Zaman Serilerinde R ile Veri Analizi; Anomalilerin Ayrıştırılması, Tahminleme Yapılması ve İlişkilerin Saptanması

May 19, 2018

Büşra Uysal 090130357

Tez Danışmanı: Doç. Dr. Atabey KAYGUN

Contents

1 Zaman Serisi Analizi **3**
- 1.1 Zaman Serisi Analizindeki Hedefler . . . 3
  - 1.1.1 Keşif Analizi: . . . 3
  - 1.1.2 Modelleme: . . . 3
  - 1.1.3 Tahmin: . . . 3
  - 1.1.4 Regresyon: . . . 3
  - 1.1.5 Süreç Kontrol: . . . 4
- 1.2 Zaman Serisi Analizindeki Anomalilerin tespiti . . . 4
- 1.3 Veri setinin yüklenmesi . . . 4
- 1.4 Zaman serilerinin açıklanması . . . 4
- 1.5 Zaman serilerinin ayrıştırılması (Decomposition) . . . 7
- 1.6 Beyaz Gürültü . . . 7
  - 1.6.1 Beyaz Gürültü Testleri . . . 10
  - 1.6.2 Shapiro-Wilks test . . . 11
  - 1.6.3 Kolmogorov-Smirnov test . . . 11
- 1.7 Durağanlık . . . 13
  - 1.7.1 Genişletilmiş Dickey-Fuller (ADF) Testi . . . 14
  - 1.7.2 KPSS (Kwiatkowski-Phillips-Schmidt-Shin) Testi . . . 14
- 1.8 Box-Jenkins Modelleri . . . 16
  - 1.8.1 Durağan Zaman Serisi Modelleri . . . 16
  - 1.8.2 Otoregresif Modeller (AR) . . . 16
  - 1.8.3 Kısmi Otokorelasyon Fonksiyonu(PACF) . . . 16
  - 1.8.4 Hareketli Ortalamalar Modeli(MA) . . . 19
  - 1.8.5 Otokorelasyon Fonksiyonu(ACF) . . . 20
  - 1.8.6 ARMA Modelleri . . . 23
  - 1.8.7 ARIMA Modelleri . . . 24
  - 1.8.8 Model kriterlerinin kullanılarak uygun ARIMA modelinin seçimi . . . 24

2 Tahminleme **32**
- 2.1 ARIMA ile Tahminleme . . . 32
- 2.2 Holt-Winters Yöntemi ile Tahminleme . . . 33
  - 2.2.1 Toplamsal Holt-Winters Yöntemi . . . 34
  - 2.2.2 Çarpımsal Holt-Winters Yöntemi . . . 34
  - 2.2.3 Uygun Tahminleme Yönteminin Seçilmesi . . . 35

3 Çok Değişkenli Zaman Serileri Analizi **37**
- 3.1 Dağıtılmış-Gecikmeli Model . . . 38
- 3.2 Kovaryans . . . 38
- 3.3 Korelasyon . . . 38
  - 3.3.1 Çapraz Korelasyon . . . 38

4 Değerlendirme **43**

5 Kaynakça **44**

1 Zaman Serisi Analizi

Bir zaman serisi, her biri belirli bir t zamanında kaydedilmekte olan X_t gözlemleri kümesidir. Kesikli zaman serileri gözlemlerin yapıldığı zamanların t değerlerinin birbirlerinden ayrık olduğunu söylemektedir. Gözlemler sabit zaman aralıklarında yapıldığı durumlar kesikli zaman serilerine örnek olarak verilebilir. Sürekli zaman serileri ise gözlemlerin sürekli bir $T[0,1]$ aralığında yapılmasıyla oluşmaktadır.[2]

Bir çok alanda bilim, mühendislik ve ticaret dalında zaman içinde sıralı olarak ölçülen veriler bulunmaktadır. Örneğin bankalar her gün faiz oranlarını ve döviz kurlarını tutar ya da meteoroloji ofisleri sıcaklık değerlerinin günlük olarak tutarlar. Bir değişken, zaman içinde belirli sabit bir aralıkta ölçüldüğünde, elde edilen veriler bir zaman serisi oluşturmaktadır.[1]

Ayrıca zaman serileri bir çok alanda karşımıza çıkan belirli aralıklarla ölçümlenmiş veri kümeleri olarak da tanımlanabilir. Temel bir veri analizinde birbirine benzer ve bağımsız dağılmış veriler mevcut iken zaman serilerinde birbiriyle ilişkili veriler bulunur. Zaman serilerindeki analizin amacı; geçmiş verilerdeki anomalileri saptamak, birbirleriyle ilişkilerini gözlemlemek ve gelecek için tahminleme yapabilmektir. Basit tanımlayıcı analizlerle verinin anlaşılması sağlanırken, kapsamlı bir analiz ile gözlenen verilerin rassal modellemesi yapılabilmektedir.[3]

Zaman serilerinin temel özellikleri trend ve mevsimsel değişimlerdir. Bunlar matematiksel fonksiyonları ile deterministik olarak modellenebilirler. Ancak, zaman serilerinin bir diğer önemli özelliği ise birbirleriyle ilişkili olmaları yani birbirlerine yakın gözlemlerin korelasyon içermesidir. Bir zaman dizisi analizindeki temel amaç, bu istatistiksel ilişkiyi ve verideki temel özellikleri uygun istatistiksel modeller ve tanımlayıcı yöntemler kullanarak açıklamaktır. Yöntemlerin uygunluğu ise daha sonrasında uygulanacak istatiksel testlerle ölçümlenebilmektedir. [1]

1.1 Zaman Serisi Analizindeki Hedefler

1.1.1 Keşif Analizi:

Zaman serileri keşif analizleri ağırlıklı olarak zaman serilerinin temel özelliklerinin (mevsimsellik, trend, korelasyon...) saptanması, grafik çizimlerinin yapılması yani deterministik ve stokastik kısımların doğru şekilde incelenebilmesi için yapılır.

1.1.2 Modelleme:

Zaman serisine uygun bir modellenmenin yapılması keşif analizinde elde edilen bilgiler yardımıyla yapılmaktadır. Düzgün modelleme yapılmayan analizler bir sonraki adımda doğru sonuçlar vermeyecektir.

1.1.3 Tahmin:

Zaman serileri analizinde sıkça kullanılan yapılardan biri olan gelecekteki gözlemlerin tahmin edilmesidir. Fakat gelecekteki verilerin tahmini zaman serisinin geçmiş ve şimdiki özelliklerinin devam ettiği varsayımına dayanır. İyi bir tahmin yapılabilmesi için keşif analizinin doğru yapılması ve modelin doğru kurulması gerekmektedir.

1.1.4 Regresyon:

Zaman serisi analizinde gözlemlerin gelecekteki tahminlerini yapmaktan başka sıkça kullanılan diğer bir yapı ise gözlemler arasındaki ilişkinin saptanmasıdır. Bu şekilde zaman serileri daha

açıklayıcı hale gelmektedir.

1.1.5 Süreç Kontrol:

Optimal yönetim ve kalite kontrol amacıyla birçok üretim veya diğer süreçler ölçümlenir. Bu genellikle rassal bir modelin uygun olduğu zaman serisi verisi ile sonuçlanır. Bu, verilerdeki sinyalin anlaşılmasını sağlamaktadır. Üretimdeki hangi dalgalanmaların normal olduğunu ve hangilerinin müdahale gerektirdiğini izlemek mümkün hale gelir.[3]

1.2 Zaman Serisi Analizindeki Anomalilerin tespiti

Anomaliler, iyi tanımlanmış normal davranış kavramına uymayan verilerdeki yapılardır. Bu yapıları bulma problemi anomallilerin tespiti olarak adlandırılır. Anomali tespitinin önemi, verideki anormalliklerin çok çeşitli uygulama alanlarında önemli ve eyleme geçirilebilir bilgilere dönüşebilir olmasıdır.

ARIMA, anomali tespiti için kullanılan bir modeldir. Sinyalleri doğru tahmin etmek ve anormallikleri bulmak için yeterince güçlü bir model olan ARIMA modelinin uygulanması için verinin trend, mevsimsellik gibi özelliklerinden arındırılması gerekmektedir.[6]

1.3 Veri setinin yüklenmesi

Veri analizinde kullanacağımız veri seti Londra'daki platin fiyatlarının dolar bazında 1990 ile 2018 yılları arasındaki günlük sabit değerleridir. Sabit fiyat, dünya çapındaki müşterilere ait siparişlerin eşleştirilmesini temsil eder.

Platin çok değerli madenler arasındadır. Çok fazla değerli olmasının nedeni ender bulunuşu ve kullanım alanlarının fazlasıyla geniş olmasıdır. Otomotiv, diş hekimliği, jet ve füze motorları, laboratuvar gibi birçok alanda kullanılmaktadır.

1.4 Zaman serilerinin açıklanması

Zaman serisi analizinde yapılması gereken ilk aşama verinin yüklenmesidir. Analizde kullanılacak paketlerin yüklenmesinin ardından, gerekli kütüphanelerde çağırılmalıdır. Bunun için aşağıdaki R kodları kullanılmaktadır.

```
In [2]: #install.packages("Quandl", repos='http://cran.us.r-project.org')
        #install.packages("corrplot", repos='http://cran.us.r-project.org')
        library(Quandl)
        library(corrplot)
        library(forecast)
        library(tseries)
        PlatinumPrices<-Quandl("LPPM/PALL")
```

Verinin yüklenmesi tamamlandıktan sonra verinin genel bir çerçevede incelenmesi yapılması gereken aşamalardan birisidir. Bunun için özetine bakılması, grafiğinin çizilmesi gerekir.

```
In [3]: summary(PlatinumPrices)
```

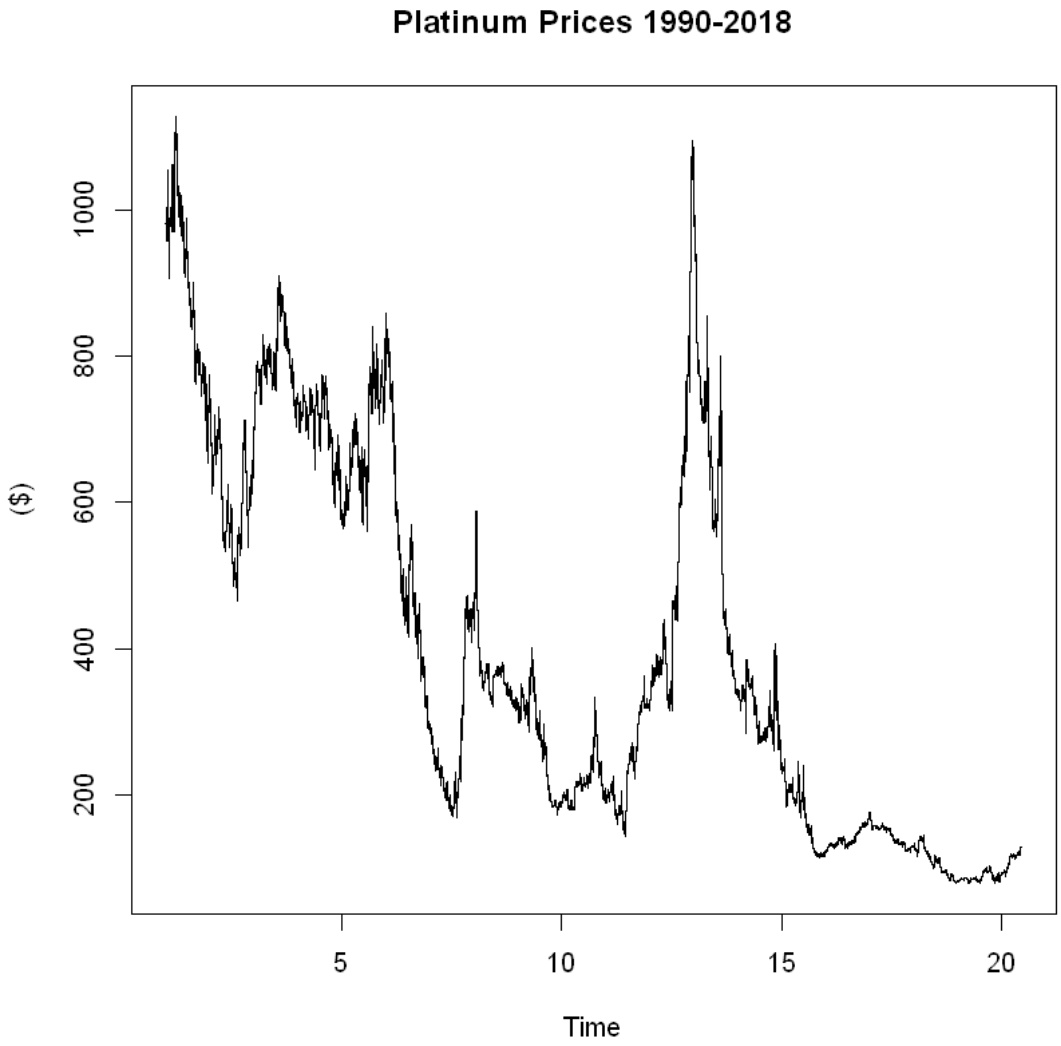
```
      Date                USD AM             EUR AM           GBP AM
 Min.   :1990-04-02   Min.   :  78.75   Min.   : 116.1   Min.   : 40.4
 1st Qu.:1997-04-11   1st Qu.: 161.56   1st Qu.: 246.3   1st Qu.: 99.7
 Median :2004-04-22   Median : 325.00   Median : 421.1   Median :179.4
 Mean   :2004-04-21   Mean   : 396.02   Mean   : 433.2   Mean   :254.6
 3rd Qu.:2011-05-04   3rd Qu.: 640.00   3rd Qu.: 577.1   3rd Qu.:414.2
 Max.   :2018-05-17   Max.   :1128.00   Max.   :1179.5   Max.   :819.5
                                        NA's   :2213
     USD PM             EUR PM           GBP PM
 Min.   :  78.25   Min.   : 122.8   Min.   : 41.0
 1st Qu.: 161.75   1st Qu.: 246.5   1st Qu.: 99.6
 Median : 325.00   Median : 419.8   Median :179.7
 Mean   : 396.25   Mean   : 433.6   Mean   :254.8
 3rd Qu.: 640.00   3rd Qu.: 578.1   3rd Qu.:415.1
 Max.   :1129.00   Max.   :1179.7   Max.   :819.2
 NA's   :53        NA's   :2248     NA's   :53
```

Yukarıdaki veri setimizdeki tüm değerlerin özeti görülmektedir. Çıkarılan özete göre, bu veri seti 1990 ile 2018 arasındaki platin fiyatlarının açılış ve kapanış fiyatlarını dolar, euro gibi para birimlerini baz alarak göstermektedir. Fakat yapılacak ilk aşama boyunca platinin dolar açılış fiyatı kullanıcak ve bu veriler arasında anomali tespiti yapılacaktır. İkinci sütunda dolar açılış fiyatlarının minimum fiyat 78.75(dolar) iken maximum 1128 (dolar) olduğu görülmektedir. Ayrıca yukarıdaki özet PlatinumPrices nesnesinin içerisinde yer alan tüm para birimleri bazında platin fiyatları içermektedir.

Zaman serileri ts sınıfının bir R nesnesi olarak saklanır. Zaman serisinin doğru analizi için PlatinumPrices nesnesinin ikinci sütununda yer alan dolar açılış fiyatları zaman serisine dönüştürülmelidir. Zaman serisi nesneleri, yukarıda verilen işlevlerin başlangıcı, sonu ve sıklığını içeren bir dizi yönteme sahiptir. Fakat zaman serileri oluşturulurken sadece frekansının yazılması da yeterli olacaktır.

Ayrıca zaman serisi analizindeki en önemli adımlardan biri verileri çizmektir; yani, grafiği oluşturmaktır. Bu işlemler aşağıdaki R komutlarıyla yapılabilir.

```
In [4]: PlatinumPricests<-ts(PlatinumPrices[,2],
                             freq=365)
        plot(PlatinumPricests,type='s',
             xlab="Time",
             ylab="($)",
             main="Platinum Prices 1990-2018")
```



Yukarıda çizilen grafikten görüldüğü üzere ele alacağımız veri seti 1990-2018 yılları arasında Platin'in dolar fiyatını içermektedir. Veri setinin grafiğini çizmek için yukarıdaki R komutu kullanılır. 2008 yılına kadar neredeyse durağan seyreden fiyatlar 2008-2010 yılları arasında bir dalgalanma olduğu çizilen grafikte görülmektedir. 2010-2015 yıllarında yüksek fiyatlarda seyretmiş ve daha sonra azalarak günümüzdeki fiyatına ulaşmaktadır.Bu dalgalanmayı içeren sinyaller veri analizi aşamalarında saptanacak ve veriden uzaklaştıralacaktır. Mevsimsellik; aylar, haftanın günleri, mevsimler, vb. gibi sürelerleilgili düzenli ve yinelenen yüksek ve düşük örüntüleri ifade etmektedir. Yukarıda görüldüğü üzere veride herhangi bir mevsimselliğe rastlanmamaktadır. Trend; ortalama olarak, ölçümlerin zamanla artma (veya azaltma) eğiliminde olması anlamına gelir. Yıllar içerisinde platin fiyatlarında bir artış olduğu görülmektedir. Aykırı değerler; regresyon çizgisinden çok uzak değerler olarak düşünülebilir.

1.5 Zaman serilerinin ayrıştırılması (Decomposition)

Zaman serilerinin çoğu trend ve mevsimsel etkiye sahiptir.

Trend : ortalama olarak, ölçümlerin zamanla artma (veya azaltma) eğiliminde olması anlamına gelir.

Mevsimsellik :yani mevsimler,aylar, haftanın günleri, gibi sürelerle ilgili düzenli ve yinelenen yüksek ve düşük örüntülerdir. Serilerin seviyesinde veya varyansta ani değişiklikler var mı?

Ayrıştırma (decomposition) modelleri ile serilerin bu özelliklerden ayrılması sağlanır. Basit bir toplamsal(additive) model aşağıdaki şekilde gösterilebilir.

$$X_t = m_t + s_t + R_t$$

Buradaki X_t zaman serisini gösterir.m_t trend bileşenini ifade ederken, s_t mevsimsel etkiyi göstermektedir. R_t genellikle ortalaması sıfır olan korelasyonlu rastgele değişkenlerin bir dizisi yani kalan terimi göstermektedir.Amaç, Rt'nin sabit bir zaman serileri süreci olduğu bir ayrıştırma bulmaktır. Böyle bir model hava yolcu rezervasyonları, işsizlik gibi bir çok modelde karşımıza çıkabilir. Ayrıca tüm bu serilerin daha yakından incelenmesi, mevsimsel etkinin ve trendin arttığında rasgele değişimin arttığını göstermektedir. Bu gibi durumlarda, çarpımsal(multiplicative) model kullanılmaktadır. Çarpımsal model aşağıdaki şekilde ifade edilebilir.

$$X_t = m_t * s_t * R_t$$

Trend m_t , mevsimsellik s_t ve kalan terimlerin R_t tahminlemesi bir çok açıdan yapılabilir. Örneğin Decompose() fonksiyonu trend, mevsimsellik değerlerini modelin toplamsal ya da çarpımsal parametresine göre ayrıştırmaktadır. Hem toplamsal(additive) hem de çarpımsal(multiplicative) model için zaman serisinin ayrıştırılması gösterilmiştir. R'nin stl () komutu, periyodik bir zaman serisinin trend, mevsimsellik ve geri kalanına ayrışmasını sağlayan başka bir yöntemdir. Tüm tahminler LOESS yumuşaklığına dayanmaktadır. Çıktısı neredeyse decompose () ile elde ettiğimiz değerlere benzer olmasına rağmen, bu fonksiyon daha güvenilir sonuçlar verebilmektedir. [3]

1.6 Beyaz Gürültü

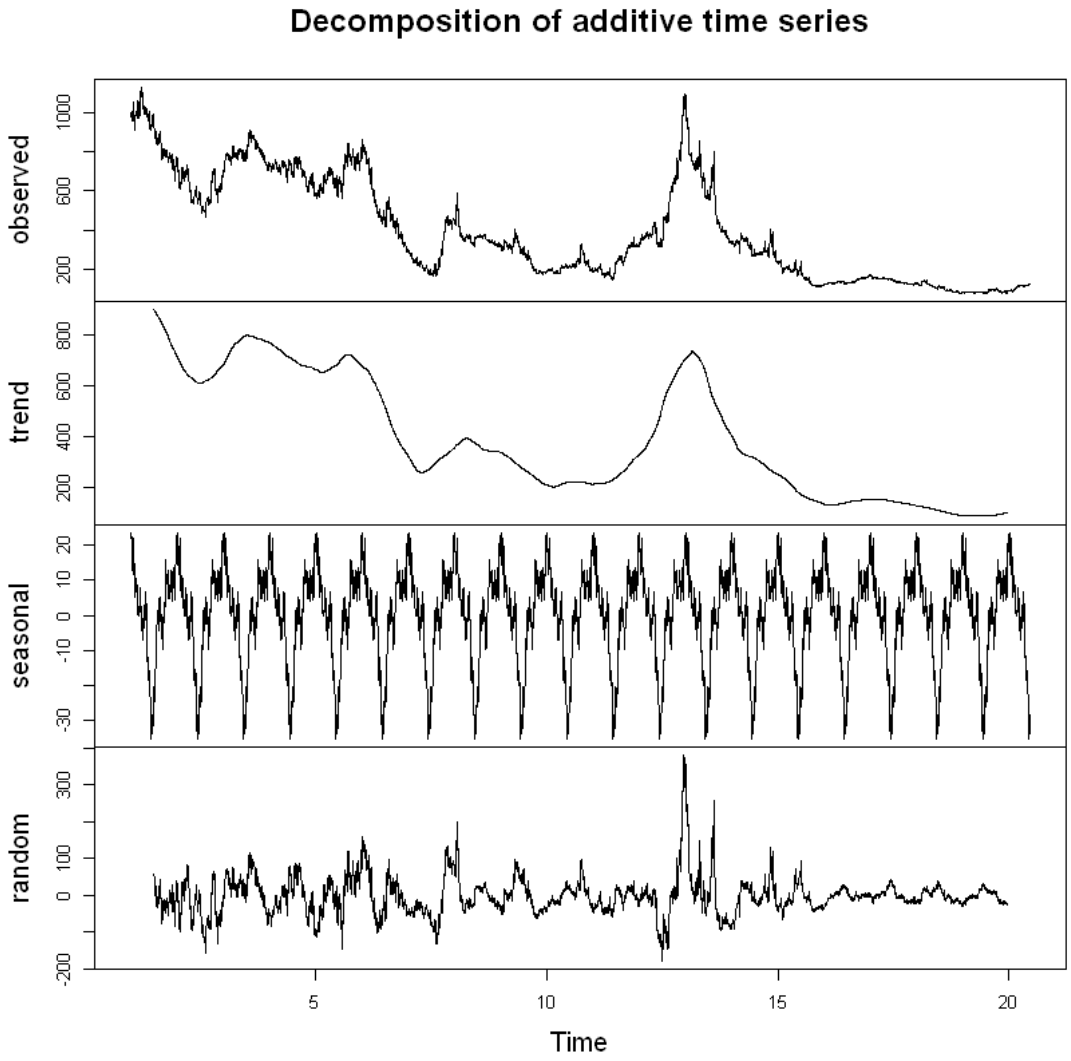
Veri setinin içindeki trend, mevsimsellik ve üssel dağılımları yukarıda anlatılan işlemlerle çıkarıldıktan sonra elimizde kalan grafiğin eğer herhangi bir aykırı değerler içermiyorsa beyaz gürültü olması beklenmektedir. Trendin ayrıştırılması için linear regresyon, mevsimselliğin ayrıştırılması için fourrier serileri ve üssel dağılımın ayrıştırılması için ise ARIMA modellerinin çıkarılması gerekmektedir. Aslında decompose() ve stl() fonksiyonları bu trend ve mevsimsellik için bu ayrıştırmayı yapmaktadır. Bu yüzden ayrışmadan elde kalan R_t random fonksiyonunu gürültü fonksiyonuna yakın olması beklenmektedir.

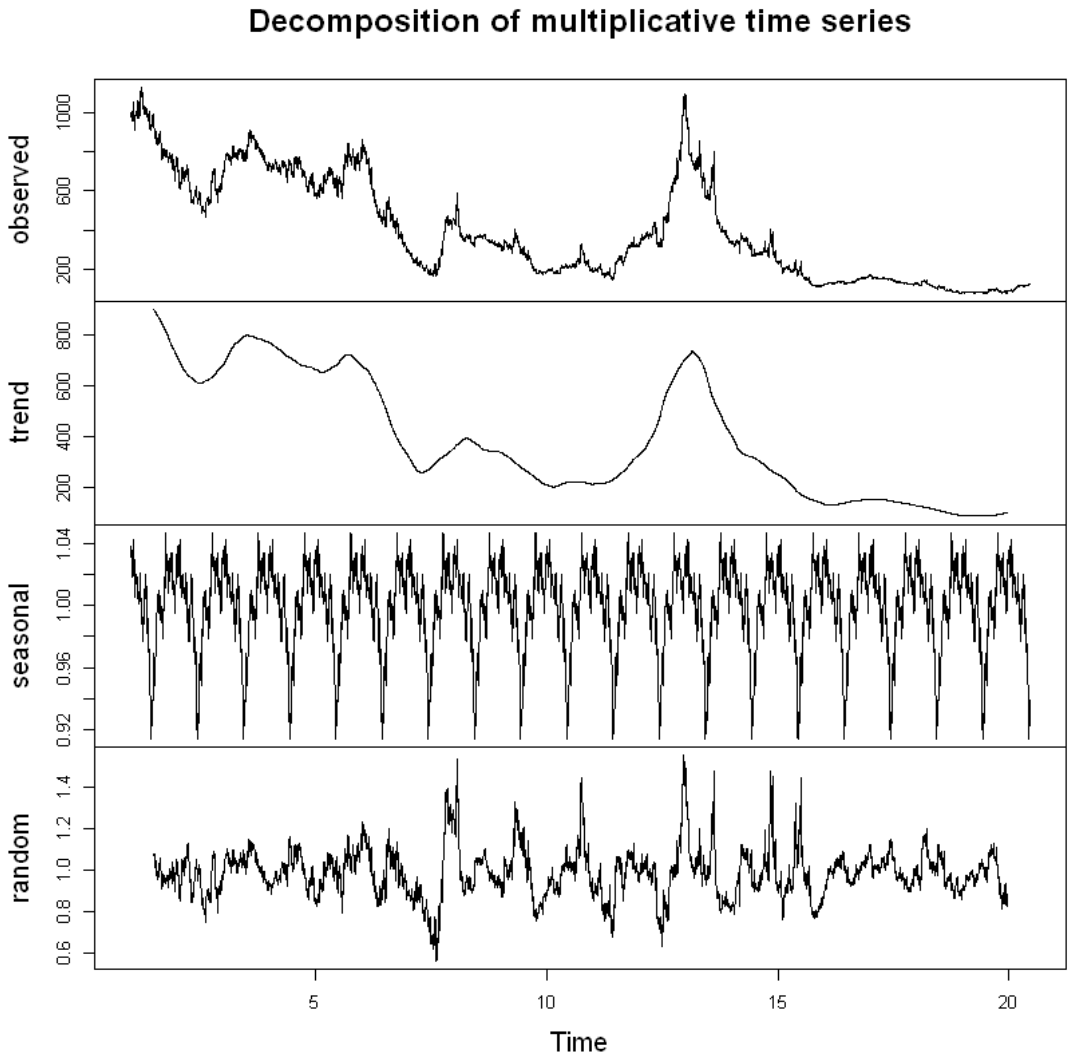
Beyaz gürültü : Bir X_t fonksiyonu bağımsız ve ortalaması sıfır ve varyansı σ^2 olarak dağılıma sahipse Beyaz Gürültü olarak adlandırılır. Bir beyaz gürültü aşağıdaki notasyon ifade edilebilir. [2]

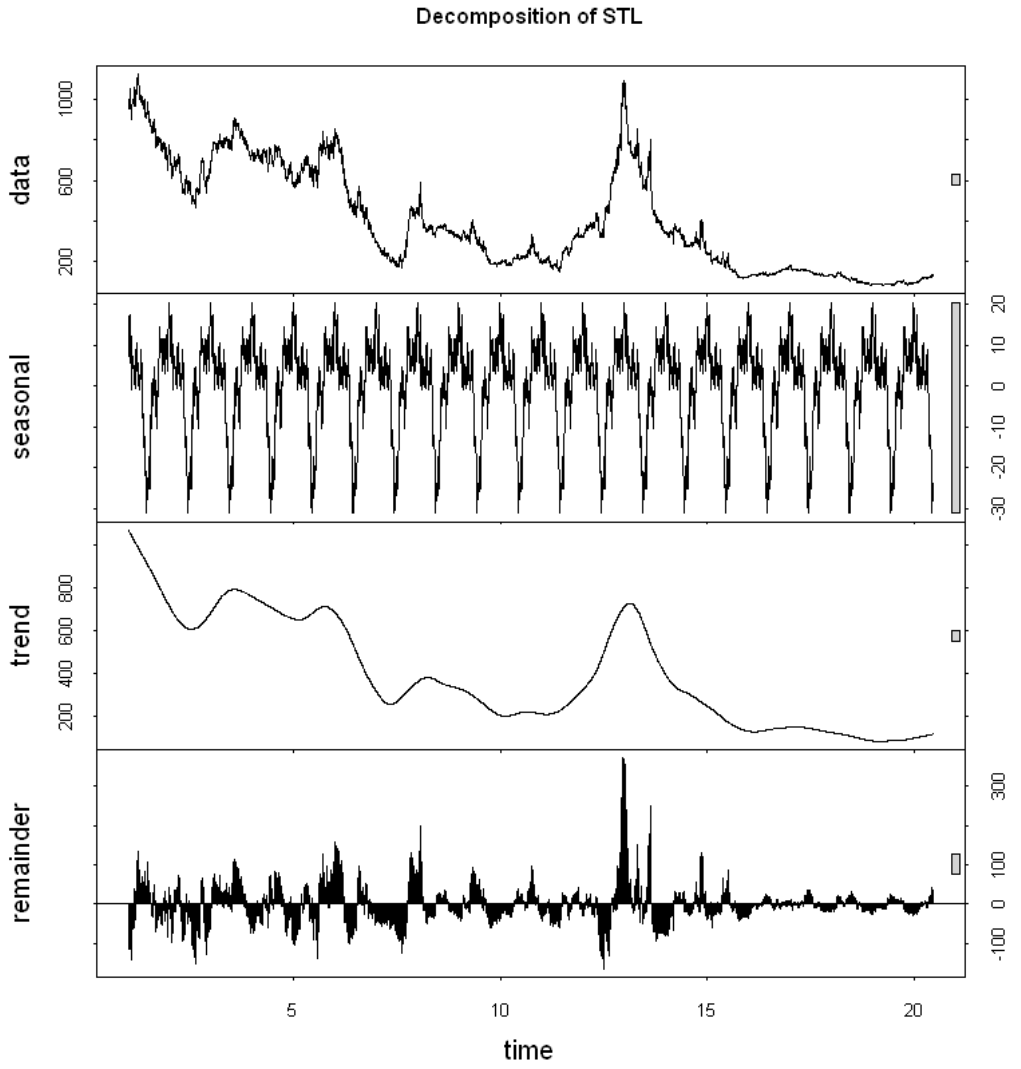
$$Xt \sim N(0, \sigma^2)$$

Aşağıda decompose() işleminin hem toplamsal, hem çarpımsal modele göre ayrıştırılması aynı zamanda stl() e göre yapılan ayrıştırmanın R komutları görülmektedir. Ayrıca bu ayrıştırmaların grafikleri oluşturulmuştur.

```
In [5]: PlatinumPricesDecomposeA<-decompose(PlatinumPricests,
                                            type = c("additive"),
                                            filter = NULL)
        plot(PlatinumPricesDecomposeA)
        PlatinumPricesDecomposeM<-decompose(PlatinumPricests,
                                            type = c("multiplicative"),
                                            filter = NULL)
        plot(PlatinumPricesDecomposeM)
        PlatinumPricesSTL<-stl(PlatinumPricests,s.window='periodic')
        plot(PlatinumPricesSTL,main='Decomposition of STL')
```




Decomposition of multiplicative time series
observed
trend
seasonal
random
Time



Zaman serisi ayrıştırmasındaki *observed* ile gösterilen verilerin gerçek grafiğidir. *Trend* bileşeni m_t yani eğilimi gösterirken, *seasonal* verideki mevsimselliği yani s_t temsil etmektedir. *Random* ise gerçek verilerden m_t ve s_t değerlerinin çıkarılması ile oluşur. Bu da R_t kalan değeri gösterir. Stl() ile yapılan ayrıştırma işleminde ise *remainder* trend ve mevsimsellik çıkarıldığındaki kalan değeri ifade etmektedir.

decompose() ve stl() ile yapılan ayrıştırmalar incelendiğinde hangisinin verinin ayrıştırmada doğru sonuçlar vereceği yani random fonksiyonlarından hangisinin beyaz gürültü fonksiyonuna daha yakın olduğu göz ile tespit edilememektedir. Bunun için normallik testlerinden yararlanılması gerekmektedir.

1.6.1 Beyaz Gürültü Testleri

Yukarıda yapılan ayrıştırma işlemlerinin amacı belirgin bir sapma göstermeyen ve özellikle belirgin bir eğilim veya mevsimselliğe sahip olmayan bir seri üretmektir.Bu aşamadan bir sonraki

adım tahmini gürültü serisini modellemektir (Yani, verilerden trend ve mevsimsel bileşenleri tahmin ederek ve çıkararak elde edilen artıklar).Eğer bu artıklar arasında bir bağımlılık yoksa, onları bağımsız rasgele değişkenlerin gözlemleri olarak görebiliriz ve onların ortalama ve varyanslarını tahmin etmek dışında başka bir modelleme yoktur. Ancak, artıklar arasında önemli bir bağımlılık varsa, o zaman bağımlılığı açıklayan gürültü için daha karmaşık bir durağan zaman serisi modelini aramamız gerekir.

R_t serisinin beyaz gürültü olup olmadığı için testler kullanılmalıdır. Eğer beyaz gürültü ise zaman serisi açıklanmış olacaktır ve içerisinde trend ve mevsimsellikten başka bir anamoliye rastlanmamıştır diyebiliriz.Fakat değil ise o zaman ARIMA gibi modellerle içerisindeki anomaliler saptanıcaktır.[2]

1.6.2 Shapiro-Wilks test

Shapiro ve Wilk tarafından geliştirilen Shapiro-Wilk testi, çoğu durumda en güçlü ve çok amaçlı testtir. Son yıllarda, SW testi, çok çeşitli alternatif testlere kıyasla iyi güç özellikleri nedeniyle normalliğin tercih edilen testi haline gelmiştir.W test istatistiği dağılımın normal olup olmadığına karar verilmesi için kullanılır. W test istatistiği aşağıdaki notasyon ile ifade edilir.

$$W = \frac{\sum_i (a_i x_i)^2}{\sum_i (x_i - \bar{x})^2}$$

x_i veri setindeki değerleri, $\bar{x}$,ortalamayı ve n, gözlemlerin sayısını temsil eder.x_t herhangi bir veri seti olmak üzere normalliğinin testi için kullanılan Shapiro-Wil test için Shapiro.test(x_t) R komutu kullanılır. Eğer p değeri, istenilen α değerinden küçük ise o zaman normallik hipotezini reddedilir. Test istatistiği 0 <W $\leq$ 1 arasındadır. W değerinin 1'e yakın değerler için normallik hipotezi reddedilmez. Daha küçük W için reddedilecektir. Yani W değeri büyük olan R_t beyaz gürültüye daha çok benzemektedir. Test sadece n $\geq$ 3 değerleri için geçerlidir ve R uygulaması n 5,000'e kadar izin vermektedir.

1.6.3 Kolmogorov-Smirnov test

KS testi ilk olarak Kolmogorov tarafından önerilmiştir ve daha sonra Smirnov tarafından geliştirilmiştir. Bu test verilerin kümülatif dağılımını beklenen kümülatif normal dağılım ile karşılaştırır. Yani iki örneklemi karşılaştırmak için kullanılmaktadır.Test istatistiği D normalliğe karar vermek için kullanılır. D test istatistiği için küçük değerleri, normalliği ifade etmektedir. D test istatistiği aşağıdaki notasyon ile ifade edilir.

$$D = \max[F_x(u) - F_y(u)]$$

Fakat bizim test etmemiz gereken bir örneklem olduğu için R'da kullanılan ks.test() fonskiyonunda ikinci örneklem için 'pnorm' yani örnek bir kümülatif dağılımı temsil etmektedir. [3]

Shapiro-Wilks test ile 3 ile 5000 tane veri test edilebilirken Kolmogorov-Smirnov testi ile daha çok veri test edilebilir. Aşağıda ayrıştırmalardan kalan veri setleri için normallik testleri uygulanmıştır. Shapiro-Wilk testi 5000 veri ile çalıştığı için mevcut veri setimizin içerisinden örneklem oluşturarak daha doğru sonuçlar elde edilecektir.

```
In [6]: SamplePlatinumPricesDecomposeA<-sample(PlatinumPricesDecomposeA$random,5000)
        SamplePlatinumPricesDecomposeM<-sample(PlatinumPricesDecomposeM$random,5000)
        SamplePlatinumPricesDecomposeSTL<-sample(PlatinumPricesSTL$time.series[,3],5000)
        shapiro.test(SamplePlatinumPricesDecomposeA)
```

```
        shapiro.test(SamplePlatinumPricesDecomposeM)
        shapiro.test(SamplePlatinumPricesDecomposeSTL)
        ks.test(PlatinumPricesDecomposeA$random,
                'pnorm',
                mean=0,
                sd=sqrt(var(na.omit(PlatinumPricesDecomposeA$random))))
        ks.test(PlatinumPricesDecomposeM$random,
                'pnorm',
                mean=0,
                sd=sqrt(var(na.omit(PlatinumPricesDecomposeM$random))))
        ks.test(PlatinumPricesSTL$time.series[,3],
                'pnorm',
                mean=0,
                sd=sqrt(var(na.omit(PlatinumPricesSTL$time.series[,3])))))
```

```
Shapiro-Wilk normality test

data:  SamplePlatinumPricesDecomposeA
W = 0.88658, p-value < 2.2e-16
```

```
Shapiro-Wilk normality test

data:  SamplePlatinumPricesDecomposeM
W = 0.96362, p-value < 2.2e-16
```

```
Shapiro-Wilk normality test

data:  SamplePlatinumPricesDecomposeSTL
W = 0.88739, p-value < 2.2e-16
```

```
One-sample Kolmogorov-Smirnov test

data:  PlatinumPricesDecomposeA$random
D = 0.10989, p-value < 2.2e-16
alternative hypothesis: two-sided
```

```
One-sample Kolmogorov-Smirnov test

data:  PlatinumPricesDecomposeM$random
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
```

```
One-sample Kolmogorov-Smirnov test

data:  PlatinumPricesSTL$time.series[, 3]
D = 0.11444, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Shapiro-Wilk testinin çıktılarına göre en büyük W değeri çarpımsal model kullanılarak yapılan ayrıştırmanın 5000 verilik örnekleminden kalan verilerinde görülmektedir. Yani shapiro-wilk testine göre çarpımsal modelin R_t değeri normalliğe daha yakındır. Fakat Kolmogorov-Smirnov testi ise D değeri küçük olan toplamsal ve stl ile yapılan ayrıştırmadan kalan R_t'nin normalliğe daha yakın olduğunu göstermektedir. Buradan anlaşılacağı üzere herhangi bir ayrıştırmanın en iyi olacağı sonucuna varılamamıştır. Yani shapiro-wilk testi ile Kolmogorov-Smirnov testi birbirleriyle çelişki çıktılara sahiptir. Bundan dolayı bir sonraki adımda yapılacak olan Box-Jenkins modelleri hem toplamsal hem de çarpımsal modellerin kalanları için uygulanmalıdır. Uygun Box-Jenkins modelin tespiti için serinin durağan olup olmadığına karar verilmesi gerekir. Durağan bir seri için Otoregresyon Modelleri (AR),Hareketli Ortalama Modelleri (MA),Otoregresif Hareketli Ortalama Modelleri(ARMA) kullanılırken durağan olmayan seriler için bütünleşik hareketli otoregresif modeller (ARIMA) kullanılır. Modellerin uygulanması için öncelikli olarak serinin durağan olup olmadığının testleri yapılmalı ve durağan olması sağlanmalıdır.

1.7 Durağanlık

Bir zaman serisi $X_t, t = 0, \pm 1, ...$, eğer her h tam sayısı için $X_{t+h}, t = 0, \pm 1, ...$ zaman kaydırılmış serisi ile benzer istatistiksel özelliklere sahipse durağandır. $\{X_t\}$ bir zaman serisi aşağıdaki özelliklere sahip olsun.

Beklenen Değer φ: E(X_t),

Varyans σ^2: Var(X_t) ,

Kovaryans γ: Cov (X_t,X_{t+h})

Başka bir deyişle durağan zaman serisi verilerinin belirli bir zaman sürecinde sürekli artma veya azalmanın olmadığı, verilerin zaman boyunca bir yatay eksen boyunca saçılım gösterdiği biçimde tanımlanır. Genel bir tanımlama ile, sabit ortalama, sabit varyans ve seriye ait iki değer arasındaki farkın zamana değil, yalnızca iki zaman değeri arasındaki farka bağlı olması şeklinde ifade edilir. Bir zaman serisinin durağan olup olmadığı 2 yöntem ile saptanabilir;

1.Serilerin zaman yolu grafiğinde ve onun korelogramında otokorelasyon ve kısmi otokorelasyon katsayıları üzerinde yapılan subjektif yargılar ile

2.Birim köklerin varlığını için istatistiki testlerin kullanılması ile.

Durağanlık için kullanılan bazı testlere örnek olarak Genişletilmiş Dickey-Fuller (ADF) Testi ve KPSS Testi verilebilir.(Marcel Dettling,2014)

1.7.1 Genişletilmiş Dickey-Fuller (ADF) Testi

Genişletilmiş Dickey Fuller Testi (ADF) durağanlık için birim kök testidir. Birim kökler, zaman serisi analizinizde öngörülemeyen sonuçlara neden olabilir. Artırılmış Dickey-Fuller testi, seri korelasyon ile kullanılabilir. ADF testi, Dickey-Fuller testinden daha karmaşık modelleri ele alabilir ve aynı zamanda daha güçlüdür. Sıfır hipotezi serinin durağan olmadığını ve birim kök içerdiğini, buna karşın alternatif hipotez ise seride birim kök olmadığını ve durağan olduğu ifade eder.

1.7.2 KPSS (Kwiatkowski-Phillips-Schmidt-Shin) Testi

KPSS testinde amaç gözlenen serideki deterministik trendi arındırarak serinin durağan olmasını sağlamaktır. Bu testte kurulan birim kök hipotezi ADF testinde kurulan hipotezlerden farklıdır. Sıfır hipotezi serinin durağan olduğunu ve birim kök içermediğini, buna karşın alternatif hipotez ise seride birim kök olduğunu ve durağan olmadığını ifade eder. Boş hipotezdeki durağanlık trend durağanlıktır. Çünkü seriler trendden arındırılmışlardır. [1]

Aşağıda durağanlık tespiti için ADF ve KPSS testleri yapılmıştır.

```
In [7]: adf.test(na.omit(SamplePlatinumPricesDecomposeA))
        adf.test(na.omit(SamplePlatinumPricesDecomposeM))
        adf.test(na.omit(SamplePlatinumPricesDecomposeSTL))
        kpss.test(na.omit(SamplePlatinumPricesDecomposeA))
        kpss.test(na.omit(SamplePlatinumPricesDecomposeM))
        kpss.test(na.omit(SamplePlatinumPricesDecomposeSTL))
```

```
Warning message in adf.test(na.omit(SamplePlatinumPricesDecomposeA)):
"p-value smaller than printed p-value"


Augmented Dickey-Fuller Test

data:  na.omit(SamplePlatinumPricesDecomposeA)
Dickey-Fuller = -16.731, Lag order = 16, p-value = 0.01
alternative hypothesis: stationary



Warning message in adf.test(na.omit(SamplePlatinumPricesDecomposeM)):
"p-value smaller than printed p-value"


Augmented Dickey-Fuller Test

data:  na.omit(SamplePlatinumPricesDecomposeM)
Dickey-Fuller = -17.946, Lag order = 16, p-value = 0.01
alternative hypothesis: stationary



Warning message in adf.test(na.omit(SamplePlatinumPricesDecomposeSTL)):
"p-value smaller than printed p-value"
```

```
Augmented Dickey-Fuller Test

data:  na.omit(SamplePlatinumPricesDecomposeSTL)
Dickey-Fuller = -16.765, Lag order = 17, p-value = 0.01
alternative hypothesis: stationary



Warning message in kpss.test(na.omit(SamplePlatinumPricesDecomposeA)):
"p-value greater than printed p-value"


KPSS Test for Level Stationarity

data:  na.omit(SamplePlatinumPricesDecomposeA)
KPSS Level = 0.20715, Truncation lag parameter = 15, p-value = 0.1



Warning message in kpss.test(na.omit(SamplePlatinumPricesDecomposeM)):
"p-value greater than printed p-value"


KPSS Test for Level Stationarity

data:  na.omit(SamplePlatinumPricesDecomposeM)
KPSS Level = 0.10767, Truncation lag parameter = 15, p-value = 0.1



Warning message in kpss.test(na.omit(SamplePlatinumPricesDecomposeSTL)):
"p-value greater than printed p-value"


KPSS Test for Level Stationarity

data:  na.omit(SamplePlatinumPricesDecomposeSTL)
KPSS Level = 0.035836, Truncation lag parameter = 16, p-value = 0.1
```

ADF teste göre p değeri çok küçük çıktığı için göre serinin durağan olduğu söylenebilir. KPSS testine göre ise STL ile yapılan ayrıştırmanın durağan olmadığını göstermektedir. Fakat toplamsal ve çarpımsal modele göre yapılan ayrıştırmanın KPSS testinde p değeri büyük olduğu için sıfır hipotezi reddedilemez ve durağan oldukları sonucuna varılır.

1.8 Box-Jenkins Modelleri

Gecikmeli doğrusal ilişkiler yoluyla ortaya çıkabilen korelasyonun getirilmesi, otoregresif (AR), hareketli ortalamalar(MA) otoregresif hareketli ortalama (ARMA) modellerin önerilmesine yol açmaktadır. Ayrıca durağan olmayan modeller için Box ve Jenkins (1970) tarafından yapılan otoregresif entegre hareketli ortalama (ARIMA) modeli ortaya çıkmıştır.

1.8.1 Durağan Zaman Serisi Modelleri

1.8.2 Otoregresif Modeller (AR)

Bir zaman serisi modelinin en doğal formülasyonu, geçmiş değerleriyle doğrusal bir ilişkiye sahip olması yani serinin kendisinin herhangi bir gerilemesi ile arasında regresyon olmasıdır. Bu, otoregresif model ile yapılır ve zaman serilerinin açıklamanın en kullanılan modelidir.Otoregresif modeller, x_t serisinin mevcut değerinin, geçmiş değerlerin, x_{t-1},x_{t-2},...,x_{t-p} nin bir fonksiyonu olarak açıklanabileceği fikrine dayanmaktadır. Burada p, mevcut değeri tahmin etmek için gerekli olan geçmiş adım sayısını belirler. $AR(p)$ aşağıdaki denklemlere göre geçmiş gözlemlerin doğrusal bir kombinasyonuna dayanır:

$$X_t = \sum_i a_i X_{t-i} + w_t$$

Buradaki a_i değerleri katsayılardır. w_t terimi bir Beyaz Gürültü işleminden gelir. $AR(p)$ modelleri sadece sabit zaman serilerine uygulanabilir. Herhangi bir eğilim ve / veya mevsimsel etkilerin öncelikle kaldırılması gerekir. Verilere bir $AR(p)$ modelinin p değerinin saptanması kısmi otokorelasyon fonksiyonunun(PACF) analizine dayanır. PACF analizine göre ilk olarak makul görünen en basit modeli denenir. PACF'nin 'cut-off' çizgisini kestiği tüm değerler olası p değerleridir. İlk olarak en küçük p değerinden başlanarak diğer değerler denenir.

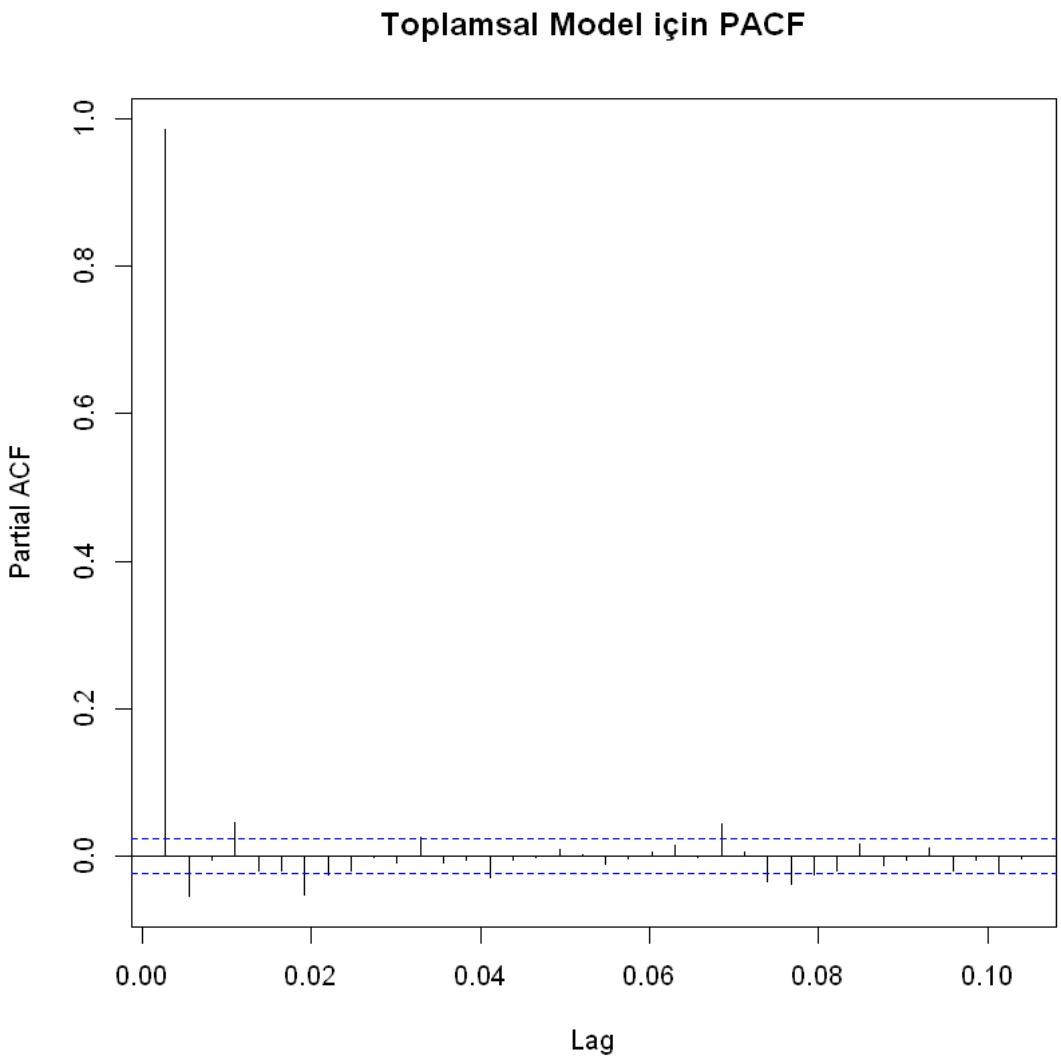
1.8.3 Kısmi Otokorelasyon Fonksiyonu(PACF)

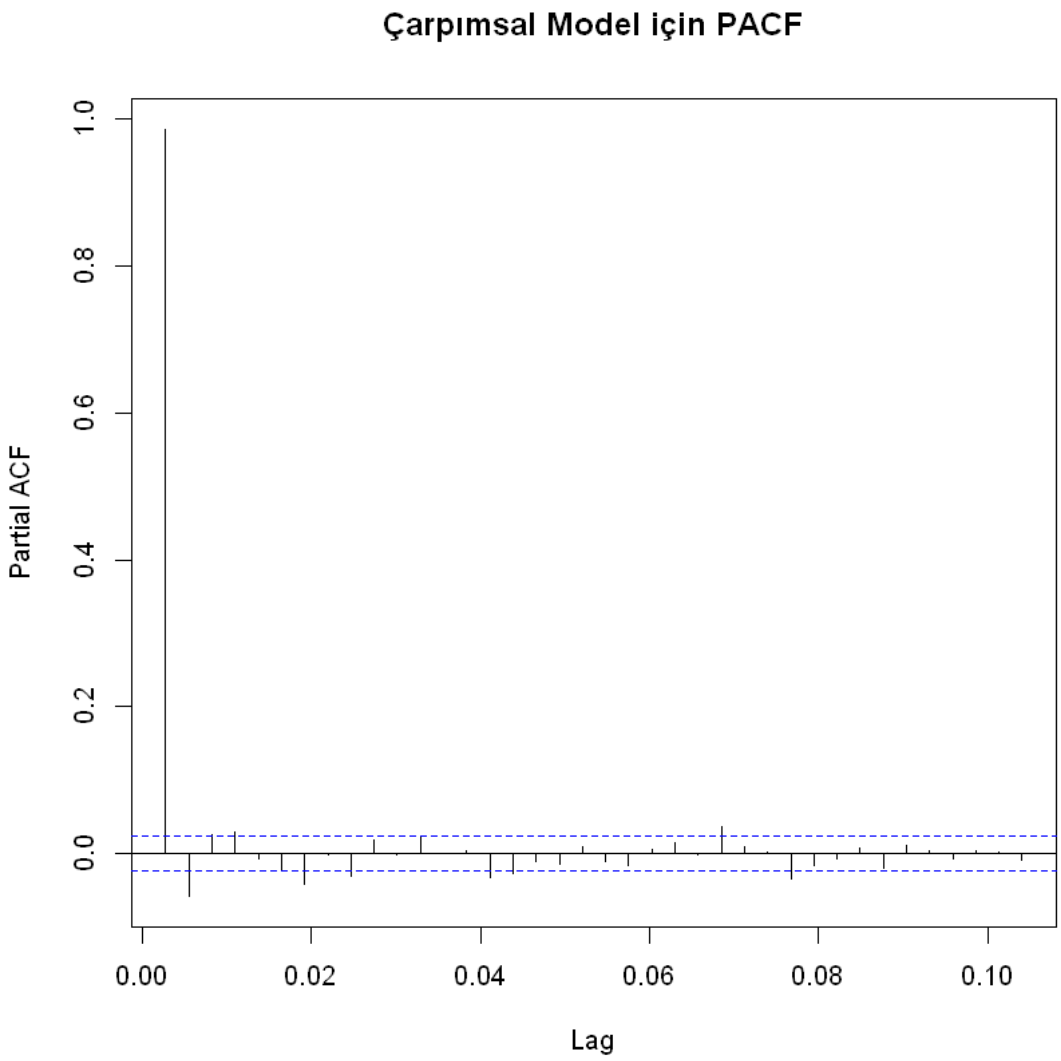
Genel olarak, kısmi bir korelasyon bir koşullu korelasyondur.Diğer değişkenler grubunun değerlerini bildiğimiz ve hesaba kattığımız varsayımı altında iki değişken arasındaki ilişkidir.Y ve x_3 arasındaki kısmi korelasyon, hem Y hem de x_3'ün x_1 ve x_2 ile ilgili olduğunu dikkate alarak belirlenen değişkenler arasındaki korelasyondur. N. dereceden Kısmi korelasyonu aşağıdaki notasyon ile tanımlayabiliriz. [3]

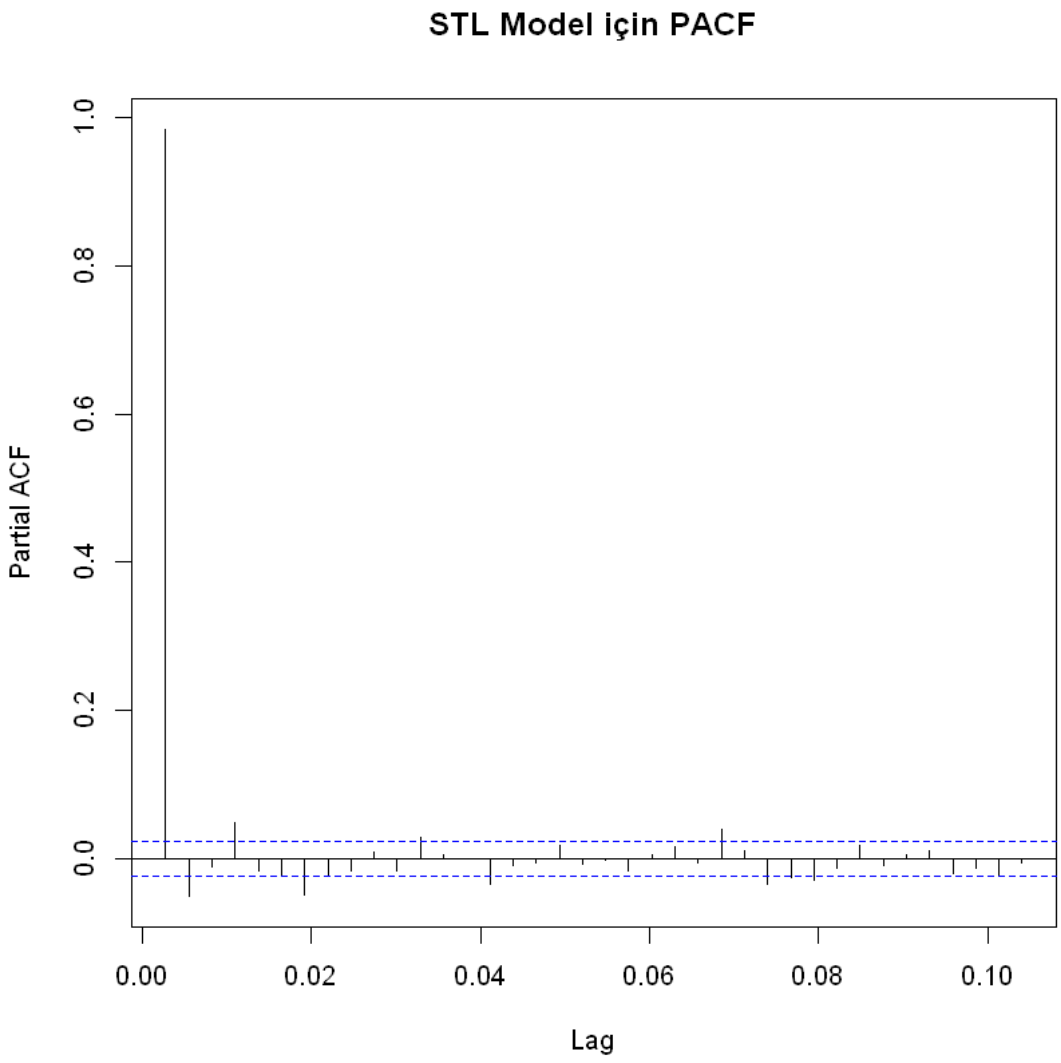
$$\frac{Kovaryans(x_t, x_{t-n}|x_{t-1}, x_{t-2}, ...x_{t-n+1})}{\sqrt{(Varyans(x_t|x_{t-1}, x_{t-2}, ...x_{t-n+1})Varyans(x_{t-n}|x_{t-1}, x_{t-2}, ...x_{t-n+1})}}$$

Aşağıda trend ve mevsimselliği ayrıştırıldıktan sonra kalan R_t fonksiyonlarının PACF korelogramları hesaplanmıştır.

```
In [8]: pacf(na.omit(PlatinumPricesDecomposeA$random),
             main='Toplamsal Model için PACF')
        pacf(na.omit(PlatinumPricesDecomposeM$random),
             main='Çarpımsal Model için PACF')
        pacf(na.omit(PlatinumPricesSTL$time.series[,3]),
             main='STL Model için PACF')
```


Toplamsal Model için PACF
Partial ACF
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.02
0.04
0.06
0.08
0.10
Lag


Çarpımsal Model için PACF
Partial ACF
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.02
0.04
0.06
0.08
0.10
Lag



Hesaplanan PACF korelogramlarına göre AR(p) modelinin p değerleri toplamsal modele göre {1,2,4,7}'dir. Çarpımsal modele göre ise {1,2,3,4}, stl modeline göre yapılan ayrıştırmaya göre ise {1,2,4,7}'dir.

1.8.4 Hareketli Ortalamalar Modeli(MA)

Hareketli ortalama (MA) modeli, mevcut beyaz gürültü teriminin ve yakın geçmişte geçen beyaz gürültü terimlerinin doğrusal bir kombinasyonu şeklinde tanımlanabilir. Birçok açıdan hareketli ortalama modellerin otoregresif modellere tamamlayıcısı olmaktadır. Yukarıda bahsedildiği üzere, MA(q) modelinin E_t yani hata terimlerinin bir kombinasyonu olduğu görülmektedir. Aşağıdaki notasyon ile ifade edilebilir;

$$x_t = E_t + \theta_1 E_{t-1} + \theta_2 E_{t-2} + \ + \theta_q E_{t-q}$$

Buradaki E_t değeri sıfır ortalama ve varyanslı beyaz gürültüdür. Verilere bir $MA(q)$ modelinin q değerinin saptanması otokorelasyon fonksiyonunun(ACF) analizine dayanır.

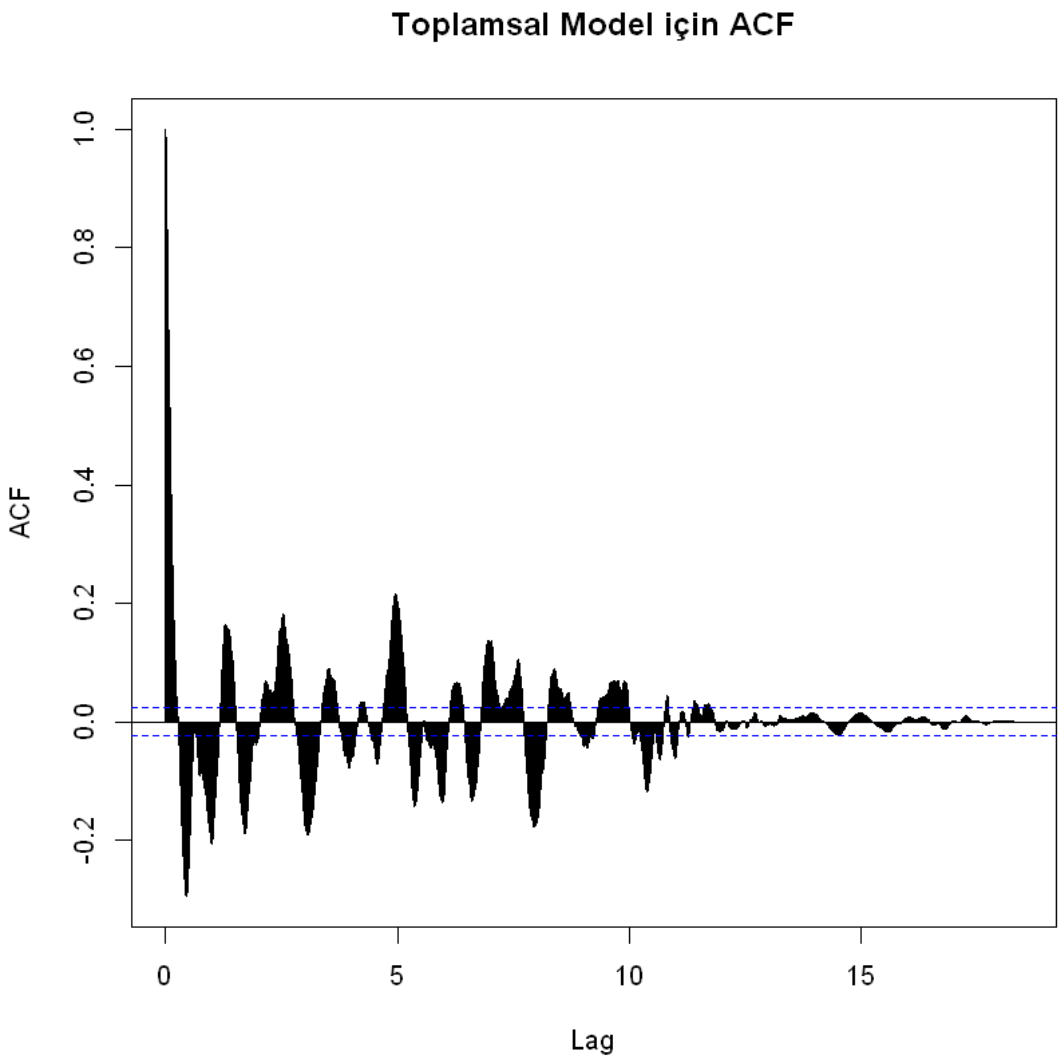
1.8.5 Otokorelasyon Fonksiyonu(ACF)

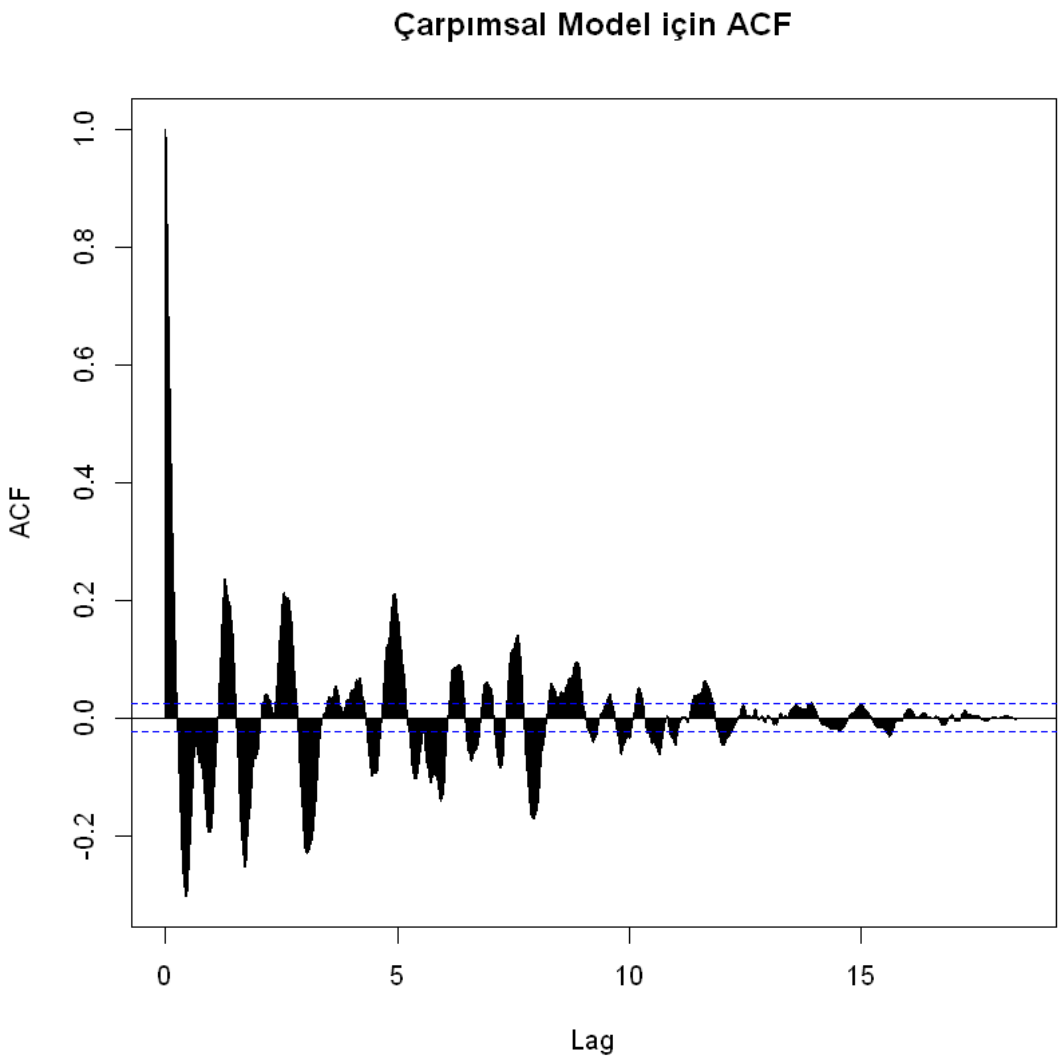
Bir zaman serisi için otokorelasyon fonksiyonu (ACF), dizinin x_t ve 1, 2, 3 ve benzeri gecikmeler için dizinin gecikmeli değerleri arasındaki korelasyonları verir. Gecikmiş değerler x_{t-1}, x_{t-2}, x_{t-3} ve benzeri şekilde yazılabilir. ACF,x_t vex_{t-1}, x_{t-2}, x_{t-3},.. arasındaki korelasyonları verir. X_t ile X_{t+k} arasındaki otokorelasyon;

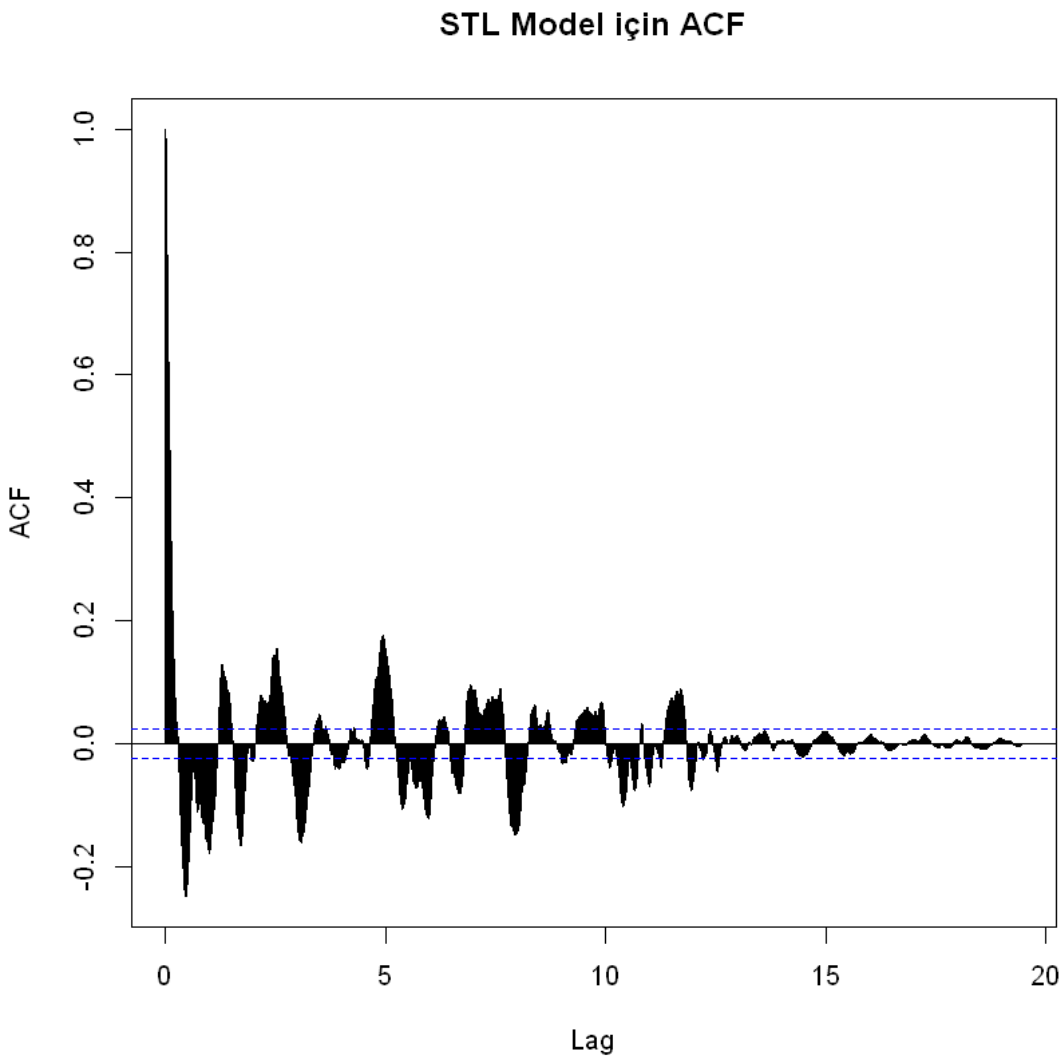
$$Cov(X_{t+k}, X_t) = \frac{Kovaryans(x_t, x_{t+k})}{\sqrt{(Varyans(x_t)Varyans(x_{t+k})}}$$

[3] Aşağıda trend ve mevsimselliği ayrıştırıldıktan sonra kalan R_t fonksiyonlarının ACF korelogramları hesaplanmıştır.

```
In [9]: acf(na.omit(PlatinumPricesDecomposeA$random),lag.max = 50000,
            main='Toplamsal Model için ACF')
        acf(na.omit(PlatinumPricesDecomposeM$random),lag.max = 50000,
            main='Çarpımsal Model için ACF')
        acf(na.omit(PlatinumPricesSTL$time.series[,3]),lag.max = 50000,
            main='STL Model için ACF')
```


Toplamsal Model için ACF
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag


Çarpımsal Model için ACF
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag



ACF korelagramına bakılarak MA(q) modeli için q parametresi tahminlemesi yapılamamaktadır.

1.8.6 ARMA Modelleri

ARMA modelleri AR(p) ve MA(q) modellerinin birlikte kullanılması durumundan ortaya çıkan model ARMA(p,q) olarak adlandırılır. Önemi, çok daha geniş bir bağımlılık yapıları yelpazesinin modellenmesinin mümkün olduğu ve bunların da karmaşık olduğu gerçeğinde yatmaktadır. Çoğu zaman, bir ARMA(p,q) sadece AR veya MA süreçleriyle bulunan p,q değerlerinden daha az sayıda parametre gerektirmektedir. Burada model hem $\{E_t\}$ kalan terimlerinin hem de $\{X_t\}$ önceki verilerin kombinasyonu şeklindedir. Bir ARMA (p,q) modeli aşağıdaki notasyon ile ifade edilebilir:

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + ... + a_p x_{t-p} + E_t + \theta_1 E_{t-1} + \theta_2 E_{t-2} + ... + \theta_q E_{t-q}$$

1.8.7 ARIMA Modelleri

Geniş bir dizi durağan olmayan seriyi içeren bu sınıfın bir genellemesi, ARIMA süreçleri, yani birçok kez farklılaştıklarında ARMA işlemlerine indirgeyen süreçler tarafından sağlanır. $\{x_t\}$ serisi bir ARMA(p,q) modelinden elde edilen bir seri olmak üzere, eğer d^{th} dereceden farkı alınmasıyla elde edilen süreç ARIMA olarak adlandırılır.

$$Y_t = (1 - B)^d X_t$$

şeklinde ifade edilir. Buradaki B, 'backshift' operatörüdür. R'da auto.arima fonksiyonu AIC, AICc veya BIC değerine göre en iyi ARIMA modelinin oluşturulmasını sağlar.[3]

1.8.8 Model kriterlerinin kullanılarak uygun ARIMA modelinin seçimi

Model seçim kriterlerine örnek olarak AIC ve BIC verilebilir.

Akaike Bilgi Kriteri (Akaike Information Criteria, AIC) Modelin kalanlarından kareler toplamı üzerinde örneklem büyüklüğü ve değişken sayısını dikkate alır ve bir düzenleme yaparak elde edilen değer sayesinde farklı modeller arasında en uygununu seçmeye yarayan bir kriterdir. Akaike tarafından 1974 yılında kazandırılmış olan AIC değeri her model için tahmin edilir ve bu değerin daha küçük olduğu modelin daha uygun olduğu ifade edilir.AIC aşağıdaki formülle hesaplanabilir.

$$AIC = -2log(L) + 2k$$

Formülde yer alan k sabit terim dahil parametre sayısı, n gözlem sayısını ve L model için olabilirlik fonksiyonu maksimize değerini göstermektedir.

Bayes bilgisi kriteri (Bayesian Information Criterion, BIC) Doğrusal regresyonda seçilmiş model problemleri için BIC model seçim kriterini kullanılır. AIC gibi farklı modeller arasında en uygununu seçmeye yarayan bir kriterdir. BIC değeri küçük olan model daha uygundur. AIC aşağıdaki formülle hesaplanabilir.

$$BIC = -2log(L) + klog(n)$$

Model seçim kriterleri olarak minimum AIC ve BIC değerleri kullanılır. En uygun model, minimum AIC ve BIC'ye göre seçilir. Aşağıda toplamsal, çarpımsal ve stl modellerine göre yapılan ayrıştırmanın kalanı ile herhangi bit işlem yapılmadan saf veriye auto.arima uygulanmıştır.

```
In [11]: auto.arima(na.omit(PlatinumPricesDecomposeM$random))
         auto.arima(na.omit(PlatinumPricesDecomposeA$random))
         auto.arima(na.omit(PlatinumPricesSTL$time.series[,3]))
         auto.arima(na.omit(PlatinumPrices[,2]))

Series: na.omit(PlatinumPricesDecomposeM$random)
ARIMA(1,0,3) with non-zero mean

Coefficients:
         ar1     ma1      ma2      ma3    mean
      0.9856  0.0591  -0.0209  -0.0290  0.9883
```

```
s.e.  0.0021  0.0124   0.0123   0.0118  0.0169

sigma^2 estimated as 0.0004002:  log likelihood=16807.26
AIC=-33602.51   AICc=-33602.5   BIC=-33561.61


Series: na.omit(PlatinumPricesDecomposeA$random)
ARIMA(2,0,2) with zero mean

Coefficients:
         ar1     ar2     ma1     ma2
      0.0955  0.8746  0.9468  0.0713
s.e.  0.0628  0.0620  0.0635  0.0125

sigma^2 estimated as 75.09:  log likelihood=-24124.78
AIC=48259.57   AICc=48259.57   BIC=48293.65


Series: na.omit(PlatinumPricesSTL$time.series[, 3])
ARIMA(2,0,2) with zero mean

Coefficients:
         ar1     ar2     ma1     ma2
      0.0997  0.8674  0.9383  0.0711
s.e.  0.0585  0.0576  0.0593  0.0122

sigma^2 estimated as 76.8:  log likelihood=-25507.05
AIC=51024.11   AICc=51024.12   BIC=51058.45


Series: na.omit(PlatinumPrices[, 2])
ARIMA(0,1,3)

Coefficients:
         ma1     ma2      ma3
      0.0458  0.0121  -0.0462
s.e.  0.0118  0.0117   0.0114

sigma^2 estimated as 81.76:  log likelihood=-25724.6
AIC=51457.2   AICc=51457.2   BIC=51484.67
```

Veri içerisindeki trend, mevsimsel etki ve üssel dağılım konseptleri çıkarıldıktan sonra elde kalan verinin beyaz gürültü olması beklenmektedir. Decompose işlemi ile hem trend hem de mevsimsellik etki veriden uzaklaştırılmıştır. Arima modelleri ise veriden üssel etkilerin çıkarılmasını sağlar.

Uyguladığımız arima modellerinden AIC,BIC değeri mutlak değerce en düşük olan multiplicative veriye uygulanan p=1, q=0,r=3 modeli seçilmiştir. Veri seti 1 gün önceye bağlı olarak ve 3 günün ortalamasını alarak ilerlemektedir. Herhangi bir fark alınmadığı için durağan bir veri

olduğu söylenebilir. Verideki durağanlık daha önceden yapılmış olan ayrıştırma işlemleriyle uzaklaştırılmıştır.

ARIMA modeli kullanılarak üstel etkilerin çıkarılması sonucu kalan seriye aşağıdaki şekilde ulaşılabilir. Ayrıca bu kalan değerin beyaz gürültü fonksiyonu olması beklenmektedir. Eğer kalan veri beyaz gürültü ise, veriye doğru model uygulanarak, gerekli anomalilerin çıkarıldığı sonucuna varılmaktadır.

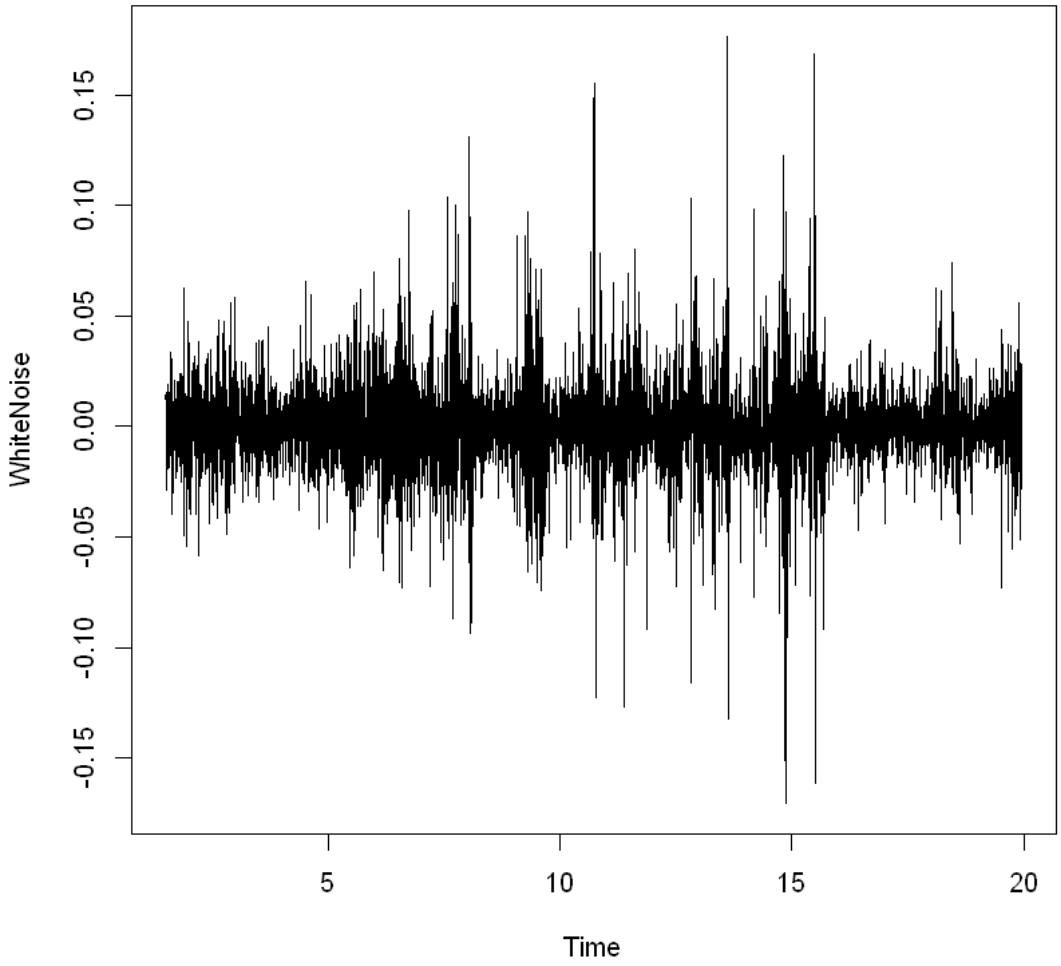
```
In [12]: PlatinumPricesArima<-arima(na.omit(PlatinumPricesDecomposeM$random),
                                    order=c(1,0,3))
         WhiteNoise<-PlatinumPricesArima$residuals
         plot(WhiteNoise)
         summary(PlatinumPricesArima)
```

```
Call:
arima(x = na.omit(PlatinumPricesDecomposeM$random), order = c(1, 0, 3))

Coefficients:
         ar1     ma1      ma2      ma3  intercept
      0.9856  0.0591  -0.0209  -0.0290     0.9883
s.e.  0.0021  0.0124   0.0123   0.0118     0.0169

sigma^2 estimated as 0.0003999:  log likelihood = 16807.26,  aic = -33602.51

Training set error measures:
                       ME       RMSE        MAE         MPE    MAPE      MASE
Training set -2.650106e-05 0.01999801 0.01345905 -0.04244688 1.35439 0.9949594
                     ACF1
Training set 2.833433e-05
```



Kalan verinin grafiğindeki bantın genişliği riski ifade etmektedir. Risk neredeyse tüm zamanlarda aynıdır. Yukarıdaki arima modelin özet çıktısı doğru modelin uygulandığını hata ölçüm parametrelerinden göstermektedir. RMSE, MAE,MPE, MAPE gibi ölçüm değerleri uygulanan model için düşük çıkmıştır. Ayrıca özet fonksiyonunu çıktısındaki ACF değeri de modelin doğru olduğunu göstermektedir. Tüm bunlar da uygulanan arima(1,0,3) modelinin aslında doğru model olduğunu ve modelden çıkarılan değerlerin doğru sonuçlar verdiğini göstermektedir. Aslında arima(1,0,3) modeli herhangi bir fark alınmadığı için arma(1,3) modelinin benzeridir.

Arima modeli uygulandıktan sonra kalanından elde edilmesi beklenen beyaz gürültü fonksiyonu olup olmadığı, normallik testleriyle aşağıdaki R kodu ile yapılabilir;

```
In [13]: SampleWN<-sample(WhiteNoise,5000)
         shapiro.test(SampleWN)
         ks.test(WhiteNoise, 'pnorm',mean=0,sd=sqrt(var(WhiteNoise)))
```

```
Shapiro-Wilk normality test

data:  SampleWN
W = 0.90633, p-value < 2.2e-16
```

```
One-sample Kolmogorov-Smirnov test

data:  WhiteNoise
D = 0.080609, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Çarpımsal modelden ARIMA(1,0,3) çıkarılmadan önce yapılan normallik testlerinin sonuçları aşağıda verilmiştir.

```
In [14]: shapiro.test(SamplePlatinumPricesDecomposeM)
         ks.test(PlatinumPricesDecomposeM$random,
                 'pnorm',
                 mean=0,
                 sd=sqrt(var(na.omit(PlatinumPricesDecomposeM$random))))
```

```
Shapiro-Wilk normality test

data:  SamplePlatinumPricesDecomposeM
W = 0.96362, p-value < 2.2e-16
```

```
One-sample Kolmogorov-Smirnov test

data:  PlatinumPricesDecomposeM$random
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Shapiro-Wilk normallik testine göre W değeri küçülmüştür. Bu da arima(1,0,3) modeli uygulandıktan sonra kalan verinin normallikten uzaklaştığını göstermektedir.Fakat modelin hala normal olduğu W test istatistiğinden görülmektedir.Kolmogorov-Smirnov testine göre zaman serisinin kalanının normal dağılıma daha çok yaklaştığı D test istatistiğinden görülmektedir. D değeri küçüldüğünden seri normalliğe daha da yaklaşmıştır. Kolmogorov testi tüm veri setini test ettiği için Shapiro-Wilk testine göre daha doğru sonuçlar vermektedir.

AIC ve BIC değerleri aynı ayrıştırma modeli ile ayrıştırılmış (örneğin ikisi de toplamsal model olarak) fakat farklı arima modelleri uygulanmış veriler üzerinde daha doğru bir seçim kriteri oluşturmaktadır. Bu yüzden yukarıdaki tüm arima modelleri için auto.arima fonksiyonun sonuçları değerlendirilmeli ve normallik testleri uygulanmalıdır. Eğer bu çıktılardan herhangi biri çarpımsal modele uyguladığımız ARIMA(1,0,3) modelininden elde edilen sonuçlarından daha iyi sonuçlara sahip ise o modelin daha uygun olduğu söylenebilir. Daha sonra hangisinin beyaz gürültüye daha yakın olduğu saptanmalıdır. Aşağıda bu işlemler gösterilmiştir.

Toplamsal Model;

```
In [16]: PlatinumPricesArimaToplamsal<-arima(na.omit(PlatinumPricesDecomposeA$random),
                                              order=c(2,0,2))
         summary(PlatinumPricesArimaToplamsal)
         WhiteNoise2<-PlatinumPricesArimaToplamsal$residuals
         SampleWN2<-sample(WhiteNoise2,5000,replace = TRUE)
         shapiro.test(SampleWN2)
         ks.test(WhiteNoise2, 'pnorm',mean=0,sd=sqrt(var(WhiteNoise2)))
```

```
Call:
arima(x = na.omit(PlatinumPricesDecomposeA$random), order = c(2, 0, 2))

Coefficients:
         ar1     ar2     ma1     ma2  intercept
      0.0006  0.9681  1.0407  0.0634    -1.2111
s.e.  0.0150  0.0148  0.0191  0.0124     7.0326

sigma^2 estimated as 75.07:  log likelihood = -24125.62,  aic = 48263.23

Training set error measures:
                     ME     RMSE      MAE     MPE     MAPE      MASE
Training set -0.0100795 8.664197 5.563736 43.3694 103.5062 0.9942941
                     ACF1
Training set 0.0004094616



Shapiro-Wilk normality test

data:  SampleWN2
W = 0.89915, p-value < 2.2e-16




One-sample Kolmogorov-Smirnov test

data:  WhiteNoise2
D = 0.1105, p-value < 2.2e-16
alternative hypothesis: two-sided
```

STL modeli;

```
In [17]: PlatinumPricesArimaSTL<-arima(na.omit(PlatinumPricesSTL$time.series[,3]),
                                       order=c(2,0,2))
         summary(PlatinumPricesArimaSTL)
         WhiteNoise3<-PlatinumPricesArimaSTL$residuals
         SampleWN3<-sample(WhiteNoise3,5000)
         shapiro.test(SampleWN3)
         ks.test(WhiteNoise3, 'pnorm',mean=0,sd=sqrt(var(WhiteNoise3)))
```

```
Call:
arima(x = na.omit(PlatinumPricesSTL$time.series[, 3]), order = c(2, 0, 2))

Coefficients:
         ar1     ar2     ma1     ma2  intercept
      0.0997  0.8674  0.9382  0.0711    -0.8504
s.e.  0.0583  0.0575  0.0591  0.0122     6.2980

sigma^2 estimated as 76.76:  log likelihood = -25507.03,  aic = 51026.07

Training set error measures:
                     ME   RMSE      MAE      MPE     MAPE      MASE
Training set 0.01563383 8.7611 5.587638 -11.8678 87.51329 0.9948034
                      ACF1
Training set -0.0009323478



Shapiro-Wilk normality test

data:  SampleWN3
W = 0.8832, p-value < 2.2e-16




One-sample Kolmogorov-Smirnov test

data:  WhiteNoise3
D = 0.11626, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Saf veri seti;

```
In [18]: PlatinumPricesArimaPure<-arima(na.omit(PlatinumPrices[,2]),
                                        order=c(0,1,3))
         summary(PlatinumPricesArimaPure)
         WhiteNoise4<-PlatinumPricesArimaPure$residuals
         SampleWN4<-sample(WhiteNoise4,5000)
         shapiro.test(SampleWN4)
         ks.test(WhiteNoise4, 'pnorm',mean=0,sd=sqrt(var(WhiteNoise4)))
```

```
Call:
arima(x = na.omit(PlatinumPrices[, 2]), order = c(0, 1, 3))

Coefficients:
         ma1     ma2      ma3
      0.0458  0.0121  -0.0462
s.e.  0.0118  0.0117   0.0114

sigma^2 estimated as 81.73:  log likelihood = -25724.6,  aic = 51457.2

Training set error measures:
                     ME     RMSE      MAE         MPE     MAPE      MASE
Training set -0.1186566 9.039685 5.439082 -0.04801586 1.339236 0.9996372
                     ACF1
Training set -0.0007836219



Shapiro-Wilk normality test

data:  SampleWN4
W = 0.86965, p-value < 2.2e-16




One-sample Kolmogorov-Smirnov test

data:  WhiteNoise4
D = 0.13969, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Ayrıştırmalara uygulanan normallik testlerine göre bu çıktılardan herhangi biri çarpımsal modele uyguladığımız ARIMA(1,0,3) modelininden elde edilen sonuçlarından daha iyi sonuca sahip değildir. Ayrıca Summary fonksiyonunun çıktısındaki hata ölçüm değerleri ile AIC değeri modellerin çarpımsal modele uygulanan Arima(1,0,3) modelinin daha doğru sonuçlar verdiğini göstermektedir.

Sonuç Yukarıda yapılan testlere göre en iyi sonucun AIC ve BIC değerine göre en iyi sonucu veren çarpımsal modele uygulanan arima(1,0,3) modeliyle yapılan ayrıştırmaya göre olduğu görülmektedir. Veri setindeki tüm anomaliler çıkarıldıktan sonra kalan verinin beyaz gürültü olması beklenir. Beyaz gürültü fonksiyonu normal dağılıma sahiptir. Kalan verinin gürültü datası olup olmadığı test edilmiş ve gürültü datası olduğu sonucuna varılmıştır. Bu da veriye uygulanan ayrıştırma modellerinin doğru bir şekilde uygulandığını göstermektedir.

2 Tahminleme

Zaman serileri analizi ile temel hedeflerden biri, verilerin gelecekteki durumunu gösteren tahminler üretmektir.Bu tahminleme extrapolasyonla yapılır. Ekstrapolasyon her zaman doğru sonuçlar vermeyebilir ve yanlış sonuçlara yol açabilir. Tahminleme de aynı şekilde doğru yapılamayabilir. Sinyal içeren seriler gürültüye göre daha güçlü olduğu seriler oldukları için tahminlemenin doğru yapılması mümkündür. Bununla birlikte, gürültülü seriler için, tahminlerde büyük bir belirsizlik vardır ve bunlar çok kısa bir aralık için en güvenilirdir. Yukarıdakilerden yola çıkarak, belirsizliklerin ana kaynağının sürecin içindeki yenilikler olduğu söylenebilir. Doğru tahminleme yapabilmek için veri üretme sürecinin zaman içinde değişmediğinden, yani geçmişte gözlemlendiği gibi gelecekte devam edeceğinden emin olunmalıdır. Doğru model uygulansa dahi parametreler arasındaki ek belirsizlikler doğru tahminlemeyi etkilemektedir.

Zaman serisi tahminlemelerine ilk olarak, durağan süreçleri tahmin etmek için kullanılan AR, MA ve ARMA süreçleriörnek olarak gösterilebilir. Trend içeren durağan olmayan seriler için ise ARIMA modeli ile tahminleme yapılabilir.

2.1 ARIMA ile Tahminleme

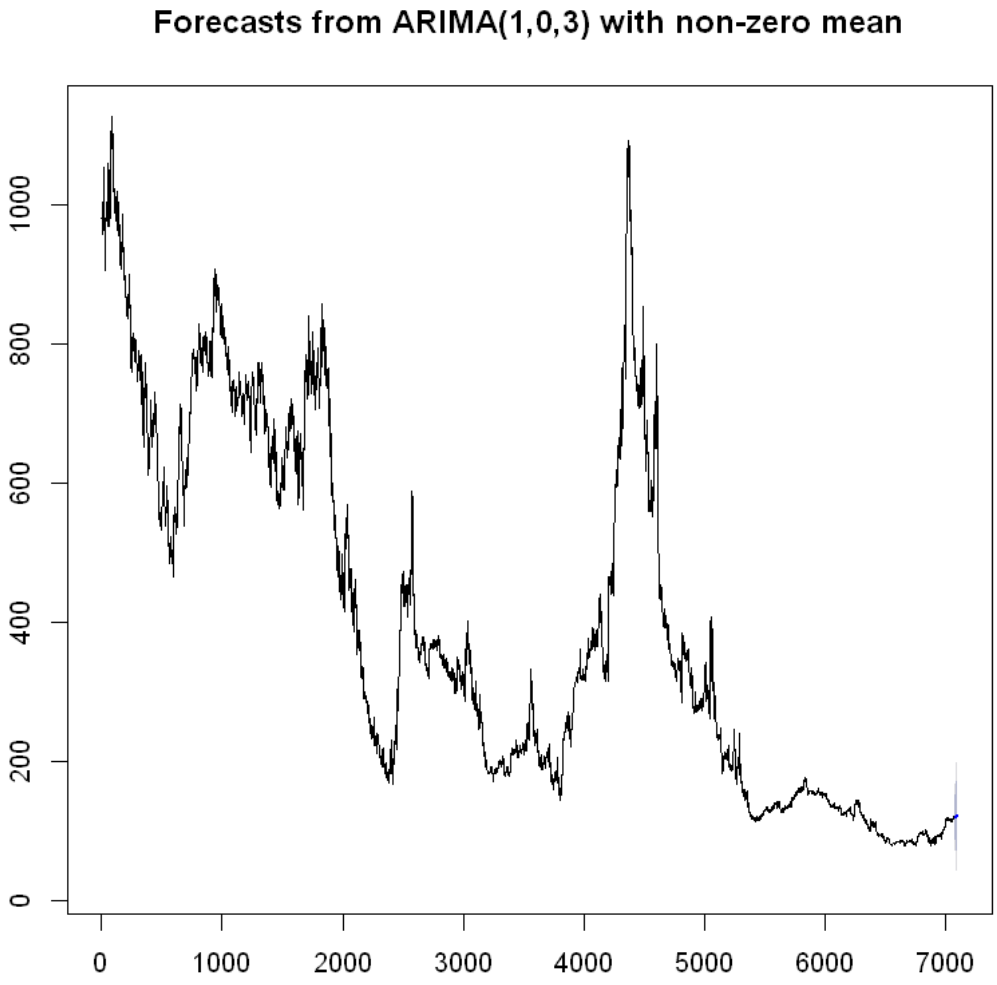
Kalan R_t yani Beyaz Gürültü gibi göründüğü bir zaman dizisindeki tahminleme ARIMA modeli ile yapılabilir. Bu koşullar altında, tahminler kolayca hesaplanabilir

Uygun ARIMA modelinin tanımlanan parametrelerin başarılı bir şekilde tahmin edildiği ve artıkların gerekli özellikleri sergilediği, yani Beyaz Gürültü gibi göründüğü bir zaman dizisi verildiğini varsayıyoruz. Bu koşullar altında, tahminler kolayca hesaplanabilir.

Bu, istediğimiz bir $ARIMA(p,q,r)$ modelinden herhangi bir tahmin üretmemizi sağlar. Bu prosedür aynı zamanda Box-Jenkins prosedürü olarak da bilinmektedir. R'da kullanılan, arima ()'dan sonra uygulanan forecast() fonksiyonunda Box-Jenkins şeması arka planda çalışmaktadır.

Verilerinin modele uyarlanması ve 30 adımlık tahminlerin üretilmesi için R komutları aşağıdaki gibidir. Arima() orderleri belli modeli oluşturur ve forecast() fonksiyonu belirlenen modele göre verilen parametre kadar tahmin üretir. Oluşturulan tahminlemenin görselleştirilmesi için aşağıdaki R kodundan yararlanılabilir.İlk aşamada yaptığımız durağan olmayan zaman serileri için ayrıştırma aşamalarından yararlanılarak trend, mevsimsellik içermeyen seriler için tahmin üretilebilecektir. Yani çarpımsal ayrıştırmanın arima(1,0,3) modeline göre tahminlemesi yapılacaktır. Son 30 veriye kadar model uygulanacak ve daha sonra 30 adet tahmin yapılıp gerçek değerlerle tahminler karşılaştırılacatır. [4]

```
In [19]: m<-length(na.omit(PlatinumPricests))
         PlatinumPricesArimaForecast<-arima(na.omit(PlatinumPricests)[1:(m-29)],
                                            order=c(1,0,3))
         tahmin <- forecast(PlatinumPricesArimaForecast,30)
         plot(tahmin)
```



Aşağıda gerçek veriler ve arima modeliyle yapılan tahmin ile yapılan veriler için ortalama hata hesaplanmıştır. Bu değer tahminlemelerin karşılaştırılması için kullanılır.

Keşif analizi yaparken bulduğumuz uygun arima modeli olan (1,0,3) modeli uygulanarak yukarıdaki tahminleme yapılmıştır.

2.2 Holt-Winters Yöntemi ile Tahminleme

Eğilim ve mevsimsellik sergileyen seriler için yapılan bir tahminleme yöntemidir. Holt-Winters mevsimsel yöntemi, tahmin denklemini ve üç yumuşatma denklemini içerir. Holt-Winters yöntemi toplamsal ve çarpımsal olarak ikiye ayrılır. Toplamsal yöntemle, mevsimsel bileşen, gözlemlenen serilerin ölçeğinde mutlak terimlerle ifade edilir ve seviye denkleminde, mevsimsel bileşenin çıkarılmasıyla seri mevsimsel olarak ayarlanır. Çarpımsal yöntemle, mevsimsel bileşen göreceli olarak ifade edilir. Seri mevsimsel bileşen tarafından bölünerek mevsimsel olarak ayarlanır.

2.2.1 Toplamsal Holt-Winters Yöntemi

Toplamsal model için kullanılan form aşağıdaki gibidir.

$$a_t = \alpha(x_t - st - p) - (1-\alpha)(a_{t-1} + b_t t - 1)$$
$$b_t = \beta(a_t - a_{t-1}) - (1-\beta)(b_{t-1})$$
$$s_t = \gamma(x_t - a_t) - (1-\gamma)(st - p)$$

Buradaki a_t,seviye, b_t eğilim, s_t mevsimsel etkidir. Seviye, eğim ve mevsimi hedefleyen üç düzeltme parametresi α,β,γ'dır.

İlk güncelleme denklemi, gözlemlediğimiz uygun mevsimsel etkinin mevcut tahminiyle, son gözlemimizin ağırlıklı ortalamasını alır ve seviyenin bir önceki adımdaki tahminini ifade eder.

İkinci güncelleme denklemi, mevcut seviye ve bir önceki seviye arasındaki farkın t-1 zamanında tahmini eğim ile ağırlıklı ortalamasını alır. Bu yalnızca mevcutsa hesaplanabilir.

Son olarak, mevsimsel terim için, aynı birim için mevsimsel terimin önceki tahmini ile gözlem ve seviye arasındaki farkın ağırlıklı ortalaması alınarak, t-p zamanında yapılan bir başka tahmin daha elde edilir.

2.2.2 Çarpımsal Holt-Winters Yöntemi

Çarpımsal model için kullanılan form aşağıdaki gibidir.

$$a_n = \alpha\frac{x_n}{s_{n-p}} - (1-\alpha)(a_{n-1} + b_{n-1})$$
$$b_n = \beta(a_n - a_{n-1}) - (1-\beta)(b_{n-1})$$
$$s_t = \gamma\frac{x_n}{a_n} - (1-\gamma)(sn - p)$$

Buradaki ilk denklem seviye denklemi, ikinci denklem eğilim ve son denklem mevsimsellik bileşenlerinin hesaplanmasında kullanılır. α,β,γ, üç düzeltme parametresidir.

R fonksiyonu HoltWinters () uygulandığında, başlangıç değerleri decompose () prosedüründen elde edilir ve uygun seviye, eğilim ve mevsimsellik için mevcut tahminleri içerir. Daha önceden yaptığımız analiz sonucuna göre veri setinin çarpımsal modelle daha iyi sonuçlar verdiği gözlenmişti. Bu yüzden Holt-Winters tahminlemesinde çarpımsal model kullanılmıştır.

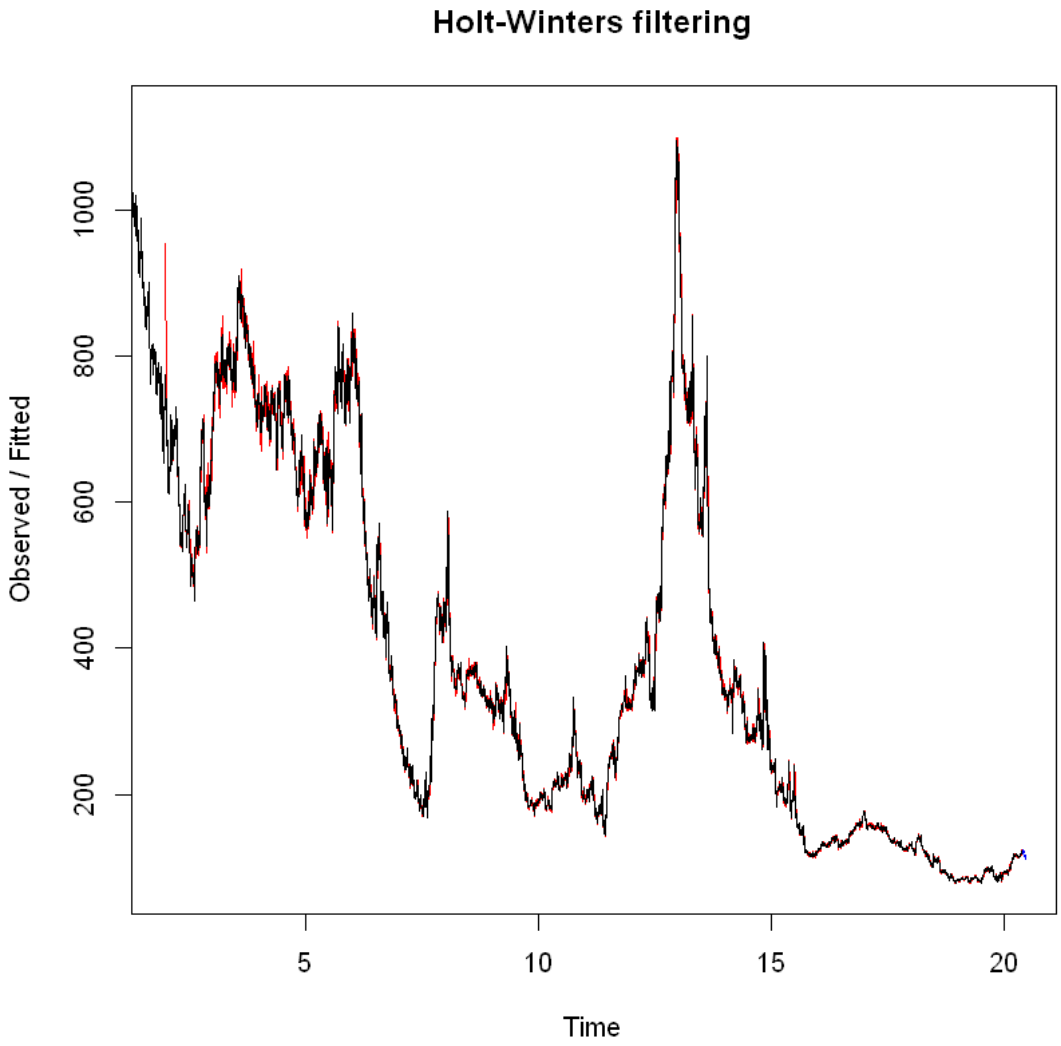
Mevsimsellik ve trend içeren bir zaman serisinin tahmininde kullanılacak Holt Winters parametreleri aşağıdaki şekilde elde edilebilir.[1]

```
In [20]: PPHW <- HoltWinters(ts(na.omit(PlatinumPricests)[1:(m-29)],
                               freq=365),
                               seasonal = "mult")
         PPHW$alpha
         PPHW$gamma
         PPHW$beta
```

alpha: 0.892454361002961
gamma: 1
beta: 0.00123075437713089

Holt-Winters yönteminin alfa,gama ve beta sayıları yukarıda gösterilmiştir. Katsayı değerleri (n zamanında), yukarıda verilen formülle bu serilerden tahmin yapmak için kullanılanlardır. 30 basamaklı bir tahmin tahmin üretmek için aşağıdaki R komutundan yararlanılabilir.

```
In [21]: plot(PPHW)
         tahminHW<-predict(PPHW, n.ahead=30)
         lines(tahminHW, col="blue", lty=4)
```



2.2.3 Uygun Tahminleme Yönteminin Seçilmesi

Yukarıdaki ortalama hata karelerine göre arima ile yapılan tahminlemenin çok daha yakın sonuçlar verdiği gözlenmiştir. Ayrıca accuracy() fonksiyonu da tahminin doğru yapılıp yapılmadığını test etmektedir. Bunun için öncelikle saf veri setinin son 30 girdisini window() fonksiyonu ile oluşturmamız gerekmektedir. Daha sonra Accuracy() fonksiyonu ile tahmin ve gerçek veriler karşılaştırılır.[5]

```
In [22]: PP1Last30 <- window(na.omit(PlatinumPricests)[(m-29):m])
         tahminHW <- window(tahminHW)
```

```
print("Accuracy Arima")
accuracy(tahmin$mean, PP1Last30)
print("Accuracy HW")
accuracy(tahminHW, PP1Last30)
```

```
[1] "Accuracy Arima"
```

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	3.046062	4.748621	4.03789	2.348815	3.186527	0.9369171	4.110629

```
[1] "Accuracy HW"
```

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	6.746535	10.71415	8.669586	5.200832	6.820946	0.9630263	9.242116

y_i, i. gözlemi, $\hat{y}_i$, i. gözlem için tahmini temsil etmek üzere hata terimi aşağıdaki şekilde ifade edilir.

$$e_i = y_i - \hat{y}_i$$

Ölçeğe bağımlı hatalar Bu hata hesaplamalarında serilerin benzer ölçekte olmaları gerekmektedir.Doğrudan e_t'ye bağlı hata ölçümü olduğundan farklı ölçeklerdeki serilerin karşılaştırılmasında kullanılamaz. En yaygın kullanılan ölçüm parametrelerinden bazıları mutlak hatalara veya karesel hatalara dayanmaktadır.

Ortalama Mutlak Hata:

$$MAE = mean(|e_i|)$$

Ortalama Kareler Hata:

$$RMSE = \sqrt{mean(e_i)^2}$$

Tahmin yöntemlerini tek bir veri kümesinde karşılaştırırken, MAE, anlaşılması ve hesaplanması kolay olduğu için daha çok tercih edilmektedir.

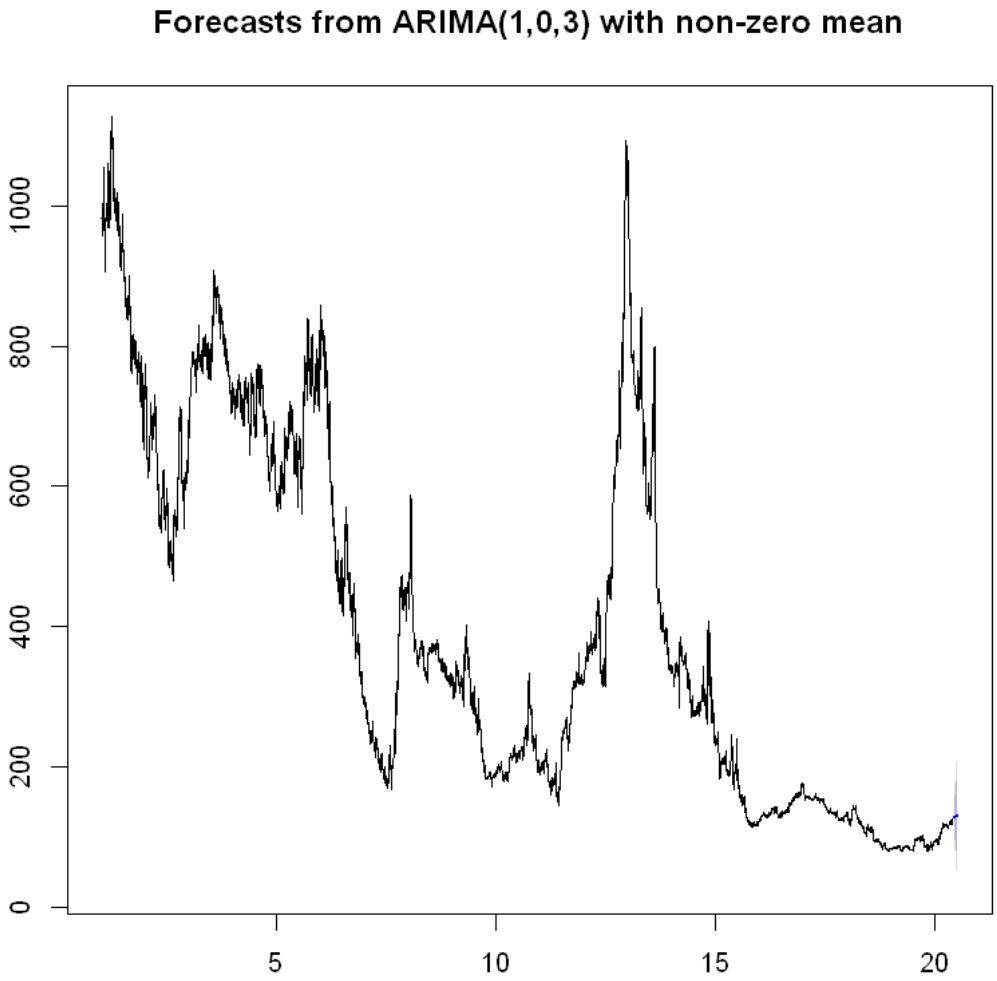
Yüzdeye bağımlı hatalar Yüzde hatası $p_t = 100\frac{e_t}{y_t}$ olarak ifade edilir. Yüzde hataları serilerin ölçeklerinden bağımsız olduklarından ölçek farklı seriler arasındaki hata ölçümü olarak kullanılmaktadır. En sık kullanılan mutlak yüzde hatası MAPE aşağıdaki şekilde ifade edilmektedir.

Ortalama Mutlak Yüzde Hatası:

$$MAPE = mean(|p_t|)$$

Accuracy() foksiyonlarının çıktısına göre Arima ile yapılan tahminlemenin MAE,RMSE ve MAPE gibi hata çıktılarına göre daha yakın sonuçlar verdiği gözlenmiştir.Buna göre Arima modeli ile tüm verilerin kullanılarak sonraki 30 gün için tahminleme yapılmıştır.

```
In [23]: PlatinumPricesArimaForecast<-arima(na.omit(PlatinumPricests),
                                            order=c(1,0,3))
         tahmin <- forecast(PlatinumPricesArimaForecast,30)
         plot(tahmin)
```



Uygun yöntem ile 30 günlük tahminleme yapıldıktan sonra başka bir analiz olan çok değişkenli zaman serilerinin analizi incelenecektir.

3 Çok Değişkenli Zaman Serileri Analizi

Veriler genellikle birden fazla değişken üzerinde toplanır. Örneğin, ekonomide, günlük döviz kurları çok çeşitli para birimleri için kullanılabilir, ya da hidrolojik çalışmalarda, hem yağış hem de nehir akış ölçümleri ilgili bir alanda alınabilir. Zamanla ölçülen değişkenler genellikle benzer özellikler sergilediğinden, değişkenleri ilişkilendirmek için regresyon kullanılabilir. Bununla birlikte, zaman serileri değişkenlerinin regresyon modelleri yanıltıcı olabilmektedir. Bu durum sahte regresyon olarak adlandırılır.Zaman serileri değişkenleri için, nedensel ilişkiyi ortaya çıkarmadan önce dikkatli olunmalıdır, çünkü zaman serisinin içerdiği anomaliler nedeniyle belirgin bir ilişki görülebilir. Örneğin artan nüfus ile artan birbiriyle ilgili olmayan iki malın miktarı arasında ilişki

gözlenebilir.

3.1 Dağıtılmış-Gecikmeli Model

Dağıtılmış-gecikmeli bir model, X'in y üzerindeki bir seferlik etkisi değil zaman içinde meydana getirdiği etkinin dinamik modelidir.

$$y_t = \alpha + \sum_{s=0}^{\infty} \beta_s x_{t-s} + u_t$$

u_t, durağan bir hata terimidir. Bireysel katsayılar β_s, gecikme ağırlıkları olarak adlandırılır ve toplu olarak gecikme dağılımını verir. X'in zamanla y'yi nasıl etkilediğini tanımlar. Denklemdeki sonsuz sayıda β katsayılarını tahmin edilmesi zordur. Pratik bir yöntem, gecikle dağılımı etkin bir şekilde sıfır olduğunda uygun olan gecikmeyi sonlu uzunluğa (q) kesmektir.Başka bir yaklaşım, denklemde gecikme dağılımının kademeli olarak sıfıra düşmesine izin veren fonksiyonel bir formun kullanılmasıdır.

3.2 Kovaryans

Kovaryans, iki değişken arasında herhangi bir ilişki olup olmadığını gösterir. Pozitif bir kovaryans değişkenler arasında pozitif bir doğrusal ilişki olduğunu ve negatif kovaryans ise negatif bir ilişkinin varlığını gösterir.Kovaryans aşağıdaki formül ile hesaplanır.

$$Cov_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

3.3 Korelasyon

Bir korelasyon, iki rastgele değişken arasındaki ilişkinin yönlü ölçüsüdür. Korelasyon iki değişken arasında tamamen simetriktir. Korelasyon analizi ile bağımsız değişken değiştiğinde,bağımlı değişkenin nasıl değişeceği saptanır. Korelasyon katsayısı -1 ile +1 arasında değerler alır. Korelasyon aşağıdaki formül ile ifade edilir. Buradaki s_x, s_y, x ve y'nin standart sapmasıdır.

$$Cor_{xy} = \frac{Cov_{xy}}{s_x s_y}$$

Çok değişkenli zaman serisi modellerinden çapraz korelasyon fonksiyonu ile açıklanacaktır.

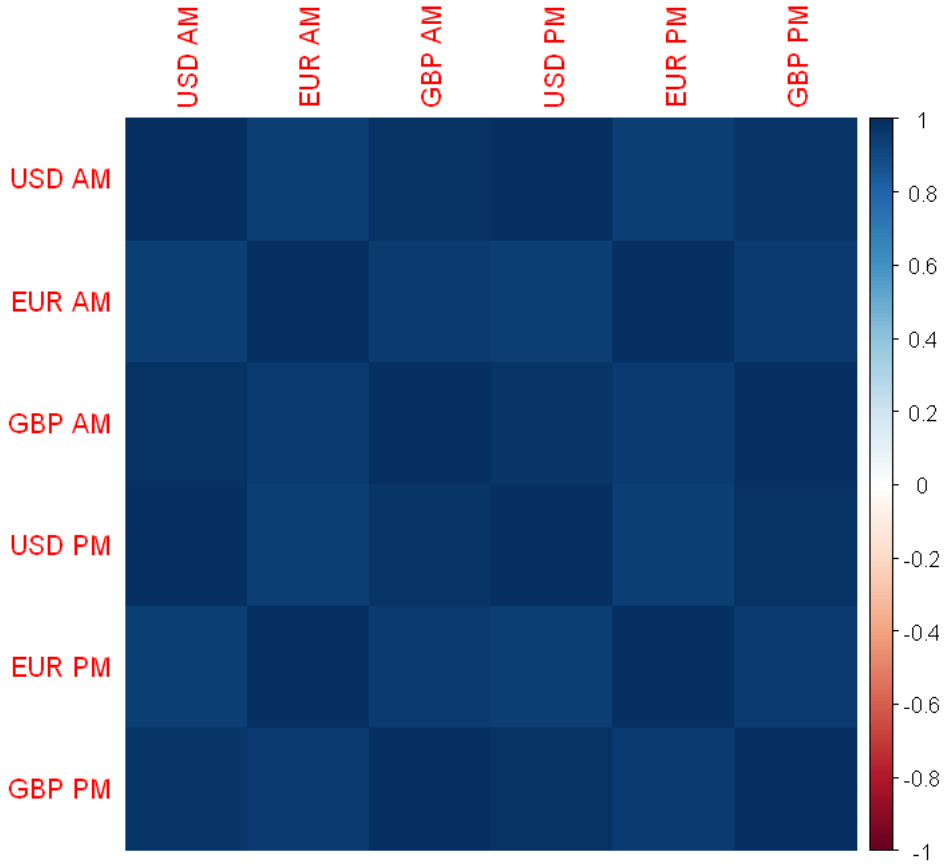
3.3.1 Çapraz Korelasyon

Çok değikenli zaman serilerinde ele alacağımız temel amaç, iki zaman dizisi arasındaki ilişkinin açıklaması ve modellenmesidir.İki zaman dizisi (y_t ve x_t) arasındaki ilişkide, y_t serisi, x_t serisinin geçmiş gecikmeleriyle ilişkili olabilir. Çapraz korelasyon fonksiyonu (CCF),y_t'nin yararlı belirleyicileri olabilecek x_t değişkeninin gecikme sürelerinin belirlenmesinde yardımcı olacaktır. R'de bulunan çapraz korelasyon fonksiyonu(CCF), x_{t+h} ve y_t arasında h = 0, $\pm$ 1, $\pm$ 2, $\pm$ 3 ve benzeri için örnek korelasyonları kümesi olarak tanımlanır. h için negatif bir değer, t'den önceki bir zamanda x değişkeni ve t zamanında y değişkeni arasında bir korelasyondur. Örneğin, h = -2'yi düşündüğümüzde CCF değeri, x_{t-2} ve y_t arasındaki korelasyonu vermektedir.

Öncelikle olarak Platin fiyatlarının genel olarak birbirleriyle aralarındaki ilişki incelenecektir. Korelasyon rastgele iki değişken arasındaki ilişkinin gücünü ve yönünü belirtmek için kullanılır. Platinum fiyatları çoklu değişkene sahip olduğu için bu ilişkiyi bir matris ile tanımlayacağız. Korelasyon matrisi adı verilecek olan bu matris, çoklu değişkenler arasındaki korelasyon katsayılarını gösterir. Bu matris yardımı ile iki değişken arasında korelasyon kolaylıkla görülebilir. Daha sonra elde edilen bu matris R'da corrplot fonksiyonu ile görselleştirilebilir.[3]

```
In [24]: cormat<-cor(na.omit(PlatinumPrices[,c(2:7)]))
         cormat
         corrplot(cormat, method = "color")
```

	USD AM	EUR AM	GBP AM	USD PM	EUR PM	GBP PM
USD AM	1.0000000	0.9404636	0.9800171	0.9997624	0.9404015	0.9796666
EUR AM	0.9404636	1.0000000	0.9559812	0.9401127	0.9997532	0.9555518
GBP AM	0.9800171	0.9559812	1.0000000	0.9799510	0.9560487	0.9997825
USD PM	0.9997624	0.9401127	0.9799510	1.0000000	0.9404793	0.9800054
EUR PM	0.9404015	0.9997532	0.9560487	0.9404793	1.0000000	0.9560112
GBP PM	0.9796666	0.9555518	0.9997825	0.9800054	0.9560112	1.0000000



Tabloda görüldüğü üzere aynı para birimi cinsinden açılış fiyatları ile kapanış fiyatları ilişkilidir.Ayrıca tabloda görüldüğü üzere herhangi bir para birimi cinsinden açılış fiyatları başka bir açılış fiyatıyla, kapanış fiyatına göre daha çok ilişkilidir. Grafikte ise tablonun görselleştirilmiş hali görülmektedir. Mavi tonları koyulaştıkça değişkenler arasındaki ilişki pozitif yönde artış göstermektedir.

Genel olarak para birimleri arasındaki ilişki incelendikten sonra dolar cinsinden açılış ve kapanış fiyatlarının arasındaki ilişki saptanacaktır. Aşağıda öncelikle aralarındaki korelasyonu inceleyeceğimiz dolar cinsinden cuma günü kapanış fiyatı ile pazartesi günü açılış fiyatlarının verimizden ayrıştırılması gerekmektedir. Bunun için aşağıdaki R kodları kullanılmaktadır.

```
In [25]: for (i in 1:7){
         PlatinumPrices[,i]<-rev(PlatinumPrices[,i])
             }
         PlatinumPricesDolarAM<-PlatinumPrices[,2]
```

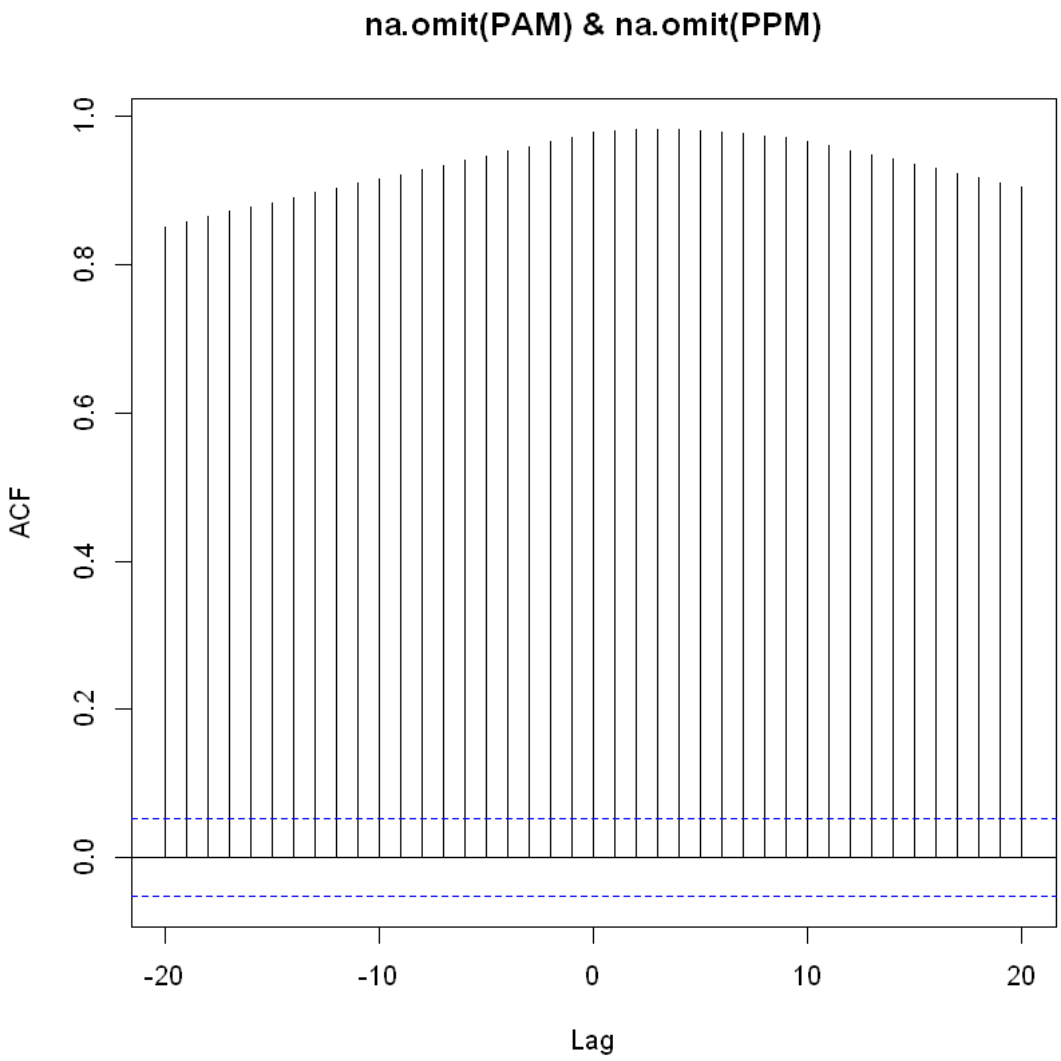
```
PlatinumPricesDolarPM<-PlatinumPrices[,5]
Monday<-c(TRUE,FALSE,FALSE,FALSE,FALSE)
Friday<-c(FALSE,FALSE,FALSE,FALSE,TRUE)
n<-length(PlatinumPricesDolarPM)
P<-rep(Monday,n/5)
C<-rep(Friday,n/5)
PAM<-subset(PlatinumPricesDolarAM,P[TRUE])
PPM<-subset(PlatinumPricesDolarPM,C[TRUE])
```

Aralarındaki korelasyona bakacağımız cuma kapanış fiyatı ile pazartesi açılış fiyatları çarpraz korelasyon R'daki ccf() fonksiyonu yardımı ile bulunabilir.

```
In [26]: ccfvalues <- ccf (na.omit(PAM), na.omit(PPM),lag=20)
         ccfvalues
```

```
Autocorrelations of series 'X', by lag

  -20   -19   -18   -17   -16   -15   -14   -13   -12   -11   -10    -9    -8
0.851 0.858 0.864 0.871 0.878 0.884 0.890 0.897 0.903 0.909 0.915 0.921 0.927
   -7    -6    -5    -4    -3    -2    -1     0     1     2     3     4     5
0.933 0.940 0.946 0.953 0.959 0.966 0.972 0.978 0.980 0.981 0.982 0.982 0.980
    6     7     8     9    10    11    12    13    14    15    16    17    18
0.978 0.976 0.974 0.970 0.965 0.960 0.954 0.947 0.942 0.936 0.930 0.923 0.917
   19    20
0.910 0.904
```



Yukarıdaki grafikten anlaşılacağı üzere en yüksek çapraz korelasyon katsayıları -5 ile 0 arasında meydana gelmektedir. Gecikmeleri tam olarak grafikten okumak zordur, bu yüzden ccf'nin listelenmesi daha doğru sonuçlar verecektir. Yukarıda yapılan listelemeye göre cuma kapanış fiyatı üç,dört gün sonraki pazartesi açılış fiyatından etkilenmektedir.

4 Değerlendirme

Bir zaman serisi, değişkenlerin zaman içinde belirli sabit bir aralıkta ölçüldüğünde, elde edilen veriler ile oluşturmaktadır. Zaman serilerindeki analizin amacı; geçmiş verilerdeki anomalileri saptamak, birbirleriyle ilişkilerini gözlemlemek ve gelecek için tahminleme yapabilmektir. Anomalilerin çıkarılmasının nedeni verideki anormalilerin çok çeşitli uygulama alanlarında önemli ve eyleme geçirilebilir bilgilere dönüşebilir olmasıdır.

Analize başlamadan önce ilk olarak veri seti uygun şekilde yüklenmiş ve zaman serisine dönüştürülmüştür. Daha sonra verinin özeti oluşturulmuş ve grafiği çizilerek verinin anlamlandırılması sağlanmıştır.

Zaman serisi analizindeki ilk aşamada verinin içindeki anomaliler saptanmış ve çıkartılmıştır. Bunun için bir dizi yöntemler uygulanmıştır. Öncelikle bu aşamada mevsimsellik ve trend gibi bileşenlerin çıkarılmıştır. Mevsimsellik ve trend bileşenlerinin veriden uzaklaştırmak için ayrıştırma işlemleri yapılmıştır. Bu ayrıştırma yöntemleri; toplamsal ve çarpımsal yöntem ile STL ile ayrıştırma yöntemidir. Bu üç ayrıştırma yöntemi veriye uygulandıktan sonra kalanın gürültü verisine uygun olması beklenir. Gürültü, sıfır ortalamalı gaussian (normal) dağılıma sahip rastgele değişkendir. Fakat bu yöntemlerden kalan verinin gürültüden farklı olduğu istatiksel yöntemlerle gösterilmiştir. Bu da veri setinin içinde hala bazı sinyallerinin olduğunu göstermektedir.Veri seti içindeki tüm sinyallerin çıkarılması için Box-Jenkins yöntemleri uygulanmıştır.

Ayrıştırma yöntemleriyle verinin içindeki üssel dağılıma sahip bazı sinyaller çıkarılmıştır. Fakat arima modeli uygulanmadan önce verinin durağanlığı istatiksel yöntemlerle test edilmiştir. Bu istatiksel yöntemler sonucunda verinin durağan olduğu sonucuna varılmıştır. Daha sonra uygun arima modelinin AIC,BIC değerlerinin küçük olan ARIMA(1,0,3) modeli olduğu saptanmıştır. ARIMA(1,0,3) modelinden kalan veriye normallik testleri uygulanmıştır ve sinyaller çıkarıldıktan sonra verinin gürültü olduğu istatiksel yöntemlerle bulunmuştur.

Analizin ikinci aşamasında tahminleme işlemi yapılmıştır. Tahminleme için arima ve Holt-Winters yöntemleri kullanılmıştır. Yapılan bu iki tahminlemenin istatiksel olarak ortalama kareler hatası ve ortalama mutlak hatası sonuçlarına göre iyi sonuçlar veren modeli olan ARIMA modeli seçilmiştir. Bu arima modeli kullanılarak önümüzdeki 30 gün için tahminleme yapılmıştır.

Analizdeki üçüncü aşamasında çok değişkenli zaman serileri için korelasyon ve kovaryans incelenmiştir. Veri setimizdeki platin fiyatlarının tüm para birimleri acısından açılış ve kapanış fiyatları arasındaki korelasyon değerleri bulunmuştur. Daha sonra cuma kapanış fiyatları ile pazartesi açılış fiyatları arasındaki korelasyon incelenmiş ve cuma kapanış fiyatı üç,dört gün sonraki pazartesi açılış fiyatından etkilendiği saptanmıştır.

5 Kaynakça

[1] Paul S.P. Cowpertwait · Andrew V. Metcalfe (2005). Introductory Time Series with R
[2] Peter J. Brockwell, Richard A. Davis (2002). Introduction to Time Series and Forecasting
[3] Dr. Marcel Dettling (2014). Applied Time Series Analysis
[4] Peter J. Brockwell, Richard A. Davis (2009). Time Series: Theory and Methods
[5] Christian Kleiber, Achim Zeileis (2008) Applied Econometrics with R
[6] A. Ian McLeod, Hao Yu, Esam Mahdi (2011) Time Series Analysis with R
[7] Robert H. Shumway, David S. Stoffer (2011) Time Series Analysis and Its Applications