ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ANKARA'NIN HAVA KİRLİLİĞİ: PERİODOGRAMLAR VE HARMONİK REGRESYON ANALİZİ

Yasin OKKAOĞLU

İSTATİSTİK ANABİLİM DALI

ANKARA
2018

Her hakkı saklıdır

TEZ ONAYI

Yasin OKKAOĞLU tarafından hazırlanan "**Ankara'nın Hava Kirliliği: Periodogramlar ve Harmonik Regresyon Analizi**" adlı tez çalışması 13/06/2018 tarihinde aşağıdaki jüri tarafından oy birliği ile Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Prof. Dr. Yılmaz AKDİ
Ankara Üniversitesi İstatistik Anabilim Dalı

Jüri Üyeleri:

Başkan: Prof. Dr. Erkan ERDİL
Orta Doğu Teknik Üniversitesi İktisat Anabilim Dalı

Üye : Prof. Dr. Yılmaz AKDİ
Ankara Üniversitesi İstatistik Anabilim Dalı

Üye : Doç. Dr. İhsan KARABULUT
Ankara Üniversitesi İstatistik Anabilim Dalı

Yukarıdaki sonucu onaylarım.

Prof. Dr. Atila YETİŞEMİYEN
Enstitü Müdürü

ETİK

Ankara Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim

13/06/2018



Yasin OKKAOĞLU

ÖZET

Yüksek Lisans Tezi

ANKARA'NIN HAVA KRİLİLİĞİ: PERİODOGRAMLAR VE HARMONİK REGRESYON ANALİZİ

Yasin OKKAOĞLU

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Yılmaz AKDİ

Çalışmada Ankara iline ait hava kirliliği indekslerinden PM10 miktarının zaman serileri analizi üzerinde durulmuştur. 1993-2017 yılları arasındaki aylık ortalama gözlem değerleri veri seti olarak göz önünde bulundurulmuştur. Gelecek dönemlerde gerçekleşecek hava kirliliği değerlerine ilişkin öngörüler bilinen zaman serileri yöntemlerine ek olarak harmonik regresyon yöntemi ile elde edilmiştir. Harmonik regresyon modeli elde edilirken verinin içerdiği periyodikliğin tespiti periodogramlar yardımıyla yapılmıştır. Durağanlık varsayımı, zaman serilerinde öngörü yapabilmek için gerekli temel varsayımlardan birisidir. Çalışmada serinin durağanlığı, verinin içerdiği periyodiklikten dolayı literatürde sıklıkla kullanılan genişletilmiş Dickey-Fuller birim kök testi ve Phillips-Perron birim kök testine ek olarak HEGY testi ve periodogram tabanlı birim kök testi ile sınanmıştır. HEGY testi sonuçlarına göre, modele kukla değişken eklenmediği durumda seri mevsimsel birim köklü çıkmıştır. Uygulanan diğer birim kök testlerine göre ise seri durağan çıkmıştır. Periodogram tabanlı birim kök testinin, HEGY testine göre kullanışlı olması ve avantajları nedeniyle bu testin sonucu göz önünde bulundurulmuştur. Her iki yöntemle de (ARIMA ve harmonik regresyon), öngörüler elde edilip çalışmanın devam ettiği 2018 yılının ilk iki ayı için gerçekleşen değerler ile karşılaştırılmıştır. Öngörü değerleri incelendiğinde, PM10 miktarının Ankara ili hava kirliliği açısından halen önemli bir tehlike oluşturduğu sonucuna ulaşılmıştır.

Haziran 2018, 110 sayfa

Anahtar Kelimeler: Zaman serileri, birim kök, durağanlık, periodogram, periyodiklik, hava kirliliği, harmonik regresyon, öngörü

ABSTRACT

Master Thesis

AIR POLLUTION OF ANKARA: PERIODOGRAMS AND HARMONIC REGRESSION ANALYSIS

Yasin OKKAOĞLU

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Prof. Dr. Yılmaz AKDİ

In this study, the amount of PM10, which is an air pollution indices, is analyzed with time series. The PM10 data is collected as monthly averages from 1993 to 2017. The forecasted values for future air pollution rates are obtained by the harmonic regression method in addition to the known time series methods. When the harmonic regression model is obtained, the periodicity of the data is determined by using periodograms. The assumption of stationarity is one of the basic assumptions needed to make forecasting in time series. In this study, since the series has a periodicities, the stationarity of series is tested by the HEGY test and periodogram based unit root test in addion to augmented Dickey-Fuller and Phillips-Perron unit root tests. According to the results of the HEGY test, when the dummy variables are not added into model, the series has seasonal unit roots. On the other hand, according to other unit root tests, the series is stationary. Because of the usefulness of the periodogram based unit root test and its advantages, the result of this test is considered. In both methods (ARIMA and harmonic regression), the forecasted values are obtained and made compare with the actual values for the first two months of 2018. When the forecasted values are examined, it is concluded that the amount of PM10 is still a significant danger in terms of air pollution in Ankara.

June 2018, 110 pages

Key Words: Time series, unit root, stationarity, periodogram, periodicities, air pollution, harmonic regression, forecasting

TEŞEKKÜR

Yüksek lisans öğrenimim boyunca ilgi ve alakasını eksik etmeyen, beni sabırla yönlendiren, zaman ayıran, bilgisini ve tecrübesini paylaşmaktan geri durmayan, saygıdeğer hocam, tez danışmanım Sayın Prof. Dr. Yılmaz AKDİ'ye (Ankara Üniversitesi İstatistik Anabilim Dalı) tezimde gösterdiği yoğun ilgi ve hoşgörüden dolayı sonsuz teşekkürlerimi sunarım.

Yüksek lisans tez savunmamda bana her türlü desteklerini veren ve gelecek dönemlerde yapılabilecek çalışmalar için bana yol gösteren hocalarım Sayın Prof. Dr. Erkan ERDİL (Ortadoğu Teknik Üniversitesi İktisat Bölümü) ve Sayın Doç. Dr. İhsan KARABULUT'a (Ankara Üniversitesi İstatistik Anabilim Dalı) teşekkürlerimi sunarım.

Gerek tez çalışmam, gerek ders dönemim, tüm yüksek lisans öğrenimim boyunca bana önemli fikirler veren üzerimde birçok emekleri olan tüm Ankara Üniversitesi İstatistik Bölümü öğretim üyesi hocalarıma çok teşekkür ederim.

Tez çalışmalarım sırasında bana karşı desteklerini her zaman hissettiğim, her zaman başarabileceğime inanan Sayın Arş. Gör. Mustafa Hilmi PEKALP (Ankara Üniversitesi İstatistik Anabilim Dalı) ve Sayın Öğr. Gör. Kamil Demirberk ÜNLÜ (Ankara Üniversitesi İstatistik Anabilim Dalı) hocalarıma teşekkür ederim.

Hayatım boyunca aldığım tüm kararlarda arkamda duran, sevgileri, bilgileri ve tecrübeleri ile yanımda olan, bana yol gösteren, haklarını asla ödeyemeyeceğim, abilerim Caner-İlker OKKAOĞLU, annem Aysel OKKAOĞLU ve babam Yusuf OKKAOĞLU'na, lisans ve yüksek lisans öğrenimim boyunca karşıma çıkan tüm engelleri aşabilmem için gerekli gücü yanımda olarak bana hissettiren, bana sonuna kadar güvenen ve yaşamımın her anını paylaşan eşim Büşra OKKAOĞLU'na en içten teşekkürlerimi ve şükranlarımı sunarım.

Yasin OKKAOĞLU
Ankara, Haziran 2018

İÇİNDEKİLER

TEZ ONAY SAYFASI
ETİK ..... **i**
ÖZET ..... **ii**
ABSTRACT ..... **iii**
TEŞEKKÜR ..... **iv**
SİMGELER DİZİNİ ..... **vii**
ŞEKİLLER DİZİNİ ..... **viii**
ÇİZELGELER DİZİNİ ..... **x**
1. GİRİŞ **1**
1.1 Lineer (Doğrusal) Regresyon **2**
1.2. Zaman Serileri **11**
1.3 Motivasyon **18**
2. DURAĞAN ZAMAN SERİSİ MODELLERİ **23**
2.1 Hareketli Ortalama (Moving Average, MA) Modelleri **24**
2.2 Otoregresif (Auto Regressive, AR) Modeller **27**
2.3 Otoregresif Hareketli Ortalama (Autoregressive Moving Average, ARMA) Modelleri **43**
2.4 Mevsimsel (Seasonal) Zaman Serisi Modelleri **46**
2.5 Öngörü **51**
3. DURAĞAN OLMAYAN ZAMAN SERİLERİ **59**
3.1 Birim Köklü Seriler **59**
3.2 Birim Kök Testleri **62**
3.2.1 Dickey-Fuller birim kök testi **63**
3.2.2 Genişletilmiş Dickey-Fuller (Augmented Dickey-Fuller, ADF) birim kök testi **67**
3.2.3 Phillips-Perron birim kök testi **68**
3.2.4 HEGY yöntemi **71**
4. UYGULAMA **77**
4.1 Periyodik Zaman Serileri Yaklaşımı **79**
4.2 PM_{10} Analizi **84**
4.2.1 Zaman boyutu (time domain) yaklaşımı **86**
4.2.2 Durağanlık **89**

4.2.3 Periyodiklik.....94
4.2.4 Harmonik regresyon modeli 97
5. SONUÇ VE TARTIŞMA.....104
KAYNAKLAR 107
ÖZGEÇMİŞ.....110

SİMGELER DİZİNİ

α	Anlamlılık Düzeyi
$\sim$	Dağılır
Φ	Kısmi Otokorelasyon
m^3	Metreküp
μm	Mikrometre
ρ	Otokorelasyon
γ	Otokovaryans
n	Örneklem Hacmi
$\bar{Y}_n$	Örneklem Ortalaması
χ_p^2	Serbestlik Derecesi p Olan Ki-Kare Dağılımı
σ^2	Varyans

Kısaltmalar

ACF	Otokorelasyon
ANOVA	Varyans Analizi
AR	Otoregresif
ARIMA	Otoregresif Bütünleşik Hareketli Ortalama
ARMA	Otoregresif Hareketli Ortalama
CO	Karbon Monoksit
MA	Hareketli Ortalama
MS	Kareler Ortalaması
NO	Nitrojen Oksit
NO_2	Nitrojen Dioksit
PACF	Kısmi Otokorelasyon
$PM_{2.5}$	Aerodinamik Çapı 2.5 μm'den Küçük Partikül Madde
PM_{10}	Aerodinamik Çapı 10 μm'den Küçük Partikül Madde
Pr	Olasılık
SAR	Mevsimsel Otoregresif
SARMA	Mevsimsel Otoregresif Hareketli Ortalama
SO_2	Sülfür Dioksit
SS	Kareler Toplamı

ŞEKİLLER DİZİNİ

Şekil 1.1 Et fiyatları ve enflasyon verisine ait saçılım ve kestirilen doğrunun grafiği 8

Şekil 1.2 Son on yıla ait aylık enflasyon oranlarının zaman serisi grafiği..... 16

Şekil 1.3 Son on yılın aylık enflasyon oranları verisine ait otokovaryans ve otokorelasyon grafikleri..... 17

Şekil 2.1 Rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri..... 26

Şekil 2.2 $\alpha_1 = 0.8$ için AR(1) serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri..... 34

Şekil 2.3 $\alpha_1 = 0.9$ için AR(1) serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri..... 34

Şekil 2.4 $\alpha_1 = 0.95$ için AR(1) serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri..... 34

Şekil 2.5 $\alpha_1 = 0.99$ için AR(1) serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri..... 35

Şekil 2.6 $\alpha_1 = 1$ için AR(1) serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri..... 35

Şekil 2.7 AR(2) modeline uygun olarak üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 39

Şekil 2.8 $\alpha_1 = 1.5$, $\alpha_2 = -0.7$ için AR(2) serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 40

Şekil 2.9 $\alpha_1 = 1.5$, $\alpha_2 = -0.6$ için AR(2) serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 40

Şekil 2.10 $\alpha_1 = 1.5$, $\alpha_2 = -0.55$ için AR(2) serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 41

Şekil 2.11 $\alpha_1 = 1.5$, $\alpha_2 = -0.51$ için AR(2) serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 41

Şekil 2.12 $\alpha_1 = 1.5$, $\alpha_2 = -0.5$ için AR(2) serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 41

Şekil 2.13 ARMA(1,2) modelline uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 45

Şekil 2.14 ARMA(2,1) modelline uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 46

Şekil 2.15 Farklı SAR modelline uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 50

Şekil 2.16 Rasgele olarak üretilen 100 birimlik zaman serisine ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri 55

Şekil 4.1 PM_{10} verisine ait zaman serisi grafiği.....84

Şekil 4.2 PM_{10} verisine ait otokorelasyon ve kısmi otokorelasyon grafikleri 85

Şekil 4.3 PM_{10} verisine ait periodogramların sıklıklara karşı grafiği 95

Şekil 4.4 PM_{10} değerleri ve (4.1)'de verilen modelden elde edilen kestirimler.....100

Şekil 4.5 PM_{10} değerleri ve harmonik modelden elde edilen kestirimler 100

Şekil 4.6 PM_{10} ortalamalarının karşılaştırılması 101

Şekil 4.7 Gerçekleşen PM_{10} değerleri ve kestirilen harmonik regresyon eğrisi 103

ÇİZELGELER DİZİNİ

Çizelge 1.1 2008:01-2016:12 dönemine ait aylık et fiyatları ve enflasyon oranları 6

Çizelge 1.2 Basit doğrusal regresyon denklemi parametrelerine ilişkin sayısal değerler 7

Çizelge 1.3 Kestirilen basit doğrusal regresyon modeline ait varyans analizi sonuçları 8

Çizelge 1.4 Son on yılın aylık enflasyon verisine ait otokovaryans ve otokorelasyon değerleri 16

Çizelge 2.1 Rasgele veriye önerilen modellere ait AIC ve SBC değerleri 56

Çizelge 2.2 Rasgele veriye önerilen $AR(2)$ modeline ait parametre tahminleri, standart hatalar ve p-değerleri 56

Çizelge 2.3 Ragele olarak üretilen veri için öngörü değerleri (101:110) 57

Çizelge 3.1 Rasgele üretilen veriye ilişkin ADF testi sonuçları 71

Çizelge 3.2 Rasgele üretilen veriye ilişkin Phillips-Perron testi sonuçları 71

Çizelge 4.1 PM_{10} miktarlarının aylara göre ortalamaları ve standart sapmaları 86

Çizelge 4.2 Önerilen modellere ilişkin AIC istatistiklerinin değerleri 86

Çizelge 4.3 $AR(13)$ modeline ilişkin PROC REG ile elde edilen parametre tahminleri 87

Çizelge 4.4 $AR(13)$ modeline ilişkin PROC ARIMA ile elde edilen parametre tahminleri 88

Çizelge 4.5 Eşitlik (4.1)'de verilen modele ilişkin PROC REG ile elde edilen parametre tahminleri 89

Çizelge 4.6 Eşitlik (4.1)'de verilen modele ilişkin PROC ARIMA ile elde edilen parametre tahminleri 89

Çizelge 4.7 PM_{10} verisine ilişkin Genişletilmiş Dickey-Fuller birim kök testi sonuçları 90

Çizelge 4.8 PM_{10} verisine ilişkin Phillips-Perron birim kök testi sonuçları 90

Çizelge 4.9 Modele kukla (dummy) değişken eklendiği durumda PM_{10} verisine ilişkin HEGY testi sonuçları 92

Çizelge 4.10 Modele kukla (dummy) değişken eklenmediği durumda PM_{10} verisine ilişkin HEGY testi sonuçları 93

Çizelge 4.11 Periodogram tabanlı birim kök testine ilişkin sonuçlar 94

Çizelge 4.12 PM_{10} verisine ilişkin en yüksek beş periodogram değeri..... 95

Çizelge 4.13 Eşitlik (4.2)'de verilen harmonik regresyon modeline ilişkin SAS PROC REG çıktıları..... 98

Çizelge 4.14 Eşitlik (4.3)'de verilen trigonometrik regresyon modeline ilişkin SAS PROC REG çıktıları..... 99

Çizelge 4.15 Elde edilen öngörü değerleri ve gerçekleşen değerler 102

1. GİRİŞ

Zaman serileri, bilimin her alanında uygulamaları bulunabilen istatistiğin bazen de iktisadın önemli uygulama alanlarından biridir. Bir zaman serisi, birim zaman aralıklarında yapılan gözlemlerin bir koleksiyonu olarak alınabilir. Bir ülkedeki yıllık ihracat ve ithalat miktarları, aylık enflasyon oranları, aylık faiz oranları, yıllık gayri safi milli hasıla (GSMH) gibi değişkenler zaman serilerine örnek olarak verilebilir. Ayrıca, bir ülkedeki yıllık ortalama buğday üretim miktarı, ortalama buğday fiyatları da zaman serilerine örnektir. Yıllık nüfus artış oranları, aylık ya da yıllık çocuk doğum-ölüm oranları, bununla birlikte, bir il ya da ülkedeki haftalık sıcaklık ölçümleri, yağış miktarları ile hava kirliliğine neden olabilecek gaz miktarı ölçümleri de zaman serilerine örnek olarak verilebilir.

Zaman serileri, zaman (time domain) ve sıklık (frequency domain) boyutu olmak üzere başlıca iki grupta toplanmaktadır. Genellikle, iktisat ve işletme gibi zaman serisi verileri için zaman boyutu, mühendislik verileri için ise sıklık boyutu teknikleri kullanılmaktadır.

Zaman serilerinde önemli amaçlardan biri, incelenen bir rasgele değişkenin gelecekte alacağı değerlerin öngörülmesidir. Örneğin, yıllık ihracat verileri ile ilgileniliyorsa, bir sonraki yılda yapılması beklenen ihracat miktarı bütçe yapıcıları için önemlidir. Zaman serilerinde kullanılan en önemli kavramlardan biri durağanlıktır. Öngörü yapabilmek için öncelikle serinin durağan olması gerekir. Eğer seri durağan değil ise, herhangi bir dönüşüm ile serinin durağanlığı sağlatıldıktan sonra istatistiki sonuç çıkarımlar yapılmalıdır. Birçok iktisadi seri durağan olmayan zaman serileridir (Akdi 2012).

Temel varsayımlarda farklılık göstermesine rağmen, zaman serileri regresyon modellerine benzer. Verilerin toplanış özelliklerine göre, zaman serileri hareketli ortalama (Moving Average, MA), otoregresif (Auroregressive, AR), otoregresif hareketli ortalama (ARMA), mevsimsel AR ve/veya MA modelleri kestirim ve öngörü modelleri olarak kullanılan başlıca zaman serisi modelleridir. MA türü seriler her zaman durağan olup, AR bileşeni içeren zaman serisi modelleri için durağanlık varsayımı

kontrol edilir. Durağan olmayan zaman serileri denildiği zaman "birim köklü" zaman serileri aklımıza gelir. Bu seriler de regresyon modellerine benzer. Bu nedenle, yine regresyon teknikleri kullanılarak parametre tahminleri yapılır. Ancak durağan olmayan zaman serilerinde tahmin edicilerin asimptotik dağılımları bilinen dağılımlardan farklı olduğu için istatistiki sonuç çıkarımlarda farklı kritik değerler kullanılır. İleriki bölümlerde bu konular ayrıntılı olarak incelenecektir. Ancak, zaman serilerinde de regresyon tekniklerinden faydalanıldığı için aşağıda ayrıntılarına girmeden regresyon kavramları kısaca özetlenecektir.

1.1 Lineer (Doğrusal) Regresyon

$(Y, X_1, X_2, \dots, X_p)$ rasgele değişkenlerinin ortak olasılık (veya olasılık yoğunluk) fonksiyonu $f(y, x_1, x_2, \dots, x_p)$ olmak üzere $X_1 = x_1, X_2 = x_2, \dots, X_p = x_p$ verildiğinde Y'nin koşullu olasılık (veya olasılık yoğunluk) fonksiyonu

$$f_{Y|X_1=x_1,X_2=x_2,\dots,X_p=x_p}(y|x_1, x_2, \dots, x_p) = \frac{f(y, x_1, x_2, \dots, x_p)}{f_{X_1,X_2,\dots,X_p}(x_1, x_2, \dots, x_p)}$$

olup $X_1 = x_1, X_2 = x_2, \dots, X_p = x_p$ verildiğinde Y'nin koşullu beklenen değeri

$$E(Y|X_1 = x_1, X_2 = x_2, \dots, X_p = x_p) = \int_{y \in D_Y} y f_{Y|X_1=x_1,X_2=x_2,\dots,X_p=x_p}(y|x_1, x_2, \dots, x_p)\, dy$$

dir. Burada D_Y, Y rasgele değişkeninin değer kümesini göstermektedir. Bu koşullu beklenen değer $x_1, x_2, \dots, x_p$'nin bir fonksiyonudur. Yani g, $\mathbb{R}^p$'den $\mathbb{R}$'ye bir fonksiyon olmak üzere ($g: \mathbb{R}^p \to \mathbb{R}$),

$$E(Y|X_1 = x_1, X_2 = x_2, \dots, X_p = x_p) = g(x_1, x_2, \dots, x_p)$$

dir. Bu koşullu beklenen değere Y'nin $X_1, X_2, \dots, X_p$ değişkenleri üzerine regresyonu denir. Bu regresyon eşitliği, e_t'ler hata terimlerini göstermek üzere,

$$Y_t = g(x_1, x_2, \dots, x_p) + e_t \, , t = 1,2, \dots, n$$

şeklinde ifade edilebilir. Burada x_i'ler açıklayıcı değişkenler olup Y ise açıklanan değişkendir. Eşitlikteki g fonksiyonu da ilişkinin yapısını gösteren ve bilinmeyen parametreler içeren bir fonksiyondur. Regresyonda amaç, g fonksiyonunu hataların kareleri minimum olacak şekilde tahmin etmektir. Burada, g fonksiyonunun yapısına göre regresyon eşitliği

i. g fonksiyonu $x_1, x_2, \dots, x_p$'lerin lineer bir fonksiyonu ise çoklu lineer regresyon
ii. g fonksiyonu $x_1, x_2, \dots, x_p$'lerin lineer olmayan bir fonksiyonu ise lineer olmayan regresyon

adını alır.

Burada, $x_{i,t}$'ler $(i = 1,2, \dots, p$ ve $t = 1,2, \dots, n)$ açıklayıcı değişkenleri, Y_t'ler $(t = 1,2, \dots, n)$ açıklanan değişkeni göstermek üzere,

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_p x_{p,t} + e_t \, , \; t = 1,2, \dots, n \qquad (1.1)$$

şeklinde verilen denklem bir doğrusal regresyon modelidir. Bu regresyon modeli, x_i'lerin rasgele olmayan bilinen değişkenler olması, hataların beklenen değerinin 0 $(E(e_t) = 0)$, varyanslarının sabit ve sonlu $(Var(e_t) = \sigma^2 < \infty)$, ilişkisiz $(k \neq s$ için $Cov(e_k, e_s))$ ve aynı dağılımlı olmaları durumunda doğrusal regresyon denklemi adını alır. Açıklayıcı değişken sayısının birden fazla olması durumunda denkleme çoklu doğrusal (ya da lineer) regresyon denklemi, tek bir açıklayıcı değişken olması durumunda ise denkleme basit doğrusal (ya da lineer) regresyon denklemi adı verilir. Burada amaç, modelde bulunan bilinmeyen β_i $(i = 0,1, \dots, p)$ ve σ^2 parametrelerinin tahmin edilmesi ve uygun modelin belirlenmesidir. Modelin belirlenmesi ve parametre tahmini için gerekli olmasa da istatistiki sonuç çıkarımı (hipotez testi, güven aralığı gibi) için normallik varsayımına ihtiyaç duyulur.

p açıklayıcı değişkenin bulunduğu eşitlik (1.1)'de verilen regresyon modeli

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\ X = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{p,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{p,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1,n} & x_{2,n} & \dots & x_{p,n} \end{bmatrix} \text{ ve } \underset{\sim}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

olmak üzere,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{p,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{p,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1,n} & x_{2,n} & \dots & x_{p,n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

şeklinde yazılabilir. Yani, p açıklayıcı değişkenin bulunduğu çoklu doğrusal regresyon denklemini $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{e}$ şeklinde bir lineer model gibi yazabiliriz. Burada amaç e_t hata terimlerinin karelerinin toplamı en küçük olacak şekilde $\underset{\sim}{\beta}$ parametre vektörünü tahmin etmektir. Elde edilen $X'X\underset{\sim}{\beta} = X'\underset{\sim}{Y}$ normal denklemlerinin çözümü ($X'X$ matrisi singüler değilse) $\underset{\sim}{\hat{\beta}} = (X'X)^{-1}X'\underset{\sim}{Y}$ olup bu çözüm $\underset{\sim}{\beta}$'nın en küçük kareler tahmin edicisidir. Ayrıca, $E(\underset{\sim}{Y}) = X\underset{\sim}{\beta}$ ve $n \times n$ boyutlu birim matris I olmak üzere, $Var(\underset{\sim}{Y}) = \sigma^2 I$ olup, $\underset{\sim}{\hat{\beta}}$ gözlemlerin lineer birleşimi, $\underset{\sim}{\hat{\beta}}$'nın bileşenleri de Y_t'lerin lineer birleşimidir. Bununla birlikte $E(\underset{\sim}{\hat{\beta}}) = \underset{\sim}{\beta}$, $Var(\underset{\sim}{\hat{\beta}}) = \sigma^2(X'X)^{-1}$ olup normallik varsayımı altında, $\underset{\sim}{Y} \sim N_P(X\underset{\sim}{\beta}, \sigma^2 I)$, $\underset{\sim}{\hat{\beta}} \sim N(\underset{\sim}{\hat{\beta}}, \sigma^2(X'X)^{-1})$'dir. Buradan da, $\underset{\sim}{\beta}$'nın bileşenleri hakkında istatistiki sonuç çıkarımlar yapılabilir. Diğer taraftan, $X'X$ matrisi singüler olmayan bir matris olduğunda $\underset{\sim}{\hat{\beta}}$ tahmin edicisi, $\underset{\sim}{\beta}$'nın en iyi lineer yansız tahmin edicisidir (BLUE, Best Linear Unbiased Estimator). Yani $\underset{\sim}{\beta}$ parametre vektörü için birçok lineer yansız tahmin edici bulunabilir. Bu şekilde elde edilen lineer yansız tahmin edicinin varyansı, diğer bütün lineer yansız tahmin edicilerin varyansından küçük ya da eşittir. Matrislerde sıralama bağıntısı tanımlı olmadığından, buradaki küçüklük "$\underset{\sim}{\tilde{\beta}}$, $\underset{\sim}{\beta}$'nın herhangi bir lineer yansız tahmin edicisi ise $[Var(\underset{\sim}{\tilde{\beta}}) - Var(\underset{\sim}{\hat{\beta}})]$ matrisi negatif olmayan tanımlıdır." anlamındadır.

Bir tane açıklayıcı değişkenin olduğu basit doğrusal regresyon modeli e_i'ler $(i = 1,2,\dots,n)$ beklenen değeri sıfır ve varyansı σ^2 olan bağımsız aynı dağılımlı rasgele değişkenler olmak üzere,

$$Y_i = \beta_0 + \beta_1 x_i + e_i\ ,\ \ i = 1,2,\dots,n$$

şeklinde verildiğinde β_0 ve β_1 parametrelerinin en küçük kareler tahmin edicileri

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{x}_n\ , \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$$

şeklindedir. Buradan, kestirim denklemi $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ şeklinde yazılır. Ayrıca,

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i\ , \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

olup σ^2'nin en küçük kareler tahmin edicisi de

$$\hat{\sigma}^2 = \frac{1}{n-(p+1)}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

dir. Burada $\hat{Y}_i$'ler $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ kestirim denklemi yardımıyla kestirilen değerlerdir,

Artık değerler (residuals) ise

$$\hat{e}_i = Y_i - \hat{Y}_i\ \ ,\ \ i = 1,2,\dots,n$$

ile hesaplanır. Kolayca görülebileceği gibi burada, $\sum_{i=1}^{n}\hat{e}_i = 0$, $\sum_{i=1}^{n}\hat{Y}_i = \sum_{i=1}^{n} Y_i$, $\sum_{i=1}^{n}\hat{e}_i x_i = 0$ ve $\sum_{i=1}^{n}\hat{e}_i \hat{y}_i = 0$ eşitliklerinin doğruluğu normal denklemlerin bir sonucudur.

Örneğin, 2008:01-2016:12 dönemine ait aylık et fiyatlarının ($Y_i, i = 1,2, \dots, 108$), aylık enflasyon oranları ($x_i, i = 1,2, \dots, 108$) üzerine basit doğrusal regresyonunu göz önüne alalım. Burada, 2008:01-2016:12 dönemine ait aylık et fiyatları ve enflasyon miktarı verileri çizelge 1.1'de verilmiş olup, analizlerde verilerin logaritmaları kullanılmıştır. Regresyon varsayımları bir an için göz ardı edildiğinde bu regresyon eşitliği,

$$Y_i = \beta_0 + \beta_1 x_i + e_i\ ,\ \ i = 1,2, \dots, 108 \tag{1.2}$$

şeklinde verilir. Burada e_i'ler birbirinden bağımsız aynı dağılımlı rasgele değişkenler ve x_i'ler bilinen rasgele olmayan değişkenler olmalıdırlar. Ancak burada açıklayıcı değişken rolündeki x_i'ler de bağımlı rasgele değişkenlerdir.

Çizelge 1.1 2008:01-2016:12 dönemine ait aylık et fiyatları ve enflasyon oranları

Aylara Göre Et Fiyatları (Dana karkas kilogram fiyatı, Türk Lirası)

	2008	2009	2010	2011	2012	2013	2014	2015	2016
Ocak	11.03	11.88	16.84	19.14	18.27	16.00	15.88	18.93	24.17
Şubat	11.37	11.80	17.92	18.60	18.08	15.90	16.06	19.39	24.18
Mart	11.34	11.88	18.28	18.29	17.95	15.91	16.13	19.78	24.35
Nisan	11.40	12.22	19.22	18.16	17.72	15.88	16.26	20.10	24.56
Mayıs	11.50	12.06	18.45	18.16	17.52	15.72	16.47	20.50	24.68
Haziran	11.90	12.70	17.45	18.06	17.43	15.56	16.82	20.91	24.94
Temmuz	11.85	13.15	17.68	18.28	17.49	15.59	17.29	21.48	25.28
Ağustos	12.17	13.81	18.35	18.28	17.3	15.76	17.54	22.16	25.51
Eylül	12.11	14.59	18.70	18.73	17.27	15.87	18.04	22.51	25.7
Ekim	12.08	15.28	19.38	19.03	17.21	15.90	18.07	22.57	25.67
Kasım	11.98	15.65	19.34	18.87	17.08	15.92	18.23	22.75	25.65
Aralık	11.92	15.83	19.15	18.77	16.76	15.95	18.40	22.64	25.72

Aylık Enflasyon Oranları (2007:12=100)

	2008	2009	2010	2011	2012	2013	2014	2015	2016
Ocak	100.8	109.96	117.89	122.72	132.94	140.05	147.57	154.6	163.81
Şubat	102.09	109.62	119.34	123.45	133.50	140.35	148.00	155.31	163.79
Mart	103.05	110.72	119.92	123.87	133.91	141.01	149.13	156.50	163.75
Nisan	104.73	110.74	120.52	124.74	135.43	141.43	150.47	158.13	164.53
Mayıs	106.22	111.38	120.16	127.16	135.22	141.58	150.87	158.69	165.11
Haziran	105.86	111.49	119.60	125.73	134.32	142.34	151.18	158.18	165.58
Temmuz	106.44	111.74	119.12	125.32	134.09	142.65	151.63	158.27	166.74
Ağustos	106.20	111.44	119.52	126.05	134.65	142.55	151.72	158.67	166.45
Eylül	106.65	111.83	120.75	126.80	135.68	143.32	151.86	159.56	166.63
Ekim	109.25	114.24	122.58	130.07	137.64	145.12	153.76	161.11	168.07
Kasım	110.08	115.51	122.61	131.80	138.02	145.13	153.94	161.78	168.59
Aralık	109.67	116.04	122.31	132.38	138.40	145.59	153.50	161.99	170.23

Eşitlik (1.2)'de verilen basit doğrusal regresyon modeli için β_0 ve β_1'in en küçük kareler tahminleri, bu tahminlerin standart hataları ve test istatistiklerinin değerleri ile bunlara karşılık gelen p-değerleri hesaplanarak, çizelge 1.2'de verilmiştir.

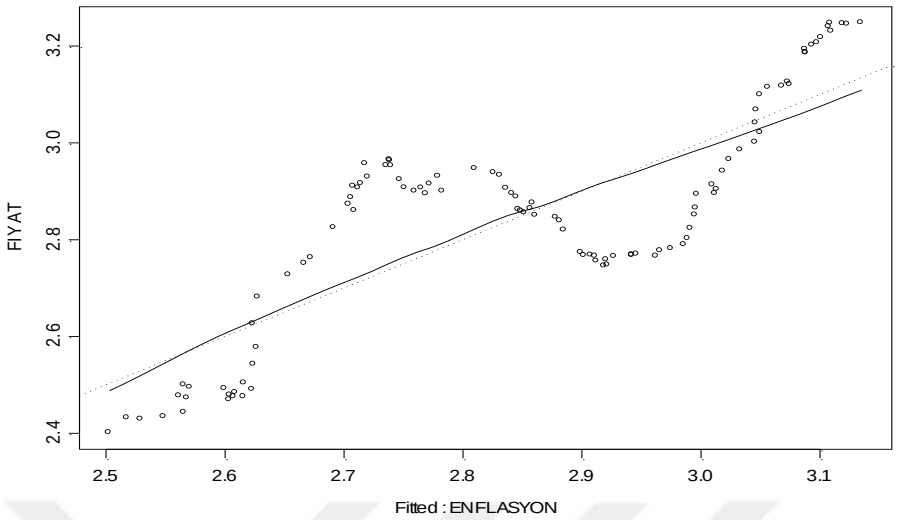
Çizelge 1.2 Basit doğrusal regresyon denklemi parametrelerine ilişkin sayısal değerler

	Tahmin	**Standart Hata**	**t-istatistiği değeri**	**Pr** $> \|t\|$
$\hat{\beta}_0$	-3.0623	0.4120	-7.4334	<0.0001
$\hat{\beta}_1$	1.2064	0.0841	14.3482	<0.0001

Buna göre kestirim denklemi de

$$\hat{Y}_i = -3.0623 + 1.2064x_i \ , \ \ i = 1,2, \dots, 108$$

şeklinde yazılır. Çizelge 1.1 incelendiğinde iki parametre için de $\Pr(> |t|)$ değerleri 0.05 değerinden küçük olduğu için parametrelerin anlamsız olduğu yokluk hipotezleri $(H_0: \beta_i = 0, i = 1,2)$, $H_0: \beta_i \neq 0, i = 1,2$ alternatiflerine karşı $\alpha = 0.05$ anlam düzeyinde reddedilir. Yani, parametreler anlamlıdır. Başka bir ifade ile enflasyon oranlarındaki değişim, et fiyatlarını anlamlı bir şekilde etkilemektedir. Kestirilen regresyon doğrusu ve verinin saçılım grafiği (scatter plot) şekil 1.1'de verilmiştir.



Şekil 1.1 Et fiyatları ve enflasyon verisine ait saçılım ve kestirilen doğrunun grafiği

Basit doğrusal regresyon modeline ait varyans analizi (ANOVA) tablosu da çizelge 1.3'de verilmiştir.

Çizelge 1.3 Kestirilen basit doğrusal regresyon modeline ait varyans analizi sonuçları

Kaynak	**sd**	**SS**	**MS**	**F**	$\boldsymbol{Pr > F}$
Regresyon	1	3.45094	3.45094	205.87	<0.0001
Hata	106	1.77685	0.01676		
Toplam	107	5.22780			

Burada F istatistiğine ait p-değeri 0.05 değerinden küçük olduğundan modelin anlamsız olduğu yokluk hipotezi H_0:"$Model\ anlamlıdır.$" alternatif hipotezine karşı $\alpha = 0.05$ anlam düzeyinde reddedilir. Yani, kurulan basit doğrusal regresyon denklemi anlamlıdır. Ayrıca, $R^2 = SSR/SST = 0.66$ olup enflasyonun oranlarının et fiyatlarını açıklama oranı yaklaşık olarak %66 dır. Başka bir ifade ile et fiyatlarındaki değişimin yaklaşık %66 sı enflasyon oranları tarafından açıklanmaktadır.

Çizelge 1.3'de oluşturulan ANOVA tablosundaki değerler ve parametrelerin en küçük kareler tahmin edicilerinin değerleri kullanılarak, model parametreleri hakkında istatistiki sonuç çıkarımlar yapılabilir. Örneğin β_1 parametresi için %95'lik güven aralığı

$$\hat{\beta}_1 \pm s(\hat{\beta}_1) t\left(\frac{\alpha}{2}, n-2\right)$$

formülü ile verilir. Hesaplanan değerler yerine konulduğunda β_1 için %95'lik güven aralığı $(1.0400, 1.3731)$ olarak elde edilmiştir. Benzer şekilde β_0 parametresi için de %95'lik güven aralığı da $(-3.8791, -2.2454)$ olarak hesaplanmıştır.

Diğer taraftan parametrelere ilişkin hipotez testleri de yapılabilir. Örneğin, $H_0: \beta_1 = 1.2$ yokluk hipotezi $H_a: \beta_1 \neq 1.2$ alternatif hipotezine karşı $\alpha = 0.05$ anlam düzeyinde test edilmek istensin. Bu durumda,

$$t_h = \frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} = \frac{1.2064 - 1.2}{0.0841} = 0.0761$$

olup, $|t_h| = 0.0761 < t(\alpha/2, n-2) = t(0.025, 106) = 1.9826$ olduğundan H_0 hipotezi $\alpha = 0.05$ anlam düzeyinde reddedilemez. Ayrıca, serbestlik derecesi 106 olan t dağılımı için p-değeri,

$$p - değeri = P(|T| > t_h) = P(|T| > 0.0761) = 2P(T > 0.0761) = 0.9394$$

olarak hesaplanmıştır. Görüldüğü gibi p-değeri oldukça yüksektir. Yani, β_1 parametresinin 1.2 değerine eşit olduğu, $\alpha = 0.05$ anlam düzeyinde söylenebilir. Oysa $H_0: \beta_1 = 1$ yokluk hipotezi $H_a: \beta_1 \neq 1$ alternatifine karşı $\alpha = 0.05$ anlam düzeyinde red edilir ($t_h = -2.3781 > t(0.025, 106) = 1.9826$).

Verilen örnekte regresyon varsayımları dikkate alınmamıştır. Regresyonda açıklayıcı değişkenler bilinen sabit değerlerdir. Oysa verilen örnekte bu varsayım sağlanmamaktadır. Bunun nedeni, aylık enflasyon oranlarının bilinen sabit değerler değil, önceki aylarda gerçekleşen enflasyon oranlarına bağlı rasgele değişkenler olmasıdır. Burada, $X_i, i = 1,2, \dots, 108$ değişkenleri bağımlı rasgele değişkenler olduğundan, birbirinden bağımsız olması gereken açıklanan $Y_i, i = 1,2, \dots, 108$ değişkeninin de bağımsızlığından bahsedilemez. Bununla birlikte, sağlanması gereken diğer bir varsayım da hata terimlerinin ilişkisiz olmasıdır. Bu varsayımın kontrolü için testi de Durbin-Watson yöntemi kullanılabilir. Hataların birinci dereceden otokorelasyona sahip olup olmadığını test etmek için Durbin-Watson test istatistiğinin değeri

$$d = \left[\sum_{i=1}^{n} \hat{e}_i^2\right]^{-1} \sum_{i=2}^{n} (\hat{e}_i - \hat{e}_{i-1})^2 = 0.027$$

olarak hesaplanmıştır. Eğer hata terimleri ilişkisiz ise d istatistiğinin değeri 2 civarında olmalıdır (Akdi 2012). Bu değer $\alpha = 0.05$ anlam düzeyi ve 100 gözlemin bulunduğu durumdaki alt sınır değer olan 1.65 (Kutner vd. 2005) değerinden küçük olduğundan hata terimlerinin birinci dereceden otekorelasyonlu olduğu söylenebilir. Ayrıca, istatistiki sonuç çıkarımı için gerekli olan hataların normal dağılıma sahip olması varsayımı Kolmogorov-Smirnov testi ile sınanmış ve test istatistiğine ait p-değeri 0.0013 olarak elde edilmiştir. Buna göre hata terimlerinin normal dağılıma sahip olduğu yokluk hipotezi de $\alpha = 0.05$ anlam düzeyinde red edilmektedir.

Görüldüğü gibi et fiyatları ile enflasyon arasında anlamlı lineer bir ilişki gözükmesine rağmen regresyon varsayımlarının neredeyse tamamı sağlanmamaktadır. Bunun en önemli nedenlerinden biri, açıklayıcı değişken rolündeki değişkenin kendi içinde ilişkili rasgele değişkenler olmasıdır. Ayrıca gözlemler birim zaman aralıklarında yapıldığı için her iki değişken (et fiyatları, enflasyon oranları) de birer zaman serisidir. Bu nedenle, yine de bazı regresyon tekniklerinin yanında zaman serileri tekniklerinin kullanılması daha uygun olacaktır.

1.2. Zaman Serileri

Zaman serisi en basit anlamda birim zaman aralıklarında yapılan gözlemlerin bir koleksiyonudur. Yani, bir zaman serisi rasgele değişkenlerin bir koleksiyonudur. Bu nedenle zaman serisini tanımlayabilmek için öncelikle olasılık uzayının tanımına ihtiyaç duyulur. Aşağıda zaman serisinin tanımını vermeden önce olasılık uzayları ile ilgili bir dizi tanım verilmiştir. Bu kavramların ayrıntılarına burada girilmeyecektir. Bu kavramlar ile ilgili ayrıntılı bilgiler herhangi bir olasılık teorisi kitabında bulunabilir (Leadbetter vd. 2014).

Tanım 1.1 Ω boş olmayan bir küme (genellikle bir deney gerçekleştirildiğindeki tüm sonuçların kümesi, örnek uzay) ve $\mathcal{U}$ da Ω'nın bazı alt kümelerinden oluşan bir sınıf olsun. $\mathcal{U}$ sınıfı,

i. $\Omega \in \mathcal{U}$
ii. A^c, A'nın tümleyenini göstermek üzere, her $A \in \mathcal{U}$ için $A^c \in \mathcal{U}$
iii. Her $A_n \in \mathcal{U}, n = 1,2,3, \ldots$ için $\bigcup_{n=1}^{\infty} A_n \in \mathcal{U}$

özelliklerini sağlıyor ise $\mathcal{U}$ sınıfına Ω üzerinde bir sigma cebir, $(\Omega, \mathcal{U})$ ikilisine bir ölçülebilir uzay ve $\mathcal{U}$'nun her bir elemanına da bir olay denir (Akdi 2014).

Tanım 1.2 $(\Omega, \mathcal{U})$ ölçülebilir bir uzay olmak üzere (Ω örnek uzay, $\mathcal{U}$ da Ω örnek uzayı üzerinde bir sigma cebir), $\mathcal{U}$ üzerinde tanımlanan

$$P: \mathcal{U} \to [0,1]$$
$$A \to P(A)$$

P küme fonksiyonu,

i. Her $A \in \mathcal{U}$ için $P(A) \geq 0$
ii. $P(\Omega) = 1$

iii. Her $A_n \in \mathcal{U}$ ve A_n'ler ayrık $(A_i \cap A_j = \emptyset, i \neq j)$ olayların bir dizisi olmak üzere,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

özelliklerini sağlıyor ise P'ye bir olasılık ölçüsü denir. $(\Omega, \mathcal{U}, P)$ üçlüsüne olasılık uzayı, $P(A)$ sayısına da A olayının olasılığı denir (Akdi 2014).

Tanım 1.3 $(\Omega, \mathcal{U}, P)$ bir olasılık uzayı olmak üzere,

$$Y: \Omega \to \mathbb{R}$$
$$\omega \to Y(\omega)$$

şeklinde tanımlanan Y fonksiyonu, her $x \in \mathbb{R}$ için $\{\omega: Y(\omega) \leq x\} \in \mathcal{U}$ özelliğine sahip ise Y fonksiyonuna bir rasgele değişken denir (Akdi 2014).

Tanım 1.4 $(\Omega, \mathcal{U}, P)$ bir olasılık uzayı ve T bir indis kümesi olmak üzere $\Omega \times T$'den reel sayılara tanımlanan

$$Y(.,.): \Omega \times T \to \mathbb{R}$$

$$(\omega, t) \to Y(\omega, t)$$

Y fonksiyonuna bir zaman serisi denir. Burada, T indis kümesi aksi ifade edilmediği sürece doğal sayılar kümesi olarak alınacaktır (Akdi 2012). Bir zaman serisi $\{Y_t, t \in T\}$ veya kısaca Y_t ile gösterilecektir.

Zaman serilerinde öngörü yapılırken ve öngörü yapılabilmesi için gereken varsayımlar kontrol edilirken istatistiksel yöntemlerin kullanılması nedeniyle zaman serileri istatistiğin bir ilgi alanıdır. Zaman serileri analizinde öngörü gibi istatistiksel sonuç

çıkarımların yapılabilmesi için durağanlık varsayımı önemlidir. İyi bir öngörü yapmanın zaman serileri analizindeki önemi büyüktür. Örneğin, iyi bir öngörü ile geçmiş zamanlardaki hava kirliliği değerleri yardımıyla gelecekte gerçekleşecek hava kirliliği değerleri öngörülerek, çevrenin ve insan sağlığının korunması açısından gerekli önlemler alınabilir.

Durağanlık kavramının zaman serilerinde önemli bir yer tutmasının nedeni, zaman serilerindeki neredeyse tüm istatistiki sonuç çıkarımlarda serinin durağanlık varsayımına ihtiyaç duyulmasıdır. Güçlü ve zayıf durağanlık olmak üzere iki çeşit durağanlıktan bahsedilir. Bunlardan güçlü durağanlık koşullarının sınanmasının zorluğundan dolayı uygulamada zayıf durağanlıkla ilgilenilir. Aşağıda ayrıntısına girmeden güçlü durağanlık kavramı açıklanmaya çalışılmıştır.

Tanım 1.5 (güçlü durağanlık) T bir indis kümesi ve $\{Y_t, t \in T\}$ bir zaman serisi olmak üzere, her $n, h \in \mathbb{N}$, her $t_1, t_2, \dots, t_n \in T$, her $t_1 + h,\ t_2 + h, \dots,\ t_n + h \in T$ ve her $y_1, y_2, \dots, y_n \in T$ olmak üzere,

$$F_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}}(y_1, y_2, \dots, y_n) = F_{Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h}}(y_1, y_2, \dots, y_n)$$

koşulu sağlanıyorsa Y_t zaman serisi güçlü durağandır denir (Akdi 2012).

Güçlü durağanlığın sağlanabilmesi için zaman serisi içerisindeki rasgele değişkenlerinin zaman boyutunda gerçekleştirilecek tüm ötelenmeleriyle elde edilecek her boyuttan dağılımlarının aynı olması gerekmektedir. Bu da hem sağlatılması hem de sağlanıp sağlanmadığına dair kontrolü zor bir varsayımdır. Bu nedenle, uygulamada aşağıda tanımı verilen zayıf durağanlık kavramı pratikliği açısından ön plana çıkmaktadır.

Tanım 1.6 (zayıf durağanlık) T bir indis kümesi ve $\{Y_t, t \in T\}$ bir zaman serisi olmak üzere,

i. $E(Y_t) = \mu$ (serinin beklenen değeri zamanın sabit bir fonksiyonu)

ii. $Cov(Y_t, Y_s) = \gamma(t-s)$ (seriyi oluşturan rasgele değişkenler arasındaki kovaryans sadece rasgele değişkenlerin ölçüldüğü zaman farkının bir fonksiyonu)

koşulları sağlanıyorsa Y_t zaman serisi zayıf durağan ya da kısaca durağandır denir (Akdi 2012).

Burada, $Cov(Y_t, Y_s)$ ile verilen ifade Y_t ve Y_s arasındaki otokovaryansı gösterir. Tanımda $s = t + h$ yazıldığında (ii) koşulu "$\gamma(h) = Cov(Y_t, Y_{t+h})$ kovaryansı sadece h'nin bir fonksiyonudur." şeklinde düzenlenebilir. Yani,

$$\gamma(h) = Cov(Y_t, Y_{t+h}) = \mathrm{E}[(Y_t - \mu)(Y_{t+h} - \mu)]$$

dir. Bu fonksiyona $\{Y_t, t \in T\}$ zaman serisinin otokovaryans fonksiyonu denir. Serinin otokovaryans fonksiyonu uygulamada veriye uygun modelin belirlenmesi ve serinin durağanlığı ile ilgili sezgisel sonuç çıkarımlar yapabilmek için kullanışlı özelliklere sahiptir. Serinin otokorelasyon fonksiyonu da otokovaryans fonksiyonunun yardımı ile

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{Cov(Y_t, Y_{t+h})}{\sqrt{Var(Y_t)Var(Y_{t+h})}}$$

şeklinde ifade edilir.

$\{Y_t, t \in T\}$ zaman serisi durağan ise serinin varyansı

$$\gamma(0) = Cov(Y_t, Y_t) = Var(Y_t) = Var(Y_{t+h})$$

dir.

Bir zaman serisine ait $Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n$ gözlem değerleri verildiğinde bu zaman serisine ait otokovaryanslar

$$\hat{\gamma}(h) = \frac{1}{n-h}\sum_{t=1}^{n-h}(Y_t - \bar{Y}_n)(Y_{t+h} - \bar{Y}_n)$$

şeklinde tahmin edilir. Bazı istatistiksel yazılımlar otokovaryans tahminlerinin hesabı için,

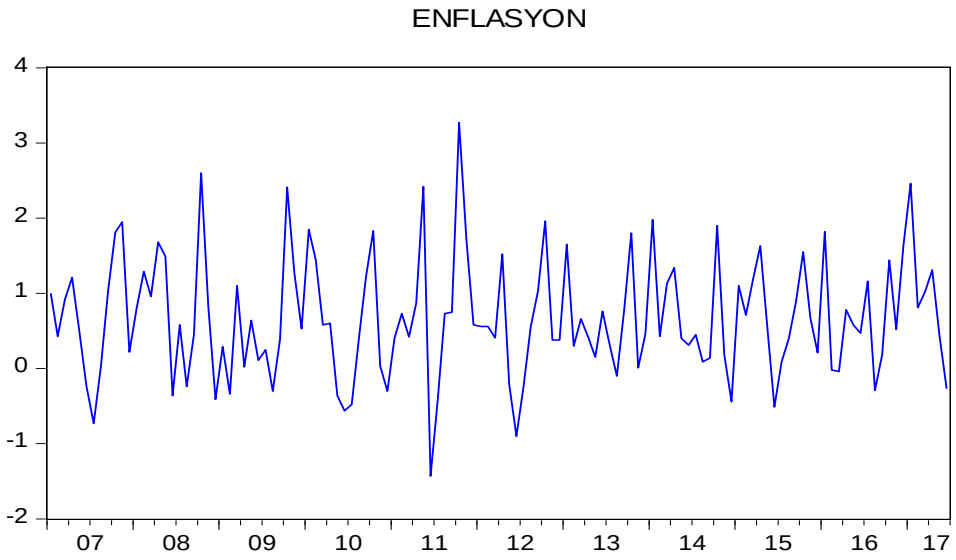
$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(Y_t - \bar{Y}_n)(Y_{t+h} - \bar{Y}_n)$$

formülünü kullanmaktadır (SAS gibi). Hesaplanan bu otokovaryans tahminlerinin yardımı ile otokorelasyonların tahminleri de

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

şeklinde hesaplanır.

Son on yıla ait (2007:01-2017:06) aylık enflasyon oranları bir zaman serisi olarak göz önüne alınabilir. Bu zaman serisinin grafiği şekil 1.2'de verilmiştir.



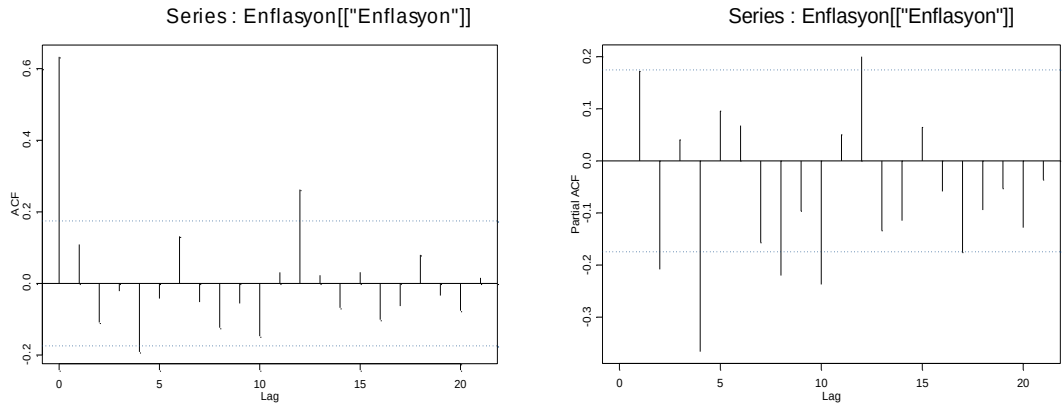
Şekil 1.2 Son on yıla ait aylık enflasyon oranlarının zaman serisi grafiği

Bu zaman serisinin otokovaryans ve otokorelasyonlarından ilk beş tanesi çizelge 1.4'de verilmiştir.

Çizelge 1.4 Son on yılın aylık enflasyon verisine ait otokovaryans ve otokorelasyon değerleri

h	0	1	2	3	4	5
$\hat{\gamma}(h)$	0.633	0.109	-0.109	-0.022	-0.192	-0.042
$\hat{\rho}(h)$	1	0.172	-0.173	-0.034	-0.303	-0.067

Serinin diğer otokovaryansları ile otokorelasyonları hesaplanarak grafikleri şekil 1.3'de verilmiştir.



(a) (b)

Şekil 1.3 Son on yılın aylık enflasyon oranları verisine ait otokovaryans ve otokorelasyon grafikleri

a. Otokovaryans grafiği, b. Otokorelasyon grafiği

$\{Y_t, t \in T\}$, otokovaryans fonksiyonu $\gamma(h)$ ve otokorelasyon fonksiyonu da $\rho(h)$ olan durağan bir zaman serisi olsun. Bu durumda serinin otokovaryans fonksiyonu

i. $\gamma(h)$ simetriktir ($\gamma(h) = \gamma(-h)$).
ii. Her $h \in \mathbb{N}$ için $|\gamma(h)| \leq \gamma(0)$'dır.
iii. $\gamma(h)$ negatif olmayan tanımlıdır.

özelliklerini sağlar. Ayrıca, serinin otokorelasyon fonksiyonunun da

i. $\rho(h)$ simetriktir $\big(\rho(h) = \rho(-h)\big)$
ii. Her $h \in \mathbb{N}$ için $|\rho(h)| \leq 1$ dir.
iii. $\rho(h)$ negatif olmayan tanımlıdır.

özelliklerini sağladığı açıktır (Brockwell ve Davis 1987).

1.3 Motivasyon

Zaman serileri, bir konu ile ilgili geçmiş ve şimdiki dönemde gerçekleşen gözlemleri kullanarak gelecek dönemler için öngörü yapmamıza olanak sağlayan, istatistiğin bazen de ekonometrinin alt bir alanıdır. Zaman serilerinin iktisadi bilimlerde kullanımı çok yaygın olsa da geleceğe dair planlama yapılması gereken tüm alanlarda kullanımı uygundur. Örneğin, aylık enflasyon miktarlarının bir zaman serisi olarak düşünülerek gelecek dönemlere ait öngörülerin tutarlı bir şekilde yapılması ilgililerin yüksek enflasyon karşısında çeşitli önlemler almasına olanak sağlar. Benzer şekilde tıp alanında, bir hastaneye belirli bir zaman aralığında (günlük, haftalık, aylık, yıllık gibi) gelen hasta sayılarının öngörülmesi hastanenin ihtiyaç duyduğu malzemelerin tahsisi açısından önemlidir. Yine, yıllık konut satışlarının bir zaman serisi olarak göz önüne alınıp gelecek dönemlere ait öngörülerin yapılması, bu alanda yapılacak yatırımlarda yetkililer için yol gösterici olur.

Zaman serilerinde öngörü yapabilmek için öncelikle geçmiş dönemlerde gerçekleşen veriler modellenir. Zaman serisi modelleri regresyon modelleri ile benzerlik göstermesine rağmen temel varsayımlarda farklılık göstermektedir. Öngörü için en çok kullanılan modeller AR (otoregresif) bileşeni içeren zaman serisi modelleridir. Verinin herhangi bir periyodik yapı içermesi durumunda ise mevsimsel zaman serisi modelleri olarak bilinen SAR (mevsimsel otoregresif) ya da SARMA (mevsimsel otoregresif hareketli ortalama) modelleri sıklıkla kullanılır.

İçerisinde periyodiklik bulunduran ve birçok çalışmaya konu olmuş en tipik mevsimsel zaman serisi örneği meteorolojik ölçümlerdir. Aynı meteorolojik ölçümlerde olduğu gibi hava kalitesi ölçümlerinde de periyodikliğe rastlanmaktadır. Bunun temel nedeni kış aylarında karbon içerikli yakıtların kullanımının artmasıdır.

Hava kirliliği, havada bulunması gereken maddelerin sınır değerlerini aşması, bulunmaması gereken maddelerin canlıları etkileyecek oranda bulunması olarak tanımlanabilir. İnsan sağlığını olumsuz yönde etkileyen hava kirliliğine neden olan birçok hava kirleticisi bulunmaktadır. Bu kirleticilerden başlıcaları PM_{10}, $PM_{2.5}$, NO

(nitrojen oksit), NO_2 (nitrojen dioksit), SO_2 (kükürt dioksit) ve CO (karbon monoksit) dir. Bu kirleticilerden PM_{10} aerodinamik çapı 10 μm'den küçük olup havada serbest şekilde bulunan partikül maddeyi belirtir. Özellikle insan sağlığını etkileyenler bu boyutta olanlardır (Cavkaytar vd. 2013). İçeriğinde doğal kaynaklı gazlar bulunabileceği gibi, bulunan çoğu gaz bileşeni insan kaynaklı sebepler ile ortaya çıkmaktadır. Doğal kaynaklı gazlardan bazıları, volkanik gazlar, deniz suyu buharı, toz fırtınaları ve orman yangınları nedeniyle ortaya çıkan gazlar olduğu gibi PM_{10}'un insan kaynaklı bileşenleri de karbon içerikli yakıtların kullanımı nedeniyle salınan gazlar, sigara dumanı ve çeşitli endüstriyel atıklar nedeniyle meydana çıkan gazlardır.

Bu çalışmada Ankara iline ait önemli hava kirleticilerden olan PM_{10} kirleticisinin modellenmesi ve gelecek dönemlere ait öngörülerinin yapılması üzerinde durulmuştur.

Hava kirliliği, insan sağlığını olumsuz yönde etkileyen önemli faktörlerden birisi olması nedeniyle gerek ülkemizde gerekse yurt dışında sürekli olarak üzerinde çalışılan ve literatürde birçok çalışmanın bulunduğu bir alan olmuştur. Bu çalışmalardan bazıları aşağıda kısaca özetlenmiştir.

Pope vd. (1992), Utah Valley'de PM_{10} kirleticisinin semptomatik ve asemptomatik çocuklara akut sağlık etkilerini araştırmışlardır. İki farklı yaklaşım ile lojistik regresyon kullandıkları çalışmada PM_{10} ve solunum yolları semptomları arasında pozitif bir ilişki tespit etmişlerdir.

Coachella Valley'de PM_{10} miktarları ve günlük ölüm oranları arasındaki ilişki araştırılmış (Ostro vd. 1999) olup Poisson regresyonu kullanılarak yapılan analizler sonucunda değişkenler arasında tutarlı bir ilişki olduğu sonucuna ulaşılmıştır.

Berktaş ve Bircan (2003), Atatürk Göğüs Hastalıkları ve Göğüs Cerrahisi Eğitim ve Araştırma Hastanesi acil servisine astım nedeniyle başvuran kişi sayısı ile PM_{10}, SO_2 miktarları ve meteorolojik değişkenler arasındaki ilişkiyi Pearson ve Spearman korelasyon ve Ki-kare testlerini kullanarak incelemişlerdir. Çalışmada acil servise astım

sebebiyle başvuran kişi sayısı ile SO_2 miktarı arasında istatistiksel olarak anlamlı ve düşük, PM_{10} ile hasta sayısı arasında anlamlı ve daha güçlü bir ilişki tespit edilmiştir.

Çiçek vd. (2004), ise Ankara Sıhhiye'den ölçümü yapılan SO_2, PM_{10}, NO, NO_2 ve CO olmak üzere beş hava kirleticisinin, bazı iklim değişkenleri üzerine çoklu regresyon ve stepwise (adımsal) regresyon yöntemlerini kullanarak hava kirliliği ile iklim koşulları arasında düşük ve orta düzeyde ilişki tespit etmişilerdir.

Koutrakis vd. (2005), Santiago, Chile' de bulunan dört farklı sit alanına ait PM_{10} ve $PM_{2.5}$ miktarları ile kategorik verilere çevrilen meteorolojik veriler arasındaki muhtemel bir ilişkinin tespiti için karma regresyon yöntemini kullanmışlardır. Meteorolojik değişkenler ve partikül madde miktarları arasında anlamlı ilişkiler tespit edilmiş olup Pazar günleri trafiğin ve hava kirliliğine yol açacak aktivitelerin azalması nedeniyle partikül madde miktarlarının azaldığı sonucuna ulaşmışlardır.

Genç vd. (2010), Ankara'daki yedi farklı yerleşim yerinden alınan PM_{10} ve SO_2 bileşenleri ile oluşturulan hava kirliliği indeksinin günlük ortalama rüzgar hızı, günlük ortalama barometrik basınç, günlük sıcaklık değişimi ve bir önceki günün hava kirliliği indeksi üzerine çoklu doğrusal regresyonunu yaparak hava kirliliği için bir kestirim denklemi önermişlerdir.

Turgut ve Deniz (2015), Box-Jenkins yaklaşımı ile Ankara, Sıhhiye istasyonundan 1 Ocak 2010 ile 31 Ekim 2014 tarihleri arasındaki haftalık PM_{10} ölçümlerini ele almışlar ve bu ölçümlerin üçüncü dereceden otoregresif (AR(3)) zaman serisi modeline uygunluğuna Akaike bilgi kriteri (AIC) , R^2, Schwarz bilgi kriteri (SIC) ve en çok olabilirlik değerlerini göz önünde bulundurarak karar vermişlerdir. Bu modeli kullanarak Kasım 2014, Aralık 2014 ve Ocak 2015 dönemlerine ait öngörülerde bulunmuşlardır. Bu çalışmada serinin durağanlığının kontrolü genişletilmiş Dickey-Fuller (ADF) birim kök testi ile yapılmıştır. Elde edilen öngörüler sonucunda PM_{10} miktarının 83 $\mu g/m^3$ değerine yaklaştığını ve çok sayıda hastanenin bulunduğu Sıhhiye'de PM_{10} miktarının azaltılmasına yönelik çalışmalar yapılmasının gerekliliği savunulmuştur.

Bu çalışmaların çoğu ortalamaya ya da regresyona dayalı istatistiki sonuçlardır. Oysa veriler birim zaman aralıklarında yapılan gözlemler olduğundan zaman serisi tekniklerinin daha açıklayıcı olduğu düşünülmektedir. Bununla beraber, ortalamaya dayalı istatistiki sonuç çıkarımlar, bağımsız aynı dağılımlı rasgele değişkenlerin gözlenen değerlerine bağlıdır. Burada ise rasgele değişkenin aldığı değerler geçmiş değerler ile ilişkilidir. Bu nedenle de ortalamaya ya da regresyona bağlı istatistiki sonuç çıkarımlar yerine zaman serisi tekniklerinin kullanılması daha anlamlıdır. Yapılan çalışmaların büyük bir kısmında hava kirliliği verileri birim zaman aralıklarında ölçülmüş değerlerdir. Daha önce de bahsedildiği gibi regresyonda açıklayıcı değişken rolündeki değişkenler bilinen, rasgele olmayan değişkenlerdir. Zaman serisi modellerinde ise açıklayıcı değişken rolündeki değişkenler de rasgele değişkenlerdir.

Çalışmada kullanılacak ölçümlerin değerleri kış aylarında artmakta yaz aylarında ise azalmaktadır. Yani, hava kirliliği verileri mevsimsellik özelliği göstermektedir. Bu nedenle verilerin analizi yapılırken bu periyodikliğin dikkate alınması gerekir. Bununla birlikte periyodik fonksiyonlar genellikle sinüs (sin) ve kosinüs (cos) fonksiyonları ile ifade edilir. Çalışmada, verilerde gözlenen periyodik değişikliklerin saptanması için bu trigonometrik fonksiyonlardan yararlanılacaktır. Yapılan çalışmalarda veriler genellikle günlük veya haftalık olarak alınmış olup bu mevsimsellik etkisi göz ardı edilmiştir. Çalışmada, PM_{10} kirleticisine ait veriler aylık (yılda 12 gözlem) alınmış olup, zaman serisi tekniklerine yer verilmiştir.

Daha önce de bahsedildiği gibi zaman serilerinde istatistiki sonuç çıkarım yapılabilmesi için durağanlık önemli bir varsayımdır. Literatürde durağanlık sınamaları için oldukça çok yöntem bulunmaktadır. Bunlar arasında, regresyon parametrelerinin en küçük kareler tahmin edicilerinin dağılımına dayanan Dickey-Fuller yöntemi ile bu yöntemin biraz daha modife edilmiş hali olan Phillips-Perron yöntemi öne çıkmaktadır. Birçok istatistiki paket program (SAS, E-Views gibi) bu test istatistiklerinin değerlerini ve kritik değerlerini doğrudan hesaplamaktadır. Ayrıca mevsimsel veriler için de Dickey vd. (1984) tarafından geliştirilen ve simetrik en küçük kareler tahmin edicilerinin dağılımına dayalı yöntem ile Hylleberg vd. (1990) tarafından önerilen ve literatürde HEGY yöntemi olarak bilinen yöntem öne çıkmaktadır. Çalışmada, incelenen verilerin

durağanlığı bu yöntemler ile sınanacaktır. Ayrıca, verideki periyodikliğin dikkate alınması için trigonometrik dönüşümler yardımı ile hesaplanan, Akdi ve Dickey (1998) tarafından geliştirilen periodogramlara dayalı durağanlık sınaması ile de verilerin durağanlığı kontrol edilecektir.

Çalışmada kullanılacak PM_{10} verileri birim zaman aralıklarında gözlenen değerler olarak ele alındığından, diğer birçok çalışmadan farklı olarak zaman serisi teknikleri kullanılacaktır. Ayrıca, veriler aylık olarak ele alınacak olup, muhtemel bir periyodikliğin varlığı göz ardı edilmeyecektir. Eğer varsa, periyodikliğin tespiti periodogramlar yardımıyla yapılacaktır. Periyodikliğin var olması halinde, serinin durağanlığı mevsimsel birim kök testi ve periodogram tabanlı birim kök testi ile de sınanacaktır. Bununla ilgili bilgiler ilerleyen bölümlerde açıklanmaya çalışılacaktır.

2. DURAĞAN ZAMAN SERİSİ MODELLERİ

Zaman serilerinde amaç, birim zaman aralıklarında gözlenen verilere bağlı olarak gelecekteki değerlerin öngörülmesidir. Bunun için hemen hemen bütün istatistiki problemlerde de karşılaşılabileceği gibi, gözlem değerleri kullanılarak bir zaman serisi modeli oluşturulması gerekir. Bu bölümde uygulamada sıkça karşılaşılan ve hemen hemen bütün zaman serileri ile ilgili kitaplarda (Hamilton 1994, Fuller 1996 gibi) bulunabilen bazı zaman serileri modelleri kısaca tanıtılacaktır. Bu modellerden bazıları, MA ile ifade edilen hareketli ortalama modelleri (Moving Average), AR olarak bilinen otoregresif zaman serisi modelleri (Autoregressive) ve mevsimsel zaman serisi modelleridir.

Daha önce de bahsedilen otokorelasyon fonksiyonu zaman serilerinde kullanışlı bir araçtır. Bununla birlikte, yine otokorelasyonlar ile de hesaplanabilen kısmi otokorelasyon fonksiyonu da özellikle otoregresif zaman serilerinin model derecelerinin belirlenmesinde önemlidir. Bu fonksiyon, ileride otoregresif modeller özetlenirken daha detaylı olarak açıklanacaktır.

Hareketli ortalama modellerinde otokorelasyonlar belirli bir yerden sonra sıfır değerlerinin alır. Belirlenen bu nokta hareketli ortalama modellerinin model derecelerini belirlemede kullanılır. Otoregresif zaman serilerinde ise kısmi otokorelasyonlar belirli bir yerden sonra sıfır olur. Otokorelasyonlar ise mutlak değerce azalır. Belirlenen bu nokta ise otoregresif zaman serisi modellerinin model derecesinin belirlenmesinde kullanılır. Otokorelasyonlar bazen periyodik bir dalgalanma gösterir. Otokorelasyonlardaki bu dalgalanma verilerin mevsimsel olarak modellenmesine bir göstergedir. Bu dalga uzunluğu ise mevsimsel zaman serisi modelinin model derecesinin belirlenmesinde kullanılır. Bazen otokorelasyonlar ve kısmi otokorelasyonlar böyle bir davranış göstermezler. Bu durumda da verilerin ARMA (veya SARMA) olarak adlandırılan otoregresif hareketli ortalama (veya mevsimsel otoregresif hareketli ortalama) modelleri ile modellenir. Bu modellerin model derecelerinin belirlenmeleri, ilgili modeller özetlenirken daha detaylı olarak açıklanacaktır.

2.1 Hareketli Ortalama (Moving Average, MA) Modelleri

Önce, bütün zaman serisi modellerinde karşımıza çıkan ve $e_t \sim WN(0, \sigma^2)$ şeklinde ifade edilen beyaz gürültü serisini hatırlayalım.

Tanım 2.1 T bir indis kümesi olmak üzere $\{e_t, t \in T\}$ bir zaman serisi, $\forall t \in T$ için $E(e_t)$=0 ve bu zaman serisine ait otovaryans fonksiyonu

$$Cov(e_t, e_{t+h}) = \gamma(h) \begin{cases} \sigma^2 & , \quad h = 0 \\ 0 & , \quad d.d \end{cases}$$

şeklinde ise bu zaman serisine beyaz gürültü (White Noise) serisi denir (Hamilton 1994).

Beyaz gürültü serisi bütün zaman serisi modellerinin önemli bir bileşenidir. Zaman serilerinde modelleme yapıldığında hatalar üzerindeki varsayımlar sınanırken genel olarak hataların beyaz gürültü serisinden gelip gelmediği kontrol edilir. Şimdi hareketli ortalama modellerini açıklayalım.

Derecesi q olan bir hareketli ortalama ($MA(q)$) zaman serisi modeli, q sonlu bir doğal sayı, $\beta_q \neq 0$ ve $\beta_0 = 1$ olmak üzere,

$$Y_t = \mu + \sum_{i=0}^{q} \beta_i e_{t-i}$$

şeklinde yazılır. Burada, modelin beklenen değeri

$$E(Y_t) = E\left(\mu + \sum_{i=0}^{q} \beta_i e_{t-i}\right) = \mu$$

varyansı ise

$$\gamma(0) = Var(Y_t) = Cov(Y_t, Y_t) = Var\left(\mu + \sum_{i=0}^{q} \beta_i e_{t-i}\right) = \sigma^2 \sum_{i=0}^{q} \beta_i^2$$

şeklinde hesaplanır. Modelin otokovaryans fonksiyonu ise $h > 0$ için,

$$\gamma(h) = Cov(Y_t, Y_{t+h}) = Cov\left(\sum_{i=0}^{q} \beta_i e_{t-i}, \sum_{j=0}^{q} \beta_j e_{t+h-j}\right)$$
$$= \sum_{i=0}^{q}\sum_{j=0}^{q} \beta_i \beta_j Cov\left(e_{t-i}, e_{t+h-j}\right) = \sigma^2 \sum_{i=0}^{q-h} \beta_i \beta_{i+h}$$

dir. Buradan $MA(q)$ modelinin otokovaryans ve otokorelasyon fonksiyonları sırası ile

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{i=0}^{q-h} \beta_i \beta_{i+h} & , \quad 0 \le h \le q \\ 0 & , \quad d.d. \end{cases}$$

$$\rho(h) = \begin{cases} \left(\sum_{i=0}^{q-h} \beta_i \beta_{i+h}\right) \Big/ \left(\sum_{i=0}^{q-h} \beta_i^2\right) & , \quad 0 \le h \le q \\ 0 & , \quad d.d. \end{cases}$$

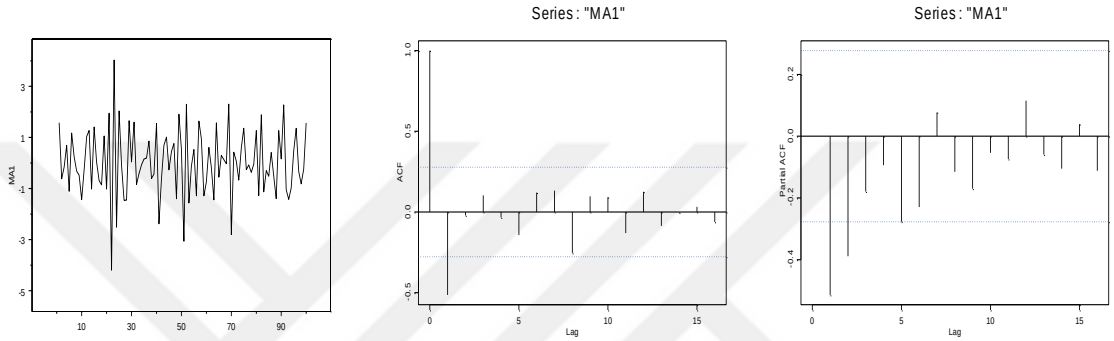
şeklindedir. Hareketli ortalama serilerinde model gecikmesinden sonraki otokovaryans değerleri ve buna bağlı olarak otokorelasyon değerleri sıfır değerini almaktadır. Bu gecikmenin sonlu olduğu durumda hareketli ortalama modelleri her zaman durağandır.

Model derecesi q olan bir hareketli ortalama ($MA(q)$) modelinde $q < \infty$ olmalıdır. Aksi halde modelin beklenen değeri ile daha yüksek dereceden momentleri (varyansı ve otokovaryansları) hesaplanamayabilir. Bu momentlerin hesaplanabilmeleri için bazı düzgünlük koşullarına (regularity conditions) ihtiyaç duyulur.

Hareketli ortalama modellerinin özelliklerini daha iyi görebilmek için farklı $MA(q)$ modellerine ilişkin rasgele üretilen 100'er birimlik zaman serisi verilerine ait zaman serisi ile otokorelasyon ve kısmi otokorelasyon fonksiyonlarının grafikleri şekil 2.1'de verilmiştir.

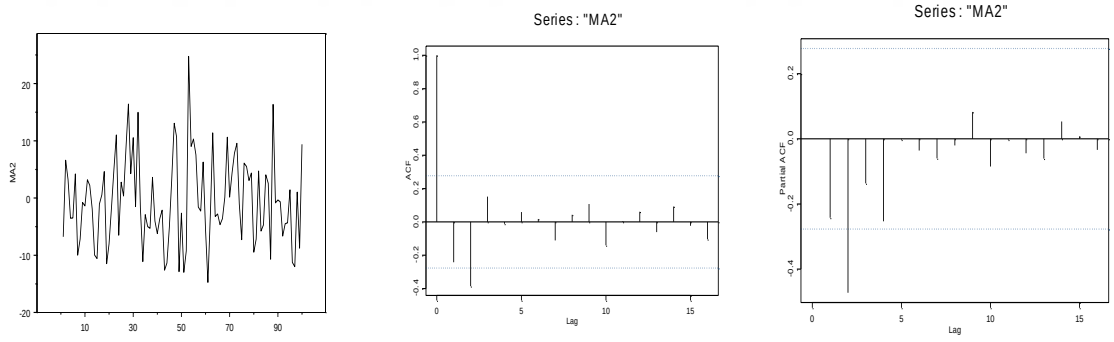
Model: $Y_t = 0.9e_{t-1} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Model: $Y_t = 0.9e_{t-1} + 0.7e_{t-2} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.1 Rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model derecesi 1 olan $Y_t = 0.9e_{t-1} + e_t$ şeklindeki $MA(1)$ modeli için üretilen 100 birimlik veriye ilişkin grafikler incelendiğinde otokorelasyonların birinci gecikmeden sonra sıfıra çok yakın değerler aldığı, kısmi otokorelasyonların ise böyle bir yapıya sahip olmadığı görülmektedir. Bununla birlikte $Y_t = 0.9e_{t-1} + 0.7e_{t-2} + e_t$ modeline uygun üretilen verilere ait grafiklerden de otokorelasyonların ikinci gecikmeden sonra sıfıra yakın değerler aldığı, kısmi otokorelasyonların ise düzenli bir yapıya sahip

olmadığı gözlenmektedir. Elimizde bu grafiklere uygun veriler olsaydı, bu verilerin $MA(1)$ ve $MA(2)$ olarak modellenmesi gerekirdi.

Hareketli ortalama modellerinde model derecesi $q = \infty$ ve $\beta_0 = 1$ olduğu durumda ise $e_t \sim WN(0, \sigma^2)$ olmak üzere modele ait beklenen değer ve otokovaryans fonksiyonu,

$$E(Y_t) = E\left(\mu + \sum_{i=0}^{\infty} \beta_i e_{t-i}\right) = \mu + E\left(\sum_{i=0}^{\infty} \beta_i e_{t-i}\right) = \mu + \sum_{i=0}^{\infty} \beta_i E(e_{t-i}) = \mu \qquad (2.1)$$

$$\begin{aligned} \gamma(h) = Cov(Y_t, Y_{t+h}) &= Cov\left(\sum_{i=0}^{\infty} \beta_i e_{t-i}, \sum_{j=0}^{\infty} \beta_j e_{t+h-j}\right) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \beta_i \beta_j Cov\left(e_{t-i}, e_{t+h-j}\right) \\ &= \sigma^2 \sum_{i=0}^{\infty} \beta_i \beta_{i+h} \end{aligned} \qquad (2.2)$$

şeklinde hesaplanır. Oysa bu beklenen değerin hesaplanabilmesi belli koşullara bağlıdır. Eğer,

$$\sum_{i=0}^{\infty} |\beta_i| < \infty \,, \sum_{i=0}^{\infty} \left|\beta_i^2\right| < \infty \qquad (2.3)$$

koşullarından biri sağlanırsa beklenen değer ve otokorelasyon fonksiyonu yukarıda hesaplandığı gibidir (Fuller 1996). Görüleceği gibi bu düzgünlük koşulları altında $q = \infty$ olması halinde $MA(q)$ modeli de durağandır.

2.2 Otoregresif (Auto Regressive, AR) Modeller

Otoregresif zaman serisi modelleri serinin kendi gecikmeleri ve beyaz gürültü serisini içeren modellerdir. Gerçek hayatta karşımıza çıkan zaman serileri genellikle otoregresif zaman serisi modelleri ile modellenir. Yani genellikle bir rasgele değişkenin gelecek

dönemlerde alacağı değerler, önceki dönemlerde gerçekleşen gözlem değerlerinden etkilenir. Örneğin, bir sonraki yılın aynı döneminde gerçekleşmesi beklenen hava kirliliği değerleri bir önceki yılın değerlerine bağlı olarak gerçekleşebilmektedir.

Bir önceki kısımda $MA(q)$ modellerinin q sonlu olduğu sürece durağan olduğu görüldü. Eğer $q = \infty$ ise $MA(q)$ modeli $\beta_0 = 1$ ve $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$Y_t = \mu + \sum_{i=0}^{\infty} \beta_i e_{t-i} \tag{2.4}$$

şeklinde yazılır. Bu modelin beklenen değeri,

$$E(Y_t) = E\left(\mu + \sum_{i=0}^{\infty} \beta_i e_{t-i}\right) = \mu + E\left(\sum_{i=0}^{\infty} \beta_i e_{t-i}\right)$$

şeklinde olacaktır. Ancak, bu beklenen değerin hesaplanabilmesi için beklenen değer operatörü ile sonsuz toplam yer değiştirilebilir olmalıdır. Bu değiştirmenin geçerli olabilmesi için de eşitlik (2.3)'de verilen koşullardan herhangi birinin sağlanması yeterlidir.

Şimdi, eşitlik (2.4)'de verilen $MA(\infty)$ modelini ($\mu = 0$) ele alalım. Burada $|\rho| < 1$ olmak üzere $\beta_i = \rho^i$ denirse model,

$$Y_t = \sum_{i=0}^{\infty} \beta_i e_{t-i} = \sum_{i=0}^{\infty} \rho^i e_{t-i}$$

olarak yazılır. Burada,

$$\sum_{i=0}^{\infty} \beta_i^2 = \frac{1}{1-\rho^2} < \infty$$

olduğundan düzgünlük koşulları sağlanır ve beklenen değer ile sonsuz toplam yer değiştirebilir. Bu durumda, eşitlik (2.1)-(2.2)'de verilen ifadelerden beklenen değer ile otokovaryans fonksiyonu,

$$E(Y_t) = 0$$

$$\gamma(h) = \sigma^2 \sum_{i=0}^{\infty} \rho^i \rho^{i+h} = \sigma^2 \rho^h \sum_{i=0}^{\infty} \rho^{2i} = \sigma^2 \frac{\rho^h}{1-\rho^2}$$

olup modelin varyansı da

$$\gamma(0) = \frac{\sigma^2}{1-\rho^2}$$

olur. Modelin beklenen değeri $\mu = 0$ olmak üzere eşitlik (2.4)'de verilen modeli

$$Y_t = \sum_{i=0}^{\infty} \rho^i e_{t-i}$$

şeklinde yazalım. Buradan

$$Y_t = e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \cdots \tag{2.5}$$

olup,

$$\rho Y_{t-1} = \rho e_{t-1} + \rho^2 e_{t-2} + \cdots \tag{2.6}$$

olarak yazılabilir. Kolayca görülebileceği gibi eşitlik (2.6)'da verilen ifade (-1) ile çarpılıp, eşitlik (2.5)'de verilen ifade ile taraf tarafa toplandığında

$$Y_t - \rho Y_{t-1} = e_t$$

elde edilir. Başka bir ifade ile $q = \infty$ ve $\beta_i = \rho^i$ olmak üzere $MA(\infty)$ modeli,

$$Y_t = \rho Y_{t-1} + e_t$$

olarak da ifade edilebilir. Bu model birinci dereceden otoregresif zaman serisi modeli olarak bilinir ve AR(1) şeklinde ifade edilir. Daha yüksek dereceden durağan otoregresif zaman serisi modelleri de $MA(\infty)$ modelleri yardımı ile elde edilir. Örneğin $AR(p)$ zaman serisi modeli $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$(Y_t - \mu) = \sum_{i=1}^{p} \alpha_i (Y_{t-i} - \mu) + e_t \tag{2.7}$$

şeklinde verilir. Bu model,

$$Y_t = \mu \left(1 - \sum_{i=1}^{p} \alpha_i\right) + \sum_{i=1}^{p} \alpha_i Y_{t-i} + e_t$$

veya

$$\alpha_0 = \mu \left(1 - \sum_{i=1}^{p} \alpha_i\right) \tag{2.8}$$

olmak üzere,

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + e_t \tag{2.9}$$

olarak da ifade edilebilir. Bu şekilde $MA(\infty)$ modelinden elde edilen AR modellerinin durağanlığı bazı koşullara bağlıdır. Yani, AR modelleri her zaman durağan değildir. Burada, eşitlik (2.9) ifadesinde verilen AR modeline karşılık gelen karakteristik denklem,

$$m^p - \sum_{i=1}^{p} \alpha_i m^{p-i} = 0 \tag{2.10}$$

olup bu denklemin bütün kökleri mutlak değerce 1'den küçükse model durağandır. Aksi halde, model durağan değildir. Durağan olmayan seriler denildiğinde ilk akla gelen birim köklü zaman serileridir. Eşitlik (2.10)'da verilen denklemin köklerinden en az biri mutlak değerce 1'e eşit ise bu tip seriler birim köklü zaman serileri olarak bilinir. Köklerin 1'den büyük olması durumu ise uygulamada karşılaşılan bir durum olmamakla birlikte ekonomik verilerin büyük bir kısmı durağan veya birim köklü serilerdir.

Otoregresif zaman serisi modellerini biraz tanıyabilmek için $AR(1)$ ve $AR(2)$ zaman serisi modellerini inceleyelim. Benzer hesaplamalar daha yüksek derecen AR modelleri için de genişletilebilir.

Önce $e_t \sim WN(0, \sigma^2)$ olmak üzere $AR(1)$ zaman serisi modelini,

$$(Y_t - \mu) = \alpha_1 (Y_{t-1} - \mu) + e_t$$

şeklinde yazalım. Bu modele karşılık gelen karakteristik denklem,

$$m - \alpha_1 = 0$$

şeklinde olup denklemin kökü $m_1 = \alpha_1$'dir. Yani, $AR(1)$ serisinin durağan olabilmesi için α_1 katsayısının mutlak değerce 1'den küçük olması gerekir. Eşitlik (2.8) kullanılarak,

$$\alpha_0 = \mu(1 - \alpha_1) \tag{2.11}$$

olarak alınırsa, verilen $AR(1)$ modeli,

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + e_t$$

şeklinde de yazılır. Bu modelde $\alpha_1 = 1$ ise eşitlik (2.11)'den $\alpha_0 = 0$ olur. Bu durumda model,

$$Y_t = Y_{t-1} + e_t$$

haline gelir. Ayrıca, serinin beklenen değeri μ modelden düşer. Yani, birim köklü zaman serilerinde beklenen değer tanımlı değildir. Şimdi, $Y_0 = 0$ olmak üzere, Y_t serisi

$$Y_t = \sum_{i=1}^{t} e_i$$

şeklinde yazılabilir. Gerçekten,

$$Y_1 = e_1$$

$$Y_2 = e_1 + e_2$$

$$\vdots$$

$$Y_t = e_1 + e_2 + \cdots + e_t$$

dir. Kolayca görüleceği gibi $E(Y_t) = 0$ olmasına rağmen, $Var(Y_t) = t\sigma^2$'dir. Yani, serinin varyansı zamana bağlı olup model durağan değildir.

Şimdi, $|\alpha_1| < 1$ durumunu inceleyelim. Yani, model $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + e_t$$

şeklinde verilmiş olsun. Kolayca görüleceği gibi serinin beklenen değeri, varyansı ve otokovaryans fonksiyonu sırası ile

$$E(Y_t) = \frac{\alpha_0}{1 - \alpha_1}$$

$$Var(Y_t) = \gamma(0) = \frac{\sigma^2}{1 - \alpha_1^2}$$

$$Cov(Y_t, Y_{t+h}) = \gamma(h) = \alpha_1^h \gamma(0)$$

şeklindedir. Burada da görüldüğü gibi serinin beklenen değer ve otokovaryans fonksiyonu zamana bağlı değildir. Serinin otokorelasyon fonksiyonu ise

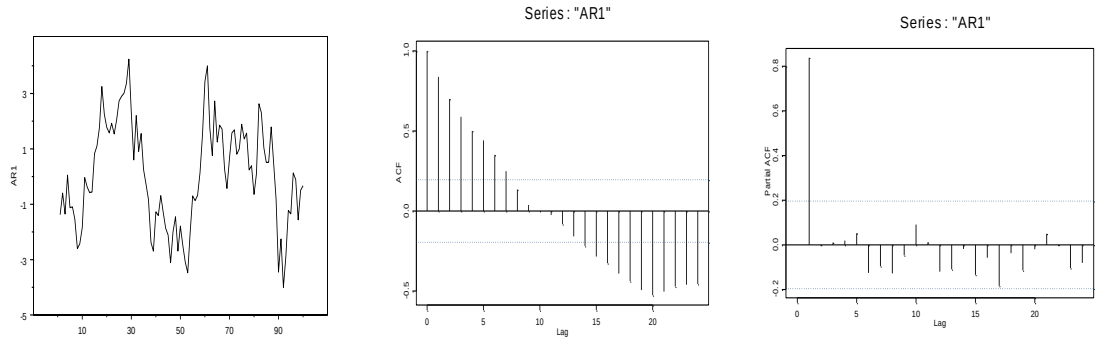
$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\alpha_1^h \gamma(0)}{\gamma(0)} = \alpha_1^h$$

olup, otokorelasyonlar mutlak değerce üstel olarak azalır. Modeldeki α_1 parametresinin değeri 1'e yaklaştıkça azalma hızı yavaştır. Yani, bir zaman serisinin otokorelasyonlarının grafiği hızlı bir şekilde azalıyorsa durağan olduğu söylenebilir. Aksi halde serinin durağanlığından şüphelenilir. Bu özellik AR kısmı içeren bütün modeller için geçerlidir.

Birinci dereceden herhangi bir AR serisinin özelliklerini görmek amacıyla farklı α_1 değerleri için $AR(1)$ modeline uygun rasgele üretilen 100 veriye ait otokorelasyon ve kısmi otokorelasyon grafikleri şekil 2.2-2.6'da verilmiştir.

Model: $Y_t = 0.8Y_{t-1} + e_t$

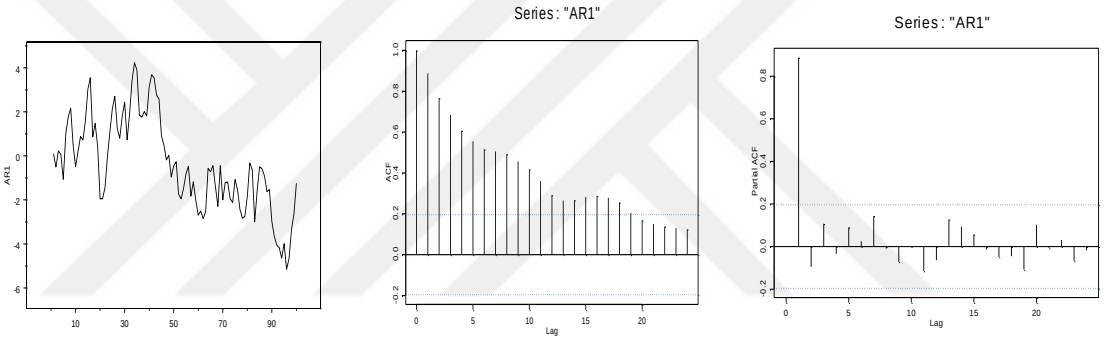
Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.2 $\alpha_1 = 0.8$ için $AR(1)$ serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model: $Y_t = 0.9Y_{t-1} + e_t$

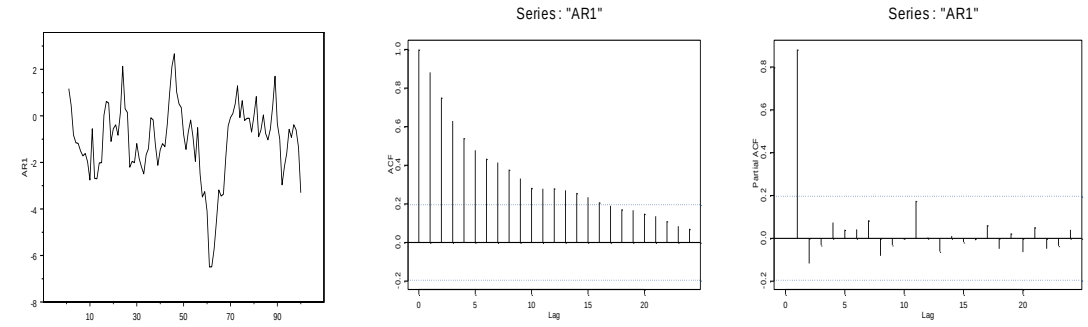
Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.3 $\alpha_1 = 0.9$ için $AR(1)$ serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model: $Y_t = 0.95Y_{t-1} + e_t$

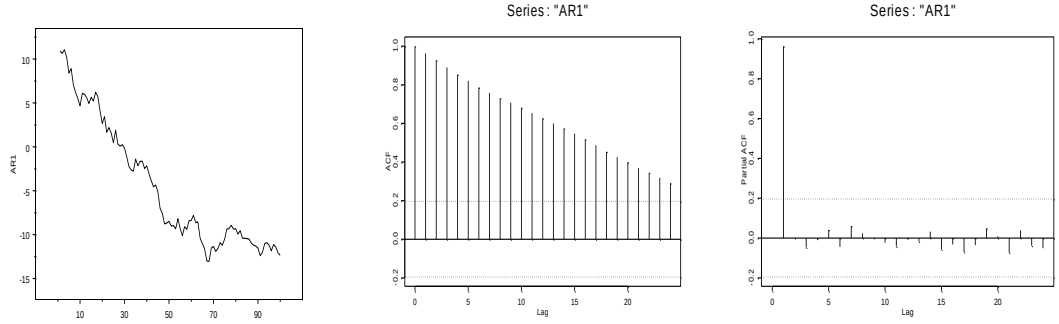
Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.4 $\alpha_1 = 0.95$ için $AR(1)$ serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model: $Y_t = 0.99Y_{t-1} + e_t$

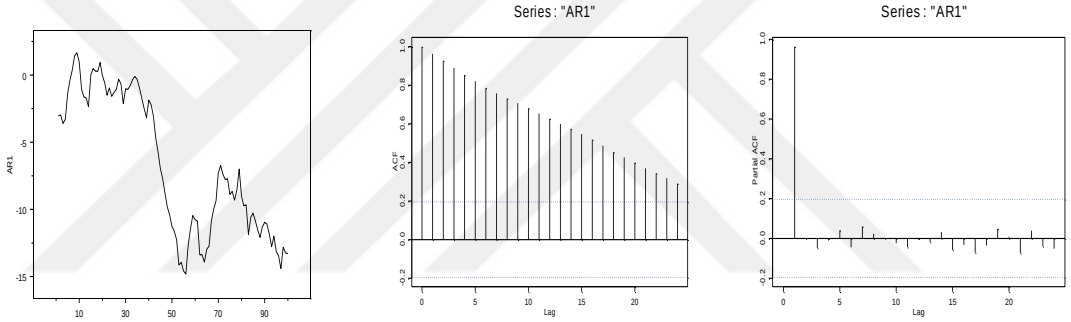
Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.5 $\alpha_1 = 0.99$ için $AR(1)$ serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model: $Y_t = Y_{t-1} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.6 $\alpha_1 = 1$ için $AR(1)$ serisine uygun rasgele üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Yukarıda grafikleri verilen $AR(1)$ modelleri incelendiğinde, α_1'in değeri 1'e yaklaştıkça otokorelasyonlardaki azalmanın yavaşladığı ve kısmi otokorelasyon fonksiyonlarının grafiklerinin ise birinci gecikmeden sonraki değerlerinin sıfıra yakın olduğu (%95 lik güven sınırları içindedir) görülmektedir.

İkinci dereceden otoregresif zaman serisi modelini ($AR(2)$) $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$(Y_t - \mu) = \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + e_t \qquad (2.12)$$

şeklinde yazalım. Modelin $\alpha_0 = \mu(\alpha_1 - \alpha_2)$ olmak üzere,

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + e_t \qquad (2.13)$$

şeklinde de yazılabileceği açıktır. Bu modele karşılık gelen karakteristik denklem,

$$m^2 - \alpha_1 m - \alpha_2 = 0$$

olup, denklemin iki kökü vardır. Bu köklerden herhangi biri mutlak değerce 1 ise model durağan değildir. Yani, birim köklüdür. O halde $m_1 = 1$ bu denklemin bir kökü ise bunun karakteristik denklemde yerine yazılması ile

$$1^2 - \alpha_1 - \alpha_2 = 0 \Leftrightarrow \alpha_1 + \alpha_2 = 1$$

elde edilir. Yani, $AR(2)$ modelinde $\alpha_1 + \alpha_2 = 1$ ise model birim köklüdür. Benzer özellik daha yüksek dereceden modeller için de geçerlidir.

Eşitlik (2.13)'de verilen AR modeli için $e_t \sim WN(0, \sigma^2)$ olmak üzere, Y_t ile Y_{t+h} arasındaki otokovaryans,

$$\begin{aligned}\gamma(h) = Cov(Y_t, Y_{t+h}) &= Cov(Y_t, \alpha_1 Y_{t+h-1} + \alpha_2 Y_{t+h-2} + e_{t+h}) \\ &= \alpha_1 Cov(Y_t, Y_{t+h-1}) + \alpha_2 Cov(Y_t, Y_{t+h-2}) + Cov(Y_t, e_{t+h}) \\ &= \alpha_1 \gamma(h-1) + \alpha_2 \gamma(h-2) + Cov(Y_t, e_{t+h})\end{aligned}$$

olup, modelin otokovaryans fonksiyonu $h > 0$ için,

$$\gamma(h) = \alpha_1 \gamma(h-1) + \alpha_2 \gamma(h-2) \qquad (2.14)$$

olarak elde edilir. Ayrıca, $h = 0$ için,

$$\gamma(0) = \alpha_1 \gamma(1) + \alpha_2 \gamma(2)$$

dir. Otokovaryanslardan ilk üç tanesi $h = 0,1,2$ için,

$$\gamma(0) = \alpha_1\gamma(1) + \alpha_2\gamma(2) + \sigma^2$$

$$\gamma(1) = \alpha_1\gamma(0) + \alpha_2\gamma(1)$$

$$\gamma(2) = \alpha_1\gamma(1) + \alpha_2\gamma(0) \tag{2.15}$$

dir. Burada $h > 0$ için $Cov(Y_t, e_{t+h}) = 0$ ve $h = 0$ için $Cov(Y_t, e_{t+h}) = \sigma^2$ olduğu açıktır. Bu denklem sistemi $AR(2)$ modeline ait Yule-Walker denklemleridir (Akdi 2012). Bu denklem sisteminde $\gamma(0)$, $\gamma(1)$ ve $\gamma(2)$ değerleri hesaplandığında diğer otokovaryanslar eşitlik (2.14) yardımı ile ardışık olarak hesaplanır. Ayrıca, $\gamma(0)$, $\gamma(1)$ ve $\gamma(2)$ otokovaryanslarının hesaplanan değerleri verildiğinde α_1 ve α_2 parametrelerinin çözümleri hesaplanan ilk üç otokovaryans türünden hesaplanır. Benzer şekilde $AR(2)$ modeline ait otokorelasyonlar da $\rho(0) = 1$ olmak üzere,

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \alpha_1\rho(h-1) + \alpha_2\rho(h-2) \tag{2.16}$$

şeklinde ardışık olarak elde edilebilir. Eşitlik (2.15)'in kullanılmasıyla elde edilen α_1 ve α_2 parametrelerinin ilk üç otokovaryans cinsinden değerleri,

$$\alpha_1 = \frac{\gamma(0)\gamma(1) - \gamma(1)\gamma(2)}{\gamma^2(0) - \gamma^2(1)}$$

$$\alpha_2 = \frac{\gamma(0)\gamma(2) - \gamma^2(1)}{\gamma^2(0) - \gamma^2(1)} \tag{2.17}$$

şeklinde hesaplanır. Bu değerler eşitlik (2.15)'de verilen birinci eşitlikte yerine konduğunda beyaz gürültü serisinin varyansı da

$$\sigma^2 = \left(1 - \frac{\alpha_1^2 + \alpha_1^2\alpha_2 + \alpha_2^2(1-\alpha_2)}{1-\alpha_2}\right)\gamma(0) \tag{2.18}$$

olur. İkinci dereceden otoregresif zaman serisi modeli için otokovaryanslardan ilk üç tanesi model parametreleri türünden,

$$\gamma(0) = \frac{(1-\alpha_2)\sigma^2}{(1+\alpha_2)((1-\alpha_2)^2-\alpha_1^2)}$$

$$\gamma(1) = \frac{\alpha_1\sigma^2}{(1+\alpha_2)((1-\alpha_2)^2-\alpha_1^2)}$$

$$\gamma(2) = \frac{(\alpha_1^2+\alpha_2(1-\alpha_2))\sigma^2}{(1+\alpha_2)((1-\alpha_2)^2-\alpha_1^2)}$$

olarak yazılabilir. Yine model parametreleri (α_1 ve α_2) verildiğinde ilk iki otokorelasyon değeri,

$$\rho(1) = \frac{\alpha_1}{1-\alpha_2}$$

$$\rho(2) = \frac{\alpha_1^2+\alpha_2(1-\alpha_2)}{(1-\alpha_2)}$$

Olup, diğer otokorelasyonlar eşitlik (2.16)'dan ardışık olarak hesaplanır. Yani, $AR(2)$ modeli için ilk iki otokorelasyon değerinin bilinmesi (veya hesaplanması) diğer otokorelasyonların hesaplanması için yeterlidir.

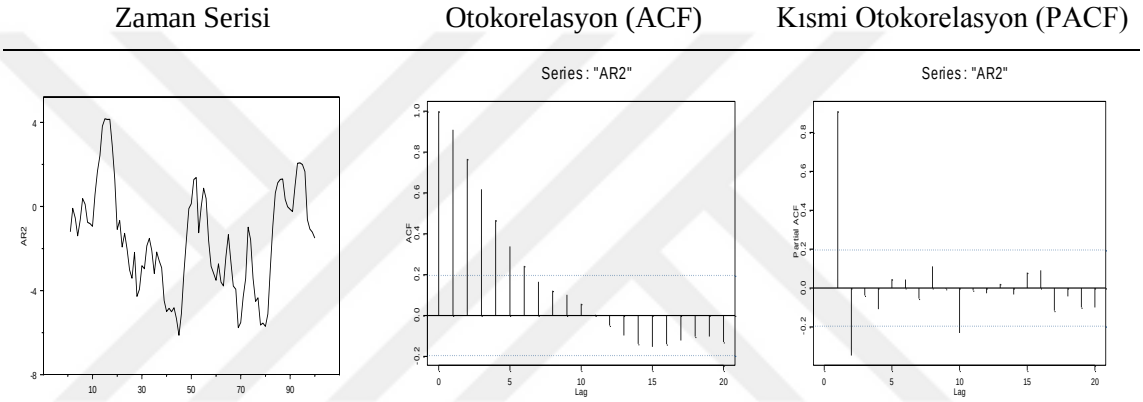
Bir zaman serisine ait otokovaryanslar,

$$\hat{\gamma}(h) = \frac{1}{n-h}\sum_{t=1}^{n-h}(Y_t-\bar{Y}_n)(Y_{t+h}-\bar{Y}_n)$$

şeklinde elde edilir. Bu tahmin değerleri eşitlik (2.17)'de yerine yazıldığında α_1 ve α_2 parametrelerinin Yule-Walker tahmin değerleri bulunur. Bu şekilde elde edilen Yule-Walker tahmin değerlerini en küçük kareler tahmin değerleri ile kıyaslamak için,

$$Y_t = 1.3Y_{t-1} - 0.42Y_{t-2} + e_t$$

modeline uygun 100 birimlik veri üretilmiştir. Verilere ilişkin zaman serisi grafikleri şekil 2.7'de verilmiştir.



Şekil 2.7 $AR(2)$ modeline uygun olarak üretilen veriye ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Bu verilere ilişkin model parametrelerine ait Yule-Walker ve en küçük kareler tahmin değerleri (Y_t'nin Y_{t-1} ve Y_{t-2} üzerine regresyonundan elde edilen tahmin değerleri)

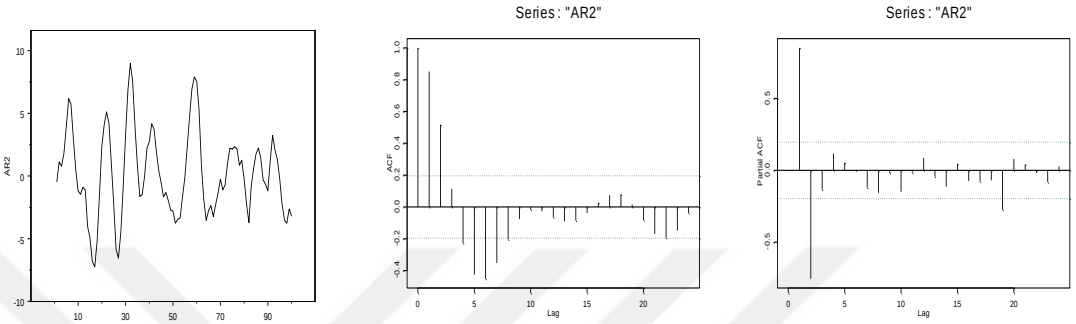
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\sigma}^2$
En Küçük Kareler	1.2404	-0.3251	1.0015
Yule-Walker	1.2249	-0.3471	0.9455

şeklinde gözlenmiştir. Burada da görüldüğü gibi en küçük kareler ve Yule-Walker tahmin değerleri birbirine çok yakındır.

Model parametrelerinin farklı değerleri için $AR(2)$ modeline uygun 100 birimlik zaman serileri üretilmiştir. Bu verilere ilişkin zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri şekil 2.8-2.12'de verilmiştir.

Model: $Y_t = 1.5Y_{t-1} - 0.7Y_{t-2} + e_t$

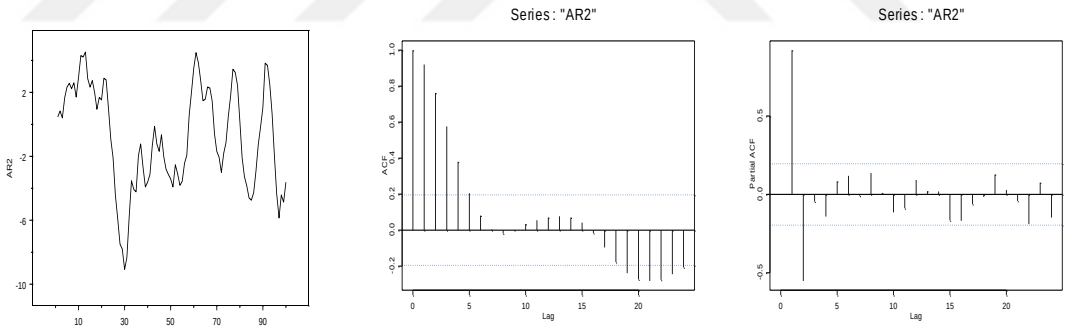
Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.8 $\alpha_1 = 1.5$, $\alpha_2 = -0.7$ için $AR(2)$ serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model: $Y_t = 1.5Y_{t-1} - 0.6Y_{t-2} + e_t$

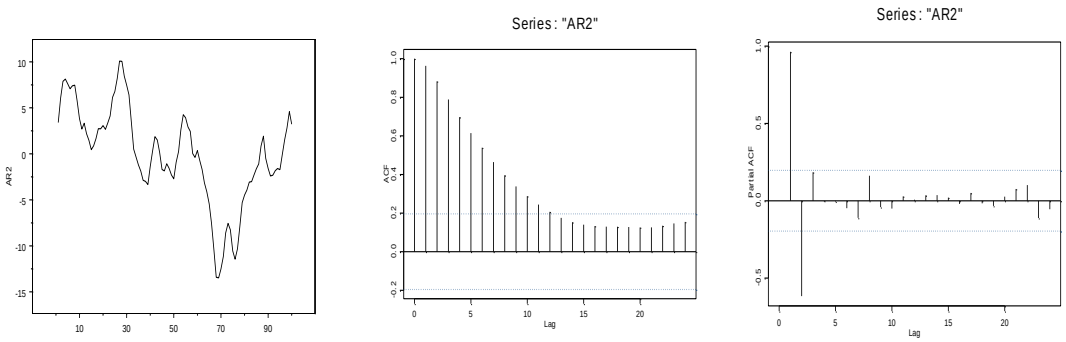
Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.9 $\alpha_1 = 1.5$, $\alpha_2 = -0.6$ için $AR(2)$ serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model: $Y_t = 1.5Y_{t-1} - 0.55Y_{t-2} + e_t$

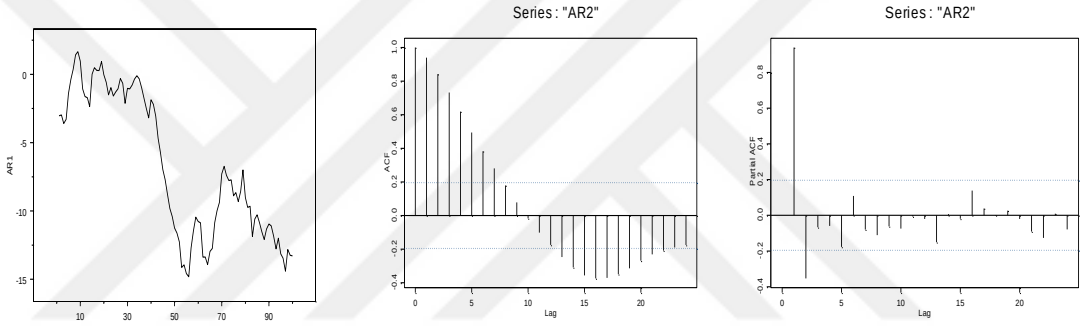
Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.10 $\alpha_1 = 1.5$, $\alpha_2 = -0.55$ için $AR(2)$ serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model: $Y_t = 1.5Y_{t-1} - 0.51Y_{t-2} + e_t$

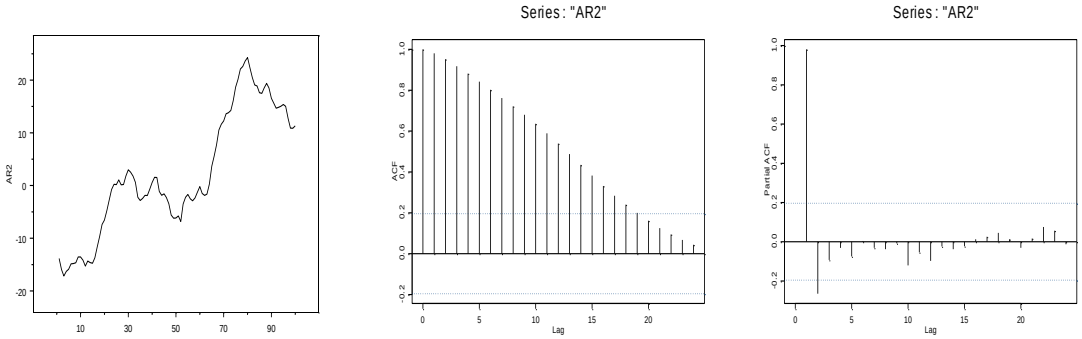
Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.11 $\alpha_1 = 1.5$, $\alpha_2 = -0.51$ için $AR(2)$ serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Model: $Y_t = 1.5Y_{t-1} - 0.5Y_{t-2} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.12 $\alpha_1 = 1.5$, $\alpha_2 = -0.5$ için $AR(2)$ serisine uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Üretilen verilere ait grafikler incelendiğinde $\alpha_1 + \alpha_2$ değeri 1'e yaklaştıkça otokorelasyonlardaki azalmanın yavaşladığı, yani serinin durağanlıktan uzaklaştığı, kısmi otokorelasyonların ise ikinci gecikmeden sonra güven düzeyinin altında değerler aldığı görülmektedir.

Bir zaman serisi modelini karakterize eden otokovaryans ve otokorelasyon fonksiyonları ve bu fonksiyonların özellikleri önceki kısımlarda incelendi. Bu fonksiyonlara ek olarak, bir zaman serisinin önemli karakteristiklerinden birisi de kısmi otokorelasyon fonksiyonudur. Otoregresif (AR) zaman serisi modelleri incelenirken rasgele üretilen verilere ait kısmi otokorelasyon değerlerinin modelin gecikme derecesinden sonra sıfır değerine yakın değerler aldığı (güven düzeyinin içerisinde yer aldığı), hareketli ortalama (MA) modellerinde ise böyle bir durumun olmadığı gözlendi.

Tanım 2.2 T bir indis kümesi olmak üzere $\{Y_t, t \in T\}$ zaman serisinin gözlem değerleri verildiğinde, Y_t'nin $Y_{t-1}, Y_{t-2}, \dots, Y_{t-h}$ üzerine regresyonu yapıldığında Y_{t-h}'nin katsayısına $\{Y_t, t \in T\}$ zaman serisinin h. kısmi otokorelasyonu denir ve $\Phi(h)$ ile gösterilir (Akdi 2012).

Tanımı itibariyle $\{Y_t, t \in T\}$ zaman serisi için, Y_t'nin Y_{t-1} üzerine regresyonundan birinci kısmi otokorelasyon Y_{t-1}'in katsayısıdır. Benzer şekilde Y_t'nin Y_{t-1} ve Y_{t-2} üzerine regresyonundan Y_{t-2}'nin katsayısı da serinin ikinci kısmi otokorelasyonudur. Herhangi bir zaman serisinin kısmi otokorelasyonlarını bu şekilde hesaplamak karmaşıklığa neden olabilir. Serinin otokorelasyonları kullanılarak da kısmi otokorelasyonlar hesaplanabilir. Model otokorelasyon fonksiyonu $\rho(h)$ olmak üzere P_h matrisi,

$$P_h = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(h-2) & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(h-3) & \rho(h-2) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \dots & \rho(1) & 1 \end{bmatrix}$$

şeklinde verilmiş olsun. Bu matrisin son sütununun $(\rho(1), \rho(2), \dots, \rho(h))'$ ile değiştirilmesi ile elde edilen matrisi de P_h^* ile gösterelim. Yani,

$$P_h^* = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(h-2) & \rho(1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(h-3) & \rho(2) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \dots & \rho(1) & \rho(h) \end{bmatrix}$$

olsun. Buradan, modelin kısmi otokorelasyon fonksiyonu,

$$\Phi(h) = \frac{\det(P_h^*)}{\det(P_h)}$$

şeklinde hesaplanır (Yaffee ve Mcgee 2000).

Kısmi otokorelasyon fonksiyonunun tanımından AR serilerinde kısmi otokorelasyonlar belirli bir yerden sonra sıfır (ya da sıfıra çok yakın) değerler almaktadır. Bu "belirli bir yer" AR modelinin derecesini ifade etmektedir. Bilindiği gibi MA modellerinde benzer özellik otokorelasyon fonksiyonu için geçerlidir. Bu özellikler verilen herhangi bir zaman serisinin türü ile model derecelerinin belirlenmesinde kullanılabilir.

2.3 Otoregresif Hareketli Ortalama (Autoregressive Moving Average, ARMA) Modelleri

Otoregresif (AR) ve hareketli ortalama (MA) modellerinde otokorelasyonlar ya da kısmi otokorelasyonlar belirli bir yerden sonra sıfırdır. Bazen, ne otokorelasyonlar ne de kısmi otokorelasyonlar belirli bir yerden sonra sıfır olur. Bu durumda, model AR ve MA modellerinin karışımı olarak bilinen $ARMA$ modelidir. Ancak, $ARMA$ modelinin model dereceleri otokorelasyon ve kısmi otokorelasyonlardan görünmeyebilir. Aşağıda $ARMA(p, q)$ modeli kısaca tanıtılacaktır.

Beklenen değeri μ olan bir $ARMA(p, q)$ modeli $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$(Y_t - \mu) = \sum_{i=1}^{p} \alpha_i (Y_{t-i} - \mu) + \sum_{j=1}^{q} \beta_j e_{t-j} + e_t \qquad (2.19)$$

şeklinde ifade edilir. Bilindiği gibi MA modelleri her zaman durağandır. Ancak AR nodellerinin durağanlığı karakteristik denklemin köklerine bağlıdır. Dereceleri p ve q olan otoregresif hareketli ortalama modelinin durağanlığı da AR kısmına karşılık gelen karakteristik denklemin köklerine bağlıdır. Yani, eşitlik (2.19)'da verilen $ARMA(p,q)$ zaman serisi modelinin durağanlığı,

$$m^p - \alpha_1 m^{p-1} - \alpha_2 m^{p-2} - \cdots - \alpha_p = 0$$

denkleminin köklerine bağlıdır. Bu denklemin bütün kökleri mutlak değerce birden küçükse $ARMA(p,q)$ modeli durağandır. Eşitlik (2.19)'da verilen $ARMA(p,q)$ modeli $\alpha_0 = \mu(1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p)$ olmak üzere,

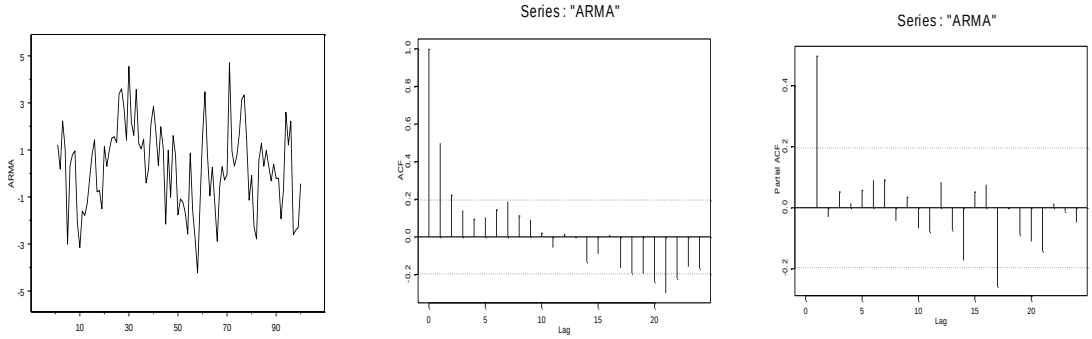
$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \cdots + \beta_q e_{t-q} + e_t$$

şeklinde de ifade edilir. Otoregresif zaman serisi modellerinde olduğu gibi denklemin köklerinden en az bir tanesi mutlak değerce 1 ise model birim köklüdür. Ayrıca köklerden en az bir tanesinin mutlak değerce 1 olması için gerek ve yeter koşul $\alpha_1 + \alpha_2 + \cdots + \alpha_p = 1$ olmasıdır. Buna göre model birim köklü ise serinin beklenen değeri (μ) modelden düşer.

Aşağıda şekil 2.13-2.14'de model parametreleri ve derecelerinin farklı değerleri için rasgele olarak üretilen 100 birimlik verilere ilişkin grafikler verilmiştir.

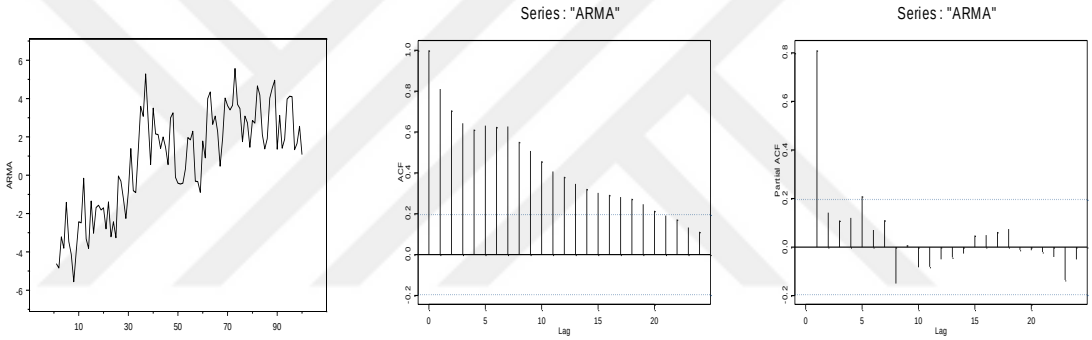
Model: $Y_t = 0.8Y_{t-1} + 0.9e_{t-1} + 0.7e_{t-1} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Model: $Y_t = Y_{t-1} + 0.9e_{t-1} + 0.7e_{t-1} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)

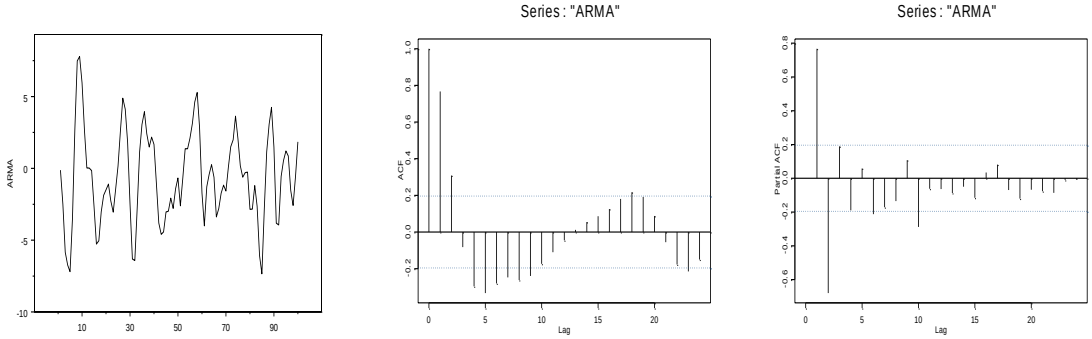


Şekil 2.13 $ARMA(1,2)$ modelline uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Yukarıdaki grafikler incelendiğinde otokorelasyonların ve kısmi otokorelasyonların her ikisinin de mutlak değerce azalmalarına rağmen bu değerlerin birçoğu güven sınırlarının dışındadır. Ayrıca, birinci modelde AR kısmının durağan olması nedeniyle otokorelasyonların hızlı bir şekilde azaldığı, ikinci modelde ise otokorelasyonlardaki azalmanın çok yavaş olduğu gözlenmektedir.

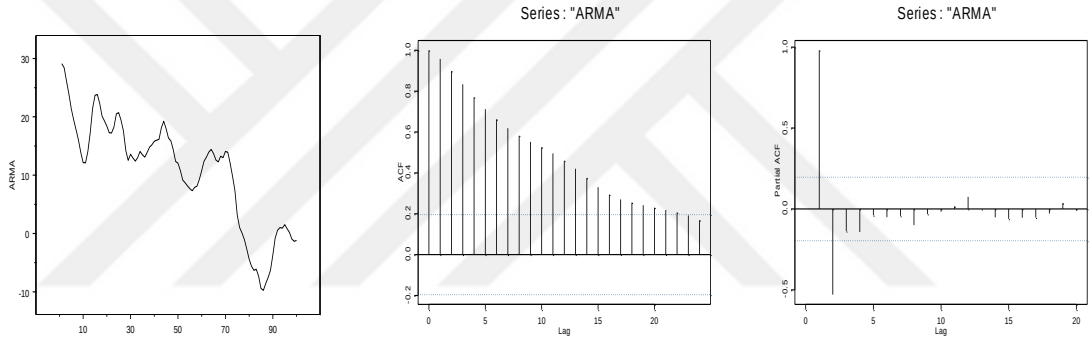
Model: $Y_t = 1.5Y_{t-1} - 0.7Y_{t-2} - 0.5e_{t-1} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Model: $Y_t = 1.5Y_{t-1} - 0.5Y_{t-2} - 0.5e_{t-1} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.14 $ARMA(2,1)$ modelline uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Şekil 2.14 incelendiğinde de benzer özellikler gözlenmektedir.

2.4 Mevsimsel (Seasonal) Zaman Serisi Modelleri

Zaman serisinin belirli periyotlarla alınan gözlemlerin bir dizisi olduğundan bahsedilmişti. Bu periyot yıl, üç ay (çeyreklik), hafta, gün, saat olabileceği gibi ay da olabilir. Aylık olarak gözlenen veriler ile ilgilenildiğinde bazen bu verilerin periyodik dalgalanmalar gösterdiği gözlenebilir. Örneğin, soğuk gazlı içeceklerin satış miktarları yaz aylarında bir artış göstermesine rağmen kış aylarında satışların düşmesi olağandır. Yine kış aylarında fosil yakıtların kullanımı nedeniyle hava kirliliğinin artması yaz aylarında ise azalması bu periyodikliğe örnek olarak verilebilir.

Aylık periyotlarla gözlenen bir zaman serisi verisi ile ilgilenildiğinde otokorelasyonların periyodik bir yapı gösterdiği gözlemlenebilir. Bu gibi durumlarda verilerin modellemesi AR, MA veya $ARMA$ modelleri yerine mevsimsel zaman serisi modelleri ($SAR, SARMA$ gibi) ile yapılır.

Mevsimsel zaman serisi modeli, d pozitif bir tamsayı olmak üzere,

$$(Y_t - \mu) = \alpha(Y_{t-d} - \mu) + e_t$$

şeklinde verilsin. Burada modelde bulunan d serinin mevsimsel gecikmesini (periyodunu) gösterir. Bu mevsimsel zaman serisi modeli $SAR_d(1)$ olarak adlandırılır. Modele karşılık gelen karakteristik denklem,

$$m^d - \alpha = 0$$

şeklinde olup, bu denklemin köklerinin mutlak değerce birden küçük olması durumunda model durağandır. Uygulamada en çok d'nin 4 olduğu durum ile karşılaşılır. Bu durumda $SAR_4(1)$ modeli

$$(Y_t - \mu) = \alpha(Y_{t-4} - \mu) + e_t$$

şeklinde ifade edilir. Bu modelin durağanlığı için gerek ve yeter şart $|\alpha| < 1$ olmasıdır. Modele ait otokovaryans fonksiyonu,

$$\gamma(h) = Cov(Y_t, Y_{t+h}) = Cov(Y_t, \alpha Y_{t+h-4} + e_{t+h}) = \alpha Cov(Y_t, Y_{t+h-4}) + Cov(Y_t, e_{t+h})$$
$$= \alpha\gamma(h-4) + Cov(Y_t, e_{t+h})$$

şeklindedir. Burada $Cov(Y_t, e_{t+h})$ ifadesi $h = 0$ olduğu durumda σ^2, diğer durumlarda ise sıfır değerini alır. O halde otokovaryanslar sırası ile $h = 0,1,2,3,4$ için,

$$\gamma(0) = \alpha\gamma(4) + \sigma^2$$

$$\gamma(1) = \alpha\gamma(3)$$

$$\gamma(2) = \alpha\gamma(2)$$

$$\gamma(3) = \alpha\gamma(1)$$

$$\gamma(4) = \alpha\gamma(0)$$

şeklindedir. Bu eşitlikler yardımıyla,

$$\gamma(1) = \gamma(2) = \gamma(3) = 0$$

ve

$$\gamma(0) = \frac{\sigma^2}{1-\alpha^2}$$

$$\gamma(4) = \alpha\frac{\sigma^2}{1-\alpha^2}$$

sonuçları elde edilir. Yine bu sonuçlardan,

$$\gamma(5) = \alpha\gamma(1) = 0$$

$$\gamma(6) = \alpha\gamma(2) = 0$$

$$\gamma(7) = \alpha\gamma(3) = 0$$

$$\gamma(8) = \alpha\gamma(4) = \alpha^2\frac{\sigma^2}{1-\alpha^2}$$

olduğu açıktır. O halde, $SAR_4(1)$ modeli için otokorelasyon fonksiyonunun genel hali,

$$\gamma(h) = \begin{cases} \dfrac{\sigma^2 \alpha^{h/4}}{1-\alpha^2} & , \quad h = 4k\,, k = 0,1,2,\ldots \\ 0 & , \quad d.d. \end{cases}$$

şeklindedir. Buna bağlı olarak otokorelasyon fonksiyonunun genel hali ise

$$\rho(h) = \begin{cases} \alpha^{h/4} & , \quad h = 4k\,, k = 0,1,2,\ldots \\ 0 & , \quad d.d. \end{cases}$$

olarak yazılır. Bir $SAR_4(1)$ modeli için kısmi otokorelasyonlar ise $\Phi(4) = \alpha$ ve $h \neq 4$ için $\Phi(h) = 0$ şeklinde hesaplanır.

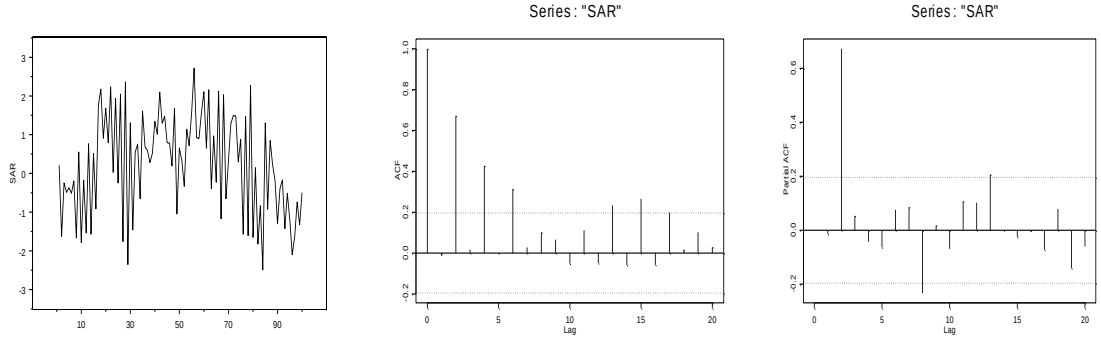
Genel olarak bir $SAR_d(p)$ modeli, d ve p pozitif tamsayı ve $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$(Y_t - \mu) = \sum_{i=1}^{p} \alpha_i (Y_{t-id} - \mu) + e_t$$

şeklinde ifade edilir. Bu modelin durağanlığı AR modellerinde olduğu gibi modele karşılık gelen karakteristik denklem yardımıyla belirlenir. Karakteristik denklemin bütün kökleri mutlak değerce birden küçükse model durağandır. Aksi halde model birim köklüdür. Mevsimsel zaman serisi modellerinin daha iyi anlaşılabilmesi için iki farklı SAR modeline uygun olarak 100'er birimlik zaman serileri üretilerek, verilere ilişkin grafikler şekil 2.15'de verilmiştir.

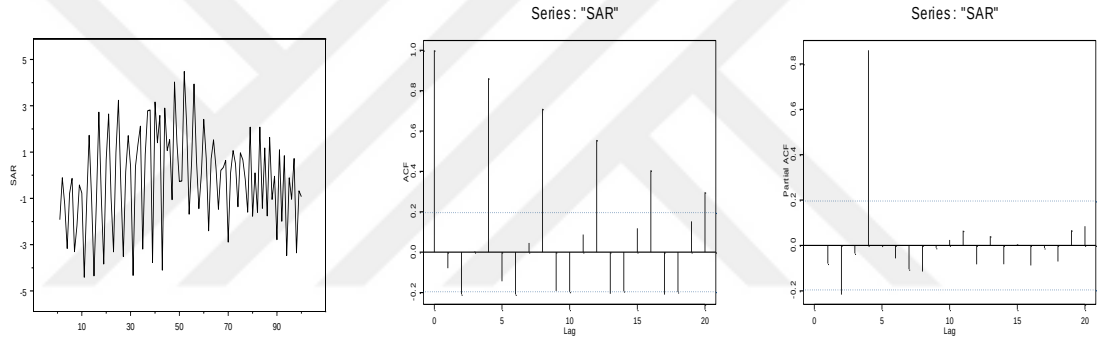
Model: $Y_t = 0.8Y_{t-2} - 0.2Y_{t-4} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Model: $Y_t = 0.9Y_{t-4} + e_t$

Zaman Serisi	Otokorelasyon (ACF)	Kısmi Otokorelasyon (PACF)



Şekil 2.15 Farklı SAR modelline uygun rasgele üretilen verilere ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Şekil 2.15'de verilen ilk model bir $SAR_2(2)$ modelidir. Bu modele ilişkin otokorelasyon fonksiyonunun grafiği incelendiğinde, otokorelasyonların periyodik bir yapı gösterdiği ve mutlak değerce hızlı bir azalma gösterdiği gözükmektedir. İkinci modelde ise $SAR_4(1)$ modeline uygun olarak üretilen veriye ait otokorelasyon değerleri mutlak değerce yavaş ve periyodik bir azalma göstermekte, kısmi otokorelasyon fonksiyonunu ise dördüncü gecikmeden sonra sıfıra çok yakın değerler almaktadır. $SAR_4(1)$ modelinde Y_{t-4}'ün katsayısının 1 olması durumunda, modelin karakteristik denkleminin kökleri $m_{1,2} = \pm 1$ ve $m_{3,4} = \pm i$ şeklinde olup, köklerin tamamı mutlak değerce bire eşit olmaktadır. Bu nedenle, Y_{t-4}'ün katsayısının 1 olduğu durumda $SAR_4(1)$ modeli durağan değildir. Verilen örnekte de Y_{t-4}'ün katsayısı 0.9 olarak alınmış olup bu değer

1'e oldukça yakındır. $SAR_4(1)$ modeli için otokorelasyonların yavaş bir azalma göstermesinin nedeninin de bu durum olduğu söylenebilir.

2.5 Öngörü

Zaman serilerinde en önemli istatistiksel sonuç çıkarımların başında öngörü (forecasting) kavramı gelir. Bu kavram genellikle tahmin (estimation) ve kestirim (prediction) kavramları ile karıştırılsa da bu kavramlardan tamamen farklıdır.

Tahmin bir kitle parametresinin herhangi bir tahmin edicisinin aldığı değere karşılık gelirken, kestirim ise bir rasgele değişkenin modellenmesinde model parametrelerinin tahmin değerlerinin yerlerine konulmasıyla elde edilen değerdir. Öngörü bu iki kavramdan da farklı olarak model parametrelerinin tahmin edilmesiyle rasgele değişkenin alacağı değerleri önceden kestirilmesine karşılık gelir. Bir zaman serisi verisi için öngörü yapmak özellikle ekonomi alanında olmakla birlikte her alanda çok önemli bir yer tutar. Örneğin, bir sonraki yılın enflasyon oranının geçmişte gerçekleşen enflasyon oranları yardımıyla öngörülmesi bir devletin ekonomi planlaması açısından önemlidir. Benzer şekilde geçmişte gerçekleşen ithalat ve ihracat miktarı verilerinin kullanılarak sonraki dönemler için bu miktarların öngörülmesi yine geliştirilecek ekonomi stratejileri için önemlidir. Ayrıca, sadece ekonomik alanlarda değil örneğin, geçmişte gerçekleşen hava kirliliği miktarları verilerinin kullanılmasıyla sonraki dönemler için gerçekleşecek hava kirliliği miktarları öngörülebilir ve bunun yardımı ile hava kirliliği konusunda çeşitli önlemler alınabilir. Buna ek olarak, zaman serisinin öngörülerini etkileyen sadece ilgilenilen değişkenin geçmişteki değerleri olmadığı da açıktır. Örneğin, enflasyon verileri için tutarlı bir öngörünün yapılabilmesi için sadece geçmişteki enflasyon verileri yeterli olmayabilir. Bunun yanında, faiz verilerine de ihtiyaç duyulabilir. Benzer şekilde, hava kirliliği verileri de öngörülürken yağış miktarları, rüzgar hızları, nem miktarları gibi verilere ihtiyaç duyulabilir. Burada önemli olan nokta neredeyse tüm istatistiksel analizlerde olduğu gibi verilere uygun modelin belirlenmesidir. Bu kısımda bir rasgele değişkenin geçmişteki değerleri yardımıyla gelecekte alacağı değerin öngörülmesi üzerinde durulacaktır.

Tanım 2.3 $Y_1, Y_2, \dots, Y_n$ rasgele değişkenleri verildiğinde Y_{n+1} için bir öngörü

$$\hat{Y}_{n+1} = E(Y_{n+1}|Y_1, Y_2, \dots, Y_n)$$

koşullu beklenen değeridir.

Yukarıda verilen $\hat{Y}_{n+1}$ öngörü istatistiği koşullu beklenen değerin tanımı gereğince,

$$E(\hat{Y}_{n+1}) = E[E(Y_{n+1}|Y_1, Y_2, \dots, Y_n)] = E(Y_{n+1})$$

olup, $\hat{Y}_{n+1}$ öngörüsü yansızdır.

Zaman serisini oluşturan $Y_1, Y_2, \dots, Y_n, Y_{n+1}$ rasgele değişkenlerinin ortak dağılımı normal ise Y_{n+1}'in öngörüsü $Y_1, Y_2, \dots, Y_n$ rasgele değişkenlerinin lineer birleşimidir. Bu nedenle, rasgele değişkenlerin normal dağılımlı olması durumunda çok değişkenli normal dağılımın özelliğinden,

$$\hat{Y}_{n+1} = E(Y_{n+1}|Y_1, Y_2, \dots, Y_n) = \sum_{i=1}^{n} \alpha_i X_i \qquad (2.20)$$

olarak yazılabilir. Burada, yine koşullu beklenen değer tanımından verilen öngörünün yansız olduğu açıktır. Ayrıca, $\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)'$ ve $\underset{\sim}{Y} = (Y_1, Y_2, \dots, Y_n)'$ olmak üzere eşitlik (2.20)'de verilen öngörü $\hat{Y}_{n+1} = \underset{\sim}{\alpha}'\underset{\sim}{Y}$ şeklinde gösterilebilir.

Gelecekteki bir değer için çok sayıda farklı öngörü önerilebilir. Bu öngörüler bazı kriterlere göre karşılaştırılarak en uygunu belirlenir. Bu kriterlerden bir tanesi hata kareler ortalamasıdır. Öngörü istatistiğinin hata kareler ortalaması,

$$MSEP(\hat{Y}_{n+1}) = E(Y_{n+1} - \hat{Y}_{n+1})^2$$

ile hesaplanır. Koşullu beklenen değer olarak tanımlanan $\hat{Y}_{n+1} = \underset{\sim}{\alpha}'\underset{\sim}{Y}$ öngörü istatistiği, lineer yansız öngörülerin tamamının içinde en küçük hata kareler ortalamasına sahip olan öngörüdür. Yani hata kareler ortalaması kriterine göre eşitlik (2.20)'de tanımlanan öngörü istatistiği en iyidir. O halde, $\hat{Y}_{n+1} = \underset{\sim}{\alpha}'\underset{\sim}{Y}$ en iyi, lineer, yansız öngörü (BLUF, Best Linear Unbiased Forecast) dür.

Verilerin normal dağılımlı bir kitleden alınmamış olduğu durumda, koşullu beklenen değer olarak tanımlanan öngörü örneklemin lineer bir birleşimi olmayabilir. Bu durumda, beklenen değerin doğrudan hesaplanması gerekir. Çoğu istatistiki sonuç çıkarımda olduğu gibi öngörüde de normallik varsayımına ihtiyaç duyulmaktadır. Normallik varsayımının sağlanmadığı durumda ise dönüşümler yapılarak normallik varsayımı sağlatılır.

Bu kısımda bazı otoregresif bileşenler içeren zaman serisi modellerinde öngörünün nasıl yapılacağı üzerinde durulacaktır. Birinci dereceden otoregresif ($AR(1)$) zaman serisi modeli için e_t'ler bağımsız, ortalaması sıfır varyansı σ^2 olan normal dağılıma sahip rasgele değişkenler olmak üzere $Y_1, Y_2, \dots, Y_n$ değişkenlerinin değerleri gözlendiğinde Y_{n+1} için bir öngörü,

$$\hat{Y}_{n+1} = E\left(Y_{n+1}|Y_{1,}Y_2, \dots, Y_n\right) = E(\alpha Y_n + e_{n+1}|Y_1, Y_2, \dots, Y_n) = \alpha Y_n$$

ile belirlenir. Benzer şekilde $(n+s)$. dönem için öngörü,

$$\hat{Y}_{n+s} = E\left(Y_{n+s}|Y_{1,}Y_2, \dots, Y_n\right) = E(\alpha Y_{n+s-1} + e_{n+s}|Y_1, Y_2, \dots, Y_n) = \alpha\hat{Y}_{n+s-1}$$

olarak hesaplanır. Bu durumda,

$$\hat{Y}_{n+2} = \alpha^2 Y_n, \hat{Y}_{n+3} = \alpha^3 Y_n, \dots, \hat{Y}_{n+s} = \alpha^s Y_n$$

şeklinde olup, öngörülerin modelin durağan olduğu durumda ($|\alpha| < 1$) serinin ortalamasına yaklaşacağı, durağan olmadığı durumda ($|\alpha| = 1$) ise sabit kalacağı açıktır. Bu nedenle, durağan olmayan zaman serisi modelleri ile öngörü yapmak istatistiksel olarak anlamsızdır. Ayrıca bir $AR(1)$ modeli kullanılarak öngörü yapılacağı durumda öngörünün alacağı değer örneklemin sadece son değerine bağlıdır. Bazı durumlarda öngörü için bir güven aralığı belirlemek önemlidir. Bunun için de öngörü hataları ve öngörü hatalarının varyansına ihtiyaç duyulur. Öngörü hataları ve hataların varyansları bir $AR(1)$ modeli için,

$$\hat{e}_{n+1} = Y_{n+1} - \hat{Y}_{n+1} = \alpha Y_n + e_{n+1} - \alpha Y_n = e_{n+1}$$

$$\begin{aligned}\hat{e}_{n+2} = Y_{n+2} - \hat{Y}_{n+2} &= \alpha Y_{n+1} + e_{n+2} - \alpha^2 Y_n = \alpha(\alpha Y_n + e_{n+1}) + e_{n+2} - \alpha^2 Y_n \\ &= \alpha^2 Y_n + \alpha e_{n+1} + e_{n+2} - \alpha^2 Y_n = \alpha e_{n+1} + e_{n+2}\end{aligned}$$

$$\hat{e}_{n+s} = Y_{n+s} - \hat{Y}_{n+s} = \sum_{i=0}^{s-1} \alpha^i e_{n+s-i}$$

$$Var\left(Y_{n+s} - \hat{Y}_{n+s}\right) = \sigma^2 \sum_{i=0}^{s-1} \alpha^{2i}$$

şeklinde belirlenir. Benzer şekilde, e_t'ler bağımsız ortalaması sıfır varyansı σ^2 olan rasgele değişkenler olmak üzere bir $AR(2)$ modeli için $(n+s)$. dönem için öngörü, öngörü hatası ve öngörü hatasının varyansının genel hali;

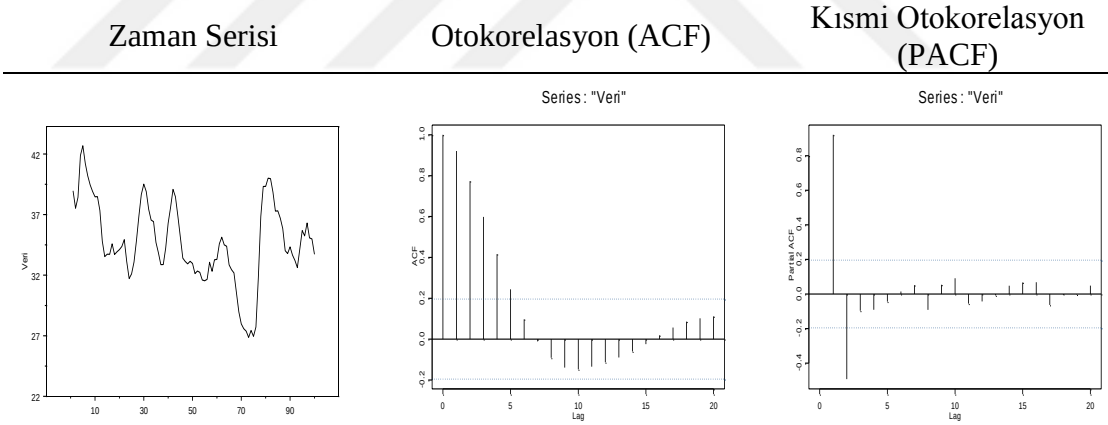
$$\begin{aligned}\hat{Y}_{n+s} = E\left(Y_{n+s}|Y_1, Y_2, \dots, Y_n\right) &= E(\alpha_1 Y_{n+s-1} + \alpha_2 Y_{n+s-2} + e_{n+s}|Y_1, Y_2, \dots, Y_n) \\ &= \alpha_1 E(Y_{n+s-1}|Y_1, Y_2, \dots, Y_n) + \alpha_2 E(Y_{n+s-2} + |Y_1, Y_2, \dots, Y_n) \\ &= \alpha_1 \hat{Y}_{n+s-1} + \alpha_2 \hat{Y}_{n+s-2}\end{aligned}$$

$$\hat{e}_{n+s} = Y_{n+s} - \hat{Y}_{n+s} = \sum_{i=0}^{s-1} w_i e_{n+s-i} \qquad (2.21)$$

$$Var\left(Y_{n+s} - \hat{Y}_{n+s}\right) = \sigma^2 \sum_{i=0}^{s-1} w_i^2 \qquad (2.22)$$

şeklindedir. Eşitlik (2.21)-(2.22)'de verilen w_i sabitleri model parametreleri olan α_1 ve α_2'nin fonksiyonlarıdır. Burada da öngörüler örneklemin son iki değerine bağlı olarak hesaplanmaktadır. Genel olarak otoregresif zaman serisi modellerinde öngörüleri hesaplamak kolaydır. Buna karşılık, otoregresif hareketli ortalama ($ARMA$) zaman serisi modellerinde ise öngörüler yine koşullu beklenen değer yardımı ile hesaplanabilmekte ancak öngörünün alacağı değer önceki dönemlere ait hatalara da bağlı olacağından, öncelikle model kurulduktan sonra hata terimleri için kestirimler yapılarak bu kestirimler yardımıyla öngörüler belirlenebilir.

Öngörü kavramının daha iyi anlaşılabilmesi açısından 100 birimlik rasgele zaman serisi üretilmiş, bu zaman serisine ait grafikler şekil 2.16'da verilmiştir.



Şekil 2.16 Rasgele olarak üretilen 100 birimlik zaman serisine ait zaman serisi, otokorelasyon ve kısmi otokorelasyon grafikleri

Şekil 2.16'da verilen grafikler incelendiğinde otokorelasyonun fonksiyonunun değerlerinde üstel bir azalmanın olduğu, kısmi otokorelasyon fonksiyonu değerlerinin ise ikinci değerden sonra güven düzeyinin altında değerler aldığı görünmektedir. Bu nedenle, sezgisel olarak ilgilenilen zaman serisi verisinin $AR(2)$ modeline uygunluk gösterdiği söylenebilir. Ancak, öngörünün yapılabilmesi için öncelikle bir modele ihtiyaç duyulmaktadır. Bu nedenle, SAS paket programı yardımıyla (PROC ARIMA)

veriye farklı p ve q değerleri için $ARMA(p,q)$ modelleri önerilmiş olup, önerilen modellere ilişkin Akaike Bilgi Kriteri (Akaike Information Criterion, AIC) ve Schwarz Bayesian Kriteri (Schwarz Bayesian Criterion, SBC) değerleri çizelge 2.1'de verilmiştir.

Çizelge 2.1 Rasgele veriye önerilen modellere ait AIC ve SBC değerleri

$ARMA(p,q)$	$p=1,q=0$	$p=1,q=1$	$p=2,q=0$	$p=2,q=1$	$p=2,q=2$
AIC	333.80	303.76	299.30	299.94	300.43
SBC	339.01	311.57	307.12	310.36	313.45

Verilen çizelge incelendiğinde en küçük AIC veya SBC değerlerinin $AR(2)$ modeli için ortaya çıktığı gözükmektedir. AIC ve SBC değerleri en küçük olan model önerilen modeller arasındaki en iyi modeldir (Brockwell ve Davis 2002). Önerilen $AR(2)$ modeline ait parametre tahminleri ise çizelge 2.2'de verilmiştir.

Çizelge 2.2 Rasgele veriye önerilen $AR(2)$ modeline ait parametre tahminleri, standart hatalar ve p-değerleri

Parametre	Tahmin	Standart Hata	p-değeri
μ	38.984	0.924	$<.0001$
α_1	1.519	0.085	$<.0001$
α_2	-0.558	0.085	$<.0001$

Çizelge 2.2 incelendiğinde tüm parametrelerin anlamlılığının testine ilişkin p-değerlerinin 0.05'den küçük olduğu, dolayısıyla %95 güvenle bütün parametrelerin anlamlı olduğu sonucuna ulaşılır. Ayrıca, şekil 2.15'de yer alan otokorelasyon fonksiyonunun grafiği incelendiğinde otokorelasyonlardaki azalmanın hızlı olduğu bu nedenle ilgilenilen zaman serisinin durağan olabileceği sezgisel olarak söylenebilir. Benzer şekilde, çizelge 2.2 incelendiğinde α_1 ve α_2 parametrelerinin tahminlerinin toplamı birden küçük olduğundan önerilen $AR(2)$ modelinin durağan olabileceği

sezgisel olarak söylenebilir. Ancak, serinin durağanlığı hakkında kesin bir yorum yapabilmek için test edilmesi gerekmektedir. Bu testler ile ilgili ayrıntılardan ilerleyen bölümlerde bahsedilecektir.

Veri için kestirilen zaman serisi modeli olan

$$Y_t = 38.984 + 1.519X_{t-1} - 0.558X_{t-2}$$

modeli kullanılarak gelecek on döneme ait öngörü değerleri, bu değerlere ait standart hatalar ve %95'lik güven aralıkları çizelge 2.3'de verilmiştir.

Çizelge 2.3 Ragele olarak üretilen veri için öngörü değerleri (101:110)

Dönem	**Öngörü**	**Standart Hata**	**Güven Aralığı (%95)**
101	33.225	1.065	[31.138,35.312]
102	33.174	1.936	[29.378,36.969]
103	33.375	2.686	[28.109,38.640]
104	33.708	3.305	[27.230,40.186]
105	34.102	3.804	[26.646,41.559]
106	34.515	4.203	[26.276,42.753]
107	34.922	4.520	[26.062,43.781]
108	35.309	4.771	[25.957,44.661]
109	35.670	4.970	[25.923,45.412]
110	36.003	5.128	[25.953,46.053]

Çizelge 2.3 incelendiğinde öngörü değerlerine ait standart hataların dönem sayısı ilerledikçe arttığı ve bu nedenle güven aralıklarının genişlediği ve tutarlı öngörülerin yapılmasının güç hale geldiği açıktır. Ayrıca, öngörü değerleri belirli bir yerden sonra serinin ortalaması olan

$$\bar{X}_n = \sum_{t=1}^{100} X_t = 34.729$$

değerinden uzaklaştığı ve bu nedenle serinin durağanlığından şüphe edilebileceği söylenebilir. Durağanlığın sınanması için ise ilerleyen bölümlerde verilecek olan bazı birim kök testleri kullanılabilir.

İyi bir öngörü yapabilmenin ön koşullardan birisi de ilgilenilen zaman serisine ait en uygun modelin belirlenmesidir. Bu modelin belirlenmesinde birçok istatistiğin değeri göz önüne alınabilir. Öngörü yapılabilmesi için ise önerilen modelin durağan olması gerekmektedir. Bu bölümde bazı durağan zaman serisi modelleri ayrıntısına girilmeden özetlendi. Ancak, her zaman ilgilenilen zaman serisine uygun modelin durağan olması söz konusu değildir. Özellikle iktisadi verilerde durağan olmayan (birim köklü) serilere sıklıkla rastlanmaktadır. Bu nedenle durağanlığın sınanması, sağlanmadığı durumlarda ise çeşitli dönüşümler yardımıyla sağlatılması, iyi bir öngörü için önemlidir. Durağanlığın nasıl sınanacağı ve durağan olmayan zaman serileri ile ilgili bilgiler bir sonraki bölümde ayrıntısına girilmeden verilecektir.

3. DURAĞAN OLMAYAN ZAMAN SERİLERİ

Zaman serileri analizinde istatistiki sonuç çıkarım için temel varsayım ilgilenilen zaman serisinin durağan olmasıdır. Zaman serisi modelinin AR bileşeni içerdiği durumlarda durağan olabilmesi AR kısmına karşılık gelen karakteristik denklemin tüm köklerinin mutlak değerce birden küçük olmasına bağlıdır. Köklerden herhangi birinin mutlak değerce bire eşit olduğu durumlarda ise seri birim köklüdür. Bu gibi durumlarda istatistiki sonuç çıkarımı için çeşitli yöntemlerle öncelikle seri durağan hale getirilir, daha sonra durağan olan bu seri yardımıyla analizler gerçekleştirilir. O halde, zaman serilerinde istatistiki sonuç çıkarımı için öncelikle modelin durağanlığı test edilir. Modelin durağanlığı birim kök testleri yardımıyla sınanır. Literatürde en çok kullanılan birim kök testleri Dickey-Fuller ve Philliphs-Perron birim kök testleridir. Zaman serisinin mevsimsellik içerdiği durumlarda ise mevsimsel birim kökün tespiti için Dickey vd. (1984) tarafından geliştirilen ve literatürde DHF testi olarak bilinen birim kök testi ile Hylleberg vd. (1990) tarafından geliştirilen ve HEGY testi olarak adlandırılan birim kök testi sıklıkla kullanılmaktadır. Serinin mevsimsel birim kök içerdiği durumlarda, birim kök tekrar edeceği için durağanlaştırma işlemi yapılırken tekrar eden birim kök sayısı dikkate alınırsa hatalı istatistiki sonuç çıkarımlar yapılabilir. Dolayısıyla mevsimsel birim kök olduğunda, seriyi durağanlaştırmak için mevsimsel fark alma yöntemi kullanılır.

Önceki bölümlerde bazı durağan zaman serisi modelleri ve bu modellere ait bazı özellikler incelendi. Bu bölümde ise literatürde sıklıkla karşılaşılan ve kritik değerleri bazı istatistiksel paket programlarda doğrudan hesaplanabilen birim kök testleri ve durağan olmayan modeller ile periodogramlara bağlı olarak geliştirilen birim kök testi üzerinde durulacaktır.

3.1 Birim Köklü Seriler

Önceki bölümlerde verilen durağanlık tanımı gereğince bir zaman serisinin durağan olmamasının iki nedeni olabilir. Bunlardan birincisi serinin ortalamasının zamana bağlı olarak değişmesi iken, ikincisi ise serinin otokovaryans fonksiyonunun zamana bağlı

olarak değişmesidir. Serinin ortalaması zamana bağlı olarak değişiyorsa seri deterministik trend içerir. Deterministik trend içeren bir seriyi durağan hale getirme işlemi bazı analitik işlemler yardımıyla basit bir şekilde yapılabilir. Serinin otokovaryans fonksiyonu zamana bağlı olarak değişiyor ise bu durumda seri stokastik trend içerir. Serinin stokastik trendden kurtarılabilmesi, deterministik trende göre daha karmaşıktır. Genellikle bu işlem fark alma, dönüşüm yapma gibi yöntemlerle gerçekleştirilir. Birim köklü zaman serileri stokastik trend içeren zaman serileridir. Bir zaman serisi $\{Y_t, t \in T\}$ ve $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$Y_t = \mu + Y_{t-1} \tag{3.1}$$

$$Y_t = Y_{t-1} + e_t \tag{3.2}$$

$$Y_t = \mu + Y_{t-1} + e_t \tag{3.3}$$

modellerini göz önüne alalım. Burada μ sabit (genellikle zaman serisinin beklenen değeri) ve $Z_t = e_1 + e_2 + \cdots + e_t$ olmak üzere yukarıdaki modeller sırasıyla,

$$Y_t = Y_0 + t\mu \tag{3.4}$$

$$Y_t = Y_0 + Z_t \tag{3.5}$$

$$Y_t = Y_0 + Z_t + t\mu \tag{3.6}$$

şeklinde yazılabilir. Burada, eşitlik (3.4)'de verilen model deterministik bir trend içerirken, eşitlik (3.5)'de verilen model stokastik ve eşitlik (3.6)'da verilen model ise hem deterministik hem de stokastik bir trend içerir. Verilen bir $\{Y_t, t \in T\}$ zaman serisinde,

$$Y_1(w) \leq Y_2(w) \leq Y_3(w) \ldots \leq Y_n(w) \leq \cdots$$

eşitsizliği sağlanıyor ise seri artan bir trend içerir. Aksi halde, yani

$$Y_1(w) \geq Y_2(w) \geq Y_3(w) \ldots \geq Y_n(w) \geq \cdots$$

eşitsizliği sağlanıyor ise seri azalan bir trend içerir.

Tanım 3.1 B gecikme operatörü ($B^k Y_t = Y_{t-k}$) olmak üzere, $\nabla = 1 - B$ operatörüne fark operatörü denir. Durağan olmayan bir $\{Y_t, t \in T\}$ zaman serisi için $\nabla^d Y_t = (1-B)^d Y_t$ serisi durağan oluyorsa $\{Y_t, t \in T\}$ zaman serisine d.nci dereceden bütünleşiktir (integrated) denir (Akdi 2012).

Burada d sabiti serinin kaçıncı dereceden farkının alındığını gösterir. Bir $\{Y_t, t \in T\}$ zaman serisi d. dereceden bütünleşik ise $Y_t \sim I(d)$ ile gösterilir. Durağan zaman serileri için ise genellikle $Y_t \sim I(0)$ gösterimi kullanılır. Durağan olmayan bir zaman serisinin çeşitli dönüşümler yardımıyla durağan hale getirilemediği durumlarda fark alma işlemi uygulanır. Bu fark alma işleminin kaçıncı dereceye kadar yapılacağı ise serinin karakteristik denkleminin kaç tane kökünün mutlak değerce bire eşit olduğu ile ilgilidir.

Önceki bölümlerde de incelendiği gibi, eşitlik (3.2)'de verilen model kullanılarak öngörü yapılmaya çalışıldığında,

$$E(Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_1) = Y_{t-1}$$

ve

$$E(\nabla Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_1) = 0$$

olup, stokastik trend içeren bir model ile öngörü yapılamayacağı açıktır. Ayrıca, yine önceki bölümlerde incelendiği gibi serinin birim kök içermesi durumunda serinin beklenen değeri μ yani, zaman serisi modelindeki kesim noktası (intercept) modelden düşer.

Bir zaman serisinin durağanlığının istatistiksel sonuç çıkarımı için temel bir varsayım olduğundan önceki kısımlarda bahsedildi. Dolayısıyla bir zaman serisinin durağanlığının tespiti de önemlidir. Durağanlığın tespiti için sezgisel olarak bir yorum yapabilmek adına uygulamada genellikle otokorelasyon fonksiyonunun aldığı değerler ile ilgilenilir. Bunun nedeni, AR kısmını içeren modeller için karakteristik denklemin köklerinin 1'e yaklaştıkça serinin otokorelasyon fonksiyonundaki azalmanın yavaşlaması, 1'den uzaklaştıkça ise serinin otokorelasyon fonksiyonunun aldığı değerlerdeki artışın hızlanmasıdır. Ayrıca, yine sezgisel olarak yorum yapabilmek için ilgilenilen zaman serisi verisine önerilen modeldeki AR kısmına ait parametre değerlerinin en küçük kareler (veya herhangi başka bir şekilde) yöntemi ile belirlenmesi sonucunda, bu parametre tahminlerinin değerleri toplamı yaklaşık olarak 1 ise o zaman modelin durağanlığından şüphe edilir. Bunun nedeni ise yine önceki bölümlerde incelendiği üzere, AR kısmı içeren modellerde karakteristik denklemin köklerinden birisinin 1 olması durumu, parametrelerinin toplamlarının bire eşit olması ile gerçekleşmesiydi. Bu yöntemler bir zaman serisinin durağanlığı ile ilgili sezgisel olarak bazı bilgiler vermektedir. Serinin durağanlığının sınanması için geliştirilen birçok test yöntemi vardır. Bu testler birim kök testleri olarak adlandırılırlar. Serinin birim köklü olup olmadığına bu testler yardımı ile karar verilir. Serinin birim kök içermesi durumunda ise öncelikle serinin dönüşümü yapılarak veya farkı alınarak durağan hale getirilir ve istatistiki sonuç çıkarımlar bu seri üzerinden yapılır. Bu birim kök testlerinden literatürde sıklıkla kullanılanları bir sonraki kısımda incelenecektir.

3.2 Birim Kök Testleri

Verilen herhangi bir zaman serisi için durağanlık varsayımının istatistiksel sonuç çıkarımlar açısından ne kadar önemli olduğundan önceki bölümlerde bahsedildi. İlgilenilen zaman serisinin durağanlığının tespiti için birçok yöntem vardır. Bunlardan bazıları, serinin otokorelasyon fonksiyonunun azalma hızı, zaman serisi grafiğindeki ani değişimlerin veya zamana bağlı olarak sürekli artış azalış gösterme durumunun incelenmesi, hesaplanan öngörü değerlerinin serinin ortalamasına yaklaşma durumunun incelenmesi gibi yöntemlerdir. Ancak, bu yöntemler kullanılarak ilgilenilen serinin birim köklü bir seri olup olmadığı hakkında kesin bir yorum yapılamamaktadır. Bu

nedenle parametrelerin tahmin edicilerinin özelliklerine göre birçok değişik birim kök testi geliştirilmiştir. Bunlardan literatürde en sık rastlananları Dickey ve Fuller (1979) tarafından geliştirilen ve Dickey-Fuller birim kök testi ile adlandırılan birim kök testi ile bu testin bazı değişikliklerle modife edilmiş hali olan Phillips-Perron birim kök testidir. Bu birim kök testlerine ait kritik değerler birçok istatistiksel paket program tarafından direkt olarak verilir. Bu testlerin mevsimsellik içeren zaman serilerinde iyi sonuçlar vermemesi nedeniyle serinin mevsimsellik içerdiği durumlar için ise Dickey vd. (1984) tarafından geliştirilen mevsimsel birim kök testi ile Hylleberg vd. (1990) tarafından önerilen ve literatürde baş harfleri ile adlandırılan HEGY testi sıklıkla kullanılmaktadır. Bu testlere ait özellikler ayrıntısına girilmeden anlatılmaya çalışılacaktır.

3.2.1 Dickey-Fuller birim kök testi

Dickey-Fuller birim kök testi literatürde en çok kullanılan birim kök test yöntemidir. Bu yöntem parametrelerin, serinin birim köklü olduğu varsayımı altındaki en küçük kareler tahmin edicilerinin dağılımına bağlı olarak geliştirilmiştir.

Hata terimleri $e_t \sim WN(0, \sigma^2)$ olmak üzere, $AR(1)$ modeli,

$$(Y_t - \mu) = \rho(Y_{t-1} - \mu) + e_t \quad , \quad t = 1,2,3, \dots, n$$

şeklindedir. Burada $H_0: \rho = 1$ yokluk hipotezi altında serinin beklenen değeri olan μ parametresi modelden düşer. Bu beklenen değerin sıfır alınması ile elde edilen ρ parametresinin en küçük kareler tahmin edicisi

$$\hat{\rho}_n = \sum_{t=2}^{n} Y_t Y_{t-1} \Big/ \sum_{t=2}^{n} Y_{t-1}^2$$

dir. Bu eşitlikte Y_t yerine yazıldığında,

$$\hat{\rho}_n - \rho = \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{t=2}^{n} e_t Y_{t-1} \Big/ \frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^2\right)$$

eşitliği elde edilir. Burada $n \to \infty$ iken

$$\hat{\rho}_n \xrightarrow{P} \rho$$

ve

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{D} N(0,1-\rho^2)$$

yakınsamaları geçerlidir (Fuller 1996).

Birinci dereceden otoregresif zaman serisinin birim köklü olması durumunda, yani $\rho = 1$ durumunda ρ'nun EKK tahmin edicisi $\hat{\rho}_n$ ile gösterilmek üzere,

$$\hat{\rho}_n - 1 = \frac{1}{n}\left(\frac{1}{n}\sum_{t=2}^{n} e_t Y_{t-1} \Big/ \frac{1}{n^2}\sum_{t=2}^{n} Y_{t-1}^2\right)$$

olarak yazılır. Burada $n(\hat{\rho}_n - 1)$'in dağılımı yardımıyla serinin birim köklü olup olmadığının sınaması yapılabilir. Bu istatistiğin dağılımı ise Dickey-Fuller dağılımı olarak bilinir. Standart Wiener süreci $W(t)$ ile gösterilmek üzere $n(\hat{\rho}_n - 1)$'in dağılımı $n \to \infty$ durumunda,

$$n(\hat{\rho}_n - 1) = \left(\frac{1}{n}\sum_{t=2}^{n} e_t Y_{t-1} \Big/ \frac{1}{n^2}\sum_{t=2}^{n} Y_{t-1}^2\right) \xrightarrow{D} \frac{1}{2}\frac{[\chi_1^2 - 1]}{\int_0^1 W^2(t)dt}$$

şeklindedir. Bu test istatistiğinin hesaplanan değeri ve kritik değeri karşılaştırılarak birinci dereceden otoregresif ($AR(1)$) modele uyduğu düşünülen bir zaman serisinin birim köklü olup olmadığı sınanabilir.

Birinci dereceden bir otoregresif zaman serisi modeli ile ilgilenilsin. Bu model,

$$Y_t = \rho Y_{t-1} + e_t$$

şeklinde yazıldığında regresyon modeline benzerliği göz önüne çıkmaktadır. Regresyon analizinde ise $H_0: \rho = 1$ hipotezinin testi için t-testi kullanılmaktadır. Burada farklı olarak, t istatistiğinin dağılımı bilinen t dağılımından farklıdır. Verilen yokluk hipotezinin testi için kullanılan test istatistiği,

$$\hat{\tau}_n = \frac{\hat{\rho}_n - 1}{S_{\hat{\rho}_n}} = (\hat{\rho}_n - 1) \Big/ \sqrt{S_n^2 \left(\sum_{t=2}^{n} Y_{t-1}^2 \right)^{-1}}$$

şeklinde olup, bu dağılım literatürde Dickey-Fuller dağılımı olarak bilinir. Burada, $\hat{\rho}_n$ ρ'nun en küçük kareler tahmin edicisidir. Bu dağılıma ait kritik değerler birçok istatistiksel paket program tarafından doğrudan verilmektedir.

Beklenen değeri sıfır olmayan ($\mu \neq 0$) bir $AR(1)$ modeli ele alındığında ise model,

$$(Y_t - \mu) = \rho(Y_{t-1} - \mu) + e_t \quad , \quad t = 1,2,3, \dots, n$$

şeklinde yazılabilir. Burada, $\alpha_0 = \mu(1 - \rho)$ denildiğinde modelin,

$$Y_t = \alpha_0 + \rho Y_{t-1} + e_t \quad , \quad t = 1,2,3, \dots, n$$

şeklinde de ifade edilebileceği önceki bölümlerde gösterilmişti. Verilen bu model kesim noktası bulunan bir basit doğrusal regresyon modeline benzemektedir. Model $\rho = 1$ durumunda durağan değildir. Durağanlığının sınanması için kullanılacak olan yokluk hipotezi ($H_0: \rho = 1$) altında serinin beklenen değeri, buna bağlı olarak da modelde bulunan kesim noktası modelden düşer. Burada,

$$\bar{Y}_{(i)} = \frac{1}{n-i}\sum_{t=2}^{n} Y_{t-i} \quad , \quad i = 0,1$$

olmak üzere, ρ'nun en küçük kareler tahmin edicisi,

$$\hat{\rho}_{n,\mu} = \sum_{t=2}^{n}\left(Y_t - \bar{Y}_{(0)}\right)\left(Y_{t-1} - \bar{Y}_{(1)}\right) \Big/ \sum_{t=2}^{n}\left(Y_{t-1} - \bar{Y}_{(1)}\right)^2$$

dir. Buruda $H_0: \rho = 1$ yokluk hipotezinin $H_0: \rho < 1$ alternatif hipotezine karşı testi için kullanılacak test istatistiği ise,

$$\hat{\tau}_{n,\mu} = \frac{\hat{\rho}_{n,\mu} - 1}{S_{\hat{\rho}_{n,\mu}}} = \left(\hat{\rho}_{n,\mu} - 1\right) \Big/ \sqrt{S_n^2 \left(\sum_{t=1}^{n} (Y_t - \hat{\alpha}_0 - \hat{\rho} Y_{t-1})^2\right)^{-1}}$$

şeklinde olup, bu değer ∇ fark operatörünü göstermek üzere, ∇Y_t'nin Y_{t-1} üzerine regresyonu yapıldığında Y_{t-1}'in katsayısını gösteren parametreye ait t istatistiğidir. Bu istatistiğin dağılımı yine bilinen t dağılımı olmayıp literatürde Dickey-Fuller dağılımı olarak bilinir.

Bir zaman serisi modeli kesim noktası içerebildiği gibi bu kesim noktası ile birlikte zamana bağlı deterministik bir trend de içerebilir. Bu durumda model,

$$(Y_t - a - bt) = \rho(Y_{t-1} - a - bt) + e_t \quad , \quad t = 1,2,3, \dots, n$$

veya

$$Y_t = a(1-\rho) + bt(1-\rho) + \rho Y_{t-1} e_t \quad , \quad t = 1,2,3,\dots,n$$

şeklinde yazılabilir. Burada model yine basit doğrusal regresyon modeline benzemekle birlikte ρ'nun en küçük kareler tahmin edicisi farklıdır. Serinin birim köklü olduğu yokluk hipotezinin ($H_0: \rho = 1$) altında model yine durağan olmayıp beklenen değer ($\mu = a + bt$) modelden düşer. Bu durumda da beklenen değerin modelde yer almadığı ve aldığı durumlardakine benzer olarak $H_0: \rho = 1$ hipotezinin testi için ρ'nun tahmin edicisi $\hat{\rho}_{n,\tau}$ ile gösterilmek üzere, $n(\hat{\rho}_{n,\tau} - 1)$ istatistiğinin dağılımına ait kritik değerler kullanılır. Bu testin sonucuna göre trend içeren $AR(1)$ modelinin durağanlığına karar verilir. Burada $n(\hat{\rho}_{n,\tau} - 1)$ istatistiğinin dağılımı yine Dickey-Fuller dağılımı olarak adlandırılan dağılım türlerindendir.

3.2.2 Genişletilmiş Dickey-Fuller (Augmented Dickey-Fuller, ADF) birim kök testi

Bir önceki kısımda birinci dereceden bir otoregresif serinin durağanlığına ait Dickey-Fuller birim kök testi ile ilgili bilgiler verildi. Bu kısımda ise derecesi p olan bir AR modelinde birim kök testinin nasıl yapılacağı üzerinde durulacaktır.

Derecesi p olan bir otoregresif zaman serisi modeli,

$$(Y_t - \mu) = \sum_{i=1}^{p} \alpha_i (Y_{t-i} - \mu) + e_t \quad , \quad t = 1,2,\dots,n \qquad (3.7)$$

şeklinde verildiğinde, karakteristik denklemin köklerinden en az bir tanesi mutlak değerce bire eşit ise seri birim köklüdür. Bu nedenle serinin birim köklü olup olmadığının tespiti için karakteristik denklemin köklerinden birinin mutlak değerce bire eşit olup olmadığının test edilmesi gerekmektedir. Burada, eşitlik (3.7)'de verilen $AR(p)$ modeli $i = 1,2,\dots,p$ için,

$$\delta_1 = \sum_{i=1}^{p} \alpha_i$$

ve

$$\delta_i = -\sum_{j=i}^{p} \alpha_j$$

olmak üzere,

$$Y_t = \delta_1 Y_{t-1} + \sum_{i=2}^{p} \delta_i (Y_{t-i} - Y_{t-i+1}) + e_t \quad , \quad t = 1,2,\dots,n$$

şeklinde yazılabilir. Karakteristik denklemin köklerinden bir tanesinin mutlak değerce bire eşit olmasının gerek ve yeter koşulunun $\delta_1 = \alpha_1 + \alpha_2 + \cdots + \alpha_p$=1 eşitliğinin sağlanması olduğundan önceki bölümlerde bahsedilmişti. O halde, serinin birim köklü olduğunun testi ile $H_0: \delta_1 = 1$ hipotezinin testi aynıdır. Burada da verilen yokluk hipotezinin testi için $nc(\hat{\delta}_{1,n} - 1)$ istatistiğinin dağılımından yararlanılır. Burada verilen c herhangi bir sabit, $\hat{\delta}_{1,n}$ ise δ_1'in en küçük kareler tahmin edicisidir. Ayrıca, test için kullanılacak test istatistiğinin ($\hat{\tau}_{1,n}$) değeri de modelin beklenen değer veya trend içerdiği durumlara göre Dickey-Fuller birim kök testine benzer olarak hesaplanarak kritik değerler ile karşılaştırılarak karar verilir. Burada test istatistiğinin dağılımı yine Dickey-Fuller dağılım türlerindendir.

3.2.3 Phillips-Perron birim kök testi

Literatürde sıklıkla kullanılan birim kök testlerinden bir tanesi de Phillips ve Perron (1988) tarafından geliştirilen birim kök testi yöntemidir. Bu yöntem Dickey-Fuller birim kök testi yöntemine benzemesine karşın temel varsayımlarda farklılık göstermektedir. Dickey-Fuller birim kök testi yöntemi hata terimlerinin bağımsız ve aynı dağılımlı

rasgele değişkenler olduğu varsayımına dayanırken, Phillips-Perron birim kök testi yöntemi hata terimlerinin otokorelasyonlu olduğu durum düşünülerek geliştirilmiştir. Yani Phillips-Perron birim kök testi yönteminde T bir indis kümesi olmak üzere $\forall t \in T$ için $e_t \sim WN(0, \sigma^2)$ varsayımı sağlanmamaktadır.

Phillips ve Perron Dickey-Fuller test istatistiklerine bir düzeltme faktörü ekleyerek yeni bir yöntem önermişlerdir. Önerdikleri bu test istatistiklerine ait kritik değerler Dickey-Fuller test istatistiklerine ait kritik değerler ile aynıdır. Bu nedenle iki yöntem için karar aşamasındaki farklılık kritik değerlerden dolayı değil, test istatistiğinin hesaplanan değerlerindeki farklılıktan kaynaklanır. Uygulamada iki birim kök testi de benzer sonuçlar vermekle birlikte, Phillips-Perron birim kök testinin kritik değerleri ile Dickey-Fuller birim kök testinin kritik değerlerinin aynı olmasından dolayı birçok istatistiksel paket program bu yöntem için de kritik değerleri doğrudan vermektedir.

Birinci dereceden otoregresif ($AR(1)$) model,

$$(Y_t - \mu) = \rho(Y_{t-1} - \mu) + v_t \quad , \quad t = 1,2, \dots, n$$

şeklinde verilsin. Burada $\alpha_0 = \mu(1 - \rho)$ olmak üzere modelin,

$$Y_t = \alpha_0 + \rho Y_{t-1} + v_t \quad , \quad t = 1,2, \dots, n$$

şeklinde yazılabileceğini önceki bölümlerden biliyoruz. Yazılan bu model için, $e_t \sim WN(0, \sigma^2)$ ve $\varphi(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ olmak üzere, $v_t = \varphi(B) e_t$ olsun. Ayrıca, Y_t'nin Y_{t-1} üzerine regresyonundan elde edilen artıklar $\hat{v}_t$ ve bu artıklardan oluşan serinin otokovaryansları $\hat{\gamma}(h)$ ile gösterilmek üzere,

$$\hat{\lambda}^2 = \hat{\gamma}(0) + 2 \sum_{j=1}^{q} [1 - j/(q+1)] \hat{\gamma}(j)$$

eşitliği yazılır. Yapılan bu regresyondan elde edilen hata kareler ortalaması da

$$S_n^2 = \frac{1}{n-2}\sum_{t=1}^{n} \hat{v}_t^2$$

dir. Buna göre, $H_0: \rho = 1$ yokluk hipotezinin testi için verilen Phillips-Perron birim kök testine ilişkin test istatistiği,

$$Z_\mu = \left(\sqrt{\frac{\hat{\gamma}(0)}{\hat{\lambda}^2}}\right)\hat{\tau}_\mu - \frac{1}{\hat{\lambda}}\left[\frac{1}{2}\left(\hat{\lambda}^2 - \hat{\gamma}(0)\right)\left(ns_{\hat{\rho}}/S_n\right)\right]$$

şeklindedir. Burada $\hat{\tau}_\mu$ Dickey-Fuller birim kök testinde modelin beklenen değerinin sıfırdan farklı olduğu durum için verilen test istatistiğini, $s_{\hat{\rho}}$ değeri ise ρ'nun standart hatasını göstermektedir.

Phillips-Perron birim kök testi yöntemi de benzer şekilde yüksek dereceden modeller için ve modelin beklenen değerinin sıfır olması durumu ile modelin deterministik bir trend içeriyor olması durumu için genişletilebilir.

İkinci bölümde öngörü yapabilmek için rasgele olarak 100 birimlik zaman serisi üretilmiş olup, üretilen bu verilerin $AR(2)$ modeline uygun olduğuna AIC ve SBC değerleri yardımıyla karar verilmişti. Ayrıca serinin otokorelasyon fonksiyonunun grafiğindeki azalmanın hızlı olması nedeniyle serinin durağan olabileceği düşünülmüşse de gerek α_1 ve α_2 parametrelerinin tahminlerinin toplamının 1'e yakın olması nedeniyle gerekse de öngörü değerlerinin ilerleyen dönemler için serinin ortalamasına yaklaşmaması nedeniyle, serinin durağanlığından şüphe edilmişti. Bu durum, serinin durağanlığının tespitinde kesin bir karar vermek için birim kök testlerinin önemini gözler önüne sermektedir. Bu nedenle, üretilen bu rasgele veriler kullanılarak serinin durağanlığı genişletilmiş Dickey-Fuller birim kök testi ve Phillips-Perron birim kök testleri ile sınanmış olup, sonuçlar çizelge 3.1-3.2'de verilmiştir.

Çizelge 3.1 Rasgele üretilen veriye ilişkin ADF testi sonuçları

	test istatistiği değeri	**Prob.**
ADF testi kritik değerleri	-3.577	0.008
$\boldsymbol{\alpha = 0.01}$ **için kritik değer**	-3.498	
$\boldsymbol{\alpha = 0.05}$ **için kritik değer**	-2.891	
$\boldsymbol{\alpha = 0.10}$ **için kritik değer**	-2.583	

Çizelge 3.2 Rasgele üretilen veriye ilişkin Phillips-Perron testi sonuçları

	test istatistiği değeri	**Prob.**
PP testi kritik değerleri	-2.802	0.061
$\boldsymbol{\alpha = 0.01}$ **için kritik değer**	-3.498	
$\boldsymbol{\alpha = 0.05}$ **için kritik değer**	-2.891	
$\boldsymbol{\alpha = 0.10}$ **için kritik değer**	-2.583	

Çizelge 3.1 incelendiğinde ADF testi sonuçlarına göre, teste ilişkin p-değeri 0.05'den küçük olduğundan, $\alpha = 0.05$ anlam düzeyinde H_0:"$Seri\ birim\ kök lüdür.$" yokluk hipotezi H_a:"$Seri\ durağandır.$" yokluk hipotezine karşı reddedilir. Yani, serinin durağan olduğu söylenebilir. Bunun aksine, çizelge 3.2 incelendiğinde ise Phillips-Perron testine ilişkin p-değerinin 0.05'den az da olsa büyük olduğu görünmektedir. Ancak, Phillips-Perron birim kök testlerinin genellikle modelin MA kısmını içerdiği durumlarda kullanıldığı göz önünde bulundurulursa, ADF testi sonuçlarına göre serinin durağan olduğu söylenebilir.

3.2.4 HEGY yöntemi

Bir zaman serisine ait veriler bazen periyodik bir yapı gösterir. Bu gibi durumlarda genellikle mevsimsel zaman serisi modelleri ile verilerin modelleme işlemi gerçekleştirilir. Bu periyodikliğin varlığı sezgisel olarak veriye ait zaman serisi grafiğinden veya otokorelasyon grafiğinden anlaşılabilir. İlgilenilen verinin

modellenmesinde mevsimsel birim kök probleminin var olması durumunda istatistiksel sonuç çıkarımı yapılamamaktadır. Mevsimsel birim kökün tespitinde Dickey-Fuller tarafından geliştirilen birim kök testinin iyi sonuç vermemesinden dolayı mevsimsel birim kökün tespiti için birçok farklı yöntem ortaya atılmıştır.

Örneğin, bir $SAR_4(1)$ modeli ile ilgilenilsin. Bu model $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$Y_t = \rho Y_{t-4} + e_t \quad , \quad t = 1,2, \dots, n$$

şeklinde verilmiş olsun. Bu durumda seriye ait karakteristik denklem,

$$m^4 - \rho = 0$$

dır. Bu denklem dördüncü dereceden bir polinom olup dört tane kökü bulunmaktadır. Burada $|\rho| < 1$ ise model durağan aksi halde ($\rho = 1$), model birim köklüdür. Bu durumda polinomun kökleri $m_{1,2} = \pm 1$ ve $m_{3,4} = \pm i$ şeklindedir. Burada $i = \sqrt{-1}$'dir. Yani, modelin karakteristik denkleminin köklerinin tamamı mutlak değerce bire eşittir. Böyle bir modelin durağan hale getirilebilmesi için serinin dört kez farkının alınması gerekmektedir. Serinin dördüncü dereceden farkı ∇ fark operatörünü ve B gerileme (backshift) operatörünü göstermek üzere,

$$\nabla^4 Y_t = (1 - B)^4 = (1 - 4B + 6B^2 - 4B^3 + B^4) Y_t$$

şeklindedir. Böylelikle model karmaşık bir hal alır. Ayrıca, serinin dördüncü dereceden farkının alınması birçok bilgi kaybına neden olur. Bu nedenle, serinin dördüncü farkının alınması yerine,

$$\nabla_4 Y_t = (1 - B^4) Y_t = Y_t - Y_{t-4} = e_t$$

dönüşümünün yapılması modeli durağan hale getirir. Böyle bir dönüşümün yapılması ile hem bilgi kaybı azaltılmış olur hem de daha sade bir regresyon modeli ile ilgilenilir. Bu nedenle, mevsimsel birim kökün tespiti önemlidir.

Mevsimsel birim kökün tespitindeki bu zorluklar birçok farklı mevsimsel birim kök testinin ortaya atılmasına neden olmuştur. Literatürde sıklıkla kullanılan ve pratik mevsimsel birim kök tespit yöntemlerinden bir tanesi Hylleberg vd. (1990) tarafından önerilen ve HEGY testi olarak bilinen yöntemdir. Bu yöntem yardımıyla mevsimsel birim kök tespiti yapılırken $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$Y_t = \rho Y_{t-4} + e_t \quad , \quad t = 1,2, \dots, n$$

modeli için,

$$\delta(B) Z_{4,t} = \pi_1 Z_{1,t-1} + \pi_2 Z_{2,t-1} + \pi_3 Z_{3,t-2} + \pi_4 Z_{3,t-1} + e_t \tag{3.8}$$

şeklindeki yardımcı regresyon modelinden yararlanılır. Burada $Z_{i,t}$ değişkenleri $i = 1,2,3,4$ için,

$$Z_{1,t} = (1 + B + B^2 + B^3) Y_t$$

$$Z_{2,t} = -(1 - B + B^2 - B^3) Y_t$$

$$Z_{3,t} = -(1 - B^2) Y_t$$

$$Z_{4,t} = (1 - B^4) Y_t$$

şeklindedir. HEGY yöntemiyle birim köklerin tespiti için bu değişkenlerin öncelikle gecikme (backshift) operatörü yardımıyla gerekli dönüşümlerin gerçekleştirilerek belirlenmesi gerekmektedir. Daha sonra eşitlik (3.8)'de verilen yardımcı regresyon modeline ait parametre tahminleri yapılmalıdır. Bu modelde yer alan π_i parametreleri

$i = 1,2,3,4$ için mevsimsel köklerin katsayılarını göstermektedirler. Bu parametrelerin anlamlılığının testi de serinin birim kök testlerine karşılık gelmektedir. Buna göre, $H_0: \pi_1 = 0$ hipotezinin $H_a: \pi_1 < 0$ alternatifine karşı testi için t test istatistiğinin hesaplanan değeri kritik değerler ile karşılaştırılır ve hipotez reddedilemez ise seri birim köklüdür yorumu yapılır. Daha sonra $H_0: \pi_2 = 0$ hipotezinin testi benzer şekilde yapılarak hipotezin reddedilememesi durumunda seri altı aylık dönemlerde yani iki devirli birim köke sahiptir yorumu yapılır. Son olarak $H_0: \pi_2 = \pi_3 = 0$ hipotezini test etmek için F test istatistiğinin hesaplanan değeri ile kritik değer karşılaştırılır. Yokluk hipotezi reddedilemez ise seri mevsimsel birim köklüdür yorumu yapılır (Türe ve Akdi 2006). Burada, test istatistiklerinin hesaplanan değerlerinin karşılaştırılacağı kritik değerler yine Hylleberg vd. (1990) tarafından verilmiştir. Bu işlemler benzer olarak modelin ortalamasının sıfırdan farklı olduğu veya modelin deterministik bir trend içerdiği durumlar için de yapılabilir. Ayrıca, modele deterministik kukla (dummy) değişkenler eklenerek de analizler gerçekleştirilebilir. Bu durumda bir $AR(p)$ modelinin mevsimsel birim köke sahip olup olmadığının sınanması için,

$$\delta_1(B) = (1 + B + B^2 + B^3), \delta_2(B) = -(1 - B + B^2 - B^3), \delta_3(B) = -(1 - B^2)$$

ve kukla değişkenler K_1, K_2, K_3, K_4 olmak üzere,

$$Z_{4,t} = c_1 K_{1,t} + c_2 K_{2,t} + c_3 K_{3,t} + c_4 K_{4,t} + \pi_1 \delta_1(B) Y_{t-1} + \pi_2 \delta_2(B) Y_{t-1} + \pi_3 \delta_3(B) Y_{t-2} + \pi_4 \delta_3(B) Y_{t-1} + \sum_{i=1}^{p-1} \alpha_i (1 - B^4) Y_{t-i} + \theta t + e_t$$

şeklindeki model kullanılır. Gözlenen verilerin aylık olması durumunda, $e_t \sim WN(0, \sigma^2)$ olmak üzere,

$$Y_t = \rho Y_{t-12} + e_t$$

modelinin mevsimsel birim kök içerip içermediğinin kontrolü için ise

$$\delta(B)Z_{8,t} = \pi_1 Z_{1,t-1} + \pi_2 Z_{2,t-1} + \pi_2 Z_{2,t-1} + \pi_3 Z_{3,t-1} + \pi_4 Z_{3,t-2} + \pi_5 Z_{4,t-1} + \pi_6 Z_{4,t-2} + \pi_7 Z_{5,t-1} + \pi_8 Z_{5,t-2} + \pi_9 Z_{6,t-1} + \pi_{10} Z_{6,t-2} + \pi_{11} Z_{7,t-1} + \pi_{12} Z_{7,t-2} + e_t$$

yardımcı regresyon modelinden yararlanılır (Franses ve Hobijn 1997). Burada, $i = 1,2,\dots,8$ için $Z_{i,t}$ değişkenleri,

$$Z_{1,t} = (1+B)(1+B^2)(1+B^4+B^8)Y_t$$

$$Z_{2,t} = -(1-B)(1+B^2)\,(1+B^4+B^8)Y_t$$

$$Z_{3,t} = -(1-B^2)(1+B^4+B^8)Y_t$$

$$Z_{4,t} = -(1-B^4)(1-B\sqrt{3}+B^2)(1+B^2+B^4)Y_t$$

$$Z_{5,t} = -(1-B^4)\left(1+B\sqrt{3}+B^2\right)(1+B^2+B^4)Y_t$$

$$Z_{6,t} = -(1-B^4)(1-B^2+B^4)(1-B+B^2)Y_t$$

$$Z_{7,t} = -(1-B^4)(1-B^2+B^4)(1+B+B^2)Y_t$$

$$Z_{8,t} = (1-B^{12})Y_t$$

şeklindedir. Verilerin aylık olarak ele alındığı bu durumda da mevsimsel kökün varlığına verilen yardımcı regresyon modelindeki $\pi_j, j = 1,2\dots,12$ parametrelerinin anlamlılığının testleri yardımıyla bakılır. Yani, $H_0: \pi_1 = 0$ hipotezine ait test istatistiğinin hesaplanan değeri, kritik değerler ile karşılaştırılarak karar verilir. Eğer $H_0: \pi_1 = 0$ yokluk hipotezi reddedilemiyor ise seri mevsimsel köke sahiptir. Bu teste ilişkin kritik değerler Franses ve Hobijn (1997) tarafından verilmiştir.

Durağanlığın tespiti ve birim kök probleminin giderilmesi, zaman serilerinde üzerinde sıklıkla çalışılan bir durumdur. Bir zaman serisinin durağanlığı birim kök testleri ile sınanır. Bu bölümde literatürde çok sık karşılaşılan birim kök testlerinden bazıları

detaylarına girilmeden anlatıldı. Bu birim kök testlerinin yanında bir sonraki bölümde bahsedilecek olan periodogramlara dayalı birim kök testi ile de serinin birim köklü olup olmadığının testi yapılabilir.

Serinin mevsimsellik içerdiği durumda, mevsimsel birim kökün tespiti oldukça önemlidir. Bunun nedeni, mevsimsel birim kökün varlığı durumunda bilinen fark alma işleminin yerine mevsimsel fark alma işlemi uygulanmasıdır. Mevsimsel birim kökün var olup, tespit edilemediği durumda ise modelin periyodu kadar fark alma işlemi uygulanarak seri içerisinde birçok bilgi kaybı meydana getirilmektedir. Bir sonraki bölümde, bu bölümde verilen HEGY yöntemine ek olarak periodogramlara dayalı birim kök testlerinin mevsimsellik içeren serilerde uygulanabilirliğine ve HEGY yöntemine göre olan avantajlarına değinilecektir.

4. UYGULAMA

Türkiye nüfusu hızlı bir şekilde artarken son yıllarda kırsal kesimden şehirlere göç de giderek artmaktadır. Türkiye İstatistik Kurumu (TÜİK) verilerine göre, Türkiye nüfusu 1980 yılında 43905790 iken bu sayı 2000 yılında 63174483 olurken 2017 yılı itibari ile 80810525 sayısına ulaşmıştır. Türkiye demografisinde görülen en önemli değişim kentleşme oranıdır. 1927 yılında nüfusun yüzde 75.8'i kırsal (10000'den az nüfuslu), yüzde 24.2'si kentsel alanlarda yaşarken, bugün bu oran tam tersine dönmüştür. 2011 yılı itibari ile Türkiye nüfusunun yüzde 23.2'si kırsal alanda (belde ve köyler) yaşamaktadır. Aynı oran büyük şehirlerde de görülmektedir. Örneğin, Ankara'da 2000 yılında nüfus 3889199 iken 2010 yılında 4771716 olurken bu sayı 2017 itibari ile 5445026'ya ulaşmıştır. Ankara'daki nüfusun 2017 verilerine göre, 5135070'i (yaklaşık yüzde 94.3'ü) şehir merkezinde yaşamaktadır. Nüfus oranlarındaki bu artış birçok sosyal sorunun yanında çevre sorunlarını da beraberinde getirmektedir. Benzer artışlar trafikte bulunan araç sayılarında da görülmektedir. Yine TÜİK verilerine göre, Türkiye'de 2000 yılında trafiğe kayıtlı araç sayısı 8665170 iken 2010 yılında 15095603 sayısına, 2017 yılı itibarı ile de 22218945 sayısına ulaşmıştır. Yani 17 yılda trafikte bulunan araç sayısı yüzde 100'den fazla bir artış göstermiştir.

Hava kirliliği insan ve çevre sağlığını, ekolojik düzeni doğrudan etkileyen bir olumsuzluktur. Hava kirliliği verileri çeşitli nedenlerden dolayı genellikle periyodik bir yapı içermektedir. Bu nedenlerin başında kış aylarında ısınma nedeniyle kullanılan karbon içerikli yakıtlar, trafiğin artış gösterdiği özel zamanlar ve meteorolojik bazı etkenler gelmektedir.

Türkiye'nin ikinci büyük şehri ve başkenti Ankara'da hava kirliliği sürekli bir sorun olmuştur. Ankara'da son verilere göre yaklaşık beş milyondan fazla insan yaşamakta olup bu nüfusun çok önemli bir bölümü şehir merkezinde yaşamaktadır. Nüfus artışı, buna bağlı olarak trafikteki araç sayısı ve fosil yakıtların tüketiminin artışı bu sorunun en önemli kaynakları olarak gösterilebilir. Özellikle 1988'de başlayan doğalgaz kullanımı, kurşunlu benzin tüketiminin kaldırılması, doğalgaz ile çalışan toplu taşıma araçlarının kullanımına ağırlık verilmesi, artışı bir ölçüde yavaşlatsa da hava kirliliği

Ankara'da özellikle kış aylarında hissedilebilir düzeyde bir sorun olduğu açıktır. Bu çalışmada, insan sağlığına zararlı hava kirleticilerinden birisi olan, Ankara iline ait PM_{10} miktarları üzerinde durulacaktır. PM_{10}, aerodinamik çapı 10μm'den küçük havada aerosol formda bulunan partikül maddeyi göstermektedir. Dünya Sağlık Örgütü (WHO), 2005 yılında partikül madde miktarı standartlarını yıllık ortalama $20\,\mu g/m^3$ ve 24 saatlik ortalama $50\,\mu g/m^3$ olarak belirlemiştir (Anonymous 2005). Avrupa Parlamentosu ve Avrupa Konseyi ise 21 Mayıs 2008 tarihli yönergesinde yıllık ortalama PM_{10} miktarı sınır değerini $40\,\mu g/m^3$ ve 24 saatlik ortalama PM_{10} sınırını ise 50 $\mu g/m^3$ olarak belirlemiş ancak bu değerlerin yılda en fazla 35 defa aşılabileceğine izin vermiştir (Anonymous 2008). Türkiye'de ise bu değerler 2016 yılında yıllık ortalama $52\,\mu g/m^3$ ve 24 saatlik ortalama $80\,\mu g/m^3$ olarak belirlenmiş olup bu değerler 2018 yılında yıllık ortalama $44\,\mu g/m^3$ ve 24 saatlik ortalama ise $60\,\mu g/m^3$ olarak güncellenmiştir (Anonim 2018). Bu değerler dikkate alındığında, Ankara'da elde edilen değerler sınırların oldukça üzerinde olup önlem alınmasını gerektirmektedir. Çalışmada kullanılan veriler, Ocak 1993 ile Aralık 2017 dönemleri arasında, Ankara'da bulunan sekiz farklı istasyondan elde edilen aylık ortalama PM_{10} miktarlarıdır.

Bu çalışmada amaç, Ocak 1993 ile Aralık 2017 döneminde aylık olarak Ankara'da gözlenen PM_{10} miktarları için uygun bir zaman serisi modeli belirleyerek gelecek dönemler için öngörülerde bulunmaktır. Bununla birlikte, hava kirliliğinin mevsimlere hatta aylara göre değişkenlik gösterdiği göz önüne alındığında, verilerde herhangi bir periyodikliğin bulunup bulunmadığı araştırılacaktır. Verilerde görülen periyodiklik dikkate alınarak klasik zaman serisi modellerinin yanında, harmonik regresyon teknikleri kullanılarak da bir öngörü modeli elde edilecektir. Bu iki modelden elde edilen sonuçlar istatistiksel olarak karşılaştırılacak ve 2018 yılında elde edilen veriler dikkate alınarak öngörü değerleri karşılaştırılacaktır. Verilerin zaman serisi grafiklerinden ve aylık ortalamalar dikkate alındığında kış aylarında gözlenen yüksek değerler tekrarlanmakta olup yaz aylarında nispeten düşük değerler gözlenmektedir. Bu durum, verilerin mevsimsellik özelliği gösterdiğine bir işarettir. Bununla birlikte, mevsimsellikten öte, varsa bir periyodikliğin araştırılması da periodogramlar yardımıyla yapılacaktır. Verilerin içerdiği mevsimsellikten dolayı zaman serisinin durağanlığının sınaması ise genişletilmiş Dickey-Fuller birim kök testine ve Phillips-Perron birim kök

testine ek olarak, HEGY testi ve periodogramlara dayalı birim kök testi ile yapılacaktır. Çalışmada kullanılan verilerden, 1993 Ocak – 2011 Ekim ayları arasındaki ölçümler Türkiye İstatistik Kurumu'nun web sitesinden, 2011 Kasım – 2017 Aralık ayı arasındaki ölçümler ise Çevre ve Şehircilik Bakanlığı'nın Hava Kalitesi İzleme İstasyonları web sitesinden alınmıştır. Ancak 2006 Mayıs – 2006 Aralık ayları arasındaki ölçümlere resmi kaynaklardan ulaşılamamış olup, bu aylara ilişkin kayıp gözlemler o döneme kadar elde edilen gözlemler ile öngörülerek yerlerine konmuştur. Kayıp gözlem problemi bu şekilde ortadan kaldırılmıştır.

4.1 Periyodik Zaman Serileri Yaklaşımı

Reel sayılar üzerinde tanımlı herhangi bir f fonksiyonu eğer bütün $x \in \mathbb{R}$'ler için $f(x+p) = f(x)$ özelliğini sağlıyorsa, f'ye periyodu p olan periyodik bir fonksiyon denir. Kolayca görülebileceği gibi f periyodu p olan periyodik bir fonksiyon ise $f(x+2p) = f\big((x+p)+p\big) = f(x+p) = f(x)$ olduğundan p'nin bütün katları da aynı fonksiyonun periyodudur. Yani, bir fonksiyonun birçok periyodu bulunabilir. Bu periyodların en küçüğüne fonksiyonun asli periyodu (ya da kısaca periyodu) denir.

Bir zaman serisinin durağanlığın sınanmasına ilişkin literatürde oldukça fazla yöntem bulunmasına rağmen, parametrelerin en küçük kareler tahmin edicilerinin dağılımına dayanan ve Dickey-Fuller birim kök testi olarak bilinen yöntem ile yine bu dağılımın kritik değerlerini kullanan Phillips-Perron birim kök testi olarak bilinen iki yöntem pratikliği açısından öne çıkmaktadır. Zaman serilerinde MA türü seriler her zaman durağan olup, AR veya AR bileşeni içeren modeller durağan olmayabilir. Bununla birlikte, mevsimsel serilerin durağanlığının sınanmasında HEGY türü testlerin kullanılmasının daha anlamlı olduğundan da önceki bölümlerde bahsedildi. Serinin periyodikliğinden şüphelenilmiyorsa durağanlığın testi için yukarıdaki testlerden biri kullanılabilir.

Periodogramlar genellikle seride bulunan gizli periyodikliklerin ortaya çıkarılmasında kullanılır (Brockwell ve Davis 1987, Fuller 1996, Wei 2006). Bununla birlikte Akdi ve

Dickey (1998) serinin birim köklü olup olmadığının sınanması için periodogramlara dayalı bir birim kök testi önermişlerdir. Periodogramlar ile ilgili genel açıklamalar aşağıdadır.

Periyodik fonksiyonlar denildiğinde ilk trigonometrik fonksiyonlar aklımıza gelir. O halde, serinin periyodik bir bileşen içerip içermediğini sınamak için $\{Y_1, Y_2, \dots, Y_n\}$ gibi verilen herhangi bir zaman serisini

$$Y_t = \mu + Rcos(wt + \Phi) + e_t \, , t = 1,2, \dots, n$$

şeklinde yazalım. Buradaki μ, R, Φ ve w terimleri sırası ile beklenen değer, genişlik (genlik, amplitude), faz (phase) ve sıklıkları (frequency) ifade etmektedir. Bu parametrelerin tahmin edilmesi gerekir. Ayrıca, $w_k = 2\pi k/n$ olarak seçildiğinde Fourier sıklıkları elde edilir. Kosinüs fonksiyonunun özelliğinden $\alpha = Rcos(\Phi)$ ve $\beta = Rsin(\Phi)$ olmak üzere model,

$$Y_t = \alpha \cos(w_k t) + \beta \sin(w_k t) + e_t \ \, , t = 1,2, \dots, n$$

şeklinde yazılabilir. Bu modele göre $H_0: \alpha = \beta = 0$ yokluk hipotezi H_a: "$\alpha \; veya \; \beta' dan \; en \; az \; birisi \; sıfırdan \; farklı$" alternatif hipotezine karşı reddedilirse veride periyodik bileşen vardır. Bu hipotezin testi için standart F testi kullanılabilir. Ancak, w_k sıklıkları bilinmediği durumda F istatistiğinin kullanılması anlamlı değildir (Wei 2006). Bu modele göre μ, α ve β parametrelerinin en küçük kareler tahmin edicileri sırasıyla,

$$\hat{\mu} = \bar{Y}_n$$

$$a_k = \frac{2}{n} \sum_{t=1}^{n} (Y_t - \bar{Y}_n) \cos(w_k t)$$

$$b_k = \frac{2}{n}\sum_{t=1}^{n}(Y_t - \bar{Y}_n)\sin(w_k t)$$

şeklindedir. Hesaplanan a_k ve b_k değerlerine Fourier katsayıları denir. Yine trigonometrik fonksiyonların özelliklerinden,

$$\sum_{t=1}^{n}\cos(w_k t) = \sum_{t=1}^{n}\sin(w_k t) = 0$$

olup, Fourier sıklıkları ortalamaya göre değişmezdir (invariant). Bu Fourier katsayıları yardımı ile zaman serisinin w_k sıklığındaki periodogramı (periodogram ordinate),

$$I_n(w_k) = \frac{n}{2}(a_k^2 + b_k^2)$$

şeklinde hesaplanır.

Zaman serileri genellikle iki grupta incelenir. Bunlardan ilki zaman boyutu (time domain), ikincisi ise sıklık boyutudur (frequency domain). Zaman boyutunda en önemli araç serinin otokorelasyon fonksiyonudur. Sıklık boyutunda ise spektral yoğunluk fonksiyonudur. Ayrıca, serinin otokorelasyon fonksiyonu ile spektral yoğunluk fonksiyonu arasında da bir geçiş vardır (Herglotz Teoremi, Brokwell ve Davis 1987, sayfa 115). Eğer $f(w_k)$ durağan zaman serisinin spektral yoğunluk fonksiyonu ise $I_n(w_k)/f(w_k)$ istatistiğinin asimptotik dağılımı serbestlik derecesi 2 olan ki-kare (beklenen değeri 1 olan üstel) dağılımıdır. Yani, normalleştirilmiş periodogramların dağılımının olasılık yoğunluk fonksiyonu,

$$f(x) = \begin{cases} e^{-x} & , \ x > 0 \\ 0 & , \ d.y. \end{cases}$$

dir. Dolayısı ile periodogramlar spektral yoğunluk fonksiyonunun bir tahmin edicisi olarak alınabilir. Periodogramlar veride muhtemel periyodikliklerin araştırılmasında (yani yukarıdaki $H_0: \alpha = \beta = 0$ hipotezinin test edilmesinde) da kullanılmaktadır. Durağan herhangi bir zaman serisi için her bir k sıklığındaki $I_n(w_k)$ periodogram değerleri hesaplanır. $I_n(w_{(1)})$ en büyük periodogram değerini, m de $n/2$'nin tam kısmını ($m = [|n/2|]$) göstermek üzere,

$$V = I_n(w_{(1)}) \left[\sum_{k=1}^{m} I_n(w_k) \right]^{-1}$$

istatistiğini tanımlayalım. Herhangi bir periyodik bileşen yoksa ($H_0: \alpha = \beta = 0$ yokluk hipotezi altında) $P(V > c_\alpha)$ olmadığı için,

$$P(V > c_\alpha) = \alpha \cong m(1 - c_\alpha)^{m-1}$$

yaklaşık değeri kullanılmaktadır (Wei 2006). Buradan, seçilen herhangi bir α anlam düzeyi için c_α kritik değeri,

$$c_\alpha = 1 - \left(\frac{a}{m}\right)^{1/(m-1)}$$

olarak hesaplanır. Eğer $V > c_\alpha$ ise $H_0: \alpha = \beta = 0$ yokluk hipotezi H_a: "$\alpha\ veya\ \beta' dan\ en\ az\ birisi\ sıfırdan\ farklı$" alternatif hipotezine karşı reddedilir ve seride periyodik bileşen olduğu sonucuna varılır.

Verilen zaman serisi durağan ise normalleştirilmiş periodogramların dağılımının serbestlik derecesi 2 olan ki-kare dağılımı olduğundan bahsedildi. Yani, $f(w_k)$ durağan zaman serisinin spektral yoğunluk fonksiyonu olmak üzere,

$$I_n(w_k)/f(w_k) \xrightarrow{D} \chi_2^2 \quad , \quad n \to \infty$$

dir (Fuller 1996). Serinin durağan olmadığı (birim köklü) varsayımı altında her sabit w_k için,

$$T_n(w_k) = \frac{2(1-\cos(w_k))}{\hat{\sigma}_n^2} I_n(w_k) \xrightarrow{D} Z_1^2 + 3Z_2^2 \ , n \to \infty$$

şeklindedir (Akdi 1995, Akdi ve Dickey 1998). Burada Z_1 ve Z_2 bağımsız standart normal dağılımlı rasgele değişkenler olup $\hat{\sigma}_n^2$ hata teriminin varyansının bir tahmin edicisidir. Kısaca asimptotik dağılım,

$$T_n(w_k) \xrightarrow{D} \chi_1^2 + 3\chi_1^2 \ , n \to \infty$$

dir. Yine Akdi ve Dickey (1999) yöntemin mevsimsel zaman serileri için de uygulanabilir olduğunu göstermişlerdir. Yani, $T_n(w_k)$ istatistiği mevsimsel serilerin durağanlığını sınamak için de kullanılabilir. Asimptotik dağılım her sabit w_k için geçerli olmakla birlikte hipotez testlerinde genellikle w_1 sıklığı kullanılmaktadır. Dağılımın kritik değerleri ilgili çalışmada verilmiştir. Seride özellikle periyodik bir bileşen varsa, bu yöntemin diğer yöntemlere göre avantajlı olduğu söylenebilir. Bununla birlikte,

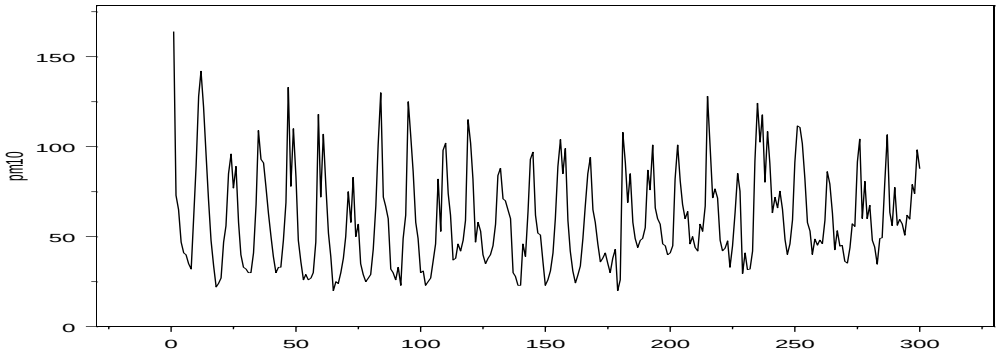
i. Yöntemde herhangi bir model varsayımı yoktur ve metot modele göre değişmezdir (invariant). Yani periodogramlar herhangi bir model varsayımına bağlı olmadan hesaplanabilir.
ii. Bilindiği gibi, diğer birim kök testleri zaman serisi modeline bağlı olduğundan, öncelikle model parametrelerinin tahminine ihtiyaç duyulur. Burada ise, beyaz gürültü serisinin varyansı haricinde herhangi bir parametre tahminine ihtiyaç duyulmaz.
iii. $T_n(w_k)$ istatistiğinin dağılımı H_0: "$Seri\ birim\ köklüdür.$" yokluk hipotezi ve H_a = "$Seri\ durağandır.$" alternatif hipotezleri altında mevcut olduğundan testin güç fonksiyonu mevcuttur.
iv. Test istatistiğinin kritik değerleri örneklem hacmine bağlı değildir.

v. Periodogramlar verinin trigonometrik dönüşümleri ile hesaplandığı için eğer veride herhangi bir periyodik bileşen varsa, bu periyodikliği dikkate aldığından daha anlamlı olacağı açıktır.

gibi avantajları nedeniyle özellikle veride herhangi bir periyodik bileşenin bulunması durumunda diğer yöntemlere göre daha avantajlı olduğu düşünülebilir.

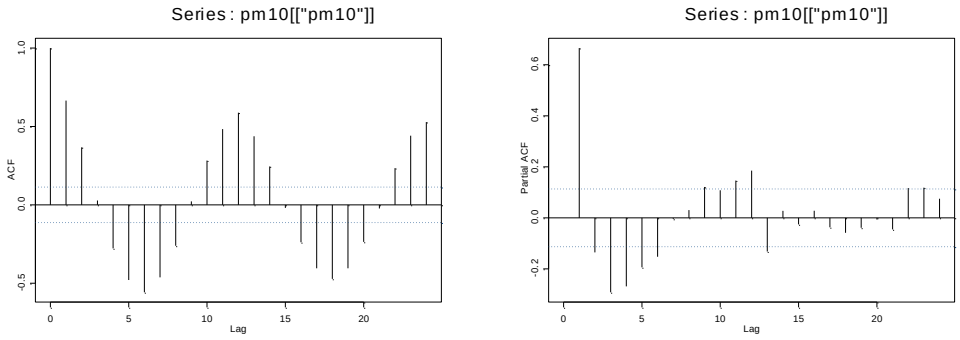
4.2 PM_{10} Analizi

Hava kirliliği günümüzün en önemli sağlık, çevre ve sosyal sorunlarından biridir. Hava kirliliğinin gelişen teknoloji ve artan nüfus ile birlikte giderek arttığı da bir gerçektir. Bu çalışmada, 1993:01-2017:12 dönemine ait Ankara'da 8 farklı istasyonda ölçülen toplam 300 aylık PM_{10} verileri analiz edilmeye çalışılmıştır. Verilere Çevre ve Şehircilik Bakanlığı internet sayfasından günlük, haftalık hatta saatlik olarak ulaşılabilmektedir. Çalışmada, 8 istasyonda elde edilen aylık verilerin ortalama değerleri kullanılmıştır. Ayrıca, 2006:05 – 2006:12 arası 8 kayıp gözlem vardır. Kayıp gözlemler, 1993:01 – 2006:5 dönemi verileri kullanılarak öngörü değerleri elde edilmiş, bu öngörü değerleri kayıp gözlemler için kullanılmıştır. Kullanılan verilerin zaman serisi grafiği şekil 4.1'de otokorelasyon ve kısmi otokorelasyon grafikleri ise şekil 4.2'de verilmiştir.



Şekil 4.1 PM_{10} verisine ait zaman serisi grafiği

Otokorelasyon (ACF) Kısmi Otokorelasyon (PACF)



Şekil 4.2 PM_{10} verisine ait otokorelasyon ve kısmi otokorelasyon grafikleri

Özellikle 90'lı yılların sonlarına doğru başlayan doğal gaz kullanımı nedeni ile PM_{10} miktarlarında önemli bir azalma görülmesine rağmen, değerler hala Avrupa Birliği, Dünya Sağlık Örgütü ve Birleşmiş Milletler standartlarının üzerindedir. Bununla birlikte aylık ortalamalar dikkate alındığında, kış aylarında ölçülen ortalama değerlerin diğer aylara göre daha fazla olduğu gözlenmektedir. En yüksek değerler Ocak, Kasım ve Aralık aylarında gözlenmektedir. En düşük değer Haziran ayında gözlenirken bunu Temmuz ve Ağustos ayları izlemektedir. Buna göre, PM_{10} değerlerinin kış aylarında yüksek, yaz aylarında ise düşük olduğu söylenebilir. Ancak, yaz aylarında gözlenen düşük değerler dahi uluslararası standartların üzerindedir. Verinin içerdiği periyodik yapı ise gerek verilen zaman serisi grafiği, gerekse otokorelasyon grafiğinden açık bir şekilde gözükmektedir.

Kış aylarında Ankara iline ait aylık ortalama PM_{10} miktarları dünya standartlarının yaklaşık olarak iki katı miktarında gözlenmektedir. Yaz aylarına ait ortalamalar incelendiğinde ise nispeten dünya standartlarına ulaşılmış gibi gözükse de ortalama değerler yine Dünya ve Avrupa standartları için sınırda bulunmaktadır. Bu durumun açık bir şekilde görülebilmesi açısından çizelge 4.1'de Ankara iline ait aylara göre ortalama PM_{10} miktarları ($\mu g/m^3$ cinsinden) ve standart sapmaları verilmiştir.

Çizelge 4.1 PM_{10} miktarlarının aylara göre ortalamaları ve standart sapmaları

Aylar	Ortalama	Standart Sapma
Ocak	**85.296**	**26.552**
Şubat	71.948	16.913
Mart	58.668	9.275
Nisan	49.168	13.799
Mayıs	39.544	10.572
Haziran	37.736	14.804
Temmuz	38.748	20.087
Ağustos	39.444	16.643
Eylül	50.340	16.920
Ekim	63.428	13.786
Kasım	**96.672**	**24.164**
Aralık	**88.188**	**24.093**

4.2.1 Zaman boyutu (time domain) yaklaşımı

Verilere uygun bir zaman serisi modeli belirleyebilmek için değişik modeller göz önüne alınmış ve en uygun model için en küçük AIC istatistiğinin değeri dikkate alınmıştır. Değişik modeller için SAS'da PROC ARIMA ile hesaplanan değerler ile bu modellere karşılık gelen varyans tahmini değerleri (parantez içerisinde) çizelge 4.2'de verilmiştir.

Çizelge 4.2 Önerilen modellere ilişkin AIC istatistiklerinin değerleri

1	2	3	4	12	13	(4)	(12)	(1,12)	(1,12,13)
2654.46 (404.91)	2651.16 (399.16)	2625.84 (365.65)	2605.44 (340.49)	2577.06 (301.86)	2573.59 (297.44)	2806.71 (672.62)	2700.64 (472.29)	2596.64 (332.83)	2587.93 (322.24)

Yukarıda verilen çizelge 4.2'de tablonun üst kısmındaki sayılar önerilen AR modelinin gecikmesini göstermektedir. Parantez içerisindeki ifadeler ise modelin içerisinde sadece o gecikmelerinin bulunduğu AR modellerini göstermektedir (örneğin, (4) ifadesi

modelde zaman serisinin sadece dördüncü gecikmesinin yer aldığı AR modelini göstermekte iken, (1,12) ifadesi modelde serinin kendisinin birinci ve on ikinci gecikmesinin yer aldığını gösterir). Buna göre AIC istatistik değeri en küçük olan $AR(13)$ modeli verilere en uygun model olarak öne çıkmaktadır. Bu nedenle, verilere

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_{13} + e_t \, , t = 1,2, \dots , n$$

şeklindeki $AR(13)$ modeli uygun olduğu varsayımı altında parametre tahminleri SAS PROC ARIMA ve SAS PROC REG ile hesaplanarak çizelge 4.3-4.4'de verilmiştir.

Çizelge 4.3 $AR(13)$ modeline ilişkin PROC REG ile elde edilen parametre tahminleri

Parametre	**Parametre Tahmini**	**Standart Hata**	**t-istatistiği değeri**	$\boldsymbol{Pr > \lvert t \rvert}$
α_0	28.994	7.454	3.890	$\mathbf{0.0001}$
α_1	0.546	0.060	9.160	$\mathbf{<.0001}$
α_2	0.062	0.066	0.960	0.337
α_3	-0.057	0.065	-0.880	0.381
α_4	-0.082	0.065	-1.250	0.211
α_5	-0.068	0.066	-1.050	0.296
α_6	-0.095	0.066	-1.450	0.148
α_7	0.029	0.066	0.450	0.653
α_8	-0.039	0.066	0.600	0.549
α_9	0.056	0.066	0.860	0.388
α_{10}	-0.029	0.066	-0.450	0.655
α_{11}	0.043	0.066	0.670	0.506
α_{12}	0.307	0.065	4.740	$\mathbf{<.0001}$
α_{13}	-0.164	0.057	-2.900	$\mathbf{0.004}$

Çizelge 4.4 $AR(13)$ modeline ilişkin PROC ARIMA ile elde edilen parametre tahminleri

Parametre	**Parametre Tahmini**	**Standart Hata**	**t-istatistiği değeri**	$\boldsymbol{Pr > \lvert t \rvert}$
μ	60.935	1.996	30.530	$\mathbf{<.0001}$
α_1	0.520	0.059	8.870	$\mathbf{<.0001}$
α_2	0.076	0.064	1.180	0.240
α_3	-0.067	0.065	-1.050	0.296
α_4	-0.081	0.065	-1.250	0.211
α_5	-0.057	0.065	-0.890	0.376
α_6	-0.096	0.065	-1.470	0.144
α_7	-0.009	0.065	-0.13	0.895
α_8	-0.024	0.065	-0.360	0.717
α_9	0.052	0.065	0.800	0.425
α_{10}	0.007	0.065	0.100	0.920
α_{11}	0.062	0.065	0.950	0.341
α_{12}	0.255	0.065	3.910	$\mathbf{0.0001}$
α_{13}	-0.136	0.059	-2.300	$\mathbf{0.022}$

Çizelge 4.3-4.4 incelendiğinde hem PROC ARIMA'da elde edilen sonuçlara göre hem de PROC REG ile elde edilen sonuçlara göre anlamlı olan parametrelere ilişkin p-değerleri işaretlenmiştir. O halde, verilere en uygun modelin

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_{12} Y_{t-12} + \alpha_{13} Y_{t-13} + e_t \, , t = 1,2, \dots , n \qquad (4.1)$$

olduğu söylenebilir. Bu model dikkate alındığında elde edilen parametre tahminleri (PROC REG ve PROC ARIMA ya göre) çizelge 4.5-4.6'da verilmiştir.

Çizelge 4.5 Eşitlik (4.1)'de verilen modele ilişkin PROC REG ile elde edilen parametre tahminleri

Parametre	**Parametre Tahmini**	**Standart Hata**	**t-istatistiği değeri**	$\boldsymbol{Pr > \lvert t \rvert}$
α_0	9.637	2.733	3.530	0.0005
α_1	0.579	0.048	11.900	$< .0001$
α_{12}	0.475	0.050	9.440	$< .0001$
α_{13}	-0.219	0.055	-3.970	$< .0001$

Çizelge 4.6 Eşitlik (4.1)'de verilen modele ilişkin PROC ARIMA ile elde edilen parametre tahminleri

Parametre	**Parametre Tahmini**	**Standart Hata**	**t-istatistiği değeri**	$\boldsymbol{Pr > \lvert t \rvert}$
μ	106.868	12.313	8.680	$< .0001$
α_1	0.667	0.045	14.550	$< .0001$
α_{12}	0.507	0.051	9.880	$< .0001$
α_{13}	-0.202	0.058	-3.490	0.0006

Bu sonuçlar göz önüne alındığında, verilere en uygun modelin, eşitlik (4.1)'de verilen model olduğu söylenebilir. Bu model dikkate alındığında, 2018 yılı için elde edilen 12 aylık öngörü değerleri, ilerleyen kısımda sunulacaktır.

4.2.2 Durağanlık

Bir önceki kısımda verilen parametrelerin anlamlılığına ilişkin testlerin tamamının geçerli olabilmesi ancak zaman serisinin durağanlığına bağlıdır. Bunun nedeni, serinin durağan olmadığı varsayımı altında parametrelerin anlamlılığının testinde kullanılan t ve modelin anlamlılığının testinde kullanılan F testlerinin bilinen t ve F testlerinin olmamasıdır. Seri durağan değil ise öngörü de dahil olmak üzere bu tip birçok istatistiki sonuç çıkarım yapılamaz. Herhangi bir zaman serisinin durağanlığını sınamak için literatürde oldukça çok yöntem bulunmakla birlikte, bunlar arasında parametrelerin en

küçük kareler tahmin edicilerinin dağılımlarına bağlı olarak geliştirilen Dickey-Fuller yöntemi hem pratikliği hem de uygulanabilirliği açısında öne çıkmaktadır. Bununla birlikte özellikle seride MA bileşeninin bulunması durumunda Phillips-Perron yöntemi de yaygın kullanılan yöntemlerdendir. Eğer seride periyodik bileşen varsa, avantajlarından dolayı periodogramlara dayalı durağanlık testi de kullanılabilir. Bu yöntem seride periyodiklik olmasa da kullanılabilir. Ayrıca, mevsimsel zaman serilerinde durağanlık testleri yukarıda açıklanan testlere göre biraz farklıdır. Mevsimsel zaman serilerinin durağanlığı için genellikle HEGY türü testler kullanılır. Kullanılan PM_{10} verisi için durağanlığa genişletilmiş Dickey-Fuller ve Phillips-Perron birim kök testlerine ek olarak periodogramlara dayalı birim kök testleri ve serinin içerdiği mevsimsellikten ötürü HEGY testi ile bakılmıştır. Genişletilmiş Dickey-Fuller birim kök testi ve Phillips-Perron birim kök testine ilişkin sonuçlar aşağıda yer alan çizelge 4.7-4.8'de verilmiştir.

Çizelge 4.7 PM_{10} verisine ilişkin Genişletilmiş Dickey-Fuller birim kök testi sonuçları

	test istatistiği değeri	**Prob.**
ADF testi kritik değerleri	-10.989	0.000
$\boldsymbol{\alpha = 0.01}$ **için kritik değer**	-3.452	
$\boldsymbol{\alpha = 0.05}$ **için kritik değer**	-2.871	
$\boldsymbol{\alpha = 0.10}$ **için kritik değer**	-2.572	

Çizelge 4.8 PM_{10} verisine ilişkin Phillips-Perron birim kök testi sonuçları

	test istatistiği değeri	**Prob.**
PP testi kritik değerleri	-6.808	0.000
$\boldsymbol{\alpha = 0.01}$ **için kritik değer**	-3.452	
$\boldsymbol{\alpha = 0.05}$ **için kritik değer**	-2.871	
$\boldsymbol{\alpha = 0.10}$ **için kritik değer**	-2.572	

Genişletilmiş Dickey-Fuller ve Phillips-Perron birim kök testlerine ilişkin sonuçlar incelendiğinde eşitlik (4.1)'de önerilen model için H_0: "$Seri\ birim\ köklüdür.$" yokluk hipotezi H_a: "$Seri\ durağandır.$" alternatif hipotezine karşı $\alpha = 0.05$ anlam düzeyinde reddedilir. Yani, gelecek dönemlere ilişkin PM_{10} miktarları öngörülebilirdir. Her iki teste ilişkin p-değeri de 0.000 olarak hesaplanmıştır. Ayrıca, AR bileşeni içeren modellerde AR kısmına ait model parametrelerinin değerlerinin toplamlarının 1 olması durumunda modelin birim köklü olduğundan önceki bölümlerde bahsedilmişti. Eşitlik (4.1)'de verilen modelin parametre tahmin değerlerine göre, PROC ARIMA'da $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 \cong 0.97168$ ve PROC REG'de ise $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 \cong 0.83605$ elde edilmiştir. Bu değerler dikkate alındığında sonuçlar farklı görünmektedir. Ancak, çizelge 4.7-4.8'de verilen sonuçlar incelendiğinde modelin durağan olduğu görülmektedir.

Mevsimsellik içeren verilerde genişletilmiş Dickey-Fuller ve Phillips-Perron birim kök testlerinin yanlış sonuç verebilmesi nedeniyle durağanlık HEGY testi ve periodogramlara dayalı birim kök testi yardımı ile de sınanmıştır. Durağanlığın sınanmasında, Aylık veriler için önerilen HEGY testi kullanılmış olup, kullanılan kritik değerler Fransis ve Hobijn (1997) tarafından verilmiştir. Bu teste ilişkin sonuçlar modele kukla (dummy) değişken eklenip eklenmeme durumuna göre iki kısımda incelenmiş olup, iki farklı durum için teste ilişkin sonuçlar çizelge 4.9-4.10'da verilmiştir. Burada HEGY testi gerçekleştirilirken modele kukla (dummy) değişken eklenmediği durumda π_0 sabit terimi göstermek üzere,

$$\begin{aligned}\delta(B)Z_{8,t} = \pi_0 &+ \pi_1 Z_{1,t-1} + \pi_2 Z_{2,t-1} + \pi_2 Z_{2,t-1} + \pi_3 Z_{3,t-1} + \pi_4 Z_{3,t-2} + \pi_5 Z_{4,t-1} \\ &+ \pi_6 Z_{4,t-2} + \pi_7 Z_{5,t-1} + \pi_8 Z_{5,t-2} + \pi_9 Z_{6,t-1} + \pi_{10} Z_{6,t-2} + \pi_{11} Z_{7,t-1} \\ &+ \pi_{12} Z_{7,t-2} + e_t\end{aligned}$$

şeklindeki yardımcı regresyon modelinden yararlanışmış olup, modele kukla (dummy) değişken eklendiği durumda ise $K_i, i = 1,2, \dots ,11$ kukla değişkenleri göstermek üzere,

$$Z_{8,t} = \pi_0 + \sum_{i=1}^{11} c_i\, K_{i,t} + \pi_1 Z_{1,t-1} + \pi_2 Z_{2,t-1} + \pi_2 Z_{2,t-1} + \pi_3 Z_{3,t-1} + \pi_4 Z_{3,t-2} + \pi_5 Z_{4,t-1} + \pi_6 Z_{4,t-2} + \pi_7 Z_{5,t-1} + \pi_8 Z_{5,t-2} + \pi_9 Z_{6,t-1} + \pi_{10} Z_{6,t-2} + \pi_{11} Z_{7,t-1} + \pi_{12} Z_{7,t-2} + \sum_{i=1}^{p-1} \nabla_{12} Y_{t-i} + e_t$$

şeklindeki yardımcı regresyon modeli kullanılmıştır. Burada, $H_0: \pi_1 = 0$ yokluk hipotezinin $H_a: \pi_1 < 0$ alternatifine karşı testi gerçekleştirilmiştir. Verilen modellerde yer alan Z değişkenleri, zaman serisinin dönüşümleri ile elde edilen seriler olup, bununla ilgili bilgiler önceki bölümlerde verilmişti.

Çizelge 4.9 Modele kukla (dummy) değişken eklendiği durumda PM_{10} verisine ilişkin HEGY testi sonuçları

	test istatistiği değeri
HEGY testi kritik değerleri	-2.849
$\boldsymbol{\alpha = 0.01}$ **için kritik değer**	-3.330
$\boldsymbol{\alpha = 0.05}$ **için kritik değer**	-2.760
$\boldsymbol{\alpha = 0.10}$ **için kritik değer**	-2.475

Modele kukla (dummy) değişken eklendiği durumda test istatistiğinin değeri -2.849 olarak elde edilmiş olup, bu değer $\alpha = 0.05$ ve $\alpha = 0.10$ anlam düzeylerine ilişkin kritik değerlerden küçük olduğundan H_0: "*Model mevsimsel birim köke sahiptir*" yokluk hipotezi reddedilir. Yani, ilgilenilen modelin durağan olduğu sonucuna ulaşılır.

Çizelge 4.10 Modele kukla (dummy) değişken eklenmediği durumda PM_{10} verisine ilişkin HEGY testi sonuçları

	test istatistiği değeri
HEGY testi kritik değerleri	-2.503
$\boldsymbol{\alpha = 0.01}$ **için kritik değer**	-3.050
$\boldsymbol{\alpha = 0.05}$ **için kritik değer**	-2.785
$\boldsymbol{\alpha = 0.10}$ **için kritik değer**	-2.500

Modele kukla (dummy) değişken eklenmediği durumda test istatistiğinin hesaplanan değeri -2.503 olarak elde edilmiştir. Buna göre, sadece $\alpha = 0.10$ anlam düzeyinde test edildiğinde model durağan olup, diğer durumlarda ise model mevsimsel birim köklü çıkmıştır.

Periodogramlara ilişkin birim kök testlerinin avantajlarından daha önce bahsedilmişti. Bu avantajlar göz önünde bulundurularak, serinin durağanlığı periodogramlar ile de sınanmıştır. SAS PROC SPECTRA'da periodogram değeri $I_n(w_1) = 6133.66$ olarak hesaplanmış ve eşitlik (4.1)'de verilen model dikkate alındığında $\hat{\sigma}_n^2 = 322.2444$ olarak elde edilmiştir. Test istatistiğinin değeri de

$$T_n(w_1) = \frac{2(1-\cos(w_1))}{\hat{\sigma}_n^2} I_n(w_1) = 0.008349022$$

dir. Test istatistiğinin kritik değerleri *R* paket programı kullanılarak, Akdi ve Dickey (1998) tarafından elde edilen değerler güncellenmiştir. Periodogramlara dayalı birim kök testine ilişkin test istatistiğinin değeri ve kritik değerler çizelge 4.11'de verilmiştir.

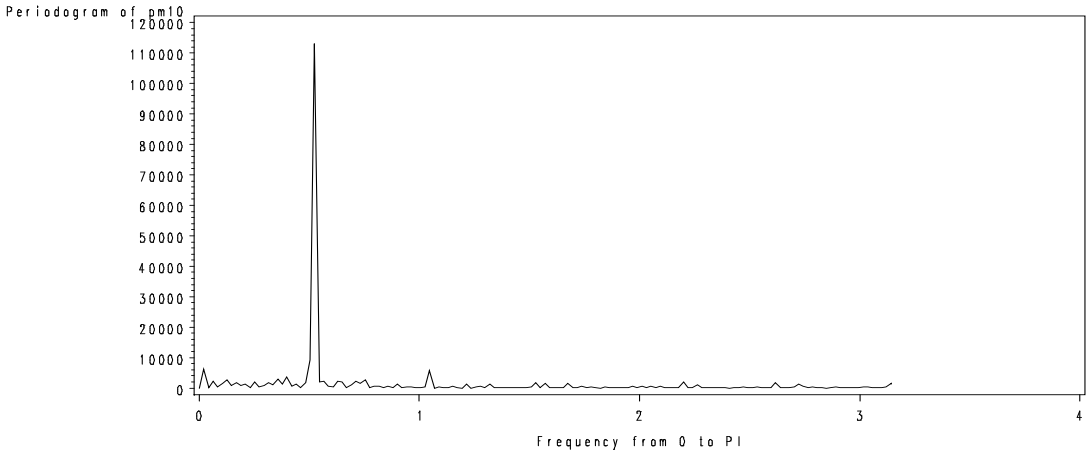
Çizelge 4.11 Periodogram tabanlı birim kök testine ilişkin sonuçlar

	test istatistiği değeri
Periodogram tabanlı birim kök testi kritik değerleri	0.0008349022
$\boldsymbol{\alpha = 0.01}$ **için kritik değer**	0.034818
$\boldsymbol{\alpha = 0.05}$ **için kritik değer**	0.178496
$\boldsymbol{\alpha = 0.10}$ **için kritik değer**	0.368089

Periodogramlara dayalı birim kök testi için test istatistiğinin hesaplanan değeri üç anlam düzeyi için de kritik değerlerden küçük olup, H_0: "$Seri\ durağan\ değildir$" yokluk hipotezi $\alpha = 0.01$, $\alpha = 0.05$ ve $\alpha = 0.10$ anlam düzeylerinin tamamı için reddedilir. Yani, gelecek dönemlere ait PM_{10} değerleri öngörülebilirdir ve bir önceki kısımda model parametreleri ile ilgili yapılan istatistiki sonuç çıkarımlar geçerlidir. HEGY testine kıyasla periodogram tabanlı birim kök testlerinin avantajları olması sebebiyle serinin durağan olduğu dikkate alınmıştır.

4.2.3 Periyodiklik

Çalışmada kullanılan PM_{10} verilerinin durağan olduğu önceki kısımda hem ADF hem de PP test yöntemi ile hem de periodogramlara dayalı yöntem ile sınanarak serinin durağan olduğu sonucu elde edilmiştir. Daha önce de belirtildiği gibi, periodogramlar seride bulunan gizli periyodikliklerin araştırılmasında da kullanılmaktadır. Aylık PM_{10} verilerine ilişkin periodogramların sıklıklara karşı grafikleri şekil 4.3'de verilmiştir.



Şekil 4.3 PM_{10} verisine ait periodogramların sıklıklara karşı grafiği

Şekik 4.3'te verilen grafikten de anlaşılabileceği gibi periodogram değerlerinden biri diğerlerine göre çok farklılık göstermektedir. Bu en büyük değer veride muhtemel bir periyodikliğin araştırılmasında kullanılabilir (Wei 2006). Verilerden elde edilen değerlerden en büyük beş periodogram değeri, bunlara karşılık gelen sıklıklar ile periyotları çizelge 4.12'de verilmiştir. Bütün sıklıklardaki periodogramların toplam değeri

$$\sum_{k=1}^{[|n/2|]} I_n(w_k) = 218213.92$$

olarak gözlenmiştir.

Çizelge 4.12 PM_{10} verisine ilişkin en yüksek beş periodogram değeri

k	**Sıklık (w_k)**	**$I_n(w_k)$**	**Periyot**
26	0.52360	113026.590	12.000
25	0.50265	9273.120	12.500
2	0.02094	6133.660	300.000
51	1.074720	5678.460	6.000
20	0.39794	3636.590	15.789

Buna göre,

$$V = I_n(w_{(1)})\left[\sum_{k=1}^{[|n/2|]} I_n(w_k)\right]^{-1} = \left[\sum_{k=1}^{[|n/2|]} I_n(w_k)\right]^{-1} max\{I_n(w_k)\}$$
$$= \frac{113026.59}{218213.92} = 0.51796228$$

olup V istatistiğine karşılık gelen kritik değerler $P(V > c_\alpha) = \alpha \cong m(1 - c_\alpha)^{m-1}$ yaklaşık değeri ile $c_\alpha = 1 - (\alpha/m)^{1/(m-1)}$ şeklinde hesaplanmıştır (Wei 2006). Burada m sayısı,

$$m = \begin{cases} (n-1)/2 & , \ n\ tek\ ise \\ (n/2) - 1 & , \ n\ çift\ ise \end{cases}$$

şeklindedir. Büyükten küçüğe sıralanan periodogram değerleri, $i = 1,2,3,4,5$ için $I_n(w_{(i)})$ ile gösterilmek üzere, diğer sıklıklarda da periyodikliğin olup olmadığına bakmak için,

$$V_i = \frac{I_n(w_{(i)})}{\sum_{k=1}^{[|n/2|]} I_n(w_k) - \sum_{k=1}^{i-1} I_n(w_{(k)})}$$

olmak üzere, $V_1 = 0.518$, $V_2 = 0.088$, $V_3 = 0.064$, $V_4 = 0.063$ ve $V_5 = 0.043$ değerleri elde edilmiştir. Buradan $\alpha = 0.01$, $\alpha = 0.05$ ve $\alpha = 0.10$ için kritik değerler

$$c_{0.01} = 0.0624972682,\ c_{0.05} = 0.0523158530,\ c_{0.10} = 0.047896912$$

olarak bulunmuştur. Eğer $V > c_\alpha$ ise veride periyodik bileşen yoktur yokluk hipotezi reddedilir ve veride periyodik bileşen olduğu sonucuna varılır. Bu sonuçlara göre, $V > c_\alpha$ olduğu için verilerin periyodik bileşen içerdiği sonucu elde edilmiştir. Ayrıca, serinin periyodu ise, $p = 2\pi/w_k = 2\pi/0.52360 = 12$ olarak elde edilmiştir. Verilerin aylık olarak toplanmasından ve hava kirliliğinin içerdiği mevsimsellikten ötürü bu

durum beklenen bir sonuçtur. Diğer taraftan, $V_1 > c_\alpha$, $V_2 > c_\alpha$, $V_3 > c_\alpha$ ve $V_4 > c_\alpha$ olmasına rağmen $V_5 < c_\alpha$'dır. Yani, beşinci en büyük periodogram değerine karşılık gelen sıklıkta bir periyodiklik yoktur. Dolayısı ile periyodik bileşenler olarak ilk dört sıklığa karşılık gelen periyodik bileşenler dikkate alınmalıdır. Gerçekleştirilen birim kök testlerinin sonuçlarına göre model durağandır. Verilerin periyodik bileşenler içerdiği de dikkate alındığında, en yüksek periodogram değerleri göz önüne alınarak (beşinci en yüksek periodogram değerine karşılık gelen periyod anlamlı değildir) bu periyotlar yardımı ile oluşturulabilecek bir regresyon modelinden bahsedilebilir.

4.2.4 Harmonik regresyon modeli

Periodogramlar yardımıyla verinin içerdiği periyodik yapının analizi bir önceki kısımda yapılmaya çalışıldı. Bu periyotlar dikkate alınarak veriler için,

$$Y_t = \mu + A_1 \cos\left(\frac{2\pi t}{12}\right) + B_1 \sin\left(\frac{2\pi t}{12}\right) + A_2 \cos\left(\frac{2\pi t}{12.5}\right) + B_2 \sin\left(\frac{2\pi t}{12.5}\right) + A_3 \cos\left(\frac{2\pi t}{300}\right) + B_3 \sin\left(\frac{2\pi t}{300}\right) + A_4 \cos\left(\frac{2\pi t}{6}\right) + B_4 \sin\left(\frac{2\pi t}{6}\right) + e_t \,, \quad t = 1,2,\ldots,300 \tag{4.2}$$

şeklinde bir regresyon modeli göz önüne alınmıştır. Bu modele göre SAS PROC REG ile elde edilen sonuçlar çizelge 4.13'de verilmiştir.

Çizelge 4.13 Eşitlik (4.2)'de verilen harmonik regresyon modeline ilişkin SAS PROC REG çıktıları

Parametre	**Parametre Tahmini**	**Standart Hata**	**t-istatistiği değeri**	$\boldsymbol{Pr > \lvert t \rvert}$
μ	59.931	0.976	61.350	$< .0001$
A_1	27.270	1.381	19.740	$< .0001$
B_1	3.139	1.381	2.270	0.0238
A_2	3.710	1.381	2.690	0.0077
B_2	-6.932	1.381	-5.020	$< .0001$
A_3	4.762	1.381	3.450	0.0007
B_3	-4.267	1.381	-3.090	0.0022
A_4	5.842	1.381	4.230	$< .0001$
B_4	-1.930	1.381	-1.400	$\mathbf{0.1634}$

Çizelge 4.13 incelendiğinde bir tek B_4 parametresi anlamsızdır ($p - değeri = 0.1634 > 0.05$). Bu sonuçlar göz önünde bulundurulduğunda ilgilenilen PM_{10} verisine uygun modelin,

$$Y_t = \mu + A_1 \cos\left(\frac{2\pi t}{12}\right) + B_1 \sin\left(\frac{2\pi t}{12}\right) + A_2 \cos\left(\frac{2\pi t}{12.5}\right) + B_2 \sin\left(\frac{2\pi t}{12.5}\right) + A_3 \cos\left(\frac{2\pi t}{300}\right) + B_3 \sin\left(\frac{2\pi t}{300}\right) + A_4 \cos\left(\frac{2\pi t}{6}\right) + e_t \,, \quad t = 1,2,\ldots,300 \tag{4.3}$$

olduğu görülmektedir. Bu modele ilişkin SAS PROC REG çıktıları çizelge 4.14'de verilmiştir. Bu model için bütün parametreler anlamlıdır (p-değeri oldukça küçüktür).

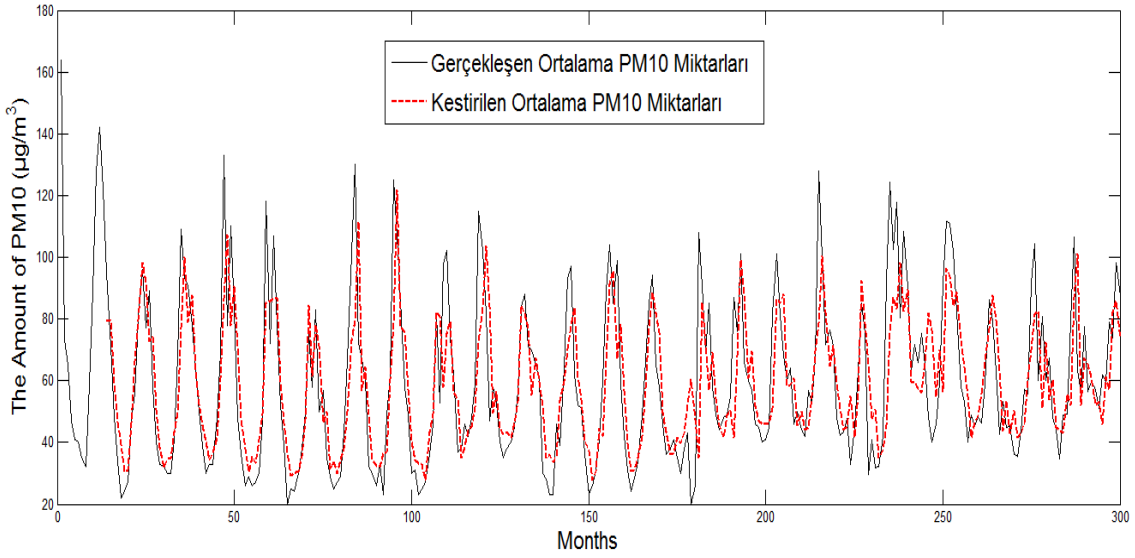
Çizelge 4.14 Eşitlik (4.3)'de verilen trigonometrik regresyon modeline ilişkin SAS PROC REG çıktıları

Parametre	**Parametre Tahmini**	**Standart Hata**	**t-istatistiği değeri**	$\boldsymbol{Pr > \lvert t \rvert}$
μ	59.931	0.976	61.250	$< .0001$
A_1	27.270	1.383	19.710	$< .0001$
B_1	3.139	1.383	2.270	0.0240
A_2	3.710	1.383	2.680	0.0078
B_2	-6.932	1.383	-5.010	$< .0001$
A_3	4.762	1.383	3.440	0.0007
B_3	-4.267	1.383	-3.080	0.0022
A_4	5.842	1.383	4.220	$< .0001$

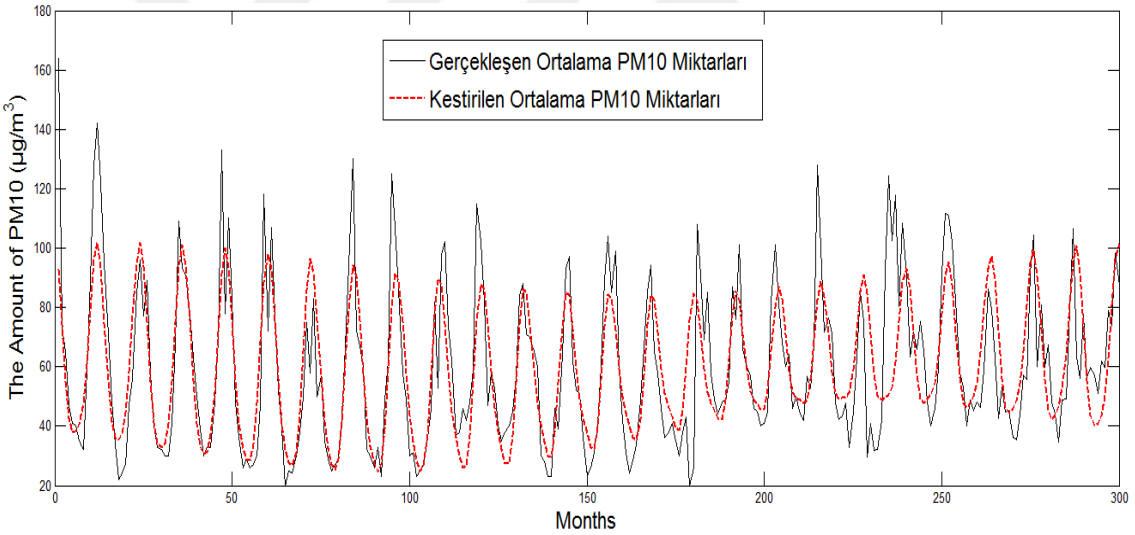
Buna göre, aylık PM_{10} verileri için,

$$\hat{Y}_t = 59.93 + 27.27\cos\left(\frac{2\pi t}{12}\right) + 3.14\sin\left(\frac{2\pi t}{12}\right) + 3.71\cos\left(\frac{2\pi t}{12.5}\right) - 6.93\sin\left(\frac{2\pi t}{12.5}\right) + 4.76\cos\left(\frac{2\pi t}{300}\right) - 4.27\sin\left(\frac{2\pi t}{300}\right) + 5.84\cos\left(\frac{2\pi t}{6}\right), t = 1,2,\dots,300$$

şeklinde bir kestirim modeli oluşturulmuştur. Bu modele göre elde edilen kestirim değerleri ile gözlenen PM_{10} değerlerinin zamana göre grafiği şekil 4.5'de verilmiştir. Ayrıca, önceki kısımlarda elde edilen eşitlik (4.1)'de verilen model ile PM_{10} değerlerinin zamana karşı grafiği de şekil 4.4'de verilmiştir.



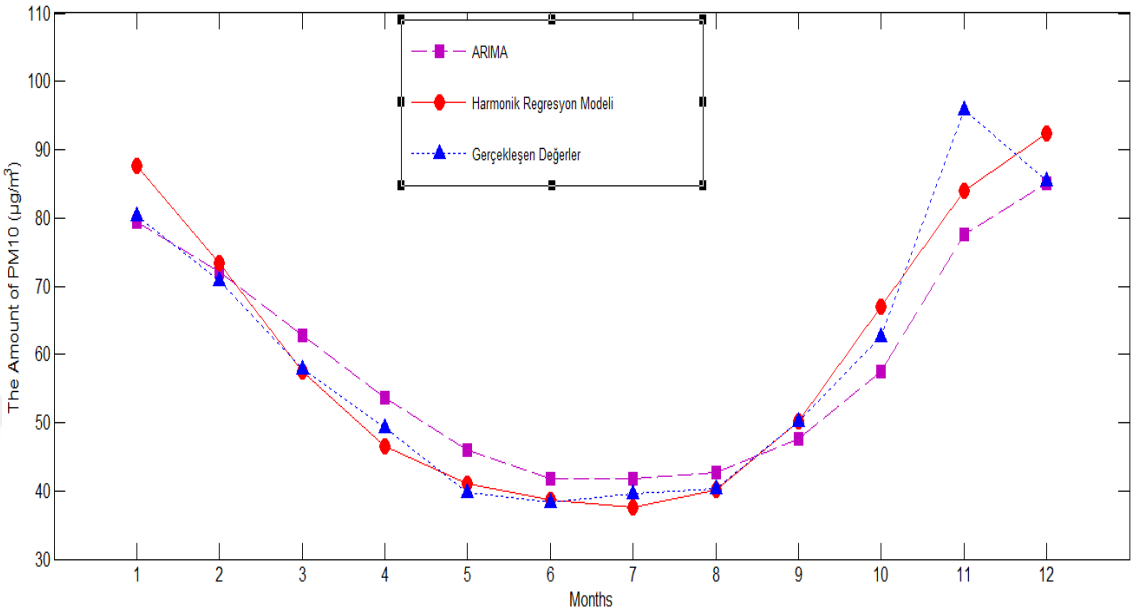
Şekil 4.4 PM_{10} değerleri ve eşitlik (4.1)'de verilen modelden elde edilen kestirimler



Şekil 4.5 PM_{10} değerleri ve harmonik modelden elde edilen kestirimler

Şekil 4.4-4.5 incelendiğinde, her iki modelde de kestirim değerlerinin gözlem değerlerine oldukça yakın olduğu görülmektedir. Otoregresif bileşen içeren model yardımıyla yapılan kestirimlerden ilk 13 tanesi modelin gecikme derecesi kadar veri kaybolduğundan şekil 4.4'de yer almamaktadır. Bu nedenle, ilk 13 verinin yer aldığı ilk iki yılın verisi göz ardı edilerek, hangi modelin daha iyi sonuçlar verdiğinin anlaşılabilmesi için şekil 4.6'da son 23 yıla ilişkin gerçekleşen aylık ortalama PM_{10}

değerleri ile ARIMA ve harmonik regresyondan elde edilen aylık ortalama kestirim değerlerinin grafiği verilmiştir.



Şekil 4.6 PM_{10} ortalamalarının karşılaştırılması

Şekil 4.6'dan da görülebileceği gibi harmonik regresyon ile elde edilen aylık ortalama kestirimler gerçek ortalama değerlere ARIMA ile elde edilen aylık ortalama kestirim değerlerine göre daha yakın görünmektedir. Burada $\bar{y}_i$ aylık gerçekleşen ortalamayı, $\bar{y}_{a,i}$ ARIMA ile elde edilen aylık kestirim ortalamalarını ve $\bar{y}_{h,i}$ harmonik regresyon ile elde edilen aylık kestirim ortalamalarını göstermek üzere, bu ortalamaların gerçekleşen ortalamalara uzaklıklarının karelerinin toplamları,

$$SS(ARIMA) = \sum_{i=1}^{12} \left(\bar{y}_i - \bar{y}_{a,i}\right)^2 = 470.18105$$

$$SS(HARMONİK) = \sum_{i=1}^{12} \left(\bar{y}_i - \bar{y}_{h,i}\right)^2 = 279.63572$$

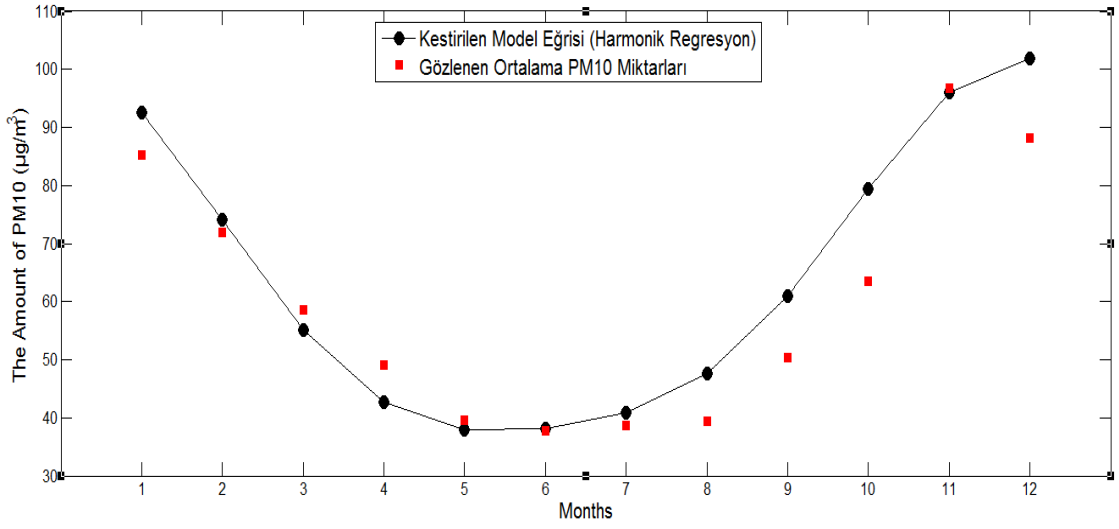
şeklindedir. Bu sonuçlara göre de harmonik regresyon ile elde edilen kestirim değerleri ARIMA ile elde edilen kestirim değerlerine göre gerçekleşen değerlere daha yakındır. Bu sonuçlar, her iki yönteme göre elde edilen kestirim değerleri dikkate alınarak hesaplanan öngörü değerlerinde daha belirgindir. Her iki yönteme göre elde edilen öngörü değerleri ve çalışmanın devam ettiği 2018 yılının ilk iki ayına ait gerçekleşen değerler çizelge 4.15'de verilmiştir.

Çizelge 4.15 Elde edilen öngörü değerleri ve gerçekleşen değerler

Ay	Yıl	Öngörü (ARIMA)	Öngörü (Harmonik Regresyon)	Gerçekleşen Değer
Ocak	2018	77.187	92.624	57.000
Şubat	2018	82.393	74.073	69.125
Mart	2018	70.925	55.019	
Nisan	2018	69.138	42.625	
Mayıs	2018	66.033	38.021	
Haziran	2018	61.326	38.148	
Temmuz	2018	65.062	40.855	
Ağustos	2018	64.198	47.607	
Eylül	2018	73.794	60.628	
Ekim	2018	73.783	79.426	
Kasım	2018	87.097	96.009	
Aralık	2018	85.751	101.915	

Çizelge 4.15'de verilen öngörü değerleri incelendiğinde harmonik regresyon modelinden elde edilen öngörü değerinin özellikle yaz aylarında (Mayıs, Haziran, Temmuz, Ağustos) küçük değerler aldığı görülmektedir. Oysa ARIMA ile elde edilen öngörü değerleri özellikle yaz aylarında gerçekleşen aylık ortalamaların oldukça üzerindedir. Aylık ortalamaların yer aldığı çizelge 4.1 incelendiğinde harmonik regresyon modelinin ARIMA modelinden elde edilen öngörülere göre daha tutarlı sonuçlar verdiği sonucuna ulaşılabilir. Ayrıca 2018 yılı Ocak ayına ait gerçekleşen değere iki modelin öngörü değerleri de uzak kalmış olmasına rağmen, Şubat ayı için ARIMA modelinden elde edilen öngörü değeri, gerçekleşen $69.125\ \mu g/m^3$ değerine

göre oldukça büyük çıkmıştır. Buna karşılık harmonik regresyondan elde edilen öngörü değeri olan 74.073 $\mu g/m^3$ değeri gerçekleşen değere ARIMA modelinden elde edilen öngörü değerine göre oldukça yakındır. Ocak ayında gerçekleşen değere iki modele ait öngörü değerinin de uzak olmasının ve Ocak ayı ortalamasına göre bu değerin oldukça düşük olmasının nedeni ise ölçümlerde yapılan hatalar veya 2018 Ocak ayında gerçekleşen özel bir neden olabileceği düşünülmüştür. Bu sonuçlar göz önünde bulundurulduğunda önerilen harmonik regresyon modelinin Ankara iline ait aylık ortalama PM_{10} miktarlarının öngörülmesinde kullanılabileceği sonucu ortaya çıkmaktadır. Aşağıda verilen şekil 4.7'de aylık ortalama değerler ve kestirilen harmonik regresyon modelinin grafiği bulunmaktadır.



Şekil 4.7 Gerçekleşen PM_{10} değerleri ve kestirilen harmonik regresyon eğrisi

Yukarıda şekil 4.7'de verilen grafik incelendiğinde kestirilen harmonik regresyon eğrisi ile gözlenen ortalama PM_{10} miktarlarının birbirlerine oldukça yakın olduğu görülmektedir. Bu nedenle ilerleyen dönemlere ait öngörüler yapılırken harmonik regresyon modelinin kullanılmasının daha iyi sonuçlar vereceği düşünülmektedir.

5. SONUÇ VE TARTIŞMA

Zaman serilerindeki temel istatistiki sonuç çıkarım öngörüdür. Öngörü yapabilmek için ise ilgilenilen serinin durağan bir yapı göstermesi gerekmektedir. Bu nedenle durağanlığın sınanması zaman serilerinde önemli bir yer tutar. Bir zaman serisinin durağanlığı birim kök testleri ile sınanır. Literatürde birçok birim kök testi bulunmatadır. Bunlardan uygulamada en sık kullanılanları ve birçok paket programın test istatistiği değerleri ile kritik değerlerini doğrudan verdiği birim kök testleri Dickey-Fuller birim kök testi ile Phillips-Perron birim kök testidir. Ancak serinin mevsimsellik içerdiği durumlarda bu testlerin iyi sonuç vermemesinden dolayı genellikle Dickey vd. (1984) tarafından önerilen ve parametrelerin simetrik en küçük kareler tahmin edicilerinin dağılımlarına dayanan mevsimsel birim kök testi ve Hylleberg vd. (1990) tarafından önerilen ve literatürde HEGY testi olarak bilinen mevsimsel birim kök testi kullanılmaktadır.

Çalışmada Ankara iline ait 1993:01 – 2017:12 dönemleri arasındaki, sekiz farklı istasyondan elde edilen aylık ortalama PM_{10} miktarları kullanılmış ve analizler gerçekleştirilmiştir. PM_{10} insan sağlığına zararlı hava kirliliği indekslerinden birisidir (Cavkaytar vd. 2013). Hava kirliliği indekslerinin birçoğunda olduğu gibi, PM_{10} miktarları da kış aylarında artış göstermekte olup, yaz aylarında ise düşük değerler almaktadır. Bunun başlıca nedenleri arasında kış aylarında kullanılan karbon içerikli yakıt tüketiminin artması ve yaz aylarında okulların tatil olması vb. nedenlerle Ankara iline ait nüfusun azalmasına bağlı olarak trafikteki araç sayısındaki azalma gösterilebilir.

Çalışmadaki temel amaç gelecek dönemlere ait PM_{10} miktarlarının öngörülmesidir. Bu nedenle öncelikle serinin durağanlığı sınanmıştır. Bu sınama işlemlerinde genişletilmiş Dickey-Fuller (ADF) ve Phillips-Perron (PP) birim kök testlerinin yanında, serinin içerdiği mevsimsellikten dolayı HEGY testi ile Akdi ve Dickey (1998) tarafından önerilen periodogramlara dayalı birim kök testi de kullanılmıştır. ADF, PP ve periodogramlara dayalı birim kök testine göre seri durağan çıkmasına rağmen, HEGY testinde modele kukla (dummy) değişken eklenmediği durumda seri mevsimsel birim

köklü çıkmıştır. Mevsimsellik içeren serilerde ADF ve PP birim kök testlerinin iyi sonuç vermediği bilinmektedir. HEGY testinin ise uygulanmasında birçok parametre tahminine ihtiyaç duyulması, periodogramlara dayalı birim kök testinin kritik değerlerinin örneklem hacmine bağlı olarak değişmemesi, model seçimine göre değişmez olması gibi avantajları göz önünde bulundurularak periodogram tabanlı birim kök testi sonuçları göz önünde bulundurulmuştur. Ayrıca periodogramlara dayalı birim kök testleri mevsimsellik içeren serilerde de birim kök tespitinde kullanılabilmektedir (Akdi ve Dickey 1999). Periodogramların hesaplanışında verilerin trigonometrik dönüşümlerinden yararlanılmasından dolayı birim kök tespitinde de bu birim kök testinin daha iyi sonuçlar verebileceği düşünülmektedir.

Çalışmada öngörünün yapılabilmesi için öncelikle AIC istatistiklerinin değerlerine göre aylık ortalama PM_{10} verileri için

$$Y_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_{12} X_{t-12} + \alpha_{13} X_{t-13} + e_t \, , t = 1,2, \dots ,300$$

şeklindeki bir ARIMA modelinin uygun olabileceği düşünülmüştür. Bu model yardımıyla gelecek dönemlere ait ortalama PM_{10} miktarları için öngörüler hesaplanmış olup bu öngörüler çizelge 4.15'de verilmiştir. Bununla birlikte, periodogramlar yardımıyla PM_{10} verisinin içerdiği periyotlar 12, 12.5, 300 ve 6 olarak elde edilmiştir. Bu periyotlar yardımıyla,

$$\begin{aligned} Y_t = \mu &+ A_1 \cos\left(\frac{2\pi t}{12}\right) + B_1 \sin\left(\frac{2\pi t}{12}\right) + A_2 \cos\left(\frac{2\pi t}{12.5}\right) + B_2 \sin\left(\frac{2\pi t}{12.5}\right) \\ &+ A_3 \cos\left(\frac{2\pi t}{300}\right) + B_3 \sin\left(\frac{2\pi t}{300}\right) + A_4 \cos\left(\frac{2\pi t}{6}\right) + e_t \, , \\ &t = 1,2, \dots ,300 \end{aligned}$$

şeklinde trigonometrik fonksiyonlar yardımıyla oluşturulan harmonik regresyon modeli ile veri modellenmiştir. Bu modele ilişkin öngörü değerleri karşılaştırma yapılabilmesi amacıyla yine çizelge 4.15'de verilmiştir. İkinci modele göre öngörü değerleri hesaplanırken, sadece hesaplanması gereken aya ilişkin olarak t değerinin girilmesi yeterlidir. Bununla birlikte, birinci modelden elde edilen öngörü değerleri

hesaplandığında, özellikle yaz aylarındaki (Mayıs-Ağustos) değerler 60-70 $\mu g/m^3$ bandında değişmektedir. Bu değerler aylık ortalamalara göre çok farklıdır. Oysa ikinci modele göre elde edilen öngörü değerleri (38-50 $\mu g/m^3$ arası) aynı aylarda elde edilen ortalamalara çok daha yakındır. Yani, ikinci modelden elde edilen öngörü değerleri birinci modele göre daha tutarlıdır. Bu değerler şekil 4.7'de verilmiştir. Ayrıca, çalışma devam ederken elde edilen 2018 yılına ait Şubat ayındaki gözlem değeri ile karşılaştırıldığında ikinci modelden elde edilen öngörü değerinin daha yakın olduğu gözlenmektedir. Ocak ayında iki model de iyi sonuç vermemesine rağmen, 2018 Ocak ayında gözlenen değerin önceki yılların değerlerine göre çok küçük olması, ölçümde hata yapılmış olabileceği veya 2018 Ocak ayında özel bir durumun gerçekleşmiş olabileceği sonucunu ortaya çıkarmıştır. Sonuç olarak, periodogramların modele göre değişmez olması ve durağanlık için test istatistiğinin kritik değerlerinin örneklem büyüklüğüne bağlı olmaması, özellikle küçük örneklemlerde durağanlığın testi için periodogram tabanlı birim kök testinin kullanılabileceği ortaya çıkmıştır. Diğer taraftan incelenen zaman serisinin periyodik bileşen içermesi durumunda öngörü değerlerinin hesaplanması için harmonik regresyona dayalı modelin kullanılması tercih edilmelidir.

Hava kirliliği indekslerinin öngörülmesi, çevre planlaması ve insan sağlığı için oldukça önemlidir. Yüksek öngörü değerleri elde edildiği durumlarda bu hava kirliliğini azaltacak ekstra önlemler alınması gerekebilir. Kurulan iki model yardımıyla (ARIMA ve harmonik regresyon) elde edilen öngörüler incelendiğinde, Ankara iline ait PM_{10} değerlerinin gelecek dönemler içinde yüksek değerler alması beklenmektedir. Özellikle kış aylarında gerçekleşmesi öngörülen değerlerin (iki model öngörüsü için de) Dünya Sağlık Örgütü ve Avrupa Birliği standartlarının oldukça üzerinde olduğu görülmüştür. Bu durum da Ankara ilindeki PM_{10} miktarını azaltmak için çalışmalar yapılması gerektiğini göstermektedir.

Ankara ili Türkiye'nin başkenti ve ikinci büyük kenti olması açısından Türkiye Cumhuriyeti Devleti için önemli bir yer tutmaktadır. Bundan sonraki gerçekleştirilecek çalışmalarda ise bu çalışmadan farklı olarak, diğer dünya metropollerinin (İstanbul, Londra, Paris, New York gibi) hava kirliliği için analizler yapılarak, Ankara iline ait sonuçlarla karşılaştırmaları yapılabilir.

KAYNAKLAR

Akdi, Y. 1995. Periodogram analysis for unit roots. Ph.D. Dissertion, North Carolina State University.

Akdi, Y. 2012. Zaman Serileri Analizi, Birim Kökler ve Kointegrasyon. Gazi Kitapevi, 373 s., Ankara.

Akdi, Y. 2014. Matematiksel İstatistiğe Giriş. Gazi Kitapevi, 552 s., Ankara.

Akdi, Y. and Dickey, D.A. 1998. Periodograms of unit root time series: distributions and tests. Communications in Statistics: Theory and Methods, 27(1), 69-87.

Akdi, Y. and Dickey, D.A. 1999. Periodograms for seasonal time series with a unit root. Journal of the Turkish Statistical Association, 2(3), 153-164.

Anonim. 2017. Hava Kalitesi İstatistik Dökümanları. Türkiye İstatistik Kurumu, Ankara.

Anonim. 2018. Ulusal Hava Kalitesi İzleme Ağı İstasyon Raporları. T.C. Çevre ve Şehircilik Bakanlığı, Ankara.

Anonymous. 2005. Air Quality Standards. European Commission, Brussels.

Anonymous. 2008. WHO Global Urban Ambient Air Pollution Database. World Health Organization, Geneva.

Berktaş, B.M. and Bircan, A. 2003. Effects of atmospheric sulphur dioxide and particulate matter concentrations on emergency room admissions due to asthma in Ankara. Tüberküloz ve Toraks Dergisi, 51(3), 231-238.

Bloomfield, P. 2000. Fourier Analysis of Time Series: An Introduction (2nd ed.). John Wiley & Sons, Inc., 275 p., Danvers.

Brockwell, P.J. and Davis, R.A. 1987. Time Series: Theory and Methods. Springer-Verlag, 519 p. , New York.

Brockwell, P.J. and Davis, R.A. 2002. Introduction to Time Series and Forecasting, Second Edition. Springer-Verlag, 421 p., New York.

Cavkaytar, O., Soyer, O.U. ve Şekerel, E.B. 2013. Türkiye'de hava kirliliğinden kaynaklanan sağlık sorunları. Hava Kirliliği Araştırmaları Dergisi, 2, 105-111.

Çiçek, İ., Türkoğlu, N. ve Gürgen, G. 2004. Ankara'da hava kirliliğinin istatistiksel analizi. Fırat Üniversitesi Sosyal Bilimler Dergisi, 14(2), 1-18.

Dickey, D.A. and Fuller, W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427-431.

Dickey, D.A., Hasza, D.P. and Fuller, W.A. 1984. Unit roots in seasonal time series. Journal of the American Statistical Association, 79(386), 355-367.

Franses, P.H. and Hobijn, B. 1997. Critical values for unit root tests in seasonal time series. Journal of Applied Statistics, 24(1), 25-47.

Fuller, W.A. 1996. Introduction to Statistical Time Series (2nd ed.). Wiley Series in Probability and Statistics, 698 p., New York.

Genc, D.D., Yesilyurt, C. and Tuncel, G. 2010. Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere. Environmental Monitoring and Assessment, 166, 11-27.

Hamilton, J.D. 1994. Time Series Analysis. Princeton University Press, 785 p., New Jersey.

Hylleberg, S., Engle R.F., and Granger, C.W.J. 1990. Seasonal integration and cointegration. Journal of Econometrics 44, 215-238.

Kaplan, Y., Saray, U. ve Azkeskin, E. 2014. Hava kirliliğine neden olan PM10 ve SO_2 maddesinin yapay sinir ağı kullanılarak tahmininin yapılması ve hata oranının hesaplanması. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 14(2), 1-6.

Koopmans, L.H. 1995. The Spectral Analysis of Time Series. Academic Press, Inc., 385 p., San Diego.

Koutrakis, P., Sax, S.N., Samat, J.A., Coull, B., Demokritou, P., Oyola, P., Garcia, J. and Gramsch, E. 2005. Analysis of PM_{10}, $PM_{2.5}$ and $PM_{2.5-10}$ concentrations in Santiago, Chile from 1989 to 2001. Journal of the Air & Waste Management Association, 55(3), 342-351.

Kurt, A., Gulbagci, B., Karaca, F. and Alagha, O. 2008. An online air pollution forecasting system using neural networks. Environment International, 34(5), 592-598.

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. 2004. Applied Linear Statistical Models. McGraw-Hill Companies, 1373 p., New York.

Leadbetter, R., Cambanis, S. and Pipiras, V. 2014. A Basic Course in Measure and Probability. Cambridge University Press, 355 p., New York.

Ostro, B.D., Hurley, S. and Lipsett, M.J. 1999. Air pollution and daily mortality in the Coachella Valley : a study of PM10 dominated by coarse particles. Environmental Research, 81(3), 231-238.

Phillips, P.C.B. and Perron, P. 1988. Testing for a unit root in time series regression. Biometrika, 75(2), 335-346.

Pope III, C.A. and Dockery, D.W. Acute health effects of PM_{10} pollution on symptomatic and asymptomatic children., American Review of Respiratory Disase, 145, 1123-1128.

Turgut, D. and Temiz, İ. 2015. Time series analysis and forecasting fora ir pollution in Ankara: a Box-Jenkins approach. Alphanumeric Journal, 3(2), 131-138.

Türe, H. ve Akdi, Y. 2006. Mevsimsel eşbütünleşme: Tüketim ve GSYİH. İktisat İşletme ve Finans, 21(242), 101-113.

Wei, W.W.S. 2006. Time Series Analysis: Univariate and Multivariate Methods (2nd. Ed.). Pearson-Addison Wesley, 599 p., Boston.

Yaffee, R.A. and McGee, M. 2000. An Introduction to Time Series and Forecasting: With Applications of SAS and SPSS. Academic Press, Inc., 528 p., New York.

Yüksek, A.G., Bircan, H., Zontul, M. ve Kaynar, O. 2007. Sivas ilinde yapay sinir ağları ile hava kalitesi modelinin oluşturulması üzerine bir uygulama. Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi, 8(1), 97-112.

ÖZGEÇMİŞ

Adı Soyadı : Yasin OKKAOĞLU
Doğum Yeri : Kütahya
Doğum Tarihi : 22/05/1993
Medeni Hali : Evli
Yabancı Dili : İngilizce

Eğitim Durumu

Lise : Çankırı Süleyman Demirel Fen Lisesi (2011)
Lisans : Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü (2015)
Yüksek Lisans: Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı (Eylül 2015 – Haziran 2018)

Çalıştığı Kurum ve Yıl

Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü Olasılık Teorisi ve Olasılık Süreçleri Anabilim Dalı, Araştırma Görevlisi, 2015-…

<u>Uluslarası Kongre Sunum</u>

Okkaoğlu, M.C., Şeşen, H., Demirkale, İ., Özdemir, M., **Okkaoğlu, Y.** ve Altay, M. 2016. Blade Type Nail Designs Cause more Varus Collaps than Screw Type Nails in the Treatment of Elderly Trochanteric Fractures, Osteosynthese International 2016, Aschaffenburg.

Okkaoğlu, Y. ve Akdi, Y. 2016. Spektral Regresyon Analizi: Türkiye'ye Ait Turizm Verilerinin Modellenmesi, 10. Uluslararası İstatistik Günleri Sempozyumu, Giresun.

Okkaoğlu, Y. ve Akdi, Y. 2018. Periodogramların Mevsimsel Hava Kirliliği Verilerine Uygulamaları, 1. Uluslararası İstatistik Matematik ve Analitik Yöntemler Kongresi, İstanbul.