketli ortalama yönteminden ayrılırlar.

Üssel düzeltme tahmininde kullanılan ifade aşağıdaki gibidir.

$$X_t = aX_t + a(1-a)X_{t-1} + \dots + a(1-a)^k X_{t-k} \dots\dots\dots (3.3)$$

Zaman serilerini meydana getiren bütün unsurları dikkate alan üssel düzeltme yöntemlerinin zaman ve para bakımından uygulama maliyeti düşüktür. Yöntemin uygulanabilmesi için çok uzun zamana gerek yoktur.

3.2.4 Uyarlayıcı Arındırma Yöntemi

Bu yöntem 1970'li yılların ilk yarısında *Whellwright ve Makridakis* tarafından geliştirilmiştir. Uyarlayıcı arındırma yönteminde zamana bağlı bir olayla ilgili tahmin modeli belirlendikten sonra, bu olayı meydana getiren unsurlarda meydana gelebilecek değişiklikleri yeniden bir tahmin modeli belirlemeye gerek bırakmadan doğrudan tahmin değerlerine yansıtma imkanı olduğundan bu yönteme ilişkin modellere "kendi kendini yenileyen modeller" denir. Bu modeller tahmin işlevinde de araştırmacının müdahalesini minimum düzeye indirir.

Uyarlayıcı arındırma tahmin yöntemine göre herhangi bir gelecek dönemin tahmin değeri, hareketli ortalamalar ve üssel düzeltme yöntemlerinde olduğu gibi geçmiş dönem gözlem değerlerinin toplamı alınarak elde edilir.

Uyarlayıcı arındırma yöntemi hareketli ortalama ve üssel düzeltme yöntemleri gibi kısa dönem tahmin amacıyla kullanılır. Bu yöntemlere dayanılarak elde edilen tahmin sonuçlarının

güvenilirliği daha fazladır. Uyarlayıcı arındırma yönteminde her gözlem değeri için ayrı ayrı hesaplanacak olan tartılar, tahmin hatalarının karelerinin toplamını minimum edecek şekilde belirlenir(*Ekern, 1976 s: 705-706*).

Yukarıda tanımlanan dört farklı tahmin yöntemi, tek değişkenli zaman serilerinin tahminine yönelik olarak kullanılan değişik yöntemlerdir. Söz konusu yöntemler, zaman serisi gözlem değerlerinin karakteristiklerine bağlı olarak değişik üstünlük ve avantajlar sağlarlar. Ancak tek değişkenli olan bu yöntemlerde tek bir tahmin fonksiyonu geliştirilip programlanır ve otomatik olarak, bilgisayar programına müdahale olmadan sürekli olarak tahmin elde edilir. Bu yöntemler zaman serilerinin ardışık gözlem değerleri arasında var olan bağımlılığı dikkate almazlar. Bu nedenle bir zaman serisi için hesaplanan ortalama değer zamanın deterministik bir fonksiyonu olduğu varsayılır. ve bu ortalamaya belirli bir dönemin hata terimi ilave edilerek o dönemin gözlem değeri elde edilir. Oysa zaman serilerinin çoğunda ardışık gözlem değerleri birbirine oldukça bağımlıdır(*Newbold, 1975 s:397*).

Bu durumda trend analizi için hareketli ortalama ve üssel düzeltme yöntemleri uygun değildir, çünkü bu yöntemler bağımlılık avantajını kullanmazlar. Halbuki *zaman serilerinde bağımlılık yapısını çok etkin bir şekilde kullanan Box-Jenkins tahmin modelleri olarak da bilinen ARIMA modeller bulunmaktadır*. Şimdi bu modeller tanıtılmaya çalışılacaktır.

3.2.5 Box-Jenkins (B.J) Tahmin Yöntemleri

Literatürde Box-Jenkins tahmin modelleri olarak da bilinen zaman serisi modelleri durağan karakterde olup olmama durumlarına göre iki grupta incelenirler. Tek değişkenli zaman serilerinin ileriye dönük tahminleri için kullanılan başarılı bir yöntemdir. Zaman serilerinin (B.J) yöntemine ilişkin modeller son 20-30 yıl içinde geliştirilmişlerdir. (B.J) grubu modeller, zamana bağlı olayların rasgele karakterde olması ve bu olaylarla ilgili zaman serilerinin stokastik süreç olduğu varsayımına dayanarak geliştirilmişlerdir. Ayrıca bu modellerde rasgele değişkenin zaman içinde ardışık olarak aldığı değerler arasında mevcut olan

otokorelasyon en etkili bir şekilde dikkate alınır. Bu nedenlerden dolayı sözkonusu modellere stokastik modeller adı verilmektedir (Özmen, 1986).

3.2.5.1 Durağan Zaman Serisi (AR(p), MA(q), ARMA(p,q)) Modelleri

Tek değişkenli zaman serilerinin ileriye dönük tahmininde kullanılan yöntemlerden biridir. Bu yöntem için en önemli varsayımlar gözlem değerlerinin eşit zaman aralıklarıyla elde edilmesi, kesikli ve durağan olmasıdır. Serinin durağanlığı üzerine konmuş varsayım, çoğu zaman bozulmaktadır. Durağan olmayan zaman serilerinin ileriye dönük tahmininde (B.J) yönteminin uygulanabilmesi için önce durağanlığı bozan sözkonusu etkenlerin ortadan kaldırılması gerekir.

3.2.5.1.1 AR(Auto-Regressif) Modeller

Bu modeller, bir zaman serisinin herhangi bir dönemindeki gözlem değerini, aynı serinin ondan önceki, belli sayıda geçmiş dönem gözlem değerine ve hata terimine bağlı olarak açıklayan modellerdir.

AR modeller, içerdikleri geçmiş dönem gözlem değeri sayısına göre isimlendirilirler. Yani AR modeli bir tane geçmiş dönem gözlem değeri içeriyorsa "1.dereceden", iki tane geçmiş dönem gözlem değeri içeriyorsa "2. dereceden" ve genel olarak p tane geçmiş dönem gözlem değeri içeriyorsa "p. dereceden" AR model olarak isimlendirilmektedir. AR(p) Modeli,

$$\chi_t = \phi_1 \chi_{t-1} + \phi_2 \chi_{t-2} + \dots + \phi_p \chi_{t-p} + \theta_t \dots \dots \dots (3.4)$$

şeklinde ifade edilir. $\chi_t = \chi_t - \mu$ küçültülmüş değerlerdir.

Burada $\chi_{t-1}, \chi_{t-2}, \dots, \chi_{t-p}$ gözlem değerlerini, $\phi_1, \phi_2, \dots, \phi_p$ modelin parametrelerini ifade etmektedir. Burada hata terimleri üzerine konmuş varsayım $N(0, S^2)$ şeklindedir. AR(p) doğrusal modeli için tahmin edilmesi gerekli parametre sayısı (p+2) tanedir.

Uygulamada sıkça kullanılan AR modelleri, birinci ve ikinci dereceden modellerdir. Bunlar AR(1) ve AR(2) biçiminde gösterilir. AP(1) Modeli,

$$\chi_t = \phi_1 \chi_{t-1} + \theta_t$$

şeklindedir, aynı şekilde AR(2) Modeli,

$$\chi_t = \Phi_1 \chi_{t-1} + \Phi_2 \chi_{t-2} + \epsilon_t$$

şeklinde gösterilir. AR modeller, fark denklemi şeklinde de yazılabilir.

$$B\chi_t = \chi_{t-1}, B^2\chi_t = \chi_{t-2}, \dots, B^p\chi_t = \chi_{t-p}$$

ifadelerinden yararlanarak,

$$\chi_t = (\Phi_1 B + \Phi_2 B^2 + \dots + \Phi_p B^p) \chi_t + \epsilon_t$$

veya

$$\epsilon_t = (1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p) \chi_t$$

biçiminde yazılabilir. Burada B, geriye öteleme işlecidir (backward shift operator).

AR modellerde durağanlık koşulu, polinomun sifıra eşitlenmesiyle bulunacak köklerin birim çemberin dışında kalmasıyla mümkündür. Eğer sözkonusu kökler birim çemberin dışında kalıyorsa, AR(p) durağan zaman serileri için kullanılabilir (Box-Jenkins, 1976).

Örneğin AR(1) modeli için durağanlık koşulu,

$$(1 - \Phi_1 B) \chi_t = \epsilon_t$$

$$|\Phi_1| < 1$$

koşulunun sağlanması gerekir. AR(2) modeli için durağanlık koşulu ise,

$$(1 - \Phi_1 B - \Phi_2 B^2) \chi_t = \epsilon_t$$

$$\Phi_2 + \Phi_1 < 1, \quad \Phi_2 - \Phi_1 < 1 \text{ ve } |\Phi_2| < 1$$

eşitsizliklerinin sağlanmasını gerektirir.

Otoregresif modelin parametrelerinin tahmini, en küçük kareler yöntemi ile yapılmaktadır. En küçük kareler fonksiyonunun tahminleyicilerini elde etmek için p'inci dereceden otoregresif modelin kurulması gerekir. Buna göre ortalaması μ olan p'inci dereceden AR(p) modeli şu şekilde yazılabilir.

$$\chi_t - \mu = \Phi_1 (\chi_{t-1} - \mu) + \Phi_2 (\chi_{t-2} - \mu) + \dots + \Phi_p (\chi_{t-p} - \mu) + \epsilon_t$$

Formülde yer alan $\chi_1, \chi_2, \dots, \chi_n$ gözlemleri, $\mu, \Phi_1, \Phi_2, \dots, \Phi_p$ parametreleri göstermektedir.

$\chi_t = \chi_t - \mu$ dönüşümü yaparsak, aynı formül,

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + e_t$$

şekline dönüşür.

Böylece parametrelerin en küçük kareler tahminleyicileri aşağıda verilen formülün minimizasyonu ile elde edilir (*Chatfield, 1980*).

$$S = \sum_{t=p+1}^n \left[(X_t - \mu) - \varphi_1 (X_{t-1} - \mu) - \dots - \varphi_p (X_{t-p} - \mu) \right]^2$$

veya

$$S = \sum_{t=p+1}^n \left[X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \dots - \varphi_p X_{t-p} \right]^2$$

AR(p=1) modelini ele alarak, bunun en küçük kareler parametre tahminlerini yazmak gerekirse model,

$$X_t - \mu = \varphi_1 (X_{t-1} - \mu) + e_t$$

iken en küçük kareler fonksiyonu,

$$S = \sum_{t=2}^n \left[(X_t - \mu) - \varphi_1 (X_{t-1} - \mu) \right]^2$$

olmakta ve bu fonksiyonu minimum yapan en küçük kareler tahminleri aşağıdaki formüllerle hesaplanmaktadır.

$$\mu = \frac{X_{(2)} - \varphi_1 X_1}{1 - \varphi_1}$$

$$\varphi_1 = \frac{\sum_{t=1}^{n-1} (X_t - \mu)(X_{t+1} - \mu)}{\sum_{t=1}^{n-1} (X_t - \mu)^2}$$

Diğer p.'inci dereceden AR modellerin parametre tahminleri de benzer şekilde yapılmaktadır. Parametrelerin bulunmasında diğer bir yöntem de otokovaryans ve otokorelasyon katsayılarının kullanılmasıdır. Bu çalışma içerisinde bu yöntemlere girilmeyecektir(*Bkz. Box-Jenkins, 1976*).

3.2.5.1.2 MA (Moving Average) Modeller

MA modelleri bir zaman serisinin herhangi bir dönemdeki gözlem değerini aynı dönemdeki hata terimi ve ondan önceki belirli sayıda dönemin hata terimine bağlı olarak açıklayan modellerdir. MA modelleri de içerdikleri geçmiş dönem hata terimi sayısına göre isimlendirilmektedir. q. dereceden MA(q) modeli,

$$\chi_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \dots \dots \dots (3.5)$$

biçimindedir. $\chi_t = \chi_t - \mu$ küçültülmüş değerdir. $\theta_0, \theta_1, \dots, \theta_q$ modelin parametreleridir. Bu parametreler, χ_t ile $e_t, e_{t-1}, \dots, e_{t-q}$ arasındaki ilişkiyi gösteren katsayılardır. q ise MA modelinin derecesini gösterir. MA modelinde hesaplanması gerekli parametre sayısı da (p+2) tanedir. Uygulamada en çok kullanılan MA modelleri 1 ve 2. dereceden modellerdir. Bu modeller sırasıyla MA(1) ve MA(2) şeklinde simgelenir. MA(1) modelinin yazılımı,

$$\chi_t = e_t - \theta_1 e_{t-1}$$

biçimindedir. Yazılımdan anlaşılacağı gibi, MA(1) modelinde bir zaman serisinin χ_t gözlem değeri t, t-1 dönemlerine ilişkin hata terimlerinin doğrusal bir bileşimidir.

MA(2) modelinin yazılımı aşağıdaki gibidir.

$$\chi_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

MA modelleri yukarıda gösterilen fark denklemi biçiminde yazılabileceği gibi, "geriye öteleme işleci" B kullanılarak da yazılabilir. Bu durumda (3.4) modeli,

$$\chi_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p) e_t$$

biçiminde yazılabilir. MA modeller için de durağanlık koşulları yerine çevrilebilirlik koşullarından söz edilir. Çünkü MA(1) modeli, sonsuz dereceden AR modeline eşittir. Bunun yanında hata terimine dayanarak bir zaman serisinin herhangi dönemine ilişkin gözlem değeri, sonsuz sayıda hata teriminin tartılı toplamları alınarak açıklanabilir. Ancak bu uygulama imkansızlığı nedeniyle MA modellerinde, zaman serilerinin herhangi bir dönemine ilişkin gözlem değeri sonlu sayıda hata teriminin tartılı toplamı alınarak açıklanmaya çalışılmaktadır (Bkz. *Box ve Jenkins, s.10; Johnson ve Montgomery, s.463*).

MA(1) modeli için "çevrilebilirlik koşulu"

$$|\theta_1| < 1$$

eşitsizliğinin sağlanması ile mümkündür. MA(2) modeli için bu koşul,

$$\theta_1 + \theta_2 < 1, \quad \theta_1 - \theta_2 < 1 \text{ ve } |\theta_2| < 1$$

eşitsizlikleri ile ifade edilir.

MA modellerde parametre tahmini oldukça güç yapılmaktadır. Örneğin MA(1) modelini ele alalım,

$$e_t = (1 - B\theta_1)^{-1} \chi_t$$

Bu modelde e_t 'nin θ_1 parametresine göre kısmi türevi,

$$\frac{de_t}{d\theta_1} = B(1 - B\theta_1)^{-2}$$

olur ve bilinmeyen q parametresinin fonksiyonudur. Bu tür modellerin parametrelerini tahmin etmek için en küçük kareler yöntemi doğrudan uygulanmaz. En küçük kareler yönteminin uygulanabilmesi için modellerin doğrusallaştırılması gerekir (Bkz. *Draper ve Smith "Applied Regression Analysis" s.458-61*).

3.2.5.1.3 ARMA(Auto-Regressif Moving Average) Modeller

ARMA modelleri durağan zaman serilerinin modellenmesinde kullanılır. AR ve MA modellerinin bir kombinasyonudur. Bu nedenle ARMA modellere "karışık modeller" denir.

Bu modellerde bir zaman serisinin herhangi bir dönemine ait gözlem değeri, ondan önceki belirli sayıda gözlem değerinin ve hata teriminin doğrusal bir bileşimi olarak ifade edilir. Eğer ARMA modeli p terimli AR ve q terimli MA modelinin bir kombinasyonu ise, $p+q$ terim içerir. ARMA(p,q) modeli aşağıdaki gibi ifade edilir.

$$\chi_t = \phi_1 \chi_{t-1} + \phi_2 \chi_{t-2} + \dots + \phi_p \chi_{t-p} + \theta_0 e_{t-1} - \dots - \theta_q e_{t-q} \dots \dots \dots (3.6)$$

Bu model için, tahmin edilmesi gerekli parametre sayısı $p+q+2$ ' dir.

ARMA modellerinin durağanlık ve çevrilebilirlik koşulunu sağlayıp sağlamadığını belirlemek amacıyla ARMA (3.6) modeli,

$$\phi(B) \chi_t = \theta(B) e_t$$

biçiminde yazılabilir. Burada $\phi(B)$ ile $\theta(B)$ sırasıyla p ve q dereceden polinomlardır. $\phi(B)$ polinomunun kökleri olan $\phi_1, \phi_2, \dots, \phi_p$ değerleri, birim çemberin dışında kalıyorsa, bu model durağanlık koşulunu, $\theta(B)$ polinomunun kökleri olan $\theta_0, \theta_1, \dots, \theta_q$ değerleri birim çemberin dışında kalıyorsa çevrilebilirlik koşulunu sağladığından söz edilebilir.

3.2.5.2 Durağan Olmayan Zaman Serisi Modelleri

Buraya kadar ele alınmış (B.J) modeller sadece durağan zaman serileri analizinde kullanılan modellerdir. Ancak uygulamada karşılaşılan serilerin çoğu, özellikle ekonomik zaman serileri durağan değildir. Bu serilerin durağanlığı trend, mevsimsel, konjonktürel dalgalanma ve tesadüfi sebepler gibi etkenler tarafından bozulur. Bu etkenlere rağmen zaman serilerinin çoğunda homojenlik görülmektedir (Box-Jenkins, 1976).

3.2.5.2.1 ARIMA (Auto-Regresif Integrated Moving Average) Modeller

Durağan olmayan seriyi durağan hale getirmek için uygun derecede fark alma işlemi yapılır. Fark alma derecesi d (=difference) ile simgelenir ve uygulamada genelde 1, en çok 2 değerini alır. Fark alma derecesi $d=0$ olduğunda, yani seri orijinal değerler itibariyle durağan ise ARIMA model aslında Bir ARMA model olmaktadır. ARIMA modelinin genel ifadesi,

$$\omega_t = \phi_1 \omega_{t-1} + \phi_2 \omega_{t-2} + \dots + \phi_p \omega_{t-p} + e_t + \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \dots \dots \dots (3.7)$$

biçimindedir.

$\omega_t = \Delta^d \chi_t$: farkı alınmış seridir.

Δ : fark işlemidir.

d : fark alma derecesidir.

ω_t : farkı alınmış seridir.

Eğer birinci farklar ($d=1$) seriyi durağan hale getiriyorsa fark operatörünün işleyişi

$$\Delta \chi_t = \omega_t = \chi_t - \chi_{t-1}$$

şeklinde gösterilir.

ARIMA(p,d,q) modelinde p veya q sıfır olabilir. Bu durumda model ya AR(d,p) veya MA(d,q) model türüne indirgenmiş olur. Uygulamada sık karşılaşılan $d=1$ veya $d>1$ koşulunu sağlayan bazı ARIMA modelleri fark denklemi şeklinde ve B operatörü kullanılarak aşağıdaki gibi verilmiştir.

1) ARIMA(0,1,1) modeli,

$$\Delta \chi_t = e_t - \theta_1 e_{t-1}$$

$$= (1 - \theta_1 B) e_t$$

eşitlikleri ile yazılır. Bu modelde $p=0$, $d=1$; $\phi(B)=1$; $\theta(B)=1 - \theta_1 B$ dir.

2) ARIMA(0,2,2) modeli,

$$\begin{aligned}\Delta^2\chi_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \\ &= (1 - \theta_1 B - \theta_2 B^2) e_t\end{aligned}$$

eşitlikleri ile gösterilir. Bu modelde $p=0$, $d=2$, $q=2$; $\varphi(B)=1$ $\theta(B)=1 - \theta_1 B - \theta_2 B^2$ dir.

3) ARIMA(1,1,1) modeli,

$$\begin{aligned}\Delta\chi_t - \Delta\chi_{t-1} &= e_t - \theta_1 e_{t-1} \\ (1 - \theta_1 B) \Delta\chi_t &= (1 - \theta_1 B) e_t\end{aligned}$$

eşitlikleri ile ifade edilir. Bu modelde $p=1$, $d=1$ $q=1$ $\varphi(B)=1 - \theta_1 B$; $\theta(B)=1 - \theta_1 B$ dir.

ARIMA modellerde durağanlık ve çevrilebilirlik koşulunun sağlanıp sağlanmadığını belirleme işlemi ARMA modellerde olduğu gibidir.

Gözlem değerlerinden oluşan zaman serilerinin birbirini izleyen yılların aynı aylarında veya dönemlerinde maksimuma ve minimuma ulaşma eğilimi mevsim dalgalanmalarını ifade etmektedir. Doğal ve sosyal nedenler sonucu ortaya çıkan ve her yıl düzenli olarak tekrar eden bu dalgalanmaları içeren serilere "Mevsimsel Zaman Serileri" adı verilir. Aylık gözlem değerlerinden meydana gelen serilerde genellikle, mevsimsel dalga uzunluğu $s=12$ ' dir. Ancak 6 aylık ($s=6$) periyoda sahip mevsimsel dalgalanmalara da rastlanabilir. Üçer aylık aralıklarla yapılan gözlem değerlerinden oluşan serilerde $s=4$ ' tür (Özmen, 1986).

3.2.5.3 Modelin Uygunluk Testi

Seçilen modelin, ele alınan seri için uygun olup olmadığının araştırılması gerekir. En az parametrelili uygun modelin seçilmesinde kullanılan testlerden örnekleme için tahmin hatalarının otokorelasyonlarından faydalanılanı uygulanabilirlik bakımından en önemlisidir. Hata otokorelasyonlarının her birini ayrı ayrı kendi standart hataları ile $1/\sqrt{n}$ 'i karşılaştırmak mümkündür. Ancak tahmin edilen otokorelasyonların standart hatasını bir ölçü olarak kullanmak, düşük derecelerdeki gecikmelerde hesaplanan otokorelasyonların sıfırdan farklılığının önemini açıkça ortaya koyamaz (Kayın, 1985).

Bu nedenle otokorelasyonları tek tek incelemek yerine belirli sayıdaki otokorelasyonları bir arada incelemek daha başarılı sonuçlar vermektedir. Bu şekilde yapılan inceleme, seçilen modelin uygunluğunu daha açıkça ortaya çıkarabilmektedir. Bu amaçla Box-Pierce tarafından geliştirilmiş olan Box-Pierce istatistiği Q şu şekilde hesaplanmaktadır(*Box-Jenkins, 1976*).

$$Q = n \sum_{k=1}^K r_k^2(\epsilon) \quad k=1,2,\dots,K$$

Formülde yer alan $r_k(\epsilon)$, örnekleme içi tahmin hatalarının çeşitli (k) gecikmelerdeki otokorelasyonlarıdır. Burada n, örnekleme hacminden fark derecesinin çıkarılması sonucu kalan değerdir.

Box-Pierce istatistiği Q, yaklaşık olarak Khi-kare dağılımlı olup, (K-p-q) serbestlik derecesine sahiptir. K hesaplanan otokorelasyon sayısı, p otoregresif, q hareketli ortalama modelindeki parametre sayılarıdır. Bu testlerde khi-kare dağılımı kullanıldığından uygulamada khi-kare testi olarak bilinir. Bu test, bazı otokorelasyonların anlamlı olarak sıfırdan farklı olup olmadıklarını ortaya koyar. Bu test ile "otokorelasyon ortalaması sıfırdan büyüktür". hipotezi test edilir. Bu hipotezin reddi, hataların rastgele dağıldığını, uygulanan yöntemin ve seçilen modelin uygun olduğunu ifade eder.

3.2.5.4 Box-Jenkins Yönteminin Üstün ve Zayıf Yönleri

3.2.5.4.1 Üstün Yönleri

- 1) B.J yönteminde uygun modelin belirlenmesi genellikle eldeki verilerin yapısı ile belirlendiği için "verilerin kendi kendine ilişkisi sağlanmış olur." Bu nedenle B.J modellerine dayanarak yapılan kısa dönem tahminlerinin diğer yöntemlere dayanarak yapılan aynı döneme ait tahminlere oranla daha güvenilir olduğu söylenebilir.
- 2) B.J yönteminde ileriye dönük tahmin amacıyla analiz edilecek bir zaman serisi için uygun model belirlenirken izlenen her aşamada bu modelin analiz edilecek seriye uygunluğunu denetleme imkanı vardır.

3) B.J yöntemine göre belirlenecek uygun modelde önemli olan parametre sayısını olabildiği kadar az tutmaktır.

4) Zaman serilerinin çoğunda ardışık gözlem değerleri birbirine bağımlıdır. B.J yöntemi zaman serilerinin bu en önemli özelliğini en etkili biçimde kullanır.

3.2.5.4.2 Zayıf Yönleri

1) B.J yöntemine dayanarak yapılan tahminler çabuk elde edilemez. Çünkü B.J yöntemi tümüyle otomatik değildir. Bu yöntemeye dayanarak ileriye dönük tahmin yapmak amacıyla yazılacak bir bilgisayar programı aşamalı (=iteratif) bir programdır. Bu tür bilgisayar programı kullanırken elle müdahaleye gerek olabilir.

2) B.J yönteminin uygulanabilmesi için deneyimli ve bilgili işgücüne gereksinim duyulur.

3) B.J yönteminin uygun model seçimi konusunda sağladığı özgürlük olanağı, tahmin yapan kişinin uygun olmayan model seçmesine neden olabilir.

4) B.J modelleriyle aynı seriyi analiz eden ve aynı seri için ileriye dönük tahmin yapan iki kişinin sayısal olarak birbirine benzer sonuçlar elde etmesi konusunda garanti yoktur.

5) Model belirlemek için çok sayıda gözlem değerine gereksinim vardır, bu sayı en az 50 olmalıdır.

Box-Jenkins yöntemlerinin bu olumsuzluklarına rağmen, bir zaman serisinin yapısını belirlediği, gözlem değerlerinin aralarındaki bağımlılığı en etkili bir şekilde kullandığı ve model belirleme aşamalarında istatistiksel testlere yer verdiği için, diğer tahmin yöntemlerine göre kısa dönem tahmin yapmada üstün bir yöntemdir(*Box ve Jenkins, 1976*).

BÖLÜM 4

4. SAPAN DEĞERLER (OUTLIERS)

4.1 Giriş

Bir kitleden çekilen örneklemede bir veya bir kaç gözlemin beklenmedik şekilde diğer gözlemlerden farklı olması mümkün, ancak istenen bir durum değildir. Kitle normal dağılmış olsa bile çekilen örneklerdeki gözlemlerden bir veya bir kaçının farklı olması beklenmelidir. Eğer, farklı olan bu gözlemler doğal rastgelelik sonucunda ortaya çıkmış iseler, bu gözlemlerin örneklem içinde tutulmaları düşünülebilir. Çünkü bazı gözlemler uç bile olsalar küçük bir olasılıkla dağılımda yer alabilirler. Ancak eğer, farklı gözlemler, ölçüm tartım veya kaydetme gibi kişi hatalarından kaynaklanmış ise o takdirde yapılması gerekli iş mutlaka gözlemleri hatalardan arındıracak yollara başvurmaktır. Bunun için, farklı olan gözlemler eğer istatistiksel yöntemler ve bazı testler sonucunda sapan gözlemler olarak nitelenirlerse, örneklemden çıkarılması, yerine yenilerinin çekilmesi, gerektiğinde ortalama yerine ortancanın kullanılması geçici de olsa bir çözüm olabilir (Chang, Tiao ve Chen, 1988).

Bazı durumlarda sapan değerlerin oluşma nedenleri araştırılmak istenebilir. Bu gelecek çalışmalar için önemlidir. Çünkü, örneklem değerlerinin çekildiği kitleden farklı bir kitleye ait olması da mümkün olabilmektedir (Tatlıdil, 1977).

Sapan değerlerin testi çalışmaları önce basit örneklem için yapılmıştır. Daha sonra bu çalışmalar, deney düzenleme, çok değişkenli istatistik analiz, regresyon analizi ve zaman serileri modellerini kapsayacak şekilde genişlemiş ve istatistiği kullanan bütün bilimlerde göz önüne alınmaya başlanmıştır. *Yapılan bilimsel çalışmalarda sapan değerlerin göz önüne alınmaması yanlış parametre kestirimleri ve geniş güven aralıkları elde edilmesine sebep olmaktadır.* Bu nedenle sapan değer analizi, bilimsel bütün çalışmalarda yapılacak istatistik analizlerin sağlıklı yürümesi bakımından temel analiz özelliği taşır.

4.2. Basit Örneklem için Sapan Değerler

Basit örneklerde sapan değer testinde kullanılan yöntemlerden bazıları bir yada iki sapan gözlem değerinin testinde kullanılırken, bazıları çok sayıda sapan gözlem değerinin testine olanak sağlamaktadır. Bunlardan yaygın olarak kullanılan yöntemler ; Thompson yöntemi, Pearson ve Chandra Sekar yöntemi, Grubbs yöntemi, Dixon Yöntemi, Ferguson yöntemi, Grubbs yöntemi II, Rosner ve diğer yöntemler şeklinde sıralanabilir. Burada basit örneklerde sapan değerlerin testleri ile ilgili Thompson yöntemi ve Pearson ve Chandra Sekar yöntemi verilecektir. Ayrıca diğer yöntemler tanımlanacaktır.

4.2.1 Thompson Yöntemi

Thompson tarafından 1935 yılında önerilen yöntem bu konudaki ilk matematiksel yöntemlerden biridir. Yöntemde $\chi_1, \chi_2, \dots, \chi_n$ gibi normal dağılımlı kitleden çekilmiş bir örnekte herhangi bir gözlem değeri χ_i 'nin sapan değer olup olmadığını saptamak için

$$T_i = \frac{(\chi_i - \bar{X})}{S} \quad i = 1, 2, \dots, n$$

gibi test istatistiği önerilmiştir. Bu istatistiğin dağılımı araştırılarak çeşitli n ve α değerleri için

$$P(|\chi_i - \bar{X}| / s > T_{\alpha}) = \alpha \quad i = 1, 2, \dots, n$$

eşitliğinden T_{α} ölçüt değerleri simülasyon yöntemi kullanılarak bulunmuştur. Bu koşullarda $|\chi_i - \bar{X}| / s > T_{\alpha}$ sonucunu veren χ_i gözlem değerinin "sapan değer" olduğuna karar verilmektedir(Wilks,1963).

4.2.2 Pearson ve Chandra Sekar Yöntemi

Pearson ve Chandra Sekar'ın 1936'da geliştirdikleri yöntem, örnekteki kuşku edilen bir en büyük ve bir en küçük gözlem değerinin sapan olup olmadıklarının testinde kullanılmaktadır. Örnekteki değerler $\chi_{(1)} < \chi_{(2)} < \dots < \chi_{(n)}$ biçiminde küçükten büyüğe sıralanmaktadır. Kuşku edilen gözlem değeri $\chi_{(n)}$ ise yukarıdaki istatistiğin aynısı olan

Örneklemeledeki deęerler $\chi_{(1)} < \chi_{(2)} < \dots < \chi_{(n)}$ biçiminde küçükten büyüęe sıralanmaktadır. Kuşkuilanılan gözlem deęeri $\chi_{(n)}$ ise yukarıdaki istatistięin aynısı olan

$$T_n = \frac{(\chi_{(n)} - \bar{x})}{s}$$

test istatistięi, kuşkuilanılan gözlem deęeri $\chi_{(1)}$ ise

$$T_1 = \frac{(\bar{x} - \chi_{(1)})}{s}$$

test istatistięi kullanılmaktadır.

Ayrıca *Pearson* ve *Chandra sekar* tarafından 1936'da önerilen yöntemi, *Grubbs* 1950 yılında geliştirek, bir en büyük veya bir en küçük gözlem deęerinin sapan deęer olup olmadığının testinde kullanılmak üzere önermiştir.

Gözlem sayısının büyük olduęu örneklemlerde standart sapma ya da varyansın hesabı işleminin güç olduęu düşüncesiyle *Dixon*, gözlemler arası uzaklığın örnekleme genişliğine bölümünden bulunan bir dizi test istatistięi önermiştir.

Ferguson tarafından 1961'de önerilen yöntemler tek yanlı testlerde % 50'den az bozulmalar için geçerli ve gizlenme etkisinden etkilenmeyen, çarpıklık katsayısı ile ve yine gizlenme etkisinden etkilenmeyen iki yanlı testlerde % 21'e kadar bozulmalar için geçerli ve basıklık katsayısı üzerine kuruludur. *Ferguson* ayrıca Monte Carlo çalışması ile tablolar geliştirmiştir.

Çok sayıda kuşku gözlem deęerlerinin aynı anda testine olanak sağlayan ilk yöntem *Grubbs* tarafından 1969'da geliştirilmiştir ve bunlar *Grubbs-Türü* test istatistikleri olarak bilinir.

Rosner'in 1975 ve 1977 yıllarındaki çalışmalarında önerdiği ve R istatistiği olarak bilinen yöntemde, normal dağılımlı kitlelerden çekilmiş örneklerdeki çok sayıda kuşkulu gözlem değerlerinin testinde kullanılan en güvenilir yöntemlerden biri olarak bilinmektedir.

Sapan değer analizi çalışmalarında, eldeki seri değerleri arasından sapan olanları bulurken, testin gücü çok önemlidir. En başarılı ve güçlü testlerin bile, sapan değeri çok olan bir seri için, etkili olamadığı, *sapan değer sayısının artmasına paralel, tarama süreçlerinin gücünde zayıflama sonucunun ortaya çıktığı tespit edilmiştir.* Bunun yanında bir seri içerisinde tek sapan değer olması durumunda, zayıf bir test yönteminin bile başarılı sonuçlar verdiği tespit edilmiştir(*Tsay, 1988*).

Aynı sonuçlar simülasyon çalışmalarıyla *Chang, Tiao ve Chen* tarafından da elde edilmiştir.

4.3. Zaman Serilerinde Sapan Değerler

Çeşitli veri analizlerinde bazı gözlemlerin diğerlerine uyum göstermediği durumlara pek sık rastlanır. Bu gözlemlere sapan değer adı verilir. *Grevler, savaşlar, bir malın piyasasında fiziksel sistemdeki beklenmeyen değişiklikler ve buna benzer durumlardan ötürü bazı gözlemler sapan değer durumuna gelir(Chang, Tiao ve Chen, 1988).*

Yukarıda belirtilen durumlar, beklenmeyen müdahale olarak tanımlanmaktadır. *Gutman ve Tiao(1978), Miller(1980) ve Chang(1980)* beklenmeyen müdahalelerin, otokorelasyonlar, kısmi otokorelasyonlar ve ARMA parametrelerini tahminlemede ciddi şekilde yanlışlığa sebep olduklarını göstermişlerdir(*Chang, Tiao ve Chen, 1988*).

Bunun sonucu olarak ta dışsal etkileri tanımlayabilmek ve gözlemlerden silmek serinin yapısını daha iyi anlayabilmek açısından önemlidir. Eğer müdahale anı biliniyorsa *Box ve Tiao(1975)* tarafından geliştirilen Müdahale Analiz Tekniği (Intervention Analysis Techniques) kullanılarak, bu etkinin ne büyüklükte olduğu belirlenebilir(*Chang, Tiao ve Chen,1988*).

Sapan deęerler, eldeki veriler için öngörülen tahmin modelinin açıklayamadığı miktar anlamına gelen hata terimiyle ilgili bir kavramdır(Cook, 1982).

Ancak bir gözlem için hesaplanan hata terimi ne kadar büyük olursa olsun, söz konusu gözlem deęerinin sapan olduğunu söylemek ancak geliştirilmiş test yöntemleri ile mümkün olmaktadır.

Bunun yanında model yanlışlığı, yapılması gerekli dönüşümlerin yapılmamış olması, ölçüm tartım yada kaydetme gibi bazı hatalar da sapan deęer sonucunu doğurmaktadır. *Bu çalışma kapsamında, üretim sürecinde olması muhtemel bir problemin ortaya çıkarılması amaç olduğundan, orjinal veriler üzerinde bir dönüşüm yapılmayacaktır.* Yukarıda sözü edilen veriler üzerinde dönüşüm yapılması yaklaşımı, parametre tahminlerinin tutarlı olmasına yönelik bir yaklaşım olarak düşünölmelidir.

Zaman serilerinde sapan deęer kavramı ilk defa Fox tarafından 1972 tarihinde çalışılmıştır. Fox, sadece AR modeller için geçerli olan ve olabilirlik oran kriteri adını verdiği tarama yöntemiyle çalışmalarını başlatmış, daha sonraları Cheng(1983), Cheng ve Tiao(1983), Fox'un geliştirmiş olduğu test yöntemine eklemeler yaparak ARIMA modeller için geçerli olan bir yöntem geliştirmiş ve bunu "iteratif yöntem" olarak isimlendirmişlerdir. Simölasyon çalışmaları, söz konusu yöntemin, dięer sapan deęer tarama yöntemleri olan ve literatürde LS, M-H, M-B, GM, GM-H, GM-B olarak bilinen yöntemlere oranla çok daha üstün ve tutarlı olduğunu göstermiştir(Chang, Tiao ve Chen, 1988).

Bu nedenle bu çalışma içerisinde yapılacak olan uygulama, iteratif yöntem ile test edilecektir.

Zaman serilerinde ortaya çıkması muhtemel iki tip sapan değer bulunmaktadır. Birinci tip sapan değer, bir gözlemlerde veya kayıtlarda yapılan büyük bir hata sonucu ortaya çıkar ve hata tek bir gözlemi etkiler. İkinci tip sapan değer durumunda birden çok gözlem etkilenir (Fox, 1972).

Bir zaman serisi içerisinde 1. tip sapan değer'in (AO) bulunması, ARMA katsayıları ve σ^2 nin tahmininde ikinci tip sapan değerlere oranla daha yanlış tahminler elde edilmesine yol açmaktadır(Ljung, 1993).

Ayrıca, Ruey S. Tsay e göre 1. Tip sapan değer (AO), bir büyük hata kaynağı olarak düşünülmelidir. Çünkü, sadece sapan değer'in bulunduğu T'inci gözlem etkilenmekte ve parametre tahminlerine olan etki çok büyük olmaktadır. Halbuki, 2. tip sapan değer durumunda T' inci pozisyonda meydana gelen sapan değer etkisi, kendinden sonraki bütün gözlemlere dağılmaktadır(Tsay, 1988).

Yapılan uzun çalışma ve gözlemler sonucu ortaya konan bir başka teşhis, eğer seriyi oluşturan veriler arasında güçlü bir otokorelasyon var ise, bu 1.tip sapan değer olarak bilinen AO' nun ortaya çıkması olasılığını güçlendirmektedir(Booth, 1987).

Gözlem değerleri, sürekli üretimin yapıldığı bir süreçten elde ediliyorsa, birinci tip sapan değer'in bulunması, ölçüm aletlerinin ayarlanması, veri tabanı üzerinde yapılmış bir hatanın onarılması ile ilgili bir işlemin yapılmasını veya benzer düzeltme işlemlerinden bir veya bir kaçının yapılmasını gerekli kılar. Bununla birlikte eğer ikinci tip sapan değer söz konusu olmuş ise, yukarıda sözü edilen ayarlamalara gerek yoktur (Muirhead,1985).

Çünkü ikinci tip hata, girdi (hammadde) 'den kaynaklanan bir hatadır(Acar ve Booth, 1987).

4.3.1 Birinci Tip Sapan Değer (Additive Outlier (AO))

Tek bir gözlemi etkileyen, bir gözlem veya kayıta yapılan hata sonucu ortaya çıkan sapan değer tipidir(Fox, 1972).

Literatürde additive outlier(AO) olarak tanımlanan model, aşağıdaki gibi gösterilmektedir

$$y_t = z_t + \delta x_t$$

Burada y_t gözlem değeri, z_t ARMA modeli ile oluşturulmuş zaman serisi gözlem değeri, δ , sapan değerinin büyüklüğü, x_t , T anında 1, diğerleri için 0 değerini alan yapay değişkeni (Dummy variable) ifade etmektedir.

4.3.2 İkinci Tip Sapan Değer (Innovation Outlier (IO))

Bulunduğu pozisyonundan itibaren, kendinden sonraki gözlemleri de etkileyen sapan değer tipidir. Gözlem sapan değeri olarak bilinir(Fox, 1972).

Literatürde Innovation Outlier (IO) olarak ta tanımlanan model, aşağıdaki gibi gösterilmektedir:

$$y_t = \frac{\theta(B)}{\Phi(B)} (e_t + \delta x_t)$$

Bununla birlikte, 1. tip sapan değer, T anındaki büyük bir hatayı ifade ederken, 2. tip sapan değer, T anında beklenmeyen şok bir etkiyi ifade eder(Ljung, 1993).

4.4 Sapan Değerlerin Tespiti

Zaman serilerinde sapan değer analizi yapmak amacıyla kullanılan yöntemlerden en etkili olan iteratif yöntem tanıtılacaktır.

4.4.1 İteratif Yöntem

Bu Yönteme göre bir süreçten elde edilen gözlemler setinde sapan değer olup olmadığını araştırmak için önce bu gözlemin bir zaman serisi oluşturduğu düşünülerek serinin uyum gösterdiği stokastik modelin parametre tahminleri elde edilir. Bu adımdan sonra iteratif yöntemle seride sapan değer aranır. Eğer seride sapan değer varsa bunların etkileri giderilir. Seride hiç sapan değer kalmayınca kadar bu işlem tekrarlanır.

Seride artık sapan değere rastlanmadığı takdirde o ana kadar hangi gözlemlere karşı gelen noktalarda sapan değere rastlanmışsa o gözlem değerleri için yeni değerler hesaplanır. Sapan değerlerin bulunduğu gözlem değerleri için yeni değerler hesaplanıp eskileri ile değiştirildikten sonra yeni seri için parametre kestirimleri elde edilir. ve iteratif prosedür tekrar edilir. Bu aşamada da sapan değere rastlanmazsa prosedüre son verilir.

4.4.2 İteratif Yöntem'in Adımları

1. Adım

Tahmin modeli kullanılarak hatalar hesap edilir ve varyansın tahmini olarak

$$\sigma^2_{e=n-1} = \sum_{t=1}^n e_t^2$$

elde edilir.

2. Adım

$$W_I = e_T, \quad W_A = \rho^2 \pi (F) e_t$$

daha açık bir şekilde yazmak gerekirse,

$$W_A = \rho^2 (1 - \pi_1 F - \pi_2 F^2 - \dots - \pi_n F^{n-1}) e_t$$

$$\rho^2 = (1 + \pi_1^2 + \pi_2^2 + \dots + \pi_{n-1}^2)^{-1}$$

F, (Forword-Shift operator), $F e_t = e_{t+1}$ olacak şekilde ileri doğru öteleme işlecidir.

$$\lambda_{1,t} = W_I / \sigma_A \quad \text{ve} \quad \lambda_{2,t} = W_A / \rho^2 \sigma_A$$

biçiminde hesaplanır.

$$\eta_t = \max \{ \lambda_{1,t}, \lambda_{2,t} \} \quad t=1, \dots, n.$$

eğer $\eta_t = \lambda_{1,t} > C$ ise, T noktasında bir IO olduğu anlaşılır, burada sapan değer etkisini tahmin etmek için ($e_t = e_t - W_I$) işlemi yapılır. Eğer, $\eta_t = \lambda_{2,t} > C$ ise T noktasında bir AO olduğu anlaşılır bu etkiyi tahmin etmek için de ($e_t = e_t - W_A \pi(B) \delta \quad t > T$). Burada C sabiti, yüksek duyarlılık için 3.0, orta duyarlılık için 3.5, düşük duyarlılık 4.0 değerine eşittir. Eğer sözkonusu noktalarda sapan değer varsa bunların etkileri giderilerek, değiştirilmiş hataların yeni varyansları hesaplanır.

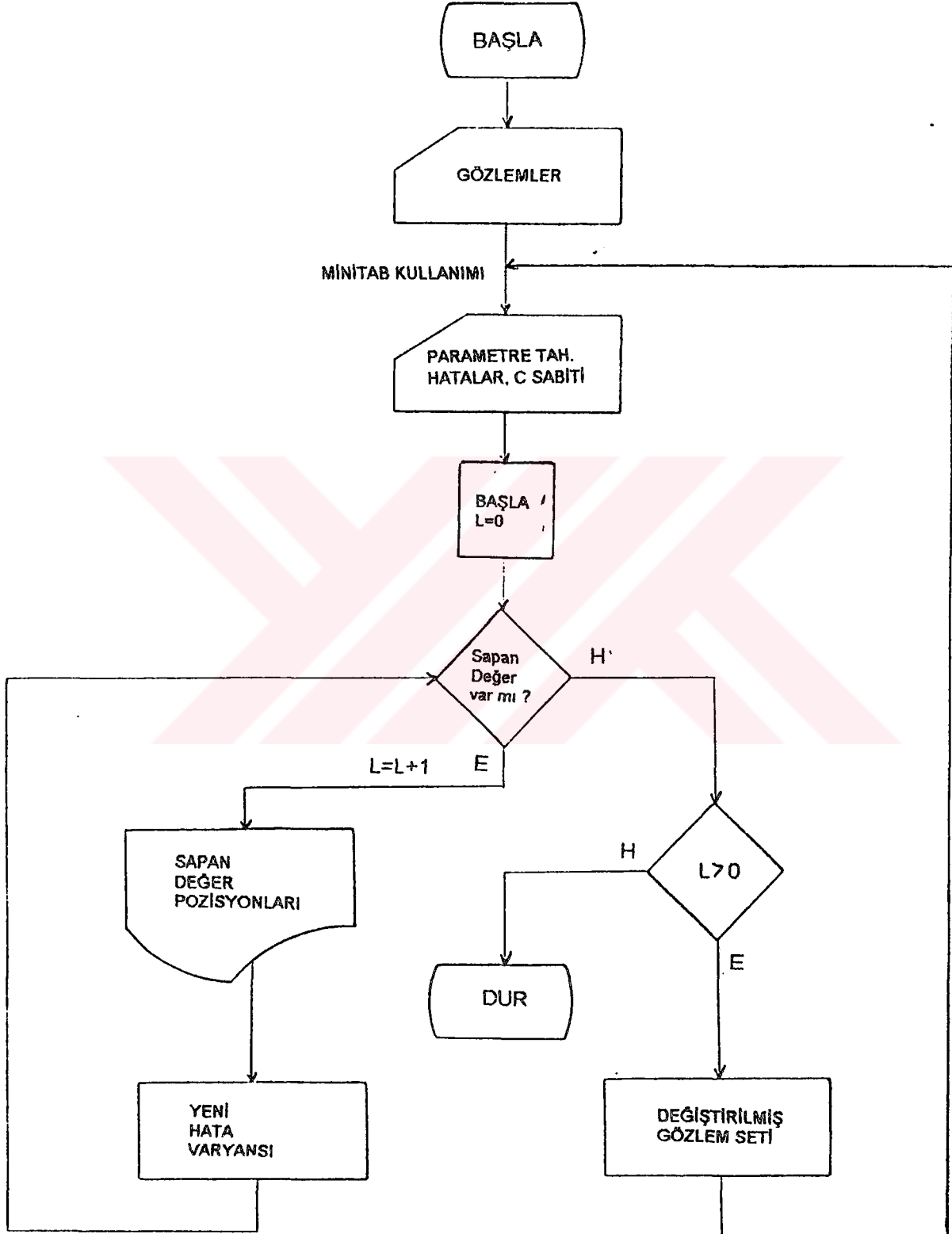
3. Adım

Eğer ikinci adımda sapan değere rastlanmışsa aynı zaman serisi parametreleri, değiştirilmiş hata terimlerine bağlı kalarak yeni hata terimleri ve varyans hesap edilir.

4. Adım

Sapan değer kalmayınca kadar 2 ve 3. adımlar tekrar edilir (*Chang, Tiao ve Chen, 1988*).

4.4.3 İteratif Yöntem'in Akış Şeması



4.4.4 İteratif Yöntem'in Performansı

İteratif prosedürün performansı ile ilgili bazı bilgiler elde etmek için simülasyon çalışmaları yapılmıştır. Önerilen süreç, sapan değer bulunması durumunda tarama ve eşanlı sapan değer tiplerini belirlemek amacıyla tasarlanmıştır. Sürecin gücünü tayin etmek için sapan değerlerin konumunu ve sapan değer tiplerini doğru bir şekilde tanımlama olasılıkları ile ilgilenmek gerekir. Bunun için

- 1) Tek ve çoklu sapan değerler
- 2) Eldeki serinin zaman serisi yapısı
- 3) Sapan değer tipleri
- 4) Sapan değer sayısı
- 5) Serinin genişliği

gibi kavramları gözönünde bulundurarak, tek ve çift sapan değer durumları için, muhtemel kombinasyonlar göz önünde bulundurularak sürecin gücü analiz edilmeye çalışılacaktır.

Tek sapan değer durumu için AR(1), $\phi=0.6$ ve $\sigma_a^2=1$ Ayrıca MA(1), $\theta = 0.6$ ve $\sigma_a^2=1$ değerleri için iki farklı sapan değer büyüklüğü için $n=50$, $n=100$ ve $n=150$ örneklem ölçümleri ve $w=3\sigma$ ve $w=5\sigma$ sapan değer büyüklükleri gözönünde bulundurularak simülasyon çalışmaları yapılmıştır. Bu çalışmalarda sapan değerler gözlemsel periyotların ortasına set edilmiştir. Yani, $n=50$ için $T=26$ (sapan değer pozisyonu), $n=100$ için $T=51$ ve $n=150$ için $T=76$ 'ya yerleştirilmiştir.

Çift sapan değer durumu için yine AR(1) modeli için $\phi=0.6$ ve $\sigma_a^2=1$ parametreleri ve sapan değerlerin bulunması muhtemel üç farklı durum (sapan değerlerden biri AO, diğeri IO, sapan değerlerin ikisi de AO veya sapan değerlerin ikisi de IO) gözönünde bulundurularak $n=50$, $n=100$ ve $n=150$ örneklem ölçümleri ve $w=3\sigma$ ve $w=5\sigma$ sapan değer büyüklükleri için simülasyon çalışmaları yapılmıştır. Burada ise sapan değerler $n=50$ örneklem ölçümü üçe bölünerek ilk ve ikinci periyotlara yerleştirilmişlerdir. Yani sapan değer pozisyonları $T_1=17$ ve $T_2=34$ olmuştur. $n=100$ ölçümü için sapan değer pozisyonları $T_1=34$ ve $T_2=66$ ve $n=150$ ölçümü için sapan değer pozisyonları $T_1=51$ ve $T_2=101$ olarak alınmıştır.

Yukarıdaki ön tasarımlar özel ölçütlere bağlı elde edilmiş zaman serisi gözlemleri olarak düşünülmelidir. Ayrıca modelin bilindiği ve örnekleme sapan değerlerinin olmadığı

varsayılmaktadır. Bunun yanında hataların yaratılması ve zaman serisi parametrelerinin tahmini, *Hillmer ve Tiao(1979)* tarafından geliştirilen tam olabilirlik tahmin metoduyla hesaplanacaktır. Tahmini parametreler ve hatalara bağlı olarak $c=4.0$ kritik değeri için tablo I, $c=3.5$ kritik değeri için tablo II ve $c=3.0$ kritik değeri Tablo III sonuçları elde edilmiştir. Daha sonra elde edilen sonuçlar alternatif ölçütlere göre karşılaştırılacaktır(*Chang, Tiao ve Chen,1988*).

Tablo 1 Sapan değerleri doğru tarama yüzdeleri ($C=4.0$ ve 1000 tekrar)

n ölçümler	$w=3\sigma$			$w=5\sigma$		
	50	100	150	50	100	150
AR, 1 AO	0.72	0.80	0.81	0.91	0.90	0.92
AR, 1 IO	0.88	0.87	0.87	0.95	0.93	0.91
MA, 1 AO	0.70	0.82	0.84	0.83	0.90	0.93
MA, 1 IO	0.97	0.92	0.88	0.99	0.96	0.97
AR, 2 AO's	0.38	0.54	0.52	0.53	0.76	0.80
1. Tip S.D	0.65	0.74	0.81	0.76	0.86	0.92
2. Tip S.D	0.64	0.73	0.78	0.74	0.88	0.88
AR, 2 IO's	0.58	0.67	0.55	0.82	0.86	0.84
1. Tip S.D	0.84	0.80	0.85	0.91	0.94	0.91
2. Tip S.D	0.77	0.88	0.83	0.92	0.91	0.92
AR, IO ve AO	0.65	0.63	0.69	0.86	0.87	0.88
1. Tip S.D	0.90	0.81	0.84	0.98	0.95	0.93
2. Tip S.D	0.72	0.81	0.81	0.88	0.92	0.93

Yukarıdaki tablo incelendiğinde, görüleceği gibi tek sapan değer durumunda iteratif Yöntem'in sapan değer saptama oranı oldukça yüksek çıkmaktadır. Ayrıca örneklem ölçümünün genişliği ve sapan değer'in büyüklüğü de yöntemin gücünü artıran diğer faktörler olarak karşımıza çıkmaktadır(*Chang, Tiao ve Chen, 1988*).

Tablo 2 Sapan değerleri doğru tarama yüzdeleri (C=3.5 ve 1000 tekrar)

n ölçümler	w=3 σ			w=5 σ		
	50	100	150	50	100	150
AR, 1 AO	0.75	0.80	0.82	0.86	0.90	0.92
AR, 1 IO	0.85	0.88	0.81	0.94	0.93	0.90
MA, 1 AO	0.75	0.83	0.84	0.84	0.91	0.93
MA, 1 IO	0.94	0.90	0.89	0.98	0.96	0.96
AR, 2 AO's	0.44	0.54	0.62	0.58	0.76	0.80
1. Tip S.D	0.70	0.75	0.81	0.78	0.86	0.92
2. Tip S.D	0.70	0.76	0.80	0.77	0.88	0.88
AR, 2 IO's	0.66	0.65	0.61	0.82	0.85	0.82
1. Tip S.D	0.83	0.81	0.84	0.91	0.94	0.91
2. Tip S.D	0.82	0.85	0.84	0.91	0.90	0.91
AR, IO AO	0.64	0.69	0.65	0.86	0.87	0.88
1. Tip S.D	0.88	0.84	0.84	0.97	0.95	0.93
2. Tip S.D	0.75	0.82	0.79	0.88	0.92	0.93

Tablo 2 dikkatle incelendiğinde, tablo 1 için yapılmış yorumların hepsi geçerli olmaktadır. Bunun yanında C (kritik değer'in) 4.00' dan 3.50'ye düşürülmesi ile, sapan değer olma ölçütü küçültülmüş, bir başka deyişle iteratif yöntemin sapan değer saptama gücü artmıştır.

Tablo 3 Sapan değerleri doğru tarama yüzdeleri ($C=3.0$ ve 1000 tekrar)

n ölçümler	w=3 σ			w=5 σ		
	50	100	150	50	100	150
AR, 1 AO	0.75	0.81	0.81	0.86	0.90	0.92
AR, 1 IO	0.83	0.82	0.77	0.93	0.93	0.90
MA, 1 AO	0.78	0.83	0.85	0.84	0.91	0.93
MA, 1 IO	0.91	0.87	0.88	0.98	0.96	0.97
AR, 2 AO's	0.50	0.59	0.63	0.60	0.76	0.80
1. Tip S.D	0.74	0.77	0.81	0.79	0.86	0.91
2. Tip S.D	0.72	0.77	0.80	0.77	0.88	0.88
AR, 2 IO's	0.61	0.63	0.65	0.81	0.85	0.82
1. Tip S.D	0.82	0.79	0.85	0.91	0.93	0.90
2. Tip S.D	0.78	0.82	0.80	0.90	0.90	0.91
AR, IO AO	0.65	0.68	0.65	0.86	0.86	0.87
1. Tip S.D	0.87	0.82	0.84	0.97	0.94	0.93
2. Tip S.D	0.76	0.83	0.79	0.89	0.91	0.93

Tablo 3 dikkatle incelendiğinde, tablo 1 ve tablo 2 için yapılmış yorumlar aynen geçerli olmakla birlikte C (kritik değer'in) 4.00' dan 3.50'ye daha sonra da 3.00 düşürülmesi, sapan değer olma ölçütünü biraz daha küçültmüş, bir başka deyişle iteratif yöntemin sapan değer tarama gücü bir kat daha artmıştır.

Kritik değerlerle ilgili olarak simülasyon çalışmalarının 1 ve 2. tip sapan değerler için gösterdiği önemli bir sonuç, test kriterinin sahip olduğu dağılımın, C kritik değerleri tarafından kuvvetli bir şekilde etkilendiğidir(Hawkins, 1979).

Burada bölüm başında Ljung(1993)' ün ileri sürdüğü gibi, serilerde birinci tip sapan değerinin (AO) bulunması, ikinci tip sapan değere oranla daha büyük varyansa, dolayısıyla yanlış parametre kestirimlerine sahip olduğu görüşü, Chang, Tiao ve Chen(1988)' in simülasyon sonuçlarıyla doğrulanmaktadır.

Burada sapan değer tarama süreçlerinin performanslarını karşılaştırmak için yapılmış simülasyon sonuçları verilecektir. Söz konusu simülasyon çalışması AR(1) modeli için yapılmış olup n (örneklem ölçümü) 50 olarak kabul edilmiştir. Simülasyon çalışması, tek, çift ve üç sapan değerinin bulunması durum için tasarlanmıştır.

Tek sapan deęerler için

Ölçümü 50 olan bir seri için, sapan deęer $T=24$ pozisyona yerleřtirilmiřtir. Her bir yöntemle iliřkin elde edilen zaman serisi parametre deęerlerine iliřkin elde edilen ortalama, standart sapma ve HKO deęerleri.

Tablo 4: Seri ierisinde ikinci tip sapan deęerin bulunması durumunda her bir yöntemle iliřkin elde edilen parametre deęerleri.

Sapan Deęer Modelleri	Elde Edilmiř Parametre Deęerleri		
	Ortalama	St Sapma	HKO
n=50 T=24 (IO)			
LS :	0.8688	0.0787	0.0071
M-H :	0.8752	0.0708	0.0056
M-B :	0.8771	0.0703	0.0055
GM-H :	0.8786	0.0711	0.0055
GM-B :	0.8783	0.0717	0.0056
İTERATİF :	0.8813	0.0673	0.0049

Tablo 5: Seri içerisinde Birinci tip sapan değer bulunması durumunda her bir yöntemle ilişkin elde edilen parametre değerleri.

n=50 T=24 (AO)		Elde Edilmiş Parametre Değerleri		
Sapan Değer Modelleri	Ortalama	St. Sapma	HKO	
LS :	0.7686	0.1155	0.0306	
M-H :	0.8163	0.1073	0.0185	
M-B :	0.8264	0.1137	0.0183	
GM-H :	0.8440	0.0938	0.0119	
GM-B :	0.8555	0.0894	0.0100	
ITERATIF :	0.8658	0.0819	0.0079	

Çift Sapan Değerler İçin

Tablo 6: Seri içerisinde birinci tip iki adet sapan değer bulunması durumunda her bir yöntemle ilişkin elde edilen parametre değerleri .

n=50 T=17, 34 (AO)		Elde Edilmiş Parametre Değerleri		
Sapan Değer Modelleri	Ortalama	St. Sapma	HKO	

Tablo 7: Seri içerisinde ikinci tip iki adet sapan değerinin bulunması durumunda her bir yöntemle ilişkin elde edilen parametre değerleri .

n=50 T=17, 34 (IO)		Elde Edilmiş Parametre Değerleri		
Sapan Değer Modelleri	Ortalama	St. Sapma	HKO	
LS :	0.8678	0.0693	0.0058	
M-H :	0.8745	0.0631	0.0046	
M-B :	0.8768	0.0629	0.0045	
GM-H :	0.8768	0.0645	0.0047	
GM-B :	0.8765	0.0698	0.0054	
İTERATİF :	0.8802	0.0635	0.0044	

Üç Sapan Değer İçin

Tablo 8: Seri içerisinde birinci tip üç adet sapan değerinin bulunması durumunda her bir yöntemle ilişkin elde edilen parametre değerleri .

n=50 T=14, 26,38 (AO)		Elde Edilmiş Parametre Değerleri		
Sapan Değer Modelleri	Ortalama	Standart Sapma	HKO	
LS :	0.6845	0.1420	0.0665	
M-H :	0.7360	0.1421	0.0470	
M-B :	0.7340	0.1507	0.0502	
GM-H :	0.7932	0.1131	0.0242	
GM-B :	0.8078	0.1084	0.0202	
İTERATİF :	0.8554	0.1000	0.0120	

Tablo 9: Seri içerisinde ikinci tip üç adet sapan değerinin bulunması durumunda her bir yöntemle ilişkin elde edilen parametre değerleri .

n=50 T=14, 26,38 (IO)		Elde Edilmiş Parametre Değerleri İçin		
Sapan Değer Modelleri	Ortalama	St. Sapma	HKO	
LS :	0.7480	0.1016	0.0334	
M-H :	0.7956	0.0970	0.0203	
M-B :	0.7961	0.1046	0.0217	
GM-H :	0.8301	0.0814	0.0115	
GM-B :	0.8393	0.0787	0.0099	
İTERATİF :	0.8743	0.0675	0.0052	

Tablo 10: Seri içerisinde ikinci tip üç adet sapan değerinin karışık pozisyonlarda bulunması durumunda her bir Yöntemle ilişkin elde edilen parametre değerleri .

n=50 pozisyonlar karışık		Elde Edilmiş Parametre Değerleri İçin		
Sapan Değer Modelleri	Ortalama	St. Sapma	HKO	
LS :	0.8110	0.1268	0.0240	
M-H :	0.8420	0.0914	0.0117	
M-B :	0.8541	0.0837	0.0091	
GM-H :	0.8592	0.0756	0.0074	
GM-B :	0.8621	0.0787	0.0076	
İTERATİF :	0.8688	0.0718	0.0061	

Tek çift ve üç sapan değerin bulunduğu zaman serilerinde, elde edilen parametre değerlerine ilişkin standart sapma ve hata kareler ortalaması değerleri, İteratif yöntem adı verilen yöntem ile en küçük olmaktadır. Bu da söz konusu yöntemin diğerlerine göre etkinliğini açık bir şekilde ortaya koymaktadır(*Chang, Tiao ve Chen, 1988*).

4.5 Sapan Değerlerin Kalite Kontrol Yorumları

4.5.1 Birinci Tip Sapan Değerler

Üretim faktörlerinin bir veya bir kaçında zaman zaman araya giren aksaklıklar şeklinde ifade edilebilir. Yani, girdi (hammadde) problemi olmamasına karşın, hedeflenen üretimin elde edilememesi durumudur. Bu, kalite kontrolde "*bir kereye özgü kötü yığın*" şeklinde ifade edilmektedir. Bunun sebepleri, 2. Bölümde kalitede değişkenlik kavramıyla ilgili özel nedenler durumu olduğu söylenebilir. Bu nedenler muhtemelen, hammaddenin sağlıklı ortamlarda muhafaza edilememesi, üretimin yapılmasını sağlayan makinalarda önceden tahmin edilemeyen bir arızanın meydana gelmesi, işçilerden birinin işini aksatması, malzeme taşıma sistemlerinde bir aksaklığın olması ve benzer hata kaynaklarının süreci etkilemesi şeklinde tanımlanmaktadır. Düzeltici işlev olarak, kötü yığın çıkarılması ve buna neden olan sistemlerde gerekli düzeltmelerin yapılması önerilmektedir(*Acar ve Booth, 1987*).

Bunların dışında anormal bir durum, 3. Tip hata durumu olarak değerlendirilmektedir. *Bu durumda üretim süreci için, sürekli proses probleminin bulunduğu anlaşılmalıdır. Bir başka deyişle, kalite programının niteliği, işletme koşulları, makina ve teçhizatın genel durumu, işçilerin fabrikadaki sosyal durumları gibi faktörler üzerinde bir kontrol mekanizması kurarak üretim sistemlerinin tüm bileşenlerinde inceleme gereklidir(Acar ve Booth, 1987).*



BÖLÜM 5

5. UYGULAMA

Uygulama amacıyla kullanılacak olan veriler, kimyasal bir üretim sürecinden iki saatte bir alınan konsantrasyon ölçümlerinden oluşmaktadır(*Box-Jenkins, 1976 s:525*).

Ayrıca Çimentaş çimento fabrikası T.A.Ş' nin 1980-1987 yılları arası yapmış olduğu aylık toplam ihracat rakamları kullanılacaktır(*Çimentaş T.A.Ş*).

Kullanılan iki farklı veri seti üzerinde hiç sapan değer kalmayınca kadar uygulanan işlemler adım adım verilecek, ayrıca analiz sonuçları ve sonuçların kalite kontrol yorumları, kimyasal üretim sürecinden elde edilen sonuçlar esas alınarak yapılacaktır.

Kimyasal üretim sürecinden Elde edilen gözlem değerleri iki saatte bir alınan konsantrasyon ölçümlerinden oluştuğundan, gözlemlerin eşit zaman aralıklarında elde edildiği anlaşılmaktadır. Bu nedenle elde edilecek zaman serisi kesikli zaman serisi olarak bilinmelidir. Ancak gözlem değerleri ölçülebilir özelliklere sahip olduğundan sürekli gözlem değerleri özelliği taşırlar.

Söz konusu veri seti için *Box* ve *Jenkins* tarafından önerilen iki modelden biri kullanılabilir(*Box ve Jenkins, 1976, s.239*).

Bu uygulamanın kapsamı içerisinde iki model de kullanılacak, bu modellere göre sonuç tabloları oluşturulacaktır.

$$\chi_t = c_{0+} \frac{(1-\theta B)}{(1-\phi B)} e_t \dots\dots\dots (5.1)$$

$$(1-B)\chi_t = (1-\theta B)e_t \dots\dots\dots (5.2)$$

Eldeki gözlem seti için (5.1) modeli kullanılarak bulunan sonuçlar MINTAB paketi kullanılarak elde edilmiştir.

Minitab çıktısı

MTB>Name c2= 'RESI2' C3 = 'FITS2' C4= 'COEF2'

MTB>ARIMA 1 0 1 C1 'RESI2' 'FITS2' 'COEF2'

MTB>Constant.

Final Estimates of Parameters

Iteration	SSE	Parameters		
0	33.2117	0.100	0.100	15.446
1	24.2840	0.250	-0.050	12.861
2	22.9740	0.400	0.080	10.287
3	21.6090	0.550	0.210	7.713
4	20.2822	0.700	0.338	5.141
5	19.3644	0.850	0.467	2.570
6	19.2640	0.891	0.525	1.869
7	19.2467	0.903	0.554	1.659
8	19.2422	0.909	0.568	1.554
9	19.2409	0.912	0.575	1.501
10	19.2406	0.914	0.579	1.474
11	19.2405	0.914	0.581	1.460
12	19.2405	0.915	0.582	1.453
13	19.2405	0.915	0.583	1.449

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
AR 1	0.9151	0.0433	21.11
MA 1	0.5828	0.0849	6.87
Constant	1.4490	0.0094	154.38
Mean	17.0656	0.1105	

No. of obs.: 197
 Residuals : SS = 19.1880 (backforecasts excluded)
 MS = 0.0989 DF = 194

Box-Pierce Statistics :

Lag	24
Chisquare	28 (DF=K-p-q=22)

22 serbestlik dereceli ki-kare tablo deęeri =33.92 'dir. Bu deęer Box-Pierce khi-kare deęeri olan 28'den byk olduęundan modelin uygunluęuna karar verilir.

<u>Sıra No</u>	<u>Gzlem Deę.</u>	<u>Kestirim Deę.</u>	<u>Hata Deę.</u>
1.	17.0	16.8767	0.12325
2.	16.6	16.9337	-0.33374
3.	16.3	16.8340	-0.53403
4.	16.1	16.6762	-0.57623
5.	17.1	16.5178	0.58219
6.	16.9	16.7578	0.14222
7.	16.8	16.8312	-0.03118
8.	17.4	16.8407	0.55928
9.	17.1	17.0457	0.05433
10.	17.0	17.0654	-0.06541
11.	16.7	17.0437	-0.34369
12.	17.4	16.9313	0.46866
13.	17.2	17.0985	0.10152
14.	17.4	17.1294	0.27058
15.	17.4	17.2139	0.18608
16.	17.0	17.2632	-0.26316
17.	17.3	17.1589	0.14106
18.	17.2	17.1979	0.00211
19.	17.4	17.1874	0.21264
20.	16.8	17.2477	-0.44768
21.	17.1	17.0835	0.01655
22.	17.4	17.0874	0.31256
23.	17.4	17.1894	0.21055
24.	17.5	17.2489	0.25110

25.	17.4	17.3168	0.08322
26.	17.6	17.3231	0.27690
27.	17.4	17.3933	0.00675
28.	17.3	17.3677	-0.06768
29.	17.0	17.3195	0.31954
30.	17.8	17.1918	0.60821
31.	17.5	17.3832	0.11681
32.	18.1	17.3950	0.70496
33.	17.5	17.6013	-0.10133
34.	17.4	17.5222	-0.12217
35.	17.4	17.4428	-0.04280
36.	17.1	17.3966	-0.29655
37.	17.6	17.2699	0.33010
38.	17.7	17.3622	0.33775
39.	17.4	17.4493	-0.04930
40.	17.8	17.4003	0.39966
41.	17.6	17.5047	0.09528
42.	17.5	17.4991	0.00090
43.	16.5 *	17.4626	-0.96259
44.	17.8	17.1090	0.69099
45.	17.3	17.3349	-0.03494
46.	17.3	17.3005	-0.00046
47.	17.1	17.2804	-0.18036
48.	17.4	17.2022	0.19781
49.	16.9	17.2563	-0.35633
50.	17.3	17.1217	0.17828
51.	17.6	17.1762	0.42380
52.	16.9	17.3076	-0.40764
53.	16.7	17.1516	-0.45163
54.	16.8	16.9942	-0.19424
55.	16.8	16.9358	-0.13575
56.	17.2	16.9017	0.29833
57.	16.8	16.9017	-0.21472
58.	17.6	16.9477	0.65231
59.	17.2	17.1745	0.02553
60.	16.6	17.1737	-0.57371
61.	17.1	16.9739	0.12612
62.	16.9	17.0236	-0.12358

63.	16.6	16.9861	-0.38608
64.	18.0*	16.8645	1.13547
65.	17.2	17.2589	-0.05893
66.	17.3	17.2229	-0.07707
67.	17.0	17.2352	-0.23518
68.	16.9	17.1426	-0.24623
69.	17.3	17.0555	0.24454
70.	16.8	17.1376	-0.33758
71.	17.3	17.0193	0.28071
72.	17.4	17.1165	0.28350
73.	17.7	17.2064	0.49361
74.	16.8	17.3585	-0.55846
75.	16.9	17.1480	-0.24802
76.	17.0	17.0586	-0.05860
77.	16.9	17.0397	-0.13972
78.	17.0	16.9955	0.00451
79.	16.6	17.0029	-0.40294
80.	16.7	16.8744	-0.17436
81.	16.8	16.8327	-0.03266
82.	16.7	16.8416	-0.14158
83.	16.4	16.8136	0.41355
84.	16.5	16.6975	0.19753
85.	16.4	16.6631	-0.26314
86.	16.6	16.6099	-0.00987
87.	16.5	16.6453	-0.14528
88.	16.7	16.6327	0.06731
89.	16.4	16.6918	-0.29182
90.	16.4	16.6266	-0.22658
91.	16.2	16.5886	-0.38856
92.	16.4	16.4999	-0.09994
93.	16.3	16.5148	-0.21476
94.	16.4	16.4902	-0.09016
95.	17.0	16.5091	0.49094
96.	16.9	16.7195	0.18054
97.	17.1	16.8088	0.29116
98.	17.1	16.9274	0.17260
99.	16.7	16.9965	-0.29649
100.	16.9	16.9038	-0.00383

101.	16.5	16.9163	-0.41629
102.	17.2	16.7906	0.40937
103.	16.4	16.9500	-0.55001
104.	17.0	16.7771	0.22295
105.	17.0	16.8756	0.12436
106.	16.7	16.9331	-0.23309
107.	16.2	16.8669	-0.66689
108.	16.6	16.6621	-0.06215
109.	16.9	16.6758	0.22425
110.	16.5	16.7834	-0.28337
111.	16.6	16.7132	-0.11317
112.	16.6	16.7055	-0.10549
113.	17.0	16.7010	0.29899
114.	17.1	16.8313	0.26858
115.	17.1	16.9405	0.15950
116.	16.7	17.0041	-0.30412
117.	16.8	16.9083	-0.10828
118.	16.3	16.8857	-0.58566
119.	16.6	16.7063	-0.10632
120.	16.8	16.7015	0.09851
121.	16.9	16.7651	0.13486
122.	17.1	16.8355	0.26453
123.	16.6	16.9429	-0.14291
124.	17.0	16.9058	0.09416
125.	17.2	16.9507	0.24931
126.	17.3	17.0433	0.25670
127.	17.2	17.1305	0.06951
128.	17.3	17.1481	0.15192
129.	17.2	17.1916	0.00844
130.	17.2	17.1837	0.01633
131.	17.5	17.1791	0.32093
132.	16.9	17.2761	-0.37608
133.	16.9	17.1332	-0.23324
134.	16.9	17.0500	-0.14999
135.	17.0	17.0015	-0.00147
136.	16.5	17.0064	-0.50643
137.	16.7	16.8432	-0.14316
138.	16.8	16.8145	-0.01448

139.	16.7	16.8310	-0.13099
140.	16.7	16.8074	-0.10738
141.	16.6	16.7936	-0.19362
142.	16.5	16.7524	-0.25237
143.	17.0	16.6951	0.30490
144.	16.7	16.8279	-0.12788
145.	16.7	16.8056	-0.10557
146.	16.9	16.7926	0.10743
147.	17.4	16.8514	0.54855
148.	17.1	17.0519	0.04808
149.	17.0	17.0691	-0.06906
150.	16.8	17.0458	-0.24582
151.	17.2	16.9658	0.23419
152.	17.2	17.0521	0.14790
153.	17.4	17.1024	0.29760
154.	17.2	17.1982	0.00183
155.	16.9	17.1875	-0.28752
156.	16.8	17.0816	-0.28162
157.	17.0	16.9867	0.01332
158.	17.4	16.9978	0.40220
159.	17.2	17.1372	0.06279
160.	17.2	17.1520	0.04800
161.	17.1	17.1606	-0.06061
162.	17.1	17.1324	-0.03240
163.	17.1	17.1160	-0.01596
164.	17.4	17.1064	0.29362
165.	17.2	17.2005	-0.00049
166.	16.9	17.1889	-0.28887
167.	16.9	17.0824	-0.18241
168.	17.0	17.0204	-0.02037
169.	16.7	17.0174	-0.31744
170.	16.9	16.9160	-0.01604
171.	17.3	16.9234	0.37659
172.	17.8	17.0606	0.73937
173.	17.8	17.3067	0.49325
174.	17.6	17.4502	0.14982
175.	17.5	17.4673	0.03269
176.	17.0	17.4441	-0.44407

177.	16.9	17.2644	-0.36437
178.	17.1	17.1264	-0.02641
179.	17.2	17.1125	0.08753
180.	17.4	17.1376	0.26242
181.	17.5	17.2187	0.28133
182.	17.9	17.2992	0.60084
183.	17.0	17.4790	-0.47899
184.	17.0	17.2847	-0.28472
185.	17.0	17.1715	-0.17150
186.	17.2	17.1055	0.09448
187.	17.3	17.1335	0.16647
188.	17.4	17.1831	0.21692
189.	17.4	17.2452	0.15481
190.	17.0	17.2814	-0.28138
191.	18.0	17.1696	0.83044
192.	18.2	17.4367	0.76331
193.	17.6	17.6588	-0.05883
194.	17.8	17.5889	0.21109
195.	17.7	17.6146	0.08538
196.	17.2	17.5964	-0.39638
197.	17.4	17.4196	-0.01959

Eldeki veriler için sapan değer analizinin yapılmasıyla elde edilen sonuçlar aşağıdaki tabloda verilmiştir.

Box-Jenkins A Serisi İçin Parametre Tahminleri ve Sapan Değer Tarama İşlemi

Tahmin Döngüsü	Parametreler				Sapan Değerler			
	C_0	ϕ	θ	$\sigma_e^2 \cdot 10^3$	Pozisyon	Tipi	W	λ
Model (5.1)								
1	17.1	0.91	0.58	98.90	64	IO	1.13	3.72
					43	AO	-0.96	-3.68
2	17.1	0.89	0.47	83.52	Sapan değere rastlanmadı.			

Aynı veri seti için (5.2) modeli kullanıldığında elde edilen sonuçlar aşağıda verilmiştir.

Box-Jenkins A Serisi İçin Parametre Tahminleri ve Sapan Değer Tarama İşlemi

Tahmin Döngüsü	Parametreler		Sapan Değerler			
	θ	$\sigma_e^2 \cdot 10^3$	Pozisyon	Tipi	W	λ
Model (5.2)						
1	0.70	100.74	64	IO	1.13	3.67
			43	AO	-0.96	-3.53
2	0.63	88.04	Sapan değere rastlanmadı.			

Şimdi ÇİMENTAŞ T.A.Ş 'nin 1980-1987 yılları arası yapılmış toplam ihracat rakamlarını kullanarak sapan değer analizi yapmaya çalışalım.

Aylar	1980	1981	1982	1983	1984
1	50.102	42.225	77.201	43.706	49.503
2	71.145	40.914	68.924	46.320	67.434
3	70.501	55.103	64.842	77.115	68.741
4	83.520	77.122	61.873	65.662	69.010
5	69.563	67.886	66.160	106.352*	72.785
6	72.458	71.705	84.226	52.389	57.894
7	60.688	67.705	87.153	74.530	60.025
8	55.644	82.760	85.334	101.974	86.433
9	76.358	91.458	64.612	66.265	78.107
10	62.741	84.560	70.985	53.068	91.288
11	76.588	60.839	67.395	73.906	78.373
12	54.032	68.323	55.755	41.513	68.427

Aylar	1985	1986	1987
1	70.711	59.683	72.633
2	52.727	66.256	86.870
3	86.862	89.786	79.721
4	81.940	101.769	103.820
5	86.384	100.492	91.222
6	63.961	79.214	91.223
7	80.405	97.618	83.663
8	85.255	77.594	76.528
9	87.404	98.737	84.026
10	83.804	104.815	100.705
11	79.582	83.386	83.752
12	77.593	73.309	84.062

Minitab Çıktısı

MTB> Name c2 = 'RESII'

MTB>ARIMA 1 0 0 C1 'RESII'

MTB> Constant.

Estimates at each iteration

Iteration	SSE	Parameters	
0	20304.8	0.100	66.903
1	18620.2	0.250	55.740
3	17892.7	0.400	44.557
4	17860.4	0.435	41.935
5	17860.0	0.438	41.631
6	17860.0	0.439	41.595
7	17860.0	0.439	41.591

Relative Change in each estimate less than 0.0010

Final Estimates of Parameters

Type	Estimate	St.Dev	t-ratio
AR 1	0.4389	0.0928	4.73
Constant	41.591	1.403	29.64
Mean	74.123	2.501	

No of obs. : 96

Residuals SS = 17770.2 (backforecasts excluded)
 MS = 189.0 DF = 94

Box-Pierce Statistics :

Lag	24
Chisquare	26.5 (DF=K-p-q=23)

23 serbestlik dereceli ki-kare tablo değeri =35.17 'dir. Bu değer, Box-Pierce khi-kare değeri olan 26.5'den büyük olduğundan modelin uygunluğuna karar verilir.

Yukarıda elde edilen parametre değerleri için iteratif yöntem kullanıldığında

Tahmin Döngüsü	Parametreler		Sapan Değerler			
	ϕ	σ_e^2	Pozisyon	Tipi	W	λ
1	0.4389	189	41	AO	28.1741	3.94
2	0.503	167	Sapan değere rastlanmadı.			

Çimentoş T.A.Ş tarafından ihracat rakamları kullanılarak elde edilen tahmin sonuçlarına bakıldığında 1993 yılı mayıs ayında beklenen ihracat rakamından 28.1741 bin ton daha fazla ihracat yapıldığı görülmektedir. Bu da beklenenden daha büyük bir rakam olduğundan sapan değer sonucu doğurmuştur. Yukarıdaki tablodan gözlenebileceği gibi sapan değerinin etkisini yok ettiğimizde AR(1) parametre değeri 0.4389 değerinden 0.503 değerine çıkmış, varyans değeri ise 189 değerinden 167 değerine gerileyerek yaklaşık % 12' lik bir küçülme meydana getirmiştir. Bu da yeni elde edilen AR(1) parametre değerinin ileriye dönük tahminlemede daha tutarlı ve etkin sonuçlar elde edileceğini göstermektedir. Söz konusu analiz sonuçları ekonomik boyutlu olduğundan daha fazla yorum yapılmayacaktır. Asıl amaç, kalite kontrolle ilişkili süreçler için sapan değer analizi olduğundan, kimyasal konsantrasyon ölçümlerinin kullanılması ile elde edilen sonuçlara dayalı yorumlar yapılacaktır.

5.1 Uygulama Sonuçları Tartışma ve Yorumlar

Sapan değer tarama yöntemlerinden bazıları, özellikle basit örneklemler için geliştirilen test yöntemleri, tek sapan değer tarama esasına göre geliştirilmiş olup, çok güçlü olmayan yöntemler olarak bilinmektedir. Bu test yöntemleri her adımda tek bir sapan değer teşhis etme esasına göre çalışan ve adimsal test yöntemleri olarak bilinen yöntemlerdir. Oysa bu gün etkin olan test yöntemleri, olması muhtemel bütün sapan değerlerin teşhis edilmesine yönelik olarak geliştirilen ve aynı anda test yöntemi olarak ta bilinen yöntemlerdir. Yukarıda karşılaştırılan test yöntemleri de bu gruba girmektedir. Her iki grup test yönteminin de bazı sakıncaları vardır. Adimsal test yöntemleri olarak bilinen 1. grup testler eğer, örnekleme bulunan sapan değerler birbirlerine pozisyon olarak çok yakın iseler, uzaklık ya da hacim ölçütlerine dayalı yöntemler kullanıldığında (Regresyon Analizinde sapan değerler için Cook yaklaşımı) sapan olan bazı gözlemlerin *gizlenme etkisiyle*(Masking Effect) sapan olmadığı bulunabilir. Aslında aynı anda test yöntemleri gizlenme etkisinin giderilmesine yönelik geliştirilmişlerdir. Birlikte test yöntemlerinde ise, örnekleme sapan değerler, ötekilerden çok farklı iseler, bu gözlem değerleri ile birlikte sapan olmayan gözlemlerin de sapan olarak bulunabilmesidir. Bu da *sürüklene etkisi* olarak bilinmektedir(*Tatlıdil, 1981*).

Kimyasal konsantrasyon ölçümlerine ilişkin gözlem değerleri üzerinde bir inceleme yapılırsa, görülebileceği gibi, sapan değer olarak bulunan değerler, eldeki veri seti içinde değer olarak en büyük veya en küçük değildir. 43. pozisyonda 1. tip sapan değer olarak bulunan 16.5 değeri ve 64. pozisyonda 2. tip sapan değer olarak bulunan 18.0 değeri kestirim değerlerinden çok farklı olduklarından sapan olarak bulunmuşlardır. Dikkat edilmesi gerekli önemli husus, sapan değer etkileri giderildikten sonra elde edilen parametre ve varyans değerlerinde meydana gelen değişimdir. Sapan değer etkisinin yok edilmesi ile parametre değerinde % 10'luk bir değişim meydana gelirken, varyans da meydana gelen küçülme % 12.70 oranında olmuştur. Bu da sapan değerlerin parametre değerleri, varyans ve güven aralıkları üzerindeki olumsuz etkisini açık bir şekilde ortaya koymaktadır. Kalite kontrol yönünden bir yorum yapmak gerekirse her iki durum için de üretim sürecinde bir kereye özgü hata sözkonusudur. 43. üretim anında sapan değerlerin elde edilmesine sebep olan muhtemel hata kaynakları,

üretimin yapılmasını sağlayan makinalarda önceden hesap edilemeyen bir arızanın ortaya çıkması, işçilerin işlerini aksatmaları, malzeme taşıma sistemlerinde bir aksaklığın ortaya çıkması, hammaddenin sağlıklı ortamlarda muhafaza edilememesi veya benzer nedenlerin bir veya bir kaçının ortaya çıkması sonucu oluşmuştur. Bir başka deyişle, kişi, makina ve ortamın etkisiyle ortaya çıkmış bir hatadan söz edilebilir. Bu durum, 2. bölümde bahsedilen kalitenin ölçüsü ve değişkenliği ile ilgili özel nedenler olarak karşımıza çıkmaktadır. 64. üretim anında sapan değer in elde edilmesine sebep olan muhtemel hata kaynakları ise, girdi hammaddesinde beklenen standartların elde edilmemiş olması veya kabul örneklemede bir düzenleme yapma gerekliliği ile ilgili bir durum olabilir. Düzeltici işlev olarak kötü yığın in çıkarılması önerilebilir.

Bu çalışma, çok küçük sapmaların bile önem arzettiği mesela ilaç üretiminin yapıldığı süreçler için kullanılabilecek bir kalite kontrol aracı olarak kabul edilebilir. Çünkü bu yöntem, üretim sürecinin doğru çalışması prensibinden hareketle, önceden beklenmeyen hatalı durumları tespit etmeye yönelik bir yöntemdir.

Bunun yanında bir üretim sürecinin kontrol altında olmaması durumunda probleme neden olabilecek bir çok olası durum karşısında her seçeneği ayrı ayrı incelemek yerine, İteratif yöntemle saptanan, sapan değer tiplerine göre problemin girdi, çıktı yada süreç problemi olup olmadığına karar verilebilir.

İteratif yöntemin bu çalışmada geliştirilen bilgisayar programı ve bir üretim sürecinden alınan gözlemlerin oluşturduğu zaman serisinin parametre kestirimlerini verecek bir bilgisayar programı, kontrol kartlarını içeren paket programlara ilave edildiği takdirde zamandan ve paradan büyük tasarruflar sağlanabilir.

SUMMARY

Outliers are defined as, the few observations, which appear to be inconsistent with the remainder group of the sample and much effects on the prediction values.

In this study, the historical information about quality control, time series, and finally outliers which are existing in simple samples and time series, are defined in introduction chapter.

The second chapter provides a definition for quality control processing.

Third chapter includes time series models, called Box-Jenkins models in time series literature, which are called AR, MA, ARMA and ARIMA models.

The main subject of this study is in the fourth chapter that are called outliers have been introduced in simple samples, and time series which can be regarded as being generated by dynamic models at known time points. Two special cases innovational outlier (IO) and additive outlier (AO) are studied. The likelihood ratio criteria for testing the existence of outliers of both types, and the criteria for distinguishing between them are defined. An iteratif procedure is proposed for detecting IO and AO in practice and for estimating the time series parameters in autoregressive-integrated-moving-average models in the presence of outliers.

At the same time, the powers of outliers in detecting outliers are investigated by simulation experiments.

The performance of proposed procedure for estimating the autoregressive coefficient of a simple AR(1) model compares favorably with robust estimation procedures proposed in the literature. Two real examples are presented.

KAYNAKLAR

- ABRAHAM, B ve CHUANG, A. 1989 : Outlier detection and time series modelling, *Technometrics*, 31,241-248.
- ABRAHAM, B ve YATAWARA, N. 1988: A score test for detection of time series outliers. *J. Time Ser. Anal.*,9, 109-119.
- ABRAHAM, B. ve YATAWARA N. 1988: A score test for detection of time series outliers, *J. Time Ser. Anal.*, Vol. 9 , No:2.
- ABRAHAM , B. ve BOX, G.E.P. 1979: Bayesian analysis of some outlier problems in time series, *Biometrika*, 66 , 2 , p : 229-236, Printed in Great Britain.
- ACAR, W. ve BOOTH, P.D.E. 1987: *Easier Quality Control: Combining problem classification with time series analysis* : Graduate School of Managment, Kent state University, Kent, oh 44242-0001.
- ALBA, E ve ZARTMAN, L. D. 1980: *Testing outliers in time series*, Anal. Time Ser, North-Holland publishing Company.
- ANDREWS, D.F. ve D. PREGIBON, 1978: Finding the outliers that matter, *J.R. Stat. Soc. Ser B*, 40: 85-93.
- ANSCOMBE, M ve TUKEY J.W., 1963: The examination and analysis of residuals, *Technometrics*, 5, 141.
- BASKAN, Ş. 1994 : Kalite kontrol ders notları.
- BAYHAN, M. 1992: Kalite kontrolünde zaman serisi analizi, *End. Müh. Dergisi*, 21:17-21.
- BOX, G.E.P ve JENKINS, G.M 1976 : *Time series analysis : Forecasting and control*, Sect 6.4.3 San Francisco : Holden-Day.
- BROWN, G. R 1970: *Smoothing Forecasting and prediction of discrete time series*, Prentice-Hall & Inc. London.
- BRUCE, A.G ve MARTIN, R.D. 1989 : Leave-k-out diagnostics for time series (with discussion), *J.R. Statist. Soc. B*,51,363-424.
- CHANG, I., TIAO, G.C., CHEN,C. 1988 : Estimation of time series parameters in the presence of outliers ., American Statistical Association and the American Society for Quality Control.

- CHAMBERS, J., MULLICK, S, K., SMITH, D.D. 1971 : How to chose the right forecasting tecnique, *Harvard business review*.
- COOK, P.D 1979: Influential observations in linear regression, *JASA*, 74: 169-174.
- CORLETT, D. ve LEWIS, T. 1976: The subjective nature of outliers rejection proceduures., *Applied statistics*, 25, no:3 s:288.
- DRAPER, N.R ve SMITH, H. 1972: *Applied regression analysis*, New-york.
- EKERN, S. 1976: Forecasting with adaptive filtering: a chiritical reexamination, *operational research Quarterly*, Vol.27, No:3, 1976.
- ERTEK, T. 1982: *Ekonometriye giriş, Araştırma Eğitim, Ekin Yayınları, İstanbul*.
- FITZMAURICE, G.M., LIPSITZ, S.R. 1995: A model for binary time series data with serial odds ratio patterns., *J.R. Statist. Soc., C*, 44, 51-61.
- FOX, A.J. 1972: Outliers in time series., *J.R. Statist. Soc. B*, 34, 350-363.
- HADI, A.S. 1992: Identifying multiple outliers in multivariate datat *J.R. Statist. Soc.,B*, 54, No:3, S:761-771.
- HAWKINS, D.M. 1978 : *Identification of outliers*, Printed in Great Britain, Chapman and Hall.
- KAYIM, H. 1989: *İstatistiksel ön tahmin yöntemleri*, HÜ.İLB.F., Yayın no: 11, Ankara, 1989.
- KOHN, R. ve ANSLEY, C.F. 1986: Estimation prediction and interpolation for ARIMA models with missing data., *JASA*, 81, 751-761.
- LJUNG, G.M., BOX, G.E.P. 1979: The likelihood function of stationary autoregressive-moving average models., *Biometrika*, 66, 265-270.
- MONTI, A.C. 1994 : A proposal for a residual autocorelation test in linear models, *Biometrika*, 81, No:4 776-781.
- MUIRHEAD, C.R. 1986: Distinguishing outlier types in time series., *J.R. Statist. Soc. B*, 48, NO:1 S:39-47.
- ÖZMEN, A. 1986: Zaman serisi analizinde Box-Jenkins yöntemi ve Banka mevduat tahmininde uygulama denemesi., *Fen Edebiyat Fakültesi yayınları No: 9.*, Eskişehir.

- PENA, D. 1987: *Measuring the importance of outliers in ARIMA models in New perspectives in theoretical and applied statistics.*, New york : Wiley.
- PENA, D., YOHAI V.J. 1995: The detection of influential subsets in linear regression by using an influence matrix., *J. R. Statisti Soc., B.*, 57., No:1., 145-157.
- PRINS, J. 1980: *Interactive multivariate time series analysis.*, *Anal. Time Ser.*, North-holland Publishing Company.
- PUKKILA, T., KOREISHA, S. ve KALLINEN, A. 1990: The identification of ARMA models., *Biometrika*, 77, s:537-548., Printed in Great Britain.
- SENTA, E. ve KOUNIAS, S. 1990: Parameter estimation and order determination of otoregressive models, *Anal. Time Ser*, North-Holland Publishing Company.
- SHARPLES, D. 1990: Identification and accomodation of outliers in general hierarchical models., *Biometrika*, 77, s:445-453.
- SHIAO, L. WU, Y. ve HOSKING, J.R.M (1993) Reallocation outliers in time series., *Applied Statistics.*, 42, No:2. S:301-313.
- ŞAHİN, M. 1982: Yönetimde bilgisayar desteği ve örnek karar modelleri, E.İ.T.İ.A Yayınları, No: 252/172, Eskişehir.
- ŞEN, A. ve ŞEHİRLİOĞLU, A.K. 1993: Sapan ve etkili gözlemlerin bulunmasında kullanılan istatistikler, *Ekonometri- İstatistik Sempozyumu*, İzmir.
- ŞENYAY, L. 1982: Zaman serisi analizlerinin temel prensipleri ve bu analiz için Genstat paket programının kullanımı üzerine bir çalışma., *Basılmamış yüksek lisans tezi.*, E.Ü. Elektronik ve Hesap Bilimleri Enstitüsü., İzmir.
- TAPLIN, H.R. 1993: Robust likelihood calculation for time series., *J.R Statist. Soc.*, b, 55, No:4 s:829-836.
- TATLIDİL, H. 1981: Doğrusal regresyonda ve çok değişkenli verilerde kuşkulu gözlemlerin testi., *Basılmamış doktora tezi.*, Hacettepe Üniversitesi., Ankara.
- TECIM, V. 1990: Zaman serisi analizlerinde sermaye piyasası uygulamaları, *Basılmamış yüksek lisans tezi.*, D. Eylül Üniv., İzmir.
- TSAY, R.S 1988: Outliers, level shifts, and variance changes in time series : *Journal of forecasting.*, 7: 1-20.

WHEELRIGHT, S.C VE MAKRIDAKIS, S. 1973: Forecasting methods for managment.,
John Wiley & Sons Inc., New york, 1973.



EK-1 İTERATİF PROSEDÜRÜN BİLGİSAYAR PROGRAMI

```

10 REM BU PROGRAM ZAMAN SERİLERİNDE SAPAN DEĞER ANALİZİ YAPAR
20 REM Bu Program Zaman Serilerinde Ortaya Çıkması Muhtemel
30 REM Sapan Değerlerin Pozisyon ve Tiplerini Ortaya Çıkarır.
40 REM Yüksek Duyarlılık için c=3.00, Orta Duyarlılık için c=3.50 ve
50 REM Düşük Duyarlılık için c=4.00 alınmalıdır.
60 DIM HATA(197), PI(197), FMBDA(2,197), Z(197)
70 DIM X(197), ZO(5), ZS(5), WA(197), WI(197)
80 LET L=0: KEY OFF:CLS
90 REM Model Parametrelerinin Okutulması
100 PRINT "ARIMA(p,d,q) Parametrelerini Okutunuz"
110 PRINT "p (AR) Parametresini Giriniz "
120 INPUT "", P
130 PRINT"d (FARK) Parametresini Giriniz"
140 INPUT "",D
150 PRINT "q (MA) parametresini Giriniz"
160 INPUT "", Q
170 PRINT "Gözlem Sayısını Giriniz"
180 INPUT "", N
190 PRINT "Gözlemleri Giriniz"
200 A$=INPUT$(1)
210 FOR K=1 TO N
220 LOCATE 10,10:PRINT K;".Gözlem"
230 LOCATE 10,25:INPUT "",Z(K)
240 CLS
250 NEXT
260 REM Hata Değerlerinin Okutulması
270 FOR K=1 TO N
280 LOCATE 10,10:PRINT K;".Hata"
290 LOCATE 10,25:INPUT "", HATA(K)
300 NEXT
310 REM Pi Ağırlıklarının Okutulması
320 FOR K=1 TO N
330 LOCATE 10,10:PRINT K;".pi Değeri"
340 LOCATE 10,25:INPUT "",PI(K)
350 NEXT
360 LET I=197
370 REM Serinin Tahmin Edilen Parametre Değerleri
380 REM Hata varyansı ve c sabitinin okutulması
390 PRINT "Tahmini Parametre Değerlerini Giriniz"
400 PRINT "AR Parametre Değerlerini Giriniz"
410 INPUT INPUT "",FI
420 PRINT "MA Parametre Değerini Giriniz"
430 INPUT "", TETA
440 PRINT "Varyans Değerini Giriniz"

```

```

450 INPUT "",VAR
460 PRINT"C Parametre Değerini Giriniz"
470 INPUT "",C
480 STHATA=VAR^0.5
490 FOR K2=1 TO 2
500 FOR K3=1 TO N
510 FMBDA(K2,K3)=0
520 NEXT
530 NEXT
540 FOR K=1 TO N
550 FMBDA(1,K)=HATA(K)/STHATA
560 IF K=1 THEN 590
570 IF ABS(FMBDA(1,K))>ENB1 THEN 590
580 GOTO 620
590 ENB1=ABS(FMBDA(1,K))
600 T1=1
610 T2=K
620 NEXT
630 REM Omegaa'nın Bulunması
640 FOR K=1 TO N-1
650 T4=N-K
660 TP1=HATA(K)
670 TP2=1
680 FOR K2=1 TO T4
690 F1=PI(K2)*HATA(K+K2)
700 FM=PI(K2)^2
710 TP1=TP1-F1
720 TP2=TP2+FM
730 NEXT
740 OMEGAA=TP1/TP2
750 REM Lambda(2,T)'lerin ve en büyük lambda(2,t)'nin bulunması
760 FMBDA(2,K)=(OMEGA*TP2^0.5)/STHATA
770 IF K=1 THEN 800
780 IF ABS(FMBDA(2,K))>ENB2 THEN 800
790 GOTO 830
800 ENB2=ABS(FMBDA(2,K))
810 T3=2
820 T4=K
830 NEXT
840 REM Lambda(1,t) ve Lambda(2,t)'ler arasından c'den büyük olanın bulunması
850 IF (ENB1>ENB2) AND ENB1>C THEN 1060
860 IF (ENB2>ENB1) AND ENB2>C THEN 890
870 PRINT "Sapan Değere Rastlanmadı l..."
880 GOTO 1240
890 REM 1. Tip sapan değer etkisinin bulunması
900 TP1=HATA(T4)
910 TP2=1

```

```

920 FOR K=1 TO N-T4
930 F1=PI(K)*HATA(T4+K)
940 FM=PI(K)^2
950 TP1=TP1-F1
960 TP2=TP2-FM
970 NEXT
980 OMEGAA=TP1/TP2
990 LET L=L+1
1000 ZS(L)=T4
1010 ZO(L)=1
1020 WA(T4)=OMEGAA
1030 PRINT "1. Tip sapan deęere rastlandı l..."
1040 HATA(K)=HATA(K)-OMEGAA*(-PI(K-T4))
1050 NEXT
1060 GOTO
1070 REM 2. Tip sapan deęer var. Hatalar yeniden hesaplanıyor .
1080 OMEGAI=HATA(T2)
1090 HATA(T2)=0
1100 L=L+1
1110 ZS(L)=T2
1120 ZO(L)=2
1130 WI(T2)=OMEGAI
1140 PRINT "2. Tip sapan deęere rastlandı l..."
1150 PRINT OMEGAI, ENB1
1160 TP=0
1170 FOR K=1 TO N
1180 THATA=HATA(K)^2
1190 TP=TP+HATA
1200 NEXT
1210 REM Yeni hata varyansının hesaplanması
1220 VAR=TP/N
1230 GOTO 480
1240 IF L=0 THEN 1360
1250 FOR I=1 TO L
1260 M=ZS(I)
1270 IF ZO(I)=1 THEN X(M)=Z(M)-WA(M)
1280 IF P=0 AND D=1 AND Q=1 THEN X(M)=Z(M)-Z(M-1)+X(M-1)-WI(M)+TETA*WI(M-1)
1290 IF P=1 AND D=1 AND Q=1 THEN ZS(I)=M:X(M)=Z(M)-FI*Z(M-1)+FI*X(M-1)+TETA*WI(M-1)+FI*X(M-1)-FI*X(M-2)
1300 IF P=1 AND D=0 AND Q=1 THEN X(M)=Z(M)-FI*Z(M-1)+FI*X(M-1)+TETA*WI(M-1)-WI(M)
1310 IF P=1 AND D=0 AND Q=0 THEN X(M)=Z(M)-WI(M)+FI*WI(M-1) ELSE X(M)=Z(M)-WI(M)+TETA*WI(M-1)
1320 PRINT M, X(M)
1330 FOR K=1 TO N
1340 PRINT X(K)
1350 NEXT

```