



## AR(1) modelinde A tipi sapan etki

Ahmet Kaya

Ege Üniversitesi, Tire Kutsan Meslek Yüksekokulu  
Bilgisayar Bölümü, 35900-Tire, İzmir  
[ahmet.kaya@ege.edu.tr](mailto:ahmet.kaya@ege.edu.tr)

### Özet

Bu çalışmada, zaman serilerinde sapan değer modelleri ve etkileri incelenmiş, A tipi sapan değer, AR(1) modeli üzerinde meydana getirdiği etki çalışılmıştır. Bu çalışmada bir simülasyon uygulaması yapılarak 0.7 parametre değerli, seri uzunlukları farklı (50, 100, 200, 500 ve 1000) 5 tür serinin merkezsel pozisyonlarına A tipi etkiler yerleştirilerek varyans değerlerinde meydana gelen artış miktarı gözlenmiş, elde edilen sonuçlara varyans analizi işlemi uygulanarak istatistiksel çıkarımlar yapılmıştır. Varyans analizi sonucunda seride ortaya çıkan her bir sapan etkinin varyans değerinde yaklaşık bir kat artışa neden olduğu görülmüş, seri uzunluklarının sapan değer değer etkisini küçülten bir faktör haline geldiği sonucuna ulaşılmıştır.

**Anahtar sözcükler:** Sapan değerler; ARMA modeller; A tipi sapan değer; B tipi sapan değer.

### Abstract

#### *Type A outlier's effect in AR(1) model*

*In this study, it is investigated the outlier models and it's effects in time series, then type A outlier effect on a AR(1) model has been studied. A simulation study has been done onto 0.7 valued AR(1) parameter, series greatness are 50, 100, 200, 500, 1000, it has been observed behavior of growing variation results by replacing type A outlier in the central points of series. And according to be taken results some inferences, variance analysis has been planned and some inferences given. After that some important results have been taken, one of them is that every outlier ascended to series have one another total error on variation for model. Another important result is that series wideness has positive contribution to minimize factor on outlier effect.*

**Keywords:** Outliers; ARMA models; Type A outlier; Type B outlier.

## 1. Giriş

Zaman serileri analizinde geleneksel yöntemler genelde durağanlık ve doğrusallık varsayımlarına dayanır. Oysa bu varsayımlara ulaşmak gerçek hayatta ulaşılması çok zor, bazen olanaksızdır. Durağanlık ve doğrusallık, yapılması düşünülen analizlerin, istatistiksel kural ve varsayımlara uygunlukları bakımından da önemlidir. Ancak, durağanlığı bozan birçok faktör bulunmaktadır. Bunlardan biri ve en önemlisi, modellere ve parametrelere kaynak teşkil eden gözlemler üzerinde meydana gelen hatalı durumlardır. Bu durum daha başlangıç aşamasında tahmin amacıyla elde edilmesi ve kullanılması düşünülen parametrelerin yanlı olmasına neden olmaktadır. Bu türden hatalı durumların ortaya çıkmasına neden olan iki faktör bulunmaktadır. Bu nedenlerden bir tanesi, doğal rastgelelik, diğeri insan faktörüdür. Bir gözlem doğal rasgelelik sonucunda beklenenden çok farklı bir biçimde ortaya çıkabilir. Örneğin, bir grev sonucunda üretimin ciddi bir biçimde sekteye uğraması, beklenmeyen ekonomik bir krizin ortaya çıkması, öngörülmesi mümkün olmayan depremlerin doğurduğu yıkıcı etkiler, konjktürel bir dalgalanmanın

meydana çıkması, insanların hücrelerinde meydana gelen beklenmeyen bir artış ya da ani ve beklenmeyen iklim hareketleri ve benzerleri biçiminde ortaya çıkmaktadır[2].

Gözlemlerin kişilerden kaynaklı olarak hatalı biçimlerde elde edilmeleri ise doğrudan doğruya özel bir durum olarak kabul edilmektedir. Bu da bir fabrikada çalışan işçilerin bir ya da birkaçının işe gelememesi nedeniyle ortaya çıkan üretim kaybı, bir ölçüm elemanının bir bilgiyi yanlış kaydetmesi, bir makinede meydana gelen arıza, elektrik ya da enerji kısıtlamasının neden olduğu üretim ya da hammadde kaybı gibi durumlar bunu ifade etmektedir[10].

## 2. Temel varsayımlar ve modeller

Zaman serisi modelinde  $\varepsilon_t$  ile gösterilen süreç ak gürültü süreci olarak tanımlanır. Bu süreç, sıfır ortalamalı rasgele değişkenlerle, sabit varyans,  $\sigma_t^2$ 'nin serisi olarak ifade edilmekte olup aşağıdaki varsayımlarla ifade edilir.

- $E(\varepsilon) = 0$ ,
- $E(\varepsilon_t^2) = \sigma_t^2$ ,
- $E(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$ .

Bir kesikli  $X_t$ , için bu bağlamda bir model tanımlanırsa, ak gürültü süreçli bir seri ve  $\eta$  fonksiyonu aşağıdaki biçimde elde edilir.

$$\eta(\dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \dots) = \varepsilon_t \quad (1)$$

Böylece (1) eşitliği,  $\eta$  fonksiyonunun doğrusal olması durumunu sağlamaktadır.

$$\sum_{k=-\infty}^{\infty} \eta X_{t-k} = \varepsilon_t \quad (2)$$

Bu eşitlik aynı zamanda doğrusal olmayan modeller için de genel bir yapıdır ancak  $X_t$ , geçmiş değerlere olduğu kadar, gelecek değerlere de bağlı olarak elde edilmektedir.

Uygulamada  $X_t$ , ancak geçmiş dönemlere bağlı olarak elde edilmektedir. Bu durumda (2) bu duruma göre yeniden formüle edildiğinde,

$$\sum_{k=0}^{\infty} \eta X_{t-k} = \varepsilon_t \quad (3)$$

biçiminde elde edilmiş olur.

Geriye doğru öteleme işleci,  $BX_t = X_{t-1}$  olarak ifade edilmektedir. Bu durumda,

$$T(B)X_t = \varepsilon_t, \quad (4)$$

$$T(l) = \sum_{k=0}^{\infty} \eta_k l^k \quad (5)$$

biçimi elde edilmiş olur. Böylece (4) eşitliği,

$$X_t = T^{-1}(B)\varepsilon_t \quad (6)$$

biçiminde elde edilmiş olur. Bu durumda  $T^{-1}(l)$  durağan serisi,  $-1 < l < 1$  biçiminde genişletilirse  $T(l)$  de durağan olacaktır. Böylece,

$$T^{-1}(l) = \alpha_0 l^0 + \alpha_1 l^1 + \alpha_2 l^2 + \dots \quad (7)$$

$$X_t = \sum_{k=0}^{\infty} \alpha_k \varepsilon_{t-k} \quad (8)$$

veya

$$X_t = \Gamma(B)\varepsilon_t \quad (9)$$

elde edilmiş olur.

Burada,  $\Gamma(z) \equiv T^{-1}(z)$  veya,  $\Gamma(z)T(z) \equiv 1$  biçiminde elde edilir.

Ayrıca (8) Eşitliği, doğrusal modeller için alternatif bir model özelliği taşır.  $X_t$ , hata terimi  $\varepsilon_t$ 'nin geçmiş dönem değerlerinin bir kombinasyonu biçiminde elde edilmiştir [9].

ARMA(p,q) modeli  $p$  parametrelili otoregresif AR(p) modeli,  $q$  parametrelili hareketli ortalamama, MA(q), modelinin bir kombinasyonu biçiminde ifade edilir [1][2].

$p$  Dereceden AR(p) Modeli,

$$X_t + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} = \varepsilon_t \quad (10)$$

$q$  Dereceden MA(q) Modeli,

$$X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (11)$$

biçiminde ifade edilir. Bu durumda ARMA(p,q) modeli,

$$X_t + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (12)$$

buradan (12) modeli,

$$\alpha(B)X_t = \beta(B)\varepsilon_t \quad (13)$$

biçiminde ifade edilir.

### 3. Sapan değerler

#### 3.1. A tipi sapan değerler

Bir gözlem seti içerisinde gözlemin sadece kendisini etkileyen sapan değer türüne, A tipi sapan değer adı verilir. Sapan değer bu tipini gösteren model aşağıdaki gibidir [3].

$$X_t = \sum_{r=1}^p \alpha_r X_{t-r} + \varepsilon_t \quad (t = p+1, \dots, n) \quad (14)$$

Burada,  $\alpha_r$  otoregresif parametreler ve  $\{\varepsilon_t\}$  bağımsız normal,  $N(0, \sigma_\varepsilon^2)$ .  $\{y_t\}$  Gözlemi aşağıda tanımlanan bir nicelik olup,

$$y_t = \begin{cases} X_t & (t \neq q) \\ X_t + \Delta & (t = q) \end{cases}$$

biçiminde tanımlanmaktadır. Burada  $\alpha_r$  değerleri  $\{X_t\}$  süreci durağan olacak biçimde elde edilmiştir. Ayrıca, regresif modelin parametresi olarak kabul edilen  $p$ ' nin bilindiği varsayılmaktadır. Burada,  $\Delta$ , A tipi sapan etkiyi gösteren bir nicelik olup,  $H_1 : \Delta \neq 0$  alternatif hipotezine karşı kurulan ve  $H_0 : \Delta = 0$  olması istenen bir değerdir. Bu tip hatalar, daha çok sürece bağlı olarak elde edilen hatalar olmayıp, insan kaynaklıdır. Mesela, bir kalite kontrol sürecinde, kontrol veya tasnif elemanın bir anlık dalgınlığı ve ihmali sonucunda bu tür hatalı durumlar ortaya çıkabilir [10]. Bu tür durumlar, çok büyük hata durumlarını ifade ederler, bu durumda gözlemler çok ciddi bir biçimde etkilenir [5]. Model içerisinde A tipi hatanın varlığı, parametre tahmin değerlerini etkilemekte ve yanlı olmalarına neden olmaktadır.

#### 3.2. B tipi sapan değerler

Bir gözlem serisi içerisinde, ortaya çıktığı pozisyondan itibaren bütün gözlemleri etkileyen sapan değer türüne, B tipi sapan değer adı verilir. Sapan değer bu tipini tanımlayan model aşağıdaki biçimde elde edilmiştir [3].

$$y_t = \sum_{r=1}^p \alpha_r y_{t-r} + \Delta_t + \varepsilon_t \quad (15)$$

Burada,

$$\Delta_t = \begin{cases} 0 & (t \neq q) \\ \Delta & (t = q) \end{cases}$$

ve  $\alpha_r$  ( $r = 1, \dots, p$ ) ve  $\{\varepsilon_t\}$  bağımsız normal  $N(0, \sigma_\varepsilon^2)$  süreçtir. Burada tanımlı,  $\Delta$ , sapan etki,  $y_q$  gözleminden başlanarak  $y_{q+1}, y_{q+2}, \dots$  bütün gözlemleri etkilemektedir.

B tipi sapan etkiler, nispeten daha az etki içermelerine rağmen, oraya çıktığı pozisyondan itibaren, azalan bir trend içerisinde bütün gözlemleri etkileme özelliğine sahiptir. Kalite kontrol süreçlerinde ise, bu bir süreç hatası durumunu ifade etmektedir [10][4]. Ayrıca bu türden olumsuzluklar beklenmeyen bir zaman içerisinde ortaya çıkmaktadır [5].

#### 4. Simülasyon çalışması

Bu çalışmada, 0.7 katsayılı AR(1) modeline uygun 50,100,200,500 ve 1000 uzunluğunda seriler yaratılarak simülasyon çalışmaları yapılmıştır. 0.7 katsayısı, serileri oluşturan gözlemlerin % 70 oranında bağımlılık gösteren elemanlardan oluştuğunu ifade etmekte, genel ve ortalama bir seri bağımlılığı olarak kabul görmektedir. Beş tür serinin merkezsiz noktaları olarak kabul edilen pozisyonlarına, 1, 2, 3, 4 ve 5 adet A tipi sapan değer yerleştirilerek model varyansında meydana gelen değişim incelenmiştir.

Sapan değerlerin yerleştirilme ilkeleri şu şekilde planlanmıştır: Tek sapan değer; 50, 100, 200, 500 ve 1000 uzunluğunda serilerin merkezi kabul edilen 26, 51, 101, 251 ve 501. inci pozisyonlarına, iki sapan değer, yine serilerin merkezi kabul edilen 17 ve 34, 33 ve 66, 67 ve 134, 166 ve 332, 333 ve 666. ıncı pozisyonlarına, bu şekilde devam edilerek seri uzunlukları ile orantılı merkezsiz pozisyonlara sapan değerler yerleştirilmiştir. Tarama süreci olarak *Chang Tiao ve Chen (1988)* tarafından geliştirilen ve etkinliği varyans analizi sonuçları ile doğrulanan yinelemeli yöntem kullanılmıştır [4].

**Çizelge 1.** Model varyansı için meydana gelen artış yüzdeleri.

Seri uzunluğu	1	2	3	4	5
50	4.89	9.36	21.36	55.72	87.36
100	4.31	8.12	18.21	45.12	70.80
200	4.07	7.89	17.20	38.76	60.13
500	3.69	7.50	15.93	30.01	49.27
1000	3.09	6.25	12.13	22.36	35.56

Çizelge-1, de yer alan gözlemler, sapan değer sayıları ve seri uzunlukları esas alınarak oluşturulmuştur. Mesela, 4.89 verisi, 50 uzunluğunda bir zaman serisi içerisine yerleştirilen 1 sapan değer, model varyansında meydana getirdiği yüzde artış değerini temsil etmektedir. Çizelgenin son değeri olan 35.56 değeri ise, 1000 uzunluğunda bir serinin merkezsiz pozisyonlarına yerleştirilen 5 adet sapan değer, model varyansında meydana getirdiği yüzde artış miktarını temsil etmektedir. Çizelge-1, ayrıntılı bir biçimde incelendiğinde seri uzunlukları büyüdüğünde, sapan olan gözlemin, model varyansında meydana getirdiği değişimin küçüldüğü görülebilir. Tam tersi bir durum olarak, seri içerisinde artan her bir sapan değer için model varyansında meydana gelen sapma bir kat artmaktadır.

**Çizelge 2.** Model varyansı için varyans analizi sonuçları.

Özellik	Sd	Hkt	Hko	F
Sapan değer sayıları	4	11235.0	2808.8	39.00
Seri uzunlukları	4	1148.7	287.2	3.99
Hata	16	1152.2	72.0	
Toplam	24	13536.0		

Çizelge-1'de elde edilen verilere iki yönlü varyans analizi uygulandığında, Çizelge-2 sonuçlarına ulaşılmıştır. Bu sonuçlara göre, sapan değer sayıları ve seri uzunlukları istatistiksel bakımdan önemli bulunmuştur. Dolayısıyla, seri içerisinde çok sayıda sapan değer bulunması, varyans değerinin artmasına ve parametre tahmin değerlerinin yanlış olmasına neden olmaktadır. Bununla birlikte seri uzunluklarının daha geniş olması, sapan değerlerin varyans değerleri üzerindeki etkisini küçültmektedir.

#### 5. Sonuç ve öneriler

Etkin parametre tahminlerine ulaşabilmek için, verilerin sağlıklı bir biçimde elde edilmesi gerekmektedir. Sağlıklı tahminlerin elde edilme koşullarından biri, örnek ölçümünün yeterli büyüklüklerde olmasıdır.

Özellikle gözlemler arası bağımlılığı esas alarak parametre tahmini yapmaya olanak tanıyan yöntemler için bazı tehditler yanında iki önemli tehdit daha bulunmaktadır, bunlar:

- Bağımlılığı bozan gözlemler varlığı (A tipi hata),
- Trendi bozan gözlemlerin varlığı (B tipi hata).

Bağımlılığı bozulması, model parametrelerinin ve parametre sayılarının yanlış olması, dolayısıyla modelin yanlış olması sonucunu doğurmaktadır. Bu bakımdan, daha analiz sürecine geçmeden, varsa bu türden hatalı durumları ortadan kaldırmak son derece önemlidir. Aksi halde tahmin aşaması ne kadar sağlıklı ve dikkatli yapılırsa yapılsın, parametrelerde bir miktar yanlışlığın ortaya çıkması kaçınılmaz hale gelmektedir. Yapılan araştırmalar, seriler üzerinde ortaya çıkan hatalı durumların, parametrelerin ve hatta yanlış etkiler çok güçlü ise, modellerin değiştiği sonucunu ortaya koymuştur [4].

Bu durumda tahmin işlemine esas teşkil eden gözlemlerin tarama süreçleri ile gözden geçirilmesi, varsa hatalı gözlemlerdeki yanlışlığın tespit edilmesi analizin sağlıklı bir biçimde tamamlanması bakımından hayati öneme haiz bir durumdur. Gözlemlerin gereğinden fazla küçük ya da büyük olmasına neden olan bu durumlar, adeta sağlıklı hücrelere bulaşmış virüslere, ya da bir bilgisayar programının sağlıklı bir biçimde çalışmasına engel olan ve virüs olarak adlandırılan bulaşmalarla çok ciddi benzerlikler göstermektedir. Özetle;

- A tipi hata, kişi ya da cihaz kaynaklı olup, şok bir durumu ifade etmektedir[6]. Ekonomik anlamda herhangi bir gün veya zamanda meydana gelen ve çok ciddi ekonomik sonuçların ortaya çıkmasına neden olan bir grev, deprem, doğal felaket ya da savaş durumunu ifade etmektedir[2].
- B tipi hata, veriler üzerinde meydana gelen doğal rasgelelik sonucunda ortaya çıkan, ortaya çıkış anından itibaren bütün gözlemlerin etkilenmesine neden olan bir durumu ifade etmektedir. Bu da ekonomik bir krizin ortaya çıktığı anı ve günü, bununla birlikte izleyen günlerde de etkisini azalan şiddette hissettirdiği durumu ifade etmektedir.
- A tipi hata, B tipi hataya oranla daha ciddi yanlışlıklara neden olmaktadır [5].
- B tipi hata, doğal bir durumu, A tipi hata, kişi kaynaklı bir hata durumunu ifade etmektedir.
- A tipi hata, veriler arası korelasyonu çok ciddi bir biçimde etkilemekte iken, B tipi hatada bu etki nispeten zayıftır. Çünkü, söz konusu etki çok sayıda gözleme bölünmekte ve toplam etki küçük olmaktadır.
- A ve B tipi etkilerin artan sayısı, varyans'ta dolayısıyla parametre tahminlerinde yanlışlığın artması anlamına gelmektedir.
- Serilerin geniş olması, hatalı durumların etkilerini azaltan bir faktördür.
- Bazı araştırmacılara göre, B tipi hata söz konusu olmuş ise, her modelde rasgelelikten doğan bir yanlışlık bulunmaktadır ve bu durumda etki giderilmemelidir. Bazı görüşlere göre de bu etkinin giderilmesi gerekir.
- Seri içerisinde sapan değer sayısı arttıkça tarama süreçleri ne kadar güçlü olursa olsun performanslarında bir zayıflama söz konusu olmaktadır [4].

## Kaynaklar

- [1] G.E.P. Box, G.M. Jenkins, (1976), *Time series analysis: Forecasting and control*, Sect 6.4.3 San Francisco: Holden-Day.
- [2] Ih Chang, G.C. Tiao, C. Chen, (1988), *Estimation of time series parameters in the presence of outliers*, American Statistical Association and the American Society for Quality Control.
- [3] A.J. Fox, (1972), Outliers in time series, *J.R. Statist. Soc. B*, 34, 350-363.

- [4] A. Kaya, (1999), "Zaman Serilerinde Sapan Değerlerin Analizi", Doktora Çalışması, Dokuz Eylül Üniversitesi. İzmir.
- [5] G.M. Ljung, G.E.P. Box, (1979), The likelihood function of stationary autoregressive-moving average models, *Biometrika*, 66, 265-270.
- [6] G.M. Ljung, (1993), On outlier detection in time series, *J.R. Statist. Soc. B*, 55: 559-567.
- [7] C.R. Muirhead, (1986), Distinguishing outlier types in time series, *J.R. Statist. Soc. B*, 48, No: 1, 39-47.
- [8] D. Pena, (1987), *Measuring the importance of outliers in ARIMA models in New perspectives in theoretical and applied statistics*, New york : Wiley.
- [9] M.B. Priestley, (1987), New Developments in Time-Series Analysis, In: *New perspectives in theoretical and applied statistics*, New york : Wiley.
- [10] M. Bayhan, (1992), Kalite kontrolünde zaman serisi analizi, *Endüstri Mühendisliği Dergisi*, 21:17-21.