

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**SAKARYA BÖLGESİ HANEHALKI ELEKTRİK
TÜKETİMİNİN DİNAMİK LINEER MODEL İLE
TAHMİNİ**

YÜKSEK LİSANS TEZİ

Emre DORUK

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ**
Tez Danışmanı : Dr. Öğr. Üyesi Tuğrul TAŞCI

Haziran 2019

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

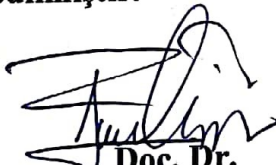
SAKARYA BÖLGESİ HANEHALKI ELEKTRİK
TÜKETİMİNİN DİNAMİK LINEER MODEL İLE
TAHMİNİ


YÜKSEK LİSANS TEZİ

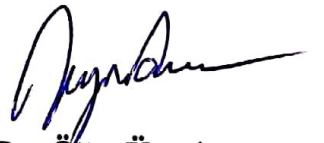
Emre DORUK

Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ

Bu tez 14.06.2019 tarihinde aşağıdaki jüri tarafından oybirliği ile kabul edilmiştir.


Doç. Dr.
Fatih TAŞPINAR
Jüri Başkanı


Doç. Dr.
Numan ÇELEBİ
Üye


Dr. Öğr. Üyesi
Tuğrul TAŞCI
Üye

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Emre DORUK

29.04.2019

TEŐEKKÜR

Yüksek lisans eğitimim boyunca değerli bilgi ve deneyimlerinden yararlandığım, her konuda bilgi ve desteğini almaktan çekinmediğim, araştırmanın planlanmasından yazılmasına kadar tüm aşamalarında yardımlarını esirgemeyen, teşvik eden saygı değer hocam Dr. Öğr. Üyesi Tuğrul TAŐCI'ya, aynı titizlikte beni yönlendiren değerli hocam Dr. Öğr. Üyesi Halil İbrahim CEBECİ'ye, son olarak beni yetiştiren, bana inanan ve maddi, manevi desteklerini esirgemeyen sevgili anneme, babama ve her zaman benim yanımda olan eşime teşekkürlerimi sunarım.

İÇİNDEKİLER

TEŞEKKÜR.....	i
SİMGELER VE KISALTMALAR LİSTESİ.....	iv
ŞEKİLLER LİSTESİ	v
TABLOLAR LİSTESİ	vii
ÖZET.....	viii
SUMMARY	ix
BÖLÜM 1.	
GİRİŞ	1
1.1. Tezin Organizasyonu	4
BÖLÜM 2.	
TEMEL KAVRAMLAR VE YÖNTEMLER.....	6
2.1. Zaman Serileri	7
2.1.1. Zaman serilerinin sınıflandırılması	7
2.1.2. Zaman serisi bileşenleri.....	10
2.1.3. Zaman serilerinin stokastik davranışları	11
2.1.4. Zaman serisi analizinin kullanım alanları	12
2.1.5. Zaman serisileri ve tahmin	14
2.2. Bayesci Yaklaşım	19
2.2.1. Kalman filtresi (KF)	22
2.2.2. Monte carlo markov zinciri (MCMC).....	23
2.2.3. GİBBS örneklemesi	25
2.3. Durum Uzay Modelleri (State-Space Models).....	27
2.3.1.Dinamik lineer model.....	29

BÖLÜM 3.	
LİTERATÜR ARAŞTIRMASI.....	31
BÖLÜM 4.	
DLM İLE ELEKTRİK TÜKETİM TAHMİNİ	36
4.1. Problemin Tanımı	36
4.1.1. Veri seti.....	37
4.2. Dinamik Lineer Modeller ile Zaman Serisinin Modellenmesi	38
4.3. Model Seçimi	43
4.3.1. Trend modeli	44
4.3.2. Sezonluk model ve mevsimsellik gösterimi	45
4.3.3. ARIMA süreçli dinamik lineer model	47
4.4. Maksimum Olabilirlik Tahmin Yöntemi.....	48
4.5. Sakarya İli Hanehalkı Elektrik Talep Tahmini İçin DLM.....	50
4.5.1. Tahmin için kullanılan dinamik lineer model varyasyonları.....	50
BÖLÜM 5.	
GERÇEKLEME, BULGULAR VE DEĞERLENDİRME	56
5.1. R Platformu ve “dlm” Uygulama Paketi	56
5.2. Modellerin Gerçeklemesi ve Çıktıları	57
5.2.1. ARIMA uygulamaları.....	60
5.2.2. SARIMA Uygulaması	64
5.2.3. Model A (Sezonluk faktör (SF) ve lineer büyüme (LB) modeli)	66
5.2.4. Model B (SF, Random Walk (RW) ve ARMA modeli).....	69
5.2.5. Model C (SF, LB ve ARMA modeli)	73
5.3. Bulgular ve Değerlendirme	76
BÖLÜM 6.	
SONUÇ VE ÖNERİLER	81
KAYNAKÇA	85
ÖZGEÇMİŞ	89

SİMGELER VE KISALTMALAR LİSTESİ

AIC	: Akaike Bilgi Kriteri
ACF1	: First-Order Autocorelation Coefficient
AR	: Otoregresif (Autoregressive)
ARMA	: Moving Average (Hareketli Ortalama)
ARIMA	: Auto Regressive Integrated Moving Average (Birleştirilmiş Otoregresif hareketli Ortalama)
BIC	: Bayes Bilgi Kriteri
DLM	: Dinamik Lineer Model
KF	: Kalman Filtresi
LB	: Lineer Büyüme
MAE	:Mean Absolute Error (Ortalama Mutlak Hata)
MAPE	:Mean Absolute Percentage Error(Ortalama Mutlak Yüzde Hata)
ME	: Mean Error (Ortalama Hata)
MCMC	: Monte Carlo Markov Chain
MLE	: Maximum Likelihood Estimation
R^2	: R Squared
RMSE	:Root mean squared error (Karekök Ortalama Hata)
RW	: Random Walk (Rassal Yürüyüş)
SF	: Sezonsal Faktör
YSA	: Yapay Sinir Ağları

ŞEKİLLER LİSTESİ

Şekil 2.1. Box-Jenkins Yaklaşımı	19
Şekil 4.1. Sakarya İli Mesken Elektrik Tüketim Verisi	37
Şekil 4.2. Dinamik Lineer Modeller ile Tahmin Süreci.....	41
Şekil 4.3. Dinamik Lineer Model ile Tahmin Aşamaları.....	43
Şekil 5.1. Trend ve Random Etki Bileşenine Ayrılmış Sakarya İli Mesken Elektrik Tüketimi	58
Şekil 5.2. DLM ile tahmin uygulamasında izlenen yol.....	59
Şekil 5.3. Durağanlaştırılmış Sakarya İli Mesken Elektrik Tüketim Verisi	61
Şekil 5.4. Seçilen Modele göre Serinin ACF Testi	61
Şekil 5.5. Seçilen Modele göre Serinin PACF Testi.....	62
Şekil 5.6. Sakarya İli Elektrik Tüketimi ve Hava Durumu Verileri ile Kurulan ARIMAX Modeline Göre Günlük Elektrik Talep Tahmini Sonucu (2017 Aralık).....	63
Şekil 5.7. 2017 Yılı Son 3 Aylık Verisi ve ARIMAX için Aralık 2017 Tahmin Değerleri	64
Şekil 5.8. SARIMA Modeline Göre Günlük Elektrik Talep Tahmini Sonucu (2017 Aralık).....	65
Şekil 5.9. 2017 Yılı Son 3 Aylık Verisi ve SARIMA Aralık 2017 Tahmin Değerleri	66
Şekil 5.10. Lineer Büyüme ve Sezonsal Faktör Modeline Göre Günlük Elektrik Talep Tahmini.....	67
Şekil 5.11. 2017 Yılı Son 3 Aylık Verisi ve Model A için Aralık 2017 Tahmin Değerleri	68
Şekil 5.12. 2017 Aralık Günlük Elektrik Talebi ve Model A DLM'ye Göre Yumuşatılmış Gözlem Değerleri.....	68

Şekil 5.13. Model A'ya Göre Günlük Elektrik Talep Tahmini Değerleri, Ortalaması ve Günlük Gerçekleşen Elektrik Tüketimi	69
Şekil 5.14. Rassal Yürüyüş Süreci, Sezon Faktör ve ARMA Bileşenine Sahip Modele Göre Günlük Elektrik Talep Tahmini	71
Şekil 5.15. 2017 Yılı Son 3 Aylık Verisi ve Model B için Aralık 2017 Tahmin Değerleri	71
Şekil 5.16. 2017 Aralık Günlük Elektrik Talebi ve Model B DLM'ye Göre Yumuşatılmış Gözlem Değerleri	72
Şekil 5.17. Model B' ye Göre 2017 Aralık Günlük Elektrik Talep Tahmini, Ortalaması ve Gözlem Değerleri	72
Şekil 5.18. Lineer Büyüme, Sezon Faktör ve ARMA Bileşenini İçeren Modele Göre Günlük Elektrik Talep Tahmini.....	74
Şekil 5.19. 2017 Yılı Son 3 Aylık Verisi ve Model C için Aralık 2017 Tahmin Değerleri	75
Şekil 5.20. 2017 Aralık Günlük Elektrik Talebi ve Model C DLM'ye Göre Yumuşatılmış Gözlem Değerleri	75
Şekil 5.21. Model C' ye Göre 2017 Aralık Günlük Elektrik Talep Tahmini, Ortalaması ve Gözlem Değerleri	76
Şekil 5.22. T1 Zaman Aralığı için Model A, Model B ve Model C Aralık 2017 Tahmin Sonuçları	77
Şekil 5.23. T2 Zaman Aralığı için Model A, Model B ve Model C Aralık 2017 Tahmin Sonuçları	78
Şekil 5.24. T3 Zaman Aralığı için Model A, Model B ve Model C Aralık 2017 Tahmin Sonuçları	79

TABLolar LİSTESİ

Tablo 3.1. Durum-Konum Modeli ve DLM Alanında Yapılan Bazı Çalışmalar....	33
Tablo 5.1. Özel işlevler için tanımlanabilen "dlm" fonksiyonları	57
Tablo 5.2. ARIMAX Model Parametreleri	62
Tablo 5.3. Tutarlılık Testi-Hata Sonuçları (ARIMAX)	63
Tablo 5.4. SARIMA Model Parametreleri	64
Tablo 5.5. Tutarlılık Testi ve Hata Sonuçları (SARIMA)	65
Tablo 5.6. Model A için Sistem ve Gözlem Matrislerinin MLE Standart Hata Oranı.....	67
Tablo 5.7. Tutarlılık ve Hata Sonuçları (Model A).....	69
Tablo 5.8. Model B için Sistem ve Gözlem Matrislerinin MLE Standart Hata Oranı.....	70
Tablo 5.9. Tutarlılık ve Hata Sonuçları (Model B).....	73
Tablo 5.10. Model C için Sistem ve Gözlem Matrislerinin MLE Standart Hata Oranı.....	74
Tablo 5.11. Tutarlılık ve Hata Sonuçları (Model C).....	76
Tablo 5.12. Gerçeklenen Uygulamaların Hata Oranları Karşılaştırması	80

ÖZET

Anahtar Kelimeler: Dinamik Lineer Model, Bayesci Yaklaşım, Durum-Uzay Modeli, Zaman Serileri Tahmini, Elektrik Tüketimi Talep Tahmini

İnsanoğlu tarihin ilk çağlarından itibaren yaşamını kolaylaştırmak ve yaşam kalitesini arttırmak için çevresindeki enerji kaynaklarını yoğun olarak kullanmıştır. Bu amaçla geçmişte yoğun olarak kullanılan kaynaklar odun, kömür ve petrol olurken günümüzde ise yaygın bir şekilde kullanımda olan elektrik enerjisi olarak karşımıza çıkmaktadır. Elektrik enerjisi yapısı gereği depolanma imkânı olmayan, üretildiğinde tüketilmesi gereken nadir karakterdeki emtialardan birisidir. Bu husus günümüz enerji şirketlerine enerji dağıtım sürekliliği sağlamak ve altyapı oluşumlarını gerçekleştirmek konusunda sıkıntı yaratmaktadır. Enerji şirketlerinin bahsedilen sorunu çözme yöntemlerinden birisi elektrik kullanıcılarının talebini ölçmektir. Kullanıcıların talebine karşılık üretilecek elektrik arzını belirlemek elektrik dağıtım şirketlerinin dağıtım maliyetlerini düşürmekle birlikte, gelecekteki talebi karşılayabilmek için altyapı yatırımlarını da planlama imkanı sunmaktadır.

Kullanıcıların gelecekteki elektrik tüketim talebini belirleme yöntemlerinden birisi, geçmiş verileri kullanarak, gelecek değerlerin zaman serileri yöntemi ile tahmin edilmesidir. Birçok zaman serisi tahmin yöntemi bulunmakla birlikte, yöntemlerin arasında bazı temel farklılıklar vardır.

Bu çalışmada tahmin ve zaman serileri konusunda Durum-Uzay modellerinin Bayesci bir yaklaşımı olan Dinamik Lineer Modeller tanıtılmaktadır. Zaman serisi kavramları ele alınarak, zaman serilerinin, Dinamik Lineer Modeller ile modellenmesi konusunda biçimsel bir uygulama yapılmıştır. Uygulamada gerçek hayattan elde edilen verilerin zaman serileri ile analizinde karşılaşılan problemleri daha iyi kavrayabilmek için SEDAŞ aracılığı ile elde edilen Sakarya iline ait hanehalkı tüketim verileri kullanılmıştır. Uygulama sonuçları zaman serilerinin modellenmesi konusunda en çok kullanılan yöntemlerden biri olan ARIMA model sonuçları ile kıyaslanarak Dinamik Lineer Modeller ile tahmin işleminin performansı ortaya konulmuştur.

ESTIMATION OF HOUSEHOLD ELECTRICITY CONSUMPTION IN SAKARYA REGION BY DYNAMIC LINEAR MODEL

SUMMARY

Keywords: Dynamic Linear Model, Bayesian Approach, State-Space Model, Time Series Estimation, Electricity Consumption Demand Forecasting

Mankind has intensely used the energy resources in the environment to make life easier and to improve the quality of life. For this purpose, while the resources used heavily in the past have been wood, coal and oil, nowadays, it is commonly used as electricity energy. Electricity is one of the rare commodities that do not be stored due to its structure and should be consumed when it is produced. This issue creates difficulties for the energy companies to ensure the continuity of energy distribution and to realize the infrastructures. One way energy companies can solve this problem is to measure the demand of electricity users. Determining the electricity supply to be produced against the demand of the users provides the opportunity to plan the infrastructure investments in order to meet the future demand while reducing the distribution costs of electricity distribution companies.

One of the methods of determining the future electricity consumption demand of the users is to estimate the future values with the time series method by using historical data. There are several time series estimation methods, but there are some fundamental differences between methods.

In this study, Dynamic Linear Models, a Bayesian approach of State-Space models, are introduced. Time series concepts were taken into consideration and a formal application was made to model time series with Dynamic Linear Models. In practice, household consumption data of Sakarya province, which was obtained by means of SEDAŞ, were used to better understand the problems encountered in the analysis of the real life data with time series. Application results are compared with ARIMA model results which are one of the most commonly used methods in modeling of time series, and the prediction performance of time series with Dynamic Linear Models is shown.

BÖLÜM 1. GİRİŞ

İnsanođlu tarihin ilk çağlarından itibaren yaşamını kolaylařtırmak ve yaşam kalitesini arttırmak için çevresindeki nesnelere en iyi şekilde faydalanmayı öğrenmiştir. Bu durum insanođlunun yerleşik hayata geçerek tarıma ilk başladığı zamanlarda dahi geçerlidir. Ancak yoğun nüfus artışı, kentleşme ve insanlığın yaşamını kolaylařtırmak ve yaşam kalitesini arttırmak için geliřtirdiđi her ürün ve yöntem beraberinde yeni ihtiyaçları da ortaya çıkarmıştır. Bu ihtiyaçların karřılanması ise insanlığın zorunlu olarak kabul etmesi gereken bazı kısıtlamaları da beraberinde getirmiştir. Bahsedilen kısıtlamalar, dünyamızda temiz su kaynaklarının korunması, iklim deđişikliklerinin saptanarak önlenmesi, tarım üretiminin nüfus artışına paralel bir şekilde arttırılması, çeşitli emtia ve enerji kaynaklarının üretimi, fabrika üretimlerinin belirli bir seviyede tutulması hatta nüfus artış hızının planlanmasına kadar genişletilebilir. Başka bir deyişle insanođlu yaşamını kolaylařtırmak ve yaşam kalitesini arttırmak için kısıtlamaları planlamak zorundadır.

Planlama işleminin ise ancak temel bir bilgi ve dayanak varsa uygulanabilir, geçerli bir kavram olur. Planlama yapılmadığında ise istenmeyen durumlarda bireysel ya da kurumsal olarak yüksek maliyetli çözümler veya düzeltilemeyen durumlar ile karşılaşılabilir. Örneğinin 1973 yılında yaşanan petrol krizinde Arap-İsrail savaşı ile birlikte Arap ülkelerinin petrol fiyatlarını arttırması, 1950'li yıllardan itibaren hızla artan tüketim ve tüketim ile aynı hızda artmayan üretim nedeniyle arz-talep dengesinin bozulması, bütün dünyayı sarsan bir olay olarak karřımıza çıkmıştır. Arz-talep dengesinin sağlanması gelecek durumların tam olarak bilinmesi de önceden öğrenilmesi ve geleceğinin planlanması ile gerçekleştirilebilir. Geleceğinin planlanması ise eldeki bilgi ve kısıtlarla birlikte geleceğinin bilgisinin öngörülmesi ile olabilir. Bu durumda eldeki veriler belirli bir düzen içinde sıraya konularak, ilişki kurarak istatistiksel ya da bilgisayarlı yöntemler ile çıkarımlarda bulunarak geleceğe dair bilgi

edinilebilir. Bilginin elde edilmesinde ki tüm bu süreç tahmin işlemi olarak da adlandırılabilir.

Tahmin işlemi insanoğlunun yukarıda bahsedilen zorunluluk karşısında engelleri aşmak için kullandığı yöntemlerden birisidir. Belirli bir arza karşı gelen talebin dengelenmesinde kritik rol oynamaktadır. Günümüzde insanlığın karşılaştığı en büyük kısıtlamalardan birisi olan enerji üretimi ve dağıtımı konusunda da tahmin işlemi, geleceğin planlanması konusunda araştırmacılar için zorunlu bir faaliyet olarak karşımıza çıkmaktadır. Tahmin yapılırken birçok farklı yöntem ve teori kullanılırken, enerji verilerinin analizi için genellikle zaman serisi analiz ve tahmin yöntemleri kullanılmaktadır.

Zaman serilerinin analizi kavramı, zaman boyunca gelişen serinin tanımlanması, modellenmesi ve tahmin edilmesi sürecini kapsar ve insanoğlunu etkileyen birçok önemli olayı tanımlama ve çözmek için kullanılan baş aktörlerden birisidir. Tahmin işlemi daha geniş bir tanımıyla; bir değişkene ait tarihsel verilerden başka bir şey olmayan bir zaman serisinde yararlanılarak gelecek dönemde alacağı değerin önceden belirlenmesi olarak nitelendirilir. Zaman serileri tahmininin ise bakıldığında istatistik, ekonomi, üretim, tıp, iklim analizleri, işletme ve verimlilik gibi farklı alanlarda uygulama alanı vardır.

Zaman serilerinin önemli olduğu diğer bir alan da bahsedildiği üzere enerji verimliliği konusudur. Günümüzde enerji üretimi, dağıtımı ve kullanımı konusunda sürekli iyileştirmeler sağlanmaktadır. Bunun başlıca başlıca sebepleri sürekli artan enerji ihtiyacı ve enerji sistemlerinin kurulum maliyetlerinin yüksek olması, piyasa serbestleşmesi ve elektronik cihazların artan kullanımı olarak sayılabilir. Enerji verimliliğinin sağlanamadığı durumlarda elektrik kesintileri, yetersiz altyapı ve kritik sistemlerde yaşanabilecek olası arızalar ile karşılaşılabilir. Ayrıca elektrik enerjisinin sürekliliğinin sağlanması birçok otoriteye göre temel hak olarak kabul edilmektedir. Artan ihtiyacı karşılayabilmek ve yeni enerji sistemlerinin kurulum ihtiyacını görebilmek için araştırmacılar, devlet ve özel şirketler enerji kullanımının gelecekteki

yönünü tayin etmek amacı ile tahminler yaparlar. Örneğin TEİAŞ, dönemsel ve uzun vadeli olarak elektrik tüketim talepleri ile ilgili bölgesel tahminler gerçekleştirir.

Ancak elektrik sektöründe, talep tahmininin doğru yapılması bazı sebeplerden dolayı kolay değildir. Elde edilmeye çalışılan tahmin değerleri içinde rassal ögeler bulunmaktadır ve değerlerin doğruya yakın tahmin edilmesi konusu rassal ögelerin de tahminini gerektirir. Bu kapsamda farklı teoriler ve yöntemler geliştirilmiştir. Bu yöntemlerin geneli zaman serisi analizi kavramına dayalı olsa da özelde farklı teorileri bünyelerinde barındırırlar.

Enerji tüketim hesabının zaman serileri ile analiz yöntemlerinden ve en çok kullanılan otoregresif-hareketli ortalama (ARIMA) modelleme yöntemidir. ARIMA modelleme yöntemleri ile zaman serisi analizi, serinin durağan olması koşulunu arar. Ancak gerçek hayata dair zaman serisi verilerinin durağanlığı genelde söz konusu olmaz. Bu durumda serinin durağanlığı fark alma işlemi ile sağlanır.

Son yıllarda ARIMA modellerinin yanında, dinamik lineer modeller ile zaman serisi analizi çalışmaları artmakta ve elektrik enerjisi tüketim analizleri de dahil olmak üzere birçok alanda kullanılmaktadır. Dinamik lineer modellere duyulan ilginin sebepleri ise, Bayesci stokastik bir yöntem olması, hesaplama zorluklarını kaldırması ve uygulama kolaylığı olarak sıralanabilir. Ayrıca hesaplama zorluğu çekilen rastlantısal ögelerin dinamik lineer modeller ile modellenmesi diğer yöntemlere nazaran daha kolaydır. Bayesci dinamik lineer modellere göre her türlü belirsizlik olasılıklarla ifade edilebilir ve ölçülebilir.

Dinamik lineer modellerinin ARIMA yöntemlerinden farklı ve en güçlü yönlerinden bir diğeri de ARIMA modellerinde zaman serisinin durağanlığının (stationarity) zorunlu olması iken DLM'de böyle bir zorunluluğun olmamasıdır. Ayrıca DLM'ler, zaman serisinin uğrayacağı rassal şoklardan dolayı nasıl değişime uğrayacaklarını da tanımlar.

DLM'ler Bayesci istatistiğe göre çalışır ve Bayes yaklaşımına göre birçok nedenden kaynaklanabilecek belirsizlik durumları karşısında tutarlı bir yapı oluşturur. Model matematiksel olarak ifade edildikten sonra eldeki verilere olasılık kuralları uygulanarak olasılıksal çıkarımlarda bulunmaktadır. Tahmin konusunda bu çıkarımlar, modelin bilinmeyen parametreleri ve gelecek dönem değerleri olabilir. Tahmin bilgisinin elde edilmesi modelin kurulmasına ve işletilmesine bağlıdır.

Dinamik lineer modeller önceki zaman adımlarında sistem durumları arasındaki geçiş ilişkisi ve gözlemlerden yararlanarak elde edilen bilginin, gelecek zaman diliminde kullanılması ve çıkarım işleminin gerçekleştirilmesini sağlayan bir işleyişe sahiptir. Bu kapsamda eğitim veri setinden çıkarılan bilgi, tahmin yapmakta kullanılır. Tahmin yöntemi olarak bilinmeyen parametreler için maksimum olabilirlik fonksiyonunu, gözlem ve sistem durumlarını tahmin etmek içinse Kalman filtresini ya da MCMC yöntemini kullanabilir.

1.1. Tezin Organizasyonu

Sakarya iline ait hanehalkı elektrik tüketim verisinin zaman serileri yöntemleri ile dinamik lineer modelleme yöntemini kullanarak modellenmesi ve gelecek değerlerin tespit edilmesinin hedeflendiği bu tez çalışmasında, öncelikli olarak zaman serilerinin tanımıyla birlikte özellikleri ele alınarak, zaman serisi modelleme yöntemlerinin tanımları ve matematiksel ifadeleri açıklanacaktır.

Ardından çalışmada kullanılan dinamik lineer modellerin temeli olan durum uzay modelleri ve Bayesci yöntemlerden Kalman Filtresi (KF), Monte Carlo Markov Zincirleri (MCMC), Gibbs Örneklemesi gibi yöntemlerin özellikleri, kullanım alanları ele alınacaktır. Bu çalışmanın temel odak noktası durum uzay modellerinin Bayesci bir yaklaşımı olan dinamik lineer modeller ile tahmin olduğundan, daha sonra Bayesci tahmin yöntemleriyle ilgili açıklamalar ile birlikte dinamik lineer modellerin tanımı, özellikleri, geçmişi, gelişimi ve yapılan çalışmalar hakkında bilgi verilecektir.

Üçüncü bölümde dinamik lineer modellemenin ortaya çıkışı, gelişiminin aşamaları ve bugüne kadar yapılan bazı çalışmalardan örnekler ile birlikte kullanım alanlarından bahsedilecektir.

Dördüncü bölümde, bu çalışmada ele alınan Dinamik Lineer Modeller ile Zaman serilerinin modellenmesi ve tahmini problemi ile ilgili tanım, veri akışı, özellikler, başlangıç koşulları verilecek, kullanılan modellerin yapısı detaylı bir şekilde ele alınacaktır.

Beşinci bölümde, elde edilen bulgular verilerek performans bilgilerinin değerlendirilmesi yapılacaktır. Son bölümde bu tez çalışmasında uygulanan yöntemin, diğer zaman serisi modelleri ile kıyaslaması yapılacak ve dinamik lineer modellerin elektrik piyasasının geniş ölçekli tahmin problemlerine nasıl uyarlanabileceği hakkında öneriler sunulacaktır.

BÖLÜM 2. TEMEL KAVRAMLAR VE YÖNTEMLER

Bir konuya ilişkin gözlem değerleri incelendiğinde değişkenler arasındaki ilişkinin yönü ve büyüklüğünü belirleme amacı kadar, bu değişkenlerin gelecekte alacakları değerleri önceden tahmin etmek büyük önem taşımaktadır. Model bilgilerindeki her bir eksiklik, çok sayıdaki değişkenin modelde ifade edilmesi zorluğu ve yöntemlerin kendi kusurları gibi faktörler tahmin değerlerinde hatalı sonuçlar elde edilmesine neden olabilir.

Araştırmacılar, herhangi bir değişkenin gelecek dönemdeki değerini tahmin etmek için 1970'li yıllara kadar yoğun olarak tek ve çok değişkenli regresyon modelleri kullanmışlardır ancak günümüzdeki tahmin araçları ve ortamları oldukça çeşitlidir. Yine 1970'li yıllarda yaşanan ekonomik bunalımlar regresyon yöntemlerine duyulan güvenin sorgulanmasına yol açmış, gözlem verilerinin incelenebilmesi ve tahmin edilmesi için farklı yaklaşımlar geliştirilmeye çalışılmıştır [1].

Bunlardan biri gözlem değerlerinin zaman serisi olarak ele alınması yaklaşımıdır. Gözlem verilerinin zaman serisi olarak incelenmesi, zaman serisini açıklayan bir model oluşturulması ile zaman serisi hakkındaki bilginin elde edilmesi sürecine dayanır. Elde edilen bilgi ile zaman serisinin gelecekteki değerleri çeşitli yöntemler ile tahmin edilebilir. Ancak bunun için kurulacak olan modelin en uygun biçimini oluşturmak gerekir.

Bu bölümde zaman serilerinin tanımı yapılmış, genel tanımı verilerek zaman serisi kavramları incelenmiştir. Ayrıca en çok kullanılan zaman serisi modelleme yöntemlerinden bazıları genel hatları ile açıklanmıştır. Bölümün sonunda bu çalışmada da tercih edilen zaman serilerinin dinamik lineer modellerle analizine yer verilmiş, genel yapısı incelenmiştir.

2.1. Zaman Serileri

İnsanoğlu için zaman kavramı, medeniyetin ilk zamanlarından beri ön plana çıkmaktadır. Günümüzde de elde edilen verilerle ilgili birçok değişkenin zamanla bağlantılı bir şekilde sunulduğu görülmektedir [2].

Değişkenlerin belirli zaman aralıklarındaki değerlerinin ardışık bir şekilde gözlenebildiği sayısal büyüklükler zaman serisi olarak adlandırılır. Bir zaman serisi bir veya birden fazla değişkeni kapsayan bir veri setidir. Bu değişkenlerin gözlem noktalarına ait verilerin zaman içinde sürekli bir biçimde gerçekleşmesi şart olmasa da düzenli aralıklarla bir değere sahip olması gereklidir [3]. Zaman serilerinin analizi ve yorumlanması ise tarım, astronomi, doğa olayları (sel, taşkın vb.), meteoroloji, üretim ve ekonomi alanlarında geleceğin tahmini için büyük önem arz etmektedir [2].

Zaman Serisi: aynı değişkenle ilgili sıralı veri değerleri dizisini temsil eder, tipik olarak ardışık zaman aralıklarında ölçülür, n gözlem sayısı olmak üzere $t, t=1, 2, \dots, n$ şeklinde gösterilir ve gözlenen verinin değeri A_t olmak üzere zaman serisi;

$$A = \{a_1, a_2, a_3, \dots, a_n\} \text{ veya } A = \{a_t\}, t = 1, 2, 3, \dots, n$$

şeklinde tanımlanır.

Değişken değerleri her an için gözlenebiliyorsa sürekli, belirli zaman aralıklarında gözlenebiliyorsa süreksiz zaman serisi olarak nitelendirilir. Sürekli zaman serilerinde zaman $T[0-1]$ gibi bir aralığa sahipken süreksiz zaman serileri $T=0$ gibi belirli bir anı işaret eder [1].

2.1.1. Zaman serilerinin sınıflandırılması

Zaman serileri gözlem değerlerinin elde edilme biçimine bağlı olarak, sürekli ve kesikli seriler, gözlem değerlerinin grafiğinin serinin ortalama değerini kesme sıklığına ya da ortalama değerden büyük sapmalar gösterip göstermediklerine göre

durağan ve durağan olmayan seriler ve son olarak dönemsel hareketlerine göre mevsimsel veya mevsimsel olmayan olarak incelenmektedir. Ayrıca zaman serileri değişken sayılarına göre ve serinin yapısına göre temelde iki bölüme ayrılır. Bu ayrışım serinin özelliklerinin iyi belirlenmesini amaçlar bu sayede yapılan analizden en doğru sonuç alınması hedeflenir [4].

a. Değişken Sayısına Göre Zaman Serileri

- Tek değişkenli zaman serisi (univariate): Eğer bir zaman serisi analizinde tek değişkenin zaman içindeki hareketi inceleniyorsa tek değişkenli zaman serisi,
- Çok değişkenli zaman serisi (multivariate): Eğer birden fazla değişkene ait verilerin zaman içindeki değişimi birlikte gözlemleniyorsa çok değişkenli zaman serisi denir.

b. Yapısına Göre Zaman Serileri

Zaman serileri yapılarına göre incelendiğinde “stokastik” ve “deterministik” zaman serileri olmak üzere ikiye ayrılır.

- Stokastik Zaman Serileri: Gelecekte serinin alabileceği değerlerin kısmı olarak serinin geçmiş değerleri tarafından tanımlanabilmesi demektir. Stokastik serilerin öngörülerini tam doğru olarak yapmak mümkün değildir. Ancak gelecekteki değerler, geçmiş değerlerin bir bilgisiyle elde edilen şartlarla bir olasılık dağılımına sahiptirler. Zaman serileri genellikle random (rastgele) değişkenler yani stokastik (olasılıklı) değişkenlerle çalışmaktadır. Stokastik özellikler değişkenin durağanlığı ile ilgilenmektedir. Stokastik süreçler bölüm (2.1.3)’te daha geniş bir biçimde ele alınmıştır.
- Deterministik Zaman Serileri: Eğer bir zaman serisi tam olarak tahmin edilebiliyorsa, deterministik (kesin) zaman serisi olarak ifade

edilmektedir. Deterministik özellikler trend, sabit ve mevsimselliğin (sezonsallığın) varlığı ile ilgilendirilir.

c. Sürekli ve Kesikli Zaman serileri

Zaman serisinin gözlem değerleri zaman içinde devamlı olarak gözlenebiliyorsa, elde edilen seriye sürekli zaman serisi denir. Sürekli zaman serileri genellikle zaman içinde eşit olmayan aralıklarda elde edilen gözlem değerlerinden oluşur. Gözlem değerleri sadece belirli zaman aralıklarında görülebiliyorsa, bu tip serilerde kesikli zaman serileri olarak nitelendirilir ve kesikli zaman serileri genellikle eşit zaman aralıklarında yapılan gözlem değerlerinden oluşur [4].

d. Durağan ve Durağan Olmayan Zaman Serileri

Zaman serisi analizlerinde en önemli kavramlardan biri de zaman serisinin durağanlığıdır. Bir zaman serisinin ortalamasında ve varyansında belirli bir düzende değişme yoksa ve düzenli periyodik değişimler meydana gelmiyorsa bu zaman serisine durağandır denilir. Diğer bir deyişle durağan zaman serileri trend, konjonktür ve mevsim etkilerinden arındırılmış serilerdir.

Bir zaman serisinin aşağıdaki koşulları sağladığı sürece tam/güçlü durağan olması beklenir.

- Zaman serilerinin ortalama değeri zaman içinde sabitse, yani trend bileşeni geçersizse
- Varyans zamanla artmıyorsa.
- Mevsimsellik etkisi yoksa.
- Gözlem değerleri kümesinin bileşik olasılık dağılımı, gözlemlerin yapıldığı süre boyunca herhangi bir değişikliğe uğramıyor ve tüm özellikleri zaman boyunca sabitse.

Bir zaman serisinin tüm özellikleri değil de sadece sıfır orijinine göre aritmetik ortalaması zamana göre değişmiyorsa birinci dereceden durağan seri, bu durağanlığa birinci dereceden durağanlık adı verilir. Sadece ortalaması, varyansı ve kovaryansı zamana göre değişmiyor ve zamandan bağımsızsa bu seriye “zayıf durağan” ya da “ikinci dereceden” durağan seri denir [3,5].

Gerçek yaşamda durağan zaman serisi örneklerinde çok az rastlanır. Zaman serileri genellikle trend, konjonktürel, mevsimsel veya rassal dalgalanmalar gösterirler. Bahsedilen etkilere maruz kalmış zaman serilerinde süreç boyunca farklı noktalarda değişik etkiler gözlenebilir. Ancak zaman serisinin analizi için geliştirilmiş yöntemlerin çoğunda bu tip zaman serilerinin zorunlu olarak durağanlaştırılması gerekir [4]. Bu çalışmanın odak noktası olan dinamik lineer modelleme yöntemi ile bu sorunun çözülebildiği görülmüştür.

e. Mevsimsel ve Mevsimsel Olmayan Zaman Serileri

Bir zaman serisi birbirini takip eden süreçlerinde, sürecin bazı dönemlerinde oluşan, diğer dönemlerden daha düşük ya da daha yüksek dalgalanmalar görülüyorsa mevsimsel seri, aksi durumda mevsimsel olmayan seri olarak nitelendirilir. Mevsimsel etkilerin görülebilmesi için zaman serilerinin yeterli sayıda gözlem değerine sahip olması gerekir [4,6].

Zaman serilerinde mevsimsel etkilerin ortaya çıkmasında iklim, insan alışkanlıkları, resmi veya dini bayramlar, özel günler ve sosyal olaylar gibi durumlar rol oynar [6].

2.1.2. Zaman serisi bileşenleri

Zaman serisini oluşturan gözlemlerin zaman içerisinde değişimini gösteren kalıpların yapısı ve seri içerisinde yer alan olağandışı gözlemler söz konusu olan serinin özelliğini ortaya koyar. Başka bir deyişle serinin içerdiği bileşenleri ortaya çıkarır. Bu bileşenler seri boyunca süren bir trendi, mevsimsel etkileri, rassal olayları ve konjonktürel bir durumu vb. belirtir [3].

- Trend(T), zaman serisi gözlem değerlerinin dönemsel olarak göstermiş olduğu artma ya da azalma şeklindeki kararlı eğilime “trend” adı verilmektedir. Doğrusal olarak ya da doğrusal olmayan bir biçimde gerçekleşebilir. Trend zamana bağlı değişken üzerindeki genel eğilime neden olan uzun dönemli etkileri açıklar.
- Mevsimsel Etkiler(S), bir yıl içerisinde tamamlanan zaman serisindeki periyodik kalıplardır. Mevsimsellik artış ya da azalış etkisini periyodik olarak gözlem değerlerine yansıtır.
- Konjonktürel Değişim(C), bir zaman serisinin dönemsel olarak ortaya çıkan değişimlerini yansıtan davranışlardır. Trend etrafındaki uzun süren dalgalanmalar da olarak tanımlanabilir.
- Rassal/Düzensiz Değişim(R), zaman serisinde belirli bir kalıba göre hareket etmeyen beklenmedik olayların etkisiyle meydana gelen dalgalanmalar olarak nitelendirilir. Deprem, siyasal olaylar, ani hava değişimleri vb. gibi değişiklikler buna örnek olarak verilebilir [7].

Bu değişkenlere göre zaman serisinin ifadesi iki şekilde formüle edilebilir [5].

- Toplamsal Model: $Y_t = T_t + C_t + S_t + R_t$
- Çarpımsal Model: $Y_t = T_t \times C_t \times S_t \times R_t$

2.1.3. Zaman serilerinin stokastik davranışları

Zaman serileri sadece zamanın deterministik bir fonksiyonu değildir yani bu olaylar sadece zaman değişkenine bağlı olarak açıklanamaz. Bir zaman serisinin analizinde, sürecin gösterdiği seyrin tam olarak açıklanabilmesi için kullanılacak olan modelin sürece etki eden bütün değişkenleri içermesi, değişkenler hakkında yeterli bilgi bulunması ve sayısal olarak ifade edilmesi beklenir ancak bu her zaman mümkün değildir.

Zamana bağılı olaylar rassal karaktere sahiptir ve rasgele değişkenlerin zamana göre sıralanmış şekilde toplanması stokastik süreç olarak adlandırılmaktadır. Bu gibi olaylarla ilgili serilerin gelecek dönemdeki ilerleyişini, bugünkü ve geçmiş dönem değerlerine dayanarak incelemek için değişik bir yaklaşım gerekir. Buna deterministik olmayan, stokastik veya istatistik yaklaşım denmektedir [4].

Stokastik süreçler kesikli olup olmamasına bağılı olarak farklı sembollerle gösterilir. Kesikli stokastik süreçler Y_t ile gösterilirken sürekli stokastik süreçler $Y(t)$ gibi gösterilir. Bir stokastik süreçte bulunan her gözlem rastgele bir şekilde çekildiğinden rassal bir değişkendir ve gözlemlerin de belirli bir olasılık dağılımına göre oluştuğu varsayılmaktadır [3].

Stokastik bir süreç için tahmin yöntemlerinden birisi t_1, t_2, \dots, t_n gibi bir veri setinin $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ birleşik olasılık dağılımını tanımlamak diğer bir yolu ise ortalama, varyans ve otokovaryans fonksiyonunu belirlemektir [6].

$$\text{Ortalama; } \mu_t = E(Y_t) \quad (2.1)$$

$$\text{Varyans; } \sigma_t^2 = \text{Var}(Y_t) \quad (2.2)$$

Y_{t_1} ile Y_{t_2} arasındaki otokovaryans fonksiyonu ise;

$$\text{Otokovaryans } \gamma_{t_1, t_2} = \text{Cov}(Y_{t_1}, Y_{t_2}) = E[(Y_{t_1} - \mu_{t_1})(Y_{t_2} - \mu_{t_2})] \quad (2.3)$$

şeklinde yazılabilir ve eğer Y_t normal dağılıma sahip ise Y_t 'nin dağılımı Gaussian süreç olarak adlandırılır. Dolayısı ile ortalama ve varyans kullanılarak Y_t 'nin özellikleri belirlenebilir [6].

2.1.4. Zaman serisi analizinin kullanım alanları

Zaman serileri analizi zamana bağılı sistemlerin tanımlanması, modeller geliştirilmesi ve bir sonunun çözümüne dair tahmin yapılması için gerekli olan yöntemlere genel

olarak verilen isimdir [8]. Zaman serisi verisi türüne (ekonomik, iklimsel, tıbbi vb.) ve şekline göre farklı uygulama alanları vardır. Örnek verecek olursak; zaman serileri ile kümeleme, sınıflama, örüntü keşfi, segmentasyon, Box-Jenkins model uygulamaları, YSA gibi uygulamalar aracılığı ile analiz edilebilir. Bunun yanında bir zaman serisi ileriye dönük tahmin, seriler arasındaki ilişki, seri unsurlarının belirlenmesi ve kontrol amaçlı olarak analiz işlemlerinden geçirilir [5].

Zaman serisi analiz uygulamaları aşağıda maddeler halinde belirtilen konularda fayda sağlamak için gerçekleştirilir.

- Aykırı (Outlier) verileri yakalama: Eldeki veriler bir seri olarak düzenlendiğinde serinin uzağında bulunan veriye aykırı veri denir. Aykırı verilerin yakalanabilmesi için o seri olarak düzenlenen verinin hareketinin belirlenebilmesi gerekir. Eldeki verilerin zaman serisi analizi ile sürecin hareketleri belirlenebilir ve hareketlerde olan farklılık yakalanabilir.
- Tahmin (Prediction): Zaman serisi analizlerinin genel olarak odak noktası süreçlerin gelecek değerlerinin tahmin edilmesidir. Zaman serileri matematiksel veya bilgisayarlı yöntemler ile modellenerek, serinin ilerleyişi hakkında bilgi edilebilir ve bu bilgi gelecek değerler yaklaşık olarak tahmin edilebilir.
- Eksik verileri Tamamlama (Inputation): Veri setlerinde bazı veriler eksik olabilir bu tip durumlarda zaman serisi tahmin yöntemleri ile tahmin edilerek bu eksiklik giderilebilir.
- Hata Düzeltme (Data scrubbing): Veri setlerinde bulunan aykırı değerlerinin diğer değerlere çeşitli yöntemlerle yaklaştırılması işleminde zaman serisi analizi kullanılmaktadır [9].

2.1.5. Zaman serisileri ve tahmin

Zaman serileri tahmini, zamanla ilişkili verinin (çoğunlukla düzgün zaman aralıklarıyla ilişkili veri noktalarının sonlu gözlem dizileri) serinin gelecekteki veri noktalarını tahmin etmek için modeller üretmek amacıyla kullanıldığı algoritmalar önermektedir. Zaman serileri analizi başlangıçta matematiksel modelleme, zaman-frekans analizi (Fourier ve dalgacık dönüşümleri) içeren araçlara dayansa da son yıllarda Yapay Sinir Ağları (YSA) ve DLM gibi klasik stokastik süreçler kuramında dayanan bilgisayar destekli programlama yöntemleri kullanılmaya başlanmıştır [2,8].

Bu yöntemlere duyulan talebin sebepleri olarak, Bayesci zaman serisi modellerinin işletiminde bazı ileri istatistik yöntemlerinin bulunması ve stokastik temelli simülasyon tekniklerinin sağladığı uygulama geliştirme kolaylığı olarak gösterilebilir.

Zaman Serisi Tahmini bilinen geçmiş olaylara dayanan ve gelecekteki olayların tahmin edilmesi için kullanılan modeli temsil eder. Burada zaman serisinin geçmişteki değerleri değişken olarak kabul edilerek gelecekteki değerlerini tahmin etmek amaçlanır.

a. Zaman Serisi Tahmin Yöntemleri

Zaman serileri tahmin yöntemlerinden bazıları; Otoregresif (Auto Regressive- AR), Hareketli Ortalama (Moving Average -MA), Basit Üstel Düzleştirme, Otoregresif-Hareketli Ortalama (Auto Regressive-Moving Average - ARMA), Regresyon ve Birleştirilmiş Otoregresif-Hareketli Ortalama (Auto Regressive Integrated Moving Average - ARIMA), Box-Jenkins tahmin modelleridir. Bu yöntemlerin en büyük bilinen dezavantajı ise zaman serisinin doğrusallığı varsayımına dayalı olarak geliştirilmeleridir. AR(p), MA(q) ve iki modelin birleşimi olan ARMA(p,q) modelleri sadece durağan süreçlere uygulanırken, ARIMA(p,d,q) modelleri durağan olmayan zaman süreçleri için kullanılmaktadır [10].

Ancak gerçek hayattaki verilerin birçoğu doğrusal olmayan özelliğe ve dış etkenlere bağlı olarak zaman zaman kırılmalara sahiptir. Bu nedenle Zaman Serileri analiz yöntemleri için YSA gibi sıkça kullanılan yöntemlere başvurulmuştur. [13] Ayrıca bu çalışmada uygulaması yapılan durum-uzay modelleri tabanlı ve kalman filtreleme metodu ile desteklenen dinamik lineer modelleme yöntemi doğrusal nitelikte olmayan zaman serilerinin tahmini için güçlü bir alternatif yöntem olma özelliğini taşımaktadır. Zaman serisi analiz ve tahmin yöntemlerinden bazıları başlıklar halinde açıklanmıştır.

b. Regresyon Analizi

Bir bağımlı değişkenin, bir veya birden fazla bağımsız değişkenle arasındaki ilişkiyi veren fonksiyonun yazılmasına regresyon analizi denir. Regresyon analizi ile değişkenler arasındaki ilişkiler araştırılır, eğer ilişki varsa bunun gücü hakkında bilgi edinilebilir. Regresyon analiz yöntemi ilk olarak Adrien Marie Legendr tarafından 1805 yılında ortaya konulmuştur. Regresyon analizi Basit ve Çoklu Regresyon olmak üzere ikiye ayrılır.

Regresyon denkleminde iki değişken bulunur. Y bağımlı değişken, X bağımsız değişken, i gözlem verisini temsil etmek üzere Basit Regresyon;

$$Y = \beta_0 + \beta_1 X_i + \mu_i \quad i = 1 \dots n \quad (2.4)$$

Çoklu Regresyon;

$$Y = \beta_0 + \beta_1 X_i + \dots + \beta_n X_n + \mu_i \quad i = 1 \dots n \quad (2.5)$$

Hata terimi μ_i ait olduğu verinin, belirlenen regresyon doğrusuna olan uzaklığını ifade eder. Varyans değeri ise normal dağılıma uymaktadır [2]. Regresyon analiziyle bir değişkenin başka bir değişken veya birkaç değişken karşısında gösterdiği farklı durumlar, sürekli bir fonksiyon şeklinde belirtilmektedir. Bu şekilde, ele alınan değişkenler arasında varsayılan ilişkinin varlığı, standart hatası, biçimi ve yönü saptanmaktadır.

Kısa dönem elektrik yük tahmininde tahmin edilecek elektrik yüküyle geçmiş yük bilgisi, hava durumu, mevsim ve gün tipi gibi değişkenler arasında ilişkiler kullanılarak regresyon denklemleri yazılabilir. Regresyonda, değişkenlerden biri bağımlı diğeri bağımsız değişken olmalıdır [11]. Burada tahmin edilecek yük bağımlı değişken olurken geçmiş yük bilgisi, hava durumu, mevsim ve gün tipi gibi değişkenler bağımsız değişken olarak adlandırılır.

c. Otoregresif Model: AR(p)

Otoregresif zaman serilerinde, serinin şimdiki değerleri geçmiş değerlerine ve beyaz gürültü sürecine bağlı olarak değişmektedir. Genel olarak p. dereceye sahip bir otoregresif zaman serisi

$$x_t = \delta + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \varepsilon_t \quad (2.6)$$

şeklinde yazılır. Burada δ kesme (veya sabit) terimidir ve stokastik bir süreç olan x_t 'nin ortalamasını gösterir. φ terimleri ise bilinmeyen otoregresif parametreleri temsil eder [6].

d. Hareketli Ortalama: MA(q)

Bir zaman serisi değişkeninin AR(q) modelinde gözlenen değeri onu geçmiş değeri ve rassal bir etki değeri ile ilgilidir. Başla bir deyişle q. dereceden bir hareketli ortalama süreci, her gözlenen x_t değeri için q değerine kadar uzanan artık terimlerin ağırlıklı ortalamasından oluşur. Hareketli ortalama MA(q) süreci için istatistiksel model;

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.7)$$

şeklinde ifade edilir. Rassal etki ε_t ortalaması sıfır ve sabit varyans değerine sahiptir. θ terimleri ise bilinmeyen parametrelerdir. Rassal etki ε_t bağımsız ve kolerasyonsuz olarak varsayılır [6].

e. Otoregresif Hareketli Ortalama Modeli [ARMA (p, q)]

Birçok durumda zaman serileri sadece AR(p) veya MA(q) yöntemleri ile tanımlanmaları yeterli olmaz. Bu nedenle, ifade edilemeyen seriler AR(p) ve MA(q) modellerinin birleşimi olan ARMA(p,q) modeli şeklinde ifade edilmeye çalışılır. Ayrıca zaman serisi modellerinde esneklik sağlanması için en az sayıda parametre kullanma ilkesini sağlamak amacıyla bazı durumlarda modelde hem hareketli ortalama hem de otoregresif parametrelerinin kullanılması birçok yarar sağlamaktadır [5,12].

Oluşturulacak bu modelin, zaman serisinin herhangi bir t dönemine ait olan x_t gözlem değeri, ondan önceki belirli sayıda $x_{t-1}, x_{t-2}, \dots, x_{t-p}$, gözlem değerleriyle ve $a_t, a_{t-1}, a_{t-2}, \dots, a_{t-q}$, hata terimlerinin doğrusal birleşiminden oluşmaktadır. ARMA (p, q) modelinin genel ifadesi; ($\theta_0 = 1$)

$$x_t = \delta + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \dots - \theta_q \alpha_{t-q} \quad (2.8)$$

şeklinde yazılır [12].

Eşitliğin $\varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p}$ kısmı değişkenlerin kendi gecikmeli değerlerini gösteren AR sürecini, $-\theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \dots - \theta_q \alpha_{t-q}$ parametreleri ise hata değerini gösteren MA kısmını oluşturur. δ ise kesme terimidir. Kesme terimi δ , x_t 'nin ortalaması ile ilgili iken hatalar α_t , $E(\alpha_t) = 0$ ve varyans $\text{Var}(\alpha_t) = \sigma_a^2$ ile korelasyonsuz rassal değişkenler olduğu varsayılır [6,13].

f. Durağan Olmayan Doğrusal Stokastik Modeller (ARIMA)

AR, MA, ARMA süreçlerinin durağan süreçler için uygulanabilirken, uygulamada karşılaşılan birçok seri durağan olmayan bir yapıya sahiptir. Bu durumda modelin kurulabilmesi için serinin durağanlaştırılması gerekir. Eğer incelenen yapı ortalamaya göre durağan olmayan bir durum gösteriyorsa, zaman serisini durağanlaştırmak için serinin bir veya daha fazla farkını alarak bir dönüştürme işlemi uygulanabilir. Başka bir deyişle Y_t yerine $\Delta^d Y_t$ alınarak ortalamasına göre durağan olmayan seri

modellenebilir. Bu model bütünleşik model olarak tanımlanır ve buradaki d , Y_t serisinin durağanlaştırılabilmesi için gereken fark işlemi sayısını göstermektedir.

Genel olarak süreci formüller ile ifade edecek olursak, durağan olmayan Y_t zaman serisi için $d=1$ durumu;

$$W_t = \Delta Y = Y_t - Y_{t-1} = Y'_t \quad (2.9)$$

şeklinde ifade edilir. Genel bir sonuç üretilirse;

$$W_t = \Delta^d Y_t \quad (2.10)$$

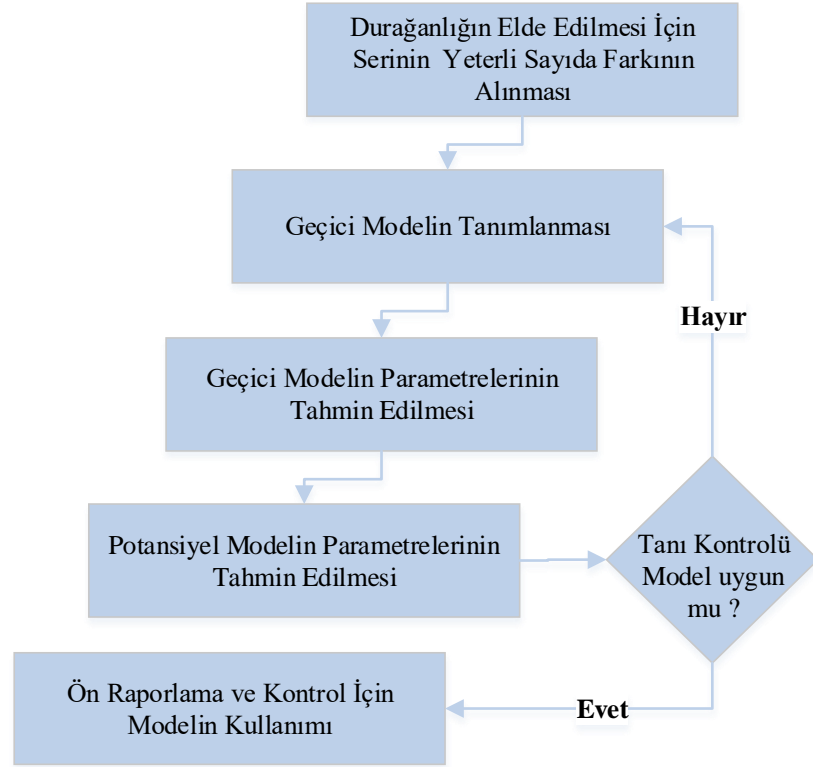
durağan bir seri ise Y_t d dereceden homojen durağan-dışıdır denilir. Durağan bir seri W_t ' nin üretilebilmesi için Y_t sürecinin farkı alındıktan sonra bir ARMA süreci gibi W_t 'yi ele almak gerekir. Eğer $W_t = \Delta^d Y_t$ olarak ifade edilebiliyorsa ve W_t bir ARMA(p,q) sürecini ifade ediyorsa Y_t (p,d,q)'uncu dereceden bir otoregresif bütünleşik hareketli ortalama süreci ya da kısaca ARIMA(p,d,q) süreci olarak ifade edilir. ARIMA(p,d,q);

$$W_t = \varphi_1 W_{t-1} + \varphi_2 W_{t-2} + \dots + \varphi_p W_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.11)$$

şeklinde gösterilir [5,6].

g. Box-Jenkins Yaklaşımı

Box-Jenkins yaklaşımı zaman serileri analizi için yaygın olarak kullanılan yöntemlerden birisidir. Yöntemin bu kadar popüler olmasının nedeni ise ele alınan herhangi bir serinin durağan olup olmaması ya da mevsimsel unsur içerip içermemesi göz önüne alınmadan bilgisayar paket programları ile çözüme kavuşturulabilmesidir. Box-Jenkins (1976) yaklaşımı zaman serileri analizinde ve önraporlama işlemlerinde genel ARIMA modelleri ile eş anlamlıdır. Aşağıdaki şekilde Box-Jenkins modelleme sürecinin genel işleyişi gösterilmiştir.



Şekil 2.1. Box-Jenkins Yaklaşımı

2.2. Bayesci Yaklaşım

Bayes teoremi 18. yüzyıl da yaşamış matematikçi ve teoloji uzmanı olan İngiliz Rahip Thomas Bayes tarafından geliştirilmiştir. Bayes metotlarının bugünkü bilinen haline gelmesi ise 19. Yüzyılda başta Laplace olmak üzere, Edgeworth, Wald, Galton ve Pearson; 20. yüzyılın başlarında Neyman ve Pearson tarafından yapılan çalışmalar ile sağlanmıştır [11].

Zamana bağlı süreçlerin Bayesci analizinin temeli optimal lineer filtrelemeye dayanır. Matematiksel olarak basit olması nedeniyle de en uygun özyinelemeli tahmin modeli ilk olarak doğrusal sistemler için geliştirilen en küçük kareler yöntemi olmuştur [14].

Bayes kuralına göre, A ve B örneklem uzayındaki iki olay olmak üzere, B olayının bilinmesi durumunda A olayının gerçekleşme olasılığı;

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0 \quad (2.12)$$

ile ifade edilir. A biliniyorken B olayının gerçekleşme olasılığı ise;

$$P(B/A) = \frac{P(B \cap A)}{P(A)}, \quad P(A) > 0 \quad (2.13)$$

şeklindedir. Bu eşitliklerden düzenlenerek eşitliğin iki tarafına konulduğunda;

$$P(A \cap B) = P(B)P(A/B) = P(A)P(B/A) \quad (2.14)$$

sonucu elde edilir. İfadeyi genelleştirmek için, içerisinde B olayının da yer aldığı aynı anda gerçekleşmesi mümkün olmayan ayrık olayların ($A_1, A_2, A_3, \dots, A_k$ örneklem uzayı olduğunu düşünürsek, B olayı biliniyorken, herhangi bir A olayının gerçekleşme olasılığı;

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}, \quad i = 1, 2, \dots, k \quad (2.15)$$

'dır. P(B) olasılık hesabı için toplam olasılık formülü uygulanırsa, bu eşitlik;

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{j=1}^k P(A_j)P(B/A_j)} \quad (2.16)$$

şeklinde ifade edilebilir [15]. Burada $P(A_i)$ değerleri önsel olasılık değerleridir ve önceden bilinmelidir.

Bayesci yaklaşım başlangıçta istatistiksel uygulamalar ile karşımıza çıksa da bununla kalmayıp, günümüzde makine öğrenmesi, yapay zekâ kavramlarının da çalışma konusu olmuş mühendislik, çevrebilim, genetik, ekonomi, üretim gibi birçok alanda kendine yer edinmiştir. Bayes teoremi bir rastlantı değişkeni için olasılık dağılımı içinde koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi gösterir [15].

Bayesci tahmin ve filtreleme metotları ise sistemi tamamen tanımlayan dinamik değişkenlerin toplanması ve ölçümlerin sisteme dahil edilmesi yöntemine dayanır. Ölçümlerdeki gürültü ise belirsizliği ifade eder. Gerçek sistem durumlarının ölçümleri elde bulunsa bile deterministik fonksiyonların durumları olmayacağından yalnızca sistemin olasılıksal durum dağılımları elde edilebilir. Sistem durumlarının zaman içerisindeki evrimi ise belli bir sistem gürültüsü tarafından etkilenen dinamik bir süreç olarak modellenir. Sistem gürültüsü, sistem dinamiğindeki belirsizliklerin modellenmesinde kullanılır. Çoğu durumda sistem gerçek stokastik değildir ancak model belirsizliklerini temsil etmek için rassallık kullanılır [14].

Klasik zaman serileri analizi ve tahmin yaklaşımları harici değişkenleri ve analiz doğrultusunda alınacak kararların kayıp/kazanç etkisini göz önüne almamaktadır. Hâlbuki kullanılan gözlem seti dışındaki harici etkiler tahmin sonuçlarına etki edebilir. Bu durum iyi bir tahmin sisteminin kurulmasını ve tutarlı sonuçlar almayı etkilemektedir [1]. Ancak bayes istatistiği her türlü belirsiz olasılığı ifade edebilme imkânına sahiptir. Bu şekilde birçok farklı çıkış noktası olan belirsiz durumların etkisi minimuma indirilerek tutarlı bir tahmin yapma olasılığı yükselir.

Bayesci tahminin temel noktası, elde edilen yeni sonuçlar ardından mevcut inancın ya da durumun değişebilmesinin mümkün olmasıdır [11]. Belirsizlik yaratan durumun matematiksel olarak ifade edilmesinin ardından elde edilen verilere olasılık kurallarının uygulanır ve olasılıksal tahminler elde edilir. Tahmin sonucu elde edilen veriler zaman serisi modelinin bir sonraki adımının değeri olabilir [1].

2.2.1. Kalman filtresi (KF)

Kalman Filtresi R.E. Kalman tarafından 1960 yılında geliştirilen doğrusal filtreleme ve tahmin (prediction) yöntemidir. Bir süreç modelinin önceki bilgileriyle birlikte giriş ve çıkış bilgilerinden sistem durumlarını belirlemek için kullanılan özyinelemeli bir tahmin tekniğine dayanır. KF, gürültülü veriler üzerinde özyinelemeli gerçek zamanlı çalışarak hataları, en az kareler eğriye sığdırma yöntemi ile filtre eder ve sistemin fiziksel karakteristiklerinin modellenmesi ile üretilen gelecek durumun matematiksel tahminine göre optimize eder. Durum-konum vektörünün ortalama ve varyansının elde edilmesi için kullanılır. Algoritmanın temelini yinelemeye dayanması sebebiyle de bilgisayar programlamasına elverişli dinamik bir sistemdir. KF' de, dinamik sisteme ait durum geçişleri lineer ve süreç ile ilgili ölçüm gürültülerinin normallik koşulları altında olduğu varsayılmıştır [1,16,17].

Kalman filtresi, parametrelerin önceden tahminini de içeren işlem hakkındaki tüm bilgiler ve bununla ilgili hata kovaryansını sağlayan bir modelle kurulur Model ait tahmin sonuçları, gözlem değerleri ile karşılaştırılır. Elde edilen fark, Kalman kazancı olarak bilinen bir değişken ile ölçeklendirilir. Daha sonra sıradaki tahminleri iyileştirmek için modele bir girdi olarak geri besleme uygulanır. Yüksek bir kazanç ile filtre gözlemleri, düşük bir kazanç ile filtre model tahminleri daha yakın olarak takip edilir. Yöntem, gerçek bilinmeyen değerlere, model tahminlerine dayanarak elde edilebilecek tahminlerden daha yakın tahminler üretmeye çalışır.

Her bir zaman adımında, Kalman Filtresi, gerçek bilinmeyen değerlerin tahminlerini belirsizlikleriyle beraber üretir. Sıradaki ölçümün sonucu gözlemlendiğinde, bu tahminler, belirsizliği düşük tahminlere daha fazla ağırlık vererek, ağırlıklı ortalama ile güncellenir [16,18].

Kalman filtresi Bayes temelli bir tahmin yöntemidir ve sonsal olasılığa Tahmin ve Güncelleme olmak üzere iki adımda öz-yinelemeli olarak yaklaşır.

Tahmin Denklemleri:

$$x_{k|k-1} = F_k \times x_{k-1|k-1} \quad (2.17)$$

$$P_{k|k-1} = F_k \times P_{k-1|k-1} \times F_k^T + Q_{k-1} \quad (2.18)$$

Güncelleme Denklemleri:

$$K_k = \times H_k^T \times \left(H_k \times P_{k|k-1} \times H_k^T + R_k \right)^{-1} \quad (2.19)$$

$$x_{k|k} = x_{k|k-1} + K_k \times \left(Z_k - H_k \times \hat{x}_{k|k-1} \right) \quad (2.20)$$

$$P_{k|k} = P_{k|k-1} - K_k \times H_k \times P_{k|k-1} \quad (2.21)$$

KF olarak uygulandığı mühendislik problemlerinin haricinde ilerleyen süreçte ekonomi ve istatistik alanında da uygulanmıştır [1]. Hedef izleme problemini olasılıksal temelde ele almış ve sinyal sorunları ile durum-konum modelleri arasındaki ilişkileri incelemiştir.

2.2.2. Monte carlo markov zinciri (MCMC)

Karmaşık problemlerde sonsal dağılımın tespit edilmesi kolay değildir. Bu tip durumların çözümü içinse yaygın olarak Monte Carlo Markov Zinciri yöntemleri kullanılmaktadır. Markov zinciri, mevcut durum, geçmiş ve gelecek durumlarının birbirinden bağımsız olduğu durumlarda stokastik bir süreçtir. Temel anlamda MCMC yaklaşımı Markov Zinciri kullanarak Monte Carlo integrasyonunun yapıldığı bir yöntemdir [19].

Monte Carlo yaklaşımı, karmaşık problemlerin integrallerinin hesaplanmasında kullanılmak üzere geliştirilmiş temeli rasgele bazı sayıların üretilmesi ve bu sayılar ile integrasyon sonucunun yaklaşık olarak bulunmasına dayanan bir yöntemdir. Monte Carlo integrasyonu ile geleneksel integrasyon yöntemlerine nazaran yüksek boyutlu

integrallerde avantaj sağlanır. Bu yöntem de amaç $\int_a^b f(x).dx$ şeklinde tanımlanan integralin çözümlenmesidir.

Çözümün ilk aşaması $\int_a^b f(x).dx$ integralinin Denklem 2.22.'de gösterildiği gibi parçalanarak ele alınmasıdır.

$$\int_a^b f(x).dx = \int_a^b g(x)p(x).dx \quad (2.22)$$

Burada $p(x)$ fonksiyonu olasılık fonksiyonu iken, $g(x)$ rastlantı değişkeninin ifade edildiği bir fonksiyondur. Temel olasılık kuramına göre eşitliğin sağ tarafında yer alan integral $E_{f(x)}[g(x)]$ 'in beklenen değerine eşit olduğu beklenebilir. Monte Carlo yöntemleriyle integral değerine yaklaşılrken, ilgili olasılık yoğunluk fonksiyonu $p(x)$ 'den bağımsız rassal örnekler alınabileceği varsayılır. Güçlü sayılar kanuna (Law of Large Numbers) göre, bilinen bir olasılık dağılımına sahip rassal değişkenler dizisinin ortalaması, Denklem 2.23.'de ifade edildiği üzere, olasılık dağılımından alınan örnek sayısı sonsuza giderken, beklenen değerine yaklaşır.

$$\frac{1}{n} \sum_{i=1}^n f(x_i) \rightarrow E[f(x)] = \int_a^b f(x)p(x).dx \quad (2.23)$$

$p(x_{0:k}|z_{1:k})$ x 'in sonsal olasılık yoğunluk fonksiyonu ve $x_{0:k}$ 'nın bir fonksiyonu olmak üzere; $x_{0:k}$ 'nın değeri Denklem 2.24.'deki gibi olacaktır.

$$E_{p(x_{0:k}|z_{1:k})}[g(x_{0:k})] = \int g(x_{0:k}) \times p(x_{0:k}|z_{1:k}) \quad (2.24)$$

Monte Carlo yöntemleri çözüme dahil edildiğinde denklemin beklenen değerine Denklem 2.25. ile yaklaşılabılır.

$$E_{p(x_{0:k}|z_{1:k})}[g(x_{0:k})] = \frac{1}{N} \sum_{i=1}^N g(x_{0:k}^{(i)}) \quad (2.25)$$

Denklem 2.25. Monte Carlo integrasyonu olarak adlandırılır. Markov Zinciri Monte Carlo yöntemleri, durağan dağılımı hedef dağılıma eşdeğer olan bir hedef dağılımından doğrudan örnekleme yapar. Böylece örneklem değerlerinden oluşturulan fonksiyonların ortalamasıyla hedef dağılımın beklenen değeri tahmin edilebilir. MCMC tabanlı yöntemlerde hedef dağılımdan örnekleme işlemini gerçekleştirecek yapı Markov Zinciri ile oluşturulurken, dağılımın integral değerine yaklaşmak için Monte Carlo integrasyonundan faydalanılır. Metropolis-Hastings ve Gibbs örnekleme yöntemleri en yaygın olarak kullanılan MCMC metotları arasındadır. Bu çalışmada Gibbs örnekleme yönteminden faydalanılmıştır [8,11].

2.2.3. GİBBS örnekleme

Amerikan fizikçi Josiah W. Gibbs'den geliştirdiği yöntem olan Gibbs örnekleme, daha sonra Geman ve Geman (1984) tarafından isimlendirilmiştir. Markov zincirleri temeline dayanan ve stokastik simülasyon uygulamalarında yaygın olarak kullanılan Gibbs örnekleme ilk kez uzaktan algılama sürecinde ortaya çıkmıştır. Geman ve Geman (1984)'nın fizik alanında görüntü analizi konusunda yaptığı bu çalışma Bayesci istatistiksel yorumlamanın uygulanabilirliğini daha da arttırmıştır. Gibbs örnekleme karmaşık Bayesian modellerinde posterior (son) dağılımları incelemek için güçlü bir iteratif metot oluşturmuştur. Genel anlamda Gibbs örnekleme algoritması, tam şartlı yoğunluk fonksiyonlarından örnekleme yaparak modelde bulunan bütün parametrelerin müşterek yoğunluk fonksiyonuna yaklaşımda bulunur.

Gibbs örnekleme, modelde bilinmeyen tüm parametreler için tam şartlı yoğunluk fonksiyonlarının (Full Conditional Densities) bilinmesini gerektirir. Tam şartlı yoğunluk fonksiyonu, modelde bütün diğer parametreler verildiğinde ilgi duyulan değişkenin yoğunluğudur. Örneğin, Gibbs örnekleme $f(a|y)$, $f(b|y)$ veya $f(a, b|y)$ 'nin dağılımlarını tahmin etmek için kullanılacaksa, bu durumda $f(a|b, y)$ ve $f(b|a, y)$ tam şartlı dağılımlarına gereksinim vardır. Gibbs örneklemesini bu olasılık

yoğunluk fonksiyonlarının herhangi birini elde etmek amacıyla kullanmak için, değişkenlerin bir tanesine rasgele bir başlangıç değeri seçilir ve daha sonra tam şartlı yoğunluk fonksiyonlarından değerler üretilir.

$$a^n \sim f(a|b^{n-1}, y) \text{ ve } b^n \sim f(b|a^{n-1}, y) \quad (2.26)$$

Burada \sim , değişkenin belirtilen dağılımın olasılığa bağlı bir değişken olduğunu belirtmektedir ve üst simge Gibbs örnekleme zincirindeki değerlerin sırasını ifade etmektedir. Eğer zincir yeteri kadar tekrarlanırsa a ve b örneklerinin dağılımı $f(a|y)$ ve $f(b|y)$ dağılımlarından olacaktır ve a, b örnek çiftleri $f(a, b|y)$ müşterek dağılımından çekilmiş olacaktır [20]. Geman ve Geman tarafından 1984 yılında ortaya konulan uzaktan algılama süreçlerine uygun Gibbs dağılımı;

$$f(x_1, \dots, x_d) \propto \left[\frac{1}{kS} E(x_1, \dots, x_d) \right] \quad (2.27)$$

şeklinde tanımlanmıştır. Burada k değeri pozitif bir sabit; S, sistem sıcaklığı; E, sistemin enerjisi olan pozitif bir fonksiyon ve x_i , sistemin i. bileşeni için ilgili olunan özelliği belirtir (i=1,2,...d).

Temel olarak bakıldığında Gibbs örnekleme yöntemi, aday noktaların kabul edilme olasılığının 1'e eşit olduğu durumda Metropolis-Hastings örneklemesinin özel bir durumunu ifade eder. Burada yapılması gereken, aday olarak kabul edilen değerleri hedef dağılıma yakınsayan bir Markov zincirinin belirlenmesidir. Gibbs örneklemeindeki önemli nokta bu yöntemin yalnızca tek değişkenli koşullu olasılık dağılımlarını ele almasıdır. Bu nedenle koşullu dağılımların simülasyonunun gerçekleştirilmesi karmaşık bileşik dağılımlardan daha kolaydır ve genellikle dağılımlar çoğu kez normal ya da diğer yaygın önsel dağılımlar gibi bilinen ve basit yapılara sahiptirler. Bu nedenle, ele alınan rastlantı değişkenlerinin bileşik olasılık dağılımından n boyutlu bir vektörü tanımlamak yerine n tane tek değişkenli koşullu olasılık dağılımından ardışık olarak rastgele seçilen n tane örnekleme yaratmak daha kolaydır [8].

Gibbs Örnekleme Algoritması;

1-) $t = 0$ alınır ve keyfi bir $\theta^{(0)} = \{\theta_1^{(0)}, \dots, \theta_k^{(0)}\}$ başlangıç değeri belirlenir.

2-) θ ' nın tüm bileşenleri

$\pi(\theta_1 | \theta_2^{(t)}, \dots, \theta_k^{(t)}, y)$ 'den $\theta_1^{(t+1)}$,

$\pi(\theta_2 | \theta_1^{(t)}, \dots, \theta_k^{(t)}, y)$ 'den $\theta_2^{(t+1)}$,

.....

$\pi(\theta_k | \theta_1^{(t)}, \dots, \theta_{k-1}^{(t)}, y)$ 'den $\theta_k^{(t+1)}$, biçiminde elde edilir.

3-) $t = t + 1$ alınır ve eğer $t < T$ (T arzu edilen örneklem genişliği) ise 2. adıma gidilir aksi durumda işlem bitirilir [21].

2.3. Durum Uzay Modelleri (State-Space Models)

1960'larda ortaya çıkan Lineer kesikli-zaman stokastik durum-uzay modelleri, hareket eden nesnelerin konumunu belirleyerek izleme ve kontrol sağlama gibi uygulamalara yönelik olarak geliştirilmiştir. Etkin bir öğrenme sürecini bünyesinde barındırır. Öğrenilen bilgilerin daha önceki bilgilere uyumlu olarak eklenmesini gerektirir [1,22].

Durum-uzay modelinin genel yapısı ve Kalman filtreleme algoritması beraber kullanılabilmesi nedeniyle zaman serileri analizinde önemli bir rol oynar. Ayrıca zaman Serileri analizi için oldukça zengin bir model içeriğine sahiptir. Bu modeller istatistik alanı başta olmak üzere, ekonomi, sinyal işleme ve mühendislik çalışmaları gibi birçok alanda kullanılmaktadır.

Bir durum uzay modeli iki denklem ifade edilir. Birinci denklem gözlem denklemi olarak isimlendirilir. Y_t gözlenebilen ($q \times 1$) boyutlu gözlem vektörü, F_t ($n \times m$) boyutlu gözlenebilen bağımsız (dışsal) değişkenler matrisi ve v_t ($m \times 1$) boyutlu gözlem gürültüsü olmak üzere gözlem denklemi;

$$Y_t = F_t \theta_t + v_t \quad , \quad v_t \sim N(0, V_t), \quad t = 1, 2, \dots, \dots, \quad (2.28)$$

şeklinde ifade edilir.

İkinci denklem durum denklemi ya da evrim denklemi olarak nitelendirilir. θ_t gözlenemeyen ($n \times 1$) boyutlu durum vektörü, G_t ise parametrelerin güncellenmesini tanımlayan bilinen ($n \times n$) boyutlu bir matris ve w_t hata terimi olmak üzere durum denklemi;

$$\theta_t = G_t \theta_{t-1} + w_t \quad , \quad w_t \sim N(0, W_t), \quad t = 1, 2, \dots, \dots, \quad (2.29)$$

w_t ve v_t ise birbirinden bağımsız, serisel ilişkisi bulunmayan, sıfır ortalamalı, sırasıyla W_t ve V_t varyans matrisli hata ve sapma terimleridir. R_t bu durumda seçim matrisi olarak adlandırılır. Değeri sıfır olmayan hata terimini belirtir [8,23,27].

Birinci dereceden (local level) durum-uzay modeli 2.28 ve 2.29 denklemlerinin en basit halidir. Tek değişkenli model olarak bilinir ve;

$$\theta_t = \mu_t, R_t = G_t = F_t = 1 \quad (2.30)$$

$$Y_t = \mu_t + v_t \quad , \quad v_t \sim N(0, V_t), \quad t = 1, 2, \dots, \dots, \quad (2.31)$$

$$\mu_{t+1} = \mu_t + w_t \quad , \quad w_t \sim N(0, W_t), \quad t = 1, 2, \dots, \dots, \quad (2.32)$$

şeklinde tanımlanır [23].

Durum uzay modelleri 3 temel varsayımda bulunur;

- a. Durum-uzay modellerinde durum vektörü, gözlem değişkeni için gerekli olan tüm bilgiyi taşır,
- b. Durum vektörünün başlangıç noktasındaki değeri θ_0 , m_0 ortalama vektörü ve C_0 varyans matrisine sahiptir,

- c. Hata terimleri serinin her anında birbirleriyle ilişkisiz olup, θ_0 başlangıç durum vektörü değişkenleri de w_t ve v_t değişkenlerinden bağımsızdır [8,27].

Herhangi bir istatistiksel modelde θ_t , gözlem değerlerinin zamansal gelişimine bağlıdır. Durum-uzay modelindeki amaç, θ_t 'yi sistem hakkında bilgilendirecek minimum sayıda öge oluşturmaktır. Herhangi bir sürecin modellenmesine ilişkin uygun model sayısı F_t , V_t , G_t , R_t ve W_t matrislerinin tanımlanma durumlarına bağlı olarak artabilmektedir. G_t , R_t ve W_t matrisleri stokastik olmayan matrislerdir ve sistem matrisleri olarak bilinirler. Ayrıca Aoki 'ye göre durum konum modellerinin kararlılık, kontrol ve gözlenebilirlik özelliklerini belirlerler.

G_t matrisi sistem denklemlerinde kararlılık özelliğini tanımlar. Güncelleme ve gözlem denklemleri ile tanımlanmış sistemin asimptotik kararlılığı, G_t sistem matrisinin özdeğer değerlerinin birden küçük olması ile belirlenir.

Durum-uzay modellerinde tahmin, düzleştirme ve parametrelerin zaman serisi noktalarındaki güncelleme işlemleri kalman filtresi tarafından sağlanmaktadır [8].

2.3.1. Dinamik lineer model

Durum uzay modellerinin önemli model sınıflarından birisi Gaussian Doğrusal durum uzay modelleridir. Aynı zamanda Dinamik Lineer modeller olarak da adlandırılır. West and Harrison (1997), Harvey (1989), Durbin and Koopman (2001) Kalman filtrelerine dayalı Bayesci durum-uzay modelleri üzerine çalışmalar yapmış ve dinamik lineer modellerinin kavram tasarımında öncü olmuşlardır [19].

Dinamik model sınıfını temel alan zaman serisi süreçlerinin matematiksel ve istatistiksel olarak modellenmesi zamana bağlı olarak değişiklik gösteren dinamik etkenleri temel alır. Dinamik terimi zamanın akışına bağlı olarak süreçte de değişim olduğunu ifade eder [27].

DLM'ler tanımlandıklarında aşağıdaki özelliklere sahip olur;

- Parametrik bir model yapısı vardır,
- Zaman serisinin her t noktası için parametreleri ve gözlemlere ilişkin olasılık bilgileri vardır,
- Parametrelerin zaman serisi süresince değişimini açıklayan ardışık bir model yapısına sahiptir,
- Öngörü değerleri olasılık dağılımları olarak elde edilir.

DLM'de kullanılan Bayesci yaklaşım ile Dinamik olmayan modellerde kullanılan Bayesci yaklaşım temelde aynıdır [8,24]. DLM'nin yapısı ve işleyişi detaylı olarak incelemesi dördüncü bölüm altında incelenecektir.

BÖLÜM 3. LİTERATÜR ARAŞTIRMASI

Zaman serileri, eski zamanlardan beri doğa bilimleri ve insan hayatını etkileyecek birçok konuda önemli rol oynamaktadır. Zaman serileri analizi ile değişkenlerin gözlemler üzerindeki etkisinin belirlenmesi ve gelecekle ilgili daha iyi tahmin için bu değişkenlerin kullanılması esastır. Örneğin Babil’li gökbilimcilerin astronomik olayları takip ve tahmin etmek için gezegenler ve yıldızların göreceli pozisyonlarında zaman serilerini kullandığı bilinmektedir [25].

Babillerin geçmişte astronomi verilerini zamanla ilişkilendirdiği bu yaklaşım, Charles Babbage ve Stanley Jevons isimli ekonomistler tarafından ele alınmış ve Warren M. Persons (1915) tarafından geliştirilmiştir. Warren P. bir zaman serisinin trend, mevsimsel, konjonktürel ve rassal bileşenlerden oluştuğunu belirtmiştir. Daha Sonra Rus istatistikçi Evgenij E. Slutsky ve bir İngiliz olan George U. Yule tarafından otoregresif süreçler ve hareketli ortalama metodunu geliştirmiş, Herman Wold (1938) da çalışmalarından bu yaklaşımları sistematize ederek genel yapılarını ortaya çıkarılmıştır.

1970’lere gelindiğinde zaman serilerinin istatistiksel analizi üzerindeki çalışmaların giderek artması ile beraber, klasik zaman serisi analiz yöntemleri terk edilerek matematiksel istatistik ve olasılık teorisinin yöntemleri kullanılmaya başlanmıştır. Bu durum zaman serileri için stokastik hareketlerin rolünün farklı olarak ele alınmasına yol açmıştır. Modern zaman serisi yaklaşımları, klasik zaman serisi yöntemlerinden farklı olarak zaman serilerinin bütün bileşenlerinin üzerinde stokastik bir etkinin olduğunu varsayar. Bilgisayar ve bilgi sistemlerinin gelişmesi ile yeni istatistiksel prosedürlerin geliştirilmesi mümkün olmuş ve zaman serileri yöntemlerinin uygulanmasında avantaj sağlayacak farklı yaklaşımlar ele alınmıştır.

Bu yaklaşımların yaygın olarak kullanılması ise ampirik çalışmalarıyla George E.P. Box ve Gwilym M. Jenkins (1970) tarafından ortaya konulan çalışma ile olmuştur. Box ve Jenkins “Zaman serilerinin tüm üretim süreçleri için mevcut bir stokastik model vardır” varsayımında bulunarak farklı bileşenler fikrinden vazgeçmiş zaman seri süreci için ortak bir stokastik model olduğunu belirtmişlerdir.

Box ve Jenkins’in yönteminde ilk olarak, belirli istatistiksel rakamlara dayanarak belirli bir modeli tanımlar. İkinci adımda, bu modelin parametreleri tahmin edilir. Üçüncü olarak, modelin özellikleri istatistiksel testlerle kontrol edilir. Model hataları belirginleşirse model değiştirilmeli ve parametreler yeniden tahmin edilmelidir. Bu prosedür, verilen kriterleri karşılayan bir model üretene kadar tekrarlanır. Bu model nihayet tahminler için kullanılabilir [25].

Zaman serilerinin modellenmesinde kullanılan klasik stokastik süreçler kuramına dayalı bu yöntemler 1970 yıllarına kadar yoğun olarak kullanılmıştır. Günümüzde kullanılan yöntemler ise Box ve Jenkins tarafından ortaya konulan yöntemlerdir. Temel olarak Wiener ve Kolmogorov teoremi sonuçlarına dayandırılan model çözümlenmesi, geçmiş gözlemler değerlerinin doğrusal fonksiyonlarından çıkarılan durağanlık varsayımı ile optimizasyon için ortalama kare hata kriterine (mean square error criterion) bağlı kılınmaktadır.

Paralel süreçte istatistiksel yöntemler ile tahmin çalışmaları devam etmiş Wiener ve Kolmogorov’un ortaya koyduğu kestirim teoremi ile akışlı süreçlerin tahmin edilmesi amaçlanmıştır. Wiener ve Kolmogorov’un teoremi üzerinde Kalman ve Busy [28] yaptıkları çalışmada durum-konum modellerini kullanarak dinamik lineer modellerin kestirimi için akışlı bir filtre düzenlemişlerdir [8].

Dinamik lineer modeller, 1960’lı yılların başında, dinamik sistemleri izlemek ve kontrol etmek için geliştirildi ancak istatistiksel literatürde öncü çalışmalar Thiele (1880)’e kadar uzanır. Dinamik lineer modellerinin daha geniş olarak durum-konum modellerinin gerçekleştirilen ilk ünlü uygulamaları APOLLO ve POLARIS havacılık ve uzay programları olsa da günümüzde biyolojiden ekonomiye, mühendislik ve kalite

kontrolünde, iklim ve çevre çalışmalarına, jeofizik biliminden genetiğe kadar geniş yelpazede kendine uygulama alanı bulmaktadır. Uygulamaların bu derece geniş bir alanda gerçekleştirilmesi Bayes çerçevesindeki modern monte Carlo yöntemlerinin kullanılarak hesaplama zorluklarını çözmeye imkanından kaynaklanmaktadır.

Zaman serilerinin modellenmesinde kullanılan bir diğer yöntem de Harrison ve Stevens'in dinamik lineer modeller üzerinde geliştirdikleri yaklaşımdır. Bu yaklaşım, bir değişkenin yerel düzeyinin, değişme oranının, değişkenin seviye atlayıp atlamadığını ya da süreksiz olduğunu belirlemek için Kalman filtreleme yöntemini kullanır. Kalman filtrelerinin Bayesci bir yorumunu içeren dinamik doğrusal modeller ise üstünde en çok çalışılan durum konum modelidir.

West and Harrison (1997), Harvey (1989), Durbin ve Koopman (2001), Rob J. Hyndman (2008), Petris G. (2009)'ın yapmış olduğu çalışmalar zaman serilerinin dinamik lineer modeller olarak ele alınması ve tahmin uygulamalarının gerçekleştirilmesine öncülük etmiştir [19]. Ayrıca Kadir Karagöz (2004) ve Hatice Y. Gürkan (2013)'ın yapmış olduğu çalışmalar zaman serilerinin dinamik lineer modelleri olarak ele alınmasına örnek olarak verilebilir. Bunlardan bazıları Tablo 3.1'de listelenmiştir.

Tablo 3.1. Durum-Konum Modeli ve DLM Alanında Yapılan Bazı Çalışmalar

Yıl	Yazarlar	Çalışma İsmi
1984	Mike West vd.	Dynamic Generalized Linear Models and Bayesian Forecasting
1994	Chang-JinKim	Dynamic linear models with Markov-switching
2004	Kadir Karagöz	Öngörü ve Zaman Serilerinde Bayesyen Yaklaşım
2011	Xiang Fei vd.	A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction
2013	Hatice Y. Gürkan	Markov Zinciri Monte Carlo Yönteminin Dinamik Doğrusal Modellere Uygulanması

Tablo 3.1. (Devam)

2015	Rob J Hyndman vd.	Monash Electricity Forecasting Model
2015	Prisca Primavera Piccoli	Identification of a dynamic linear model for the American GDP
2016	Jayhan Regner vd.	A Dynamic Approach to Forecasting Long-Term Electricity Demand in the Philippines Using Kalman Filtering Algorithm
2016	Marisol Valencia Cárdenas	Dynamic Model for the Multiproduct Inventory Optimization with Multivariate demand.
2018	Kobi Nagbe vd.	Short-Term Electricity Demand Forecasting Using a Functional State Space Model

Dinamik lineer modeller, klasik zaman serisi modellemeden farklı olarak hem metodolojik hem de hesaplama açısından avantajlara sahiptir. Bunlardan ilki sistemi deterministik olarak değil stokastik olarak ele almasıdır; açıklayacak olursak zaman serisine etki eden ve model içerisine dahil edilmeyen faktörleri, ölçüm hatalarını ve belirsizlikleri mevcut bilgiyi göz önünde bulundurarak şartlı olasılık dağılımlarını hesaplayarak sisteme dahil eder. Bu durum bayesci çıkarımın temelini oluşturur.

Zaman serilerininin ARMA modelleri ve Box-Jenkins modelleme biçimi ile incelenmesi ilk başta durum-konum modellerine göre daha kolay görünse de durum-konum modelleri ve dinamik lineer modelleme yönteminin sunmuş olduğu güçlü çerçeve yapısı, daha fazla esneklik ve daha kolay yorumlama imkânı tanır. Dinamik lineer modellerin en büyük avantajlarından birisi normal dağılıma uymayan ve doğrusal olmayan sistemlere kolaylıkla uygulanabilmesidir.

Dinamik lineer modelleme yöntemi özellikle akışlı sistemlerde büyük avantaj sağlar, verilerin art arda geldiği ve çevrimiçi çıkarım gereken durumlar buna örnektir. Bu durum depolama kapasitesinin azaltılması gereken büyük veri işleyen sistemlerde daha önemli hale gelir ve dinamik lineer modellemenin Bayesci özyinelemeli yapısı sayesinde aşılabilir bir problemdir [19].

Yukarıda belirtilen kavramlar dahilinde bu çalışmada Bayesci bir çıkarım metodunu bünyesinde barındıran durum-uzay modellerinin, dinamik lineer modelleme yöntemi incelenmiş ve uygulaması yapılarak diğer modeller ile karşılaştırılmalı olarak sunulmuştur.

BÖLÜM 4. DLM İLE ELEKTRİK TÜKETİM TAHMİNİ

4.1. Problemin Tanımı

Bu bölümde, bu tez çalışmasının temel konusu olan Dinamik Lineer Modeller ile zaman serileri analizi ve tahmininin, yoğun bir şekilde kullanımı olan ARIMA metotlarından farklı olarak, zaman serisinin durağanlığı önemsenmeksizin gerçekleştirilmesi probleminin farklı çözüm yöntemi uygulanmıştır.

Ele alınan konuda SEDAŞ aracılığıyla alınan, Sakarya iline ait elektrik tüketim verilerinin zaman serisi şeklinde ifade edilmesi ve Bayesci bir yöntem olan Dinamik Lineer Modeller ile modellenerek gelecek değerinin tahmini için data tutarlı çözüm sunabilmek ve elektrik tüketimine etki eden faktörlerin etki miktarlarını belirleyebilmektir. Bu amaçla ilk olarak eldeki verinin günümüz şartlarına uygun zaman serisi dönüşümleri sağlanmıştır. Günümüzde elektrik verileri saatlik tüketimler olarak tutulsa da belirli saat dilimlerine ayrılarak hesaplamalar gerçekleştirilmektedir.

Genel olarak gerçek yaşama dair bir zaman serisi durağan değildir. Durağan olması durumunda zaman serisi olası bütün etkiler modellenerek gelecek değerler tahmin edilebilir. Ancak bir zaman serisi trendsel, sezonsal ve konjonktürel etkilerden bunun yanında modellenmesi konusunda en çok zorluk çekilen stokastik etkilerden etkilenir. Belirgin olmayan etkilerden dolayı öngörülemeyen tahmin hataları ortaya çıkmakta ve verilen kararları etkilemektedir.

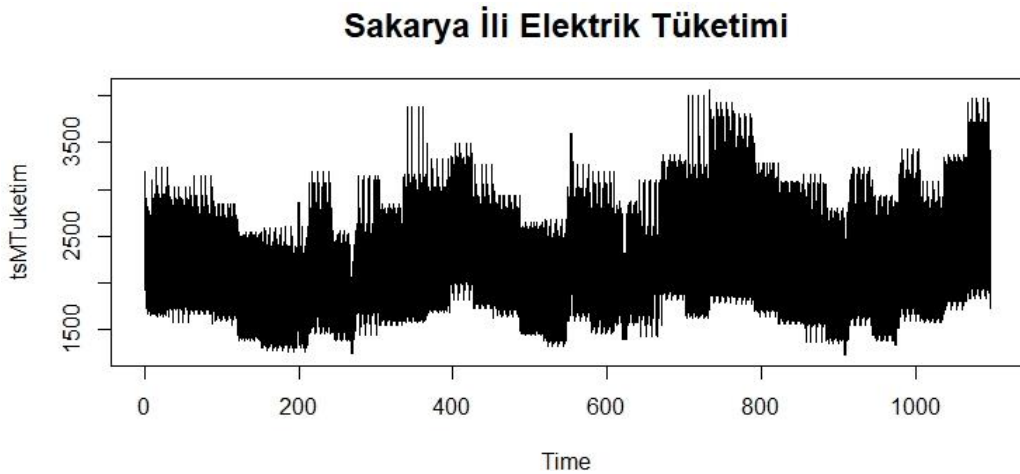
Dinamik lineer modellerin, bir zaman serisinde görülen ya da görülemeyen etkileri, Bayes olasılığına dayanan yapısı sayesinde hesaplamalara dahil ederek gerçekleştirilmesi fikrine dayanarak gelecek değer tahmini açısından daha tutarlı sonuç verebileceği düşünülmüştür.

4.1.1. Veri seti

Yapılan uygulamalarda kullanılan veri seti SEDAŞ aracılığı ile elde edilmiştir. Elde edilen veride Sakarya ili için gerçekleşen 3 yıllık (2015-2017) tüketim verileri bulunmaktadır. Bu veriler içerisinde MESKEN tüketim verileri kullanılmıştır. İlk olarak saatlik bazda tüketim şeklinde düzenlenmiş olan veri, günümüzdeki elektrik tüketim dilimleri olan ve daha önemli olduğu düşünülen T1, T2 ve T3 zaman dilimlerine uygun bir biçimde düzenlenmiştir. Verinin düzenlenmesinden sonra 3288 adet veri oluşmuştur.

T1 zaman dilimi 06:00 ile 17:00 saatleri arası tüketimi, T2 zaman dilimi 17:00 ile 22:00 saatleri arası tüketimi ve son olarak T3 zaman dilimi 22:00 ile sonraki gün 06:00 saatleri arası tüketimi ölçmektedir.

T1, T2, T3 zaman dilimleri şeklinde düzenlenen verinin, modele uygunluğu açısından zaman serisine çevrim işleminin ardından veri dinamik lineer model aracılığı ile analiz edilmiştir. Zaman serisine dönüşümü sağlanan 3288 adet veri 1096 adet zaman dilimi oluşturmuştur. 01.01.2015 tarihi, verinin görselliği açısından grafiklerde “0” başlangıç noktası olarak kabul edilmiştir. Oluşturulan 1096 adet sezondan 1066 adet veri eğitim verisi, kalan kısım (2017 Aralık) test verisi olarak ayrılmıştır. Şekil 4.3'te Sakarya ili hanehalkının 2015-2017 yılları için elektrik tüketim verileri gösterilmiştir.



Şekil 4.1. Sakarya İli Mesken Elektrik Tüketim Verisi

4.2. Dinamik Lineer Modeller ile Zaman Serisinin Modellenmesi

Regresyon modellerinin yanı sıra ARIMA gibi zaman serisi modelleri, parametreleri gözlem süreci boyunca değişmeyen ve sabit sayılar kümesi olarak ele alan statik modellerdir. Statik modellerde parametreler arasındaki sayısal ilişkinin süreç boyunca aynı olduğu varsayılır. Ancak çoğu durumda parametreler arasındaki ilişki sadece zamana bağlı olarak dahi değişebilir.

Parametrelerin arasındaki ilişkinin tam olarak belirlenebilmesi için süreçlerin ya da sistemlerin zaman içerisinde ortaya çıkan belirsizliği göz önüne alması gerekir. Hatta değişime bağlı olarak sürecin ilerleyen dönemlerinde değişkenlerin farklılaşmasından dolayı modelinde biçimini değiştirmek gerekebilir. Bu gibi durumlar denklem takımları dizisi olarak adlandırılan dinamik modelleri gerekli kılmaktadır.

Dinamik lineer modeller ilk olarak Harrison ve Stevens (1976) tarafından tanıtılmış, West ve Harrison (1997) çalışmaları ile ise detaylı olarak ele alınmıştır. Doğrusal olmayan modellere dinamik bir açılım sağlayan dinamik Bayes modelleri ise West, Harrison ve Migon (1985)'de incelenmiştir [1].

Bir DLM durum ve sistem denklemi olmak üzere iki eşitlikle ifade edilir. Sistem denklemi;

$$Y_t = F_t \theta_t + v_t \quad , \quad v_t \sim N(0, V_t), \quad t = 1, 2, \dots \dots, \quad (4.1)$$

şeklindedir ve durum denklemi ise;

$$\theta_t = G_t \theta_{t-1} + w_t \quad , \quad w_t \sim N(0, W_t), \quad t = 1, 2, \dots \dots, \quad (4.2)$$

olarak yazılır. Burada v_t , w_t birbirinden bağımsız normal dağılımlı, sıfır ortalama ve bilinen bir varyans matrisidir [8,19].

DLM'yi oluşturan denklem elemanlarını ele alacak olursak. Normal bir DLM' de denklem elemanları:

- F_t : (mxp) boyutlu t zamanında bilinen bağımlı değişkenler matrisi,
- v_t : (mx1) boyutlu gözlem hatası matrisi,
- θ_t : (px1) boyutlu süreç parametre vektörü,
- Y_t : (mx1) boyutlu süreç gözlem vektörü,
- w_t : (pxp) boyutlu bilinen sistem hatası matrisi,
- G_t : (pxp) boyutlu bilinen sabit geçiş(evrime) matrisi
- m_0 : Gözlemlerin ortalama değeri
- C_0 : Gözlemlerin varyans değeri

şeklinde nitelendirilir ve F_t bağımsız değişkenler matrisi, G_t evrim matrisi ve W_t sistem evrim hatası matrisi belirli şartlar altında model kurucu tarafından belirlenmektedir. Gözlem varyansı V_t ise çoğu zaman bilinmemektedir ve W_t göre daha büyüktür.

İlk denklem gözlem denklemi olarak isimlendirilir ve θ_t durumları için Y_t ' nin örneklem dağılımını tanımlamaktadır. Y_t terimi önceki gözlem ve parametre değerlerinden bağımsızdır. θ_t ve Y_t arasında dinamik ilişki mevcuttur. Gözlem denklemi hatası v_t başlangıç durumu için bilinmekle beraber zamanla değişme olasılığına sahiptir.

İkinci denklem evrim (sistem, durum, geçiş) denklemi olarak isimlendirilmektedir. Durum vektörü θ_t ' nin zaman içindeki gelişimini tanımlamaktadır. θ_{t-1} , G_t ve W_t sistemin bilinen girdi verileri olurken, D_t içinde bulunulan zamanda sahip olunan bilgi setini ifade etmek üzere θ_t , D_{t-1} 'den bağımsızdır. G_t , sistemin deterministik geçiş matrisidir.

Dinamik lineer modeller yeni veriler elde edildikçe mevcut bilginin güncellenmesi ilkesine dayanır ve bu durum Bayes yaklaşımı açısından elverişli bir özelliktir. Bayes yaklaşımı sayesinde evrim ve gözlem denklemlerinin olasılık yoğunluk

fonksiyonlarını birleştirmek mümkün olmaktadır. Birleştirilen olasılık yoğunluk fonksiyonundan birincil (θ_t) ve ikincil parametrelerin (G_t, W_t) son dağılımları elde edilebilir [1].

Gelecek zamanda bir noktanın tahmini yapılırken temel hedef, $(Y_1|D_0)$ için tahmin dağılımını hesaplamaktır. Bu nedenle Y_{t+1} tahmin edilirken D_t veri seti ve $(Y_{t+1}|D_t)$ tahmin dağılımı kullanılmaktadır. DLM' de zaman ilerledikçe eldeki bilgiler de değişecektir. DLM için matris yapıları şu şekildedir;

$$F_t = \text{diag}(F_t^{(1)}, \dots, F_t^{(n)}) \quad , \quad V_t = \text{diag}(V_t^{(1)}, \dots, V_t^{(n)}),$$

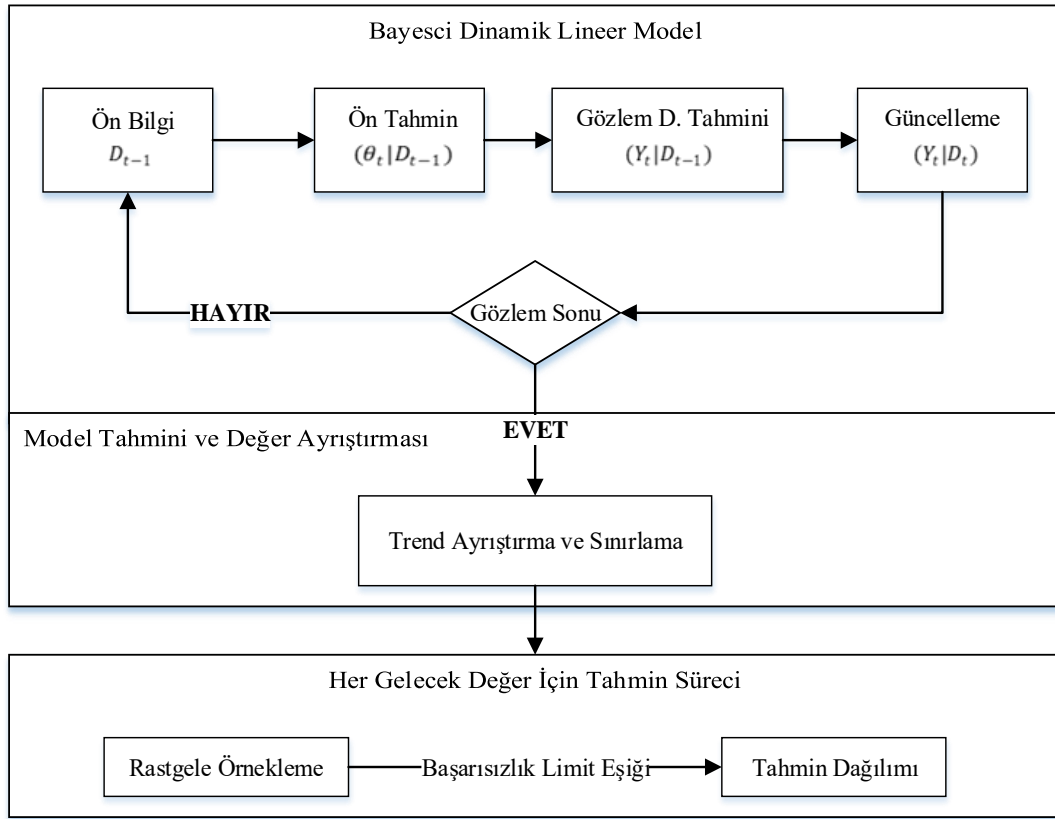
$$G_t = \text{diag}(G_t^{(1)}, \dots, G_t^{(n)}) \quad , \quad W_t = \text{diag}(W_t^{(1)}, \dots, W_t^{(n)}),$$

$$m_0 = (m_0^{(1)}, \dots, m_0^{(n)}) \quad , \quad C_0 = \text{diag}(C_0^{(1)}, \dots, C_0^{(n)})$$

Belirtilen matris yapılarının yanında trend, sezonsal faktör gibi etkilerin toplamı şeklinde modellenmiş DLM'ler için tanımlanacak matrisler F_t ve V_t ;

$$F_t = \left(F_t^{(1)} \mid \dots \mid F_t^{(n)} \right) \quad \text{ve} \quad V_t = \sum_{i=1}^n V_t^{(i)} \quad \text{şeklinde olmalıdır.}$$

Şekil 4.2.' de dinamik lineer modellerin bilgi öğrenme ve tahmin süreci akış şeklinde anlatılmıştır. Şekilde görüldüğü gibi DLM için öğrenme sürecinin kalıpları kurulan model ile belirlenmesi gerekir. Bu noktada öğrenme sürecine geçmeden önce DLM'nin yapısı belirlenmelidir. DLM'ler bünyesinde stokastik trend, sezonsal faktör ve ARMA süreçlerini barındırabilirler. Bir sonraki başlıkta bu süreçlerin modellenmesi ile ilgili bilgi verilmiş ve bu çalışmada kullanılan model ve matrisler açıklanmıştır.



Şekil 4.2. Dinamik Lineer Modeller ile Tahmin Süreci

Verilen bir DLM’de temel amaç gözlenmemiş durumlara ilişkin tahmin yapmak ve gözlemlerin bir kısmına dayanarak gelecekteki gözlemleri tahmin etmektir. Değerlendirme ve tahmin, mevcut bilgiler dikkate alınarak ilgilenilen kısmın şartlı olasılık yoğunluk dağılımları bulunarak çözülür.

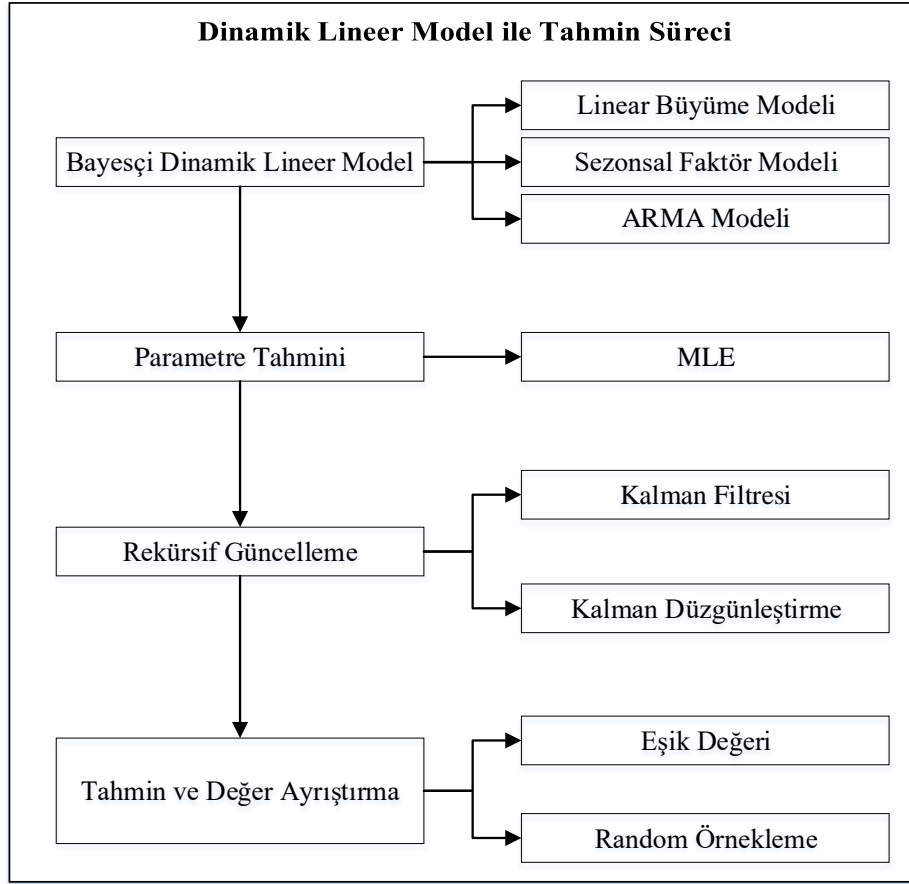
Şekil 4.3. de ise DLM ile tahmin aşamaları belirtilmiştir. Şekil incelendiğinde ilk olarak zaman serisi dinamik lineer model aracılığı ile modellenir. Bu işlemden önce zaman serisinin davranışları incelenmeli ve zaman serisinin içerdiği bileşenlere uygun olarak bir model kurulmalıdır.

Bir sonraki adımda bilinmeyen parametrelerin tahmini yer alır. Burada sistem ve gözlem hata matrislerinin değeri kurulan dinamik lineer modele göre MLE fonksiyonundan faydalanılır (bkz. 4.2.). Bilinmeyen parametrelerin tahmin edilmesinin ardından kurulan modele göre Kalman filtresinin özyinelemeli yapısından faydalanılarak zaman serisinin ilk değerinden gelecek durumuna kadar tahmin süreci

işletilir. Kalman filtresi durum vektörü üzerindeki mevcut çıkarımımızı yeni veriler ile birlikte güncellememizi sağlar ($\pi(\theta_n|y_{1:t}) \rightarrow \pi(\theta_n|y_{1:t+1})$). Elde edilen değerler yine Kalman düzleştiricisi aracılığı düzleştirilir. Bu noktada trend ve sezonsal faktör bileşenlerini içeren tahminler elde edilmiş olur.

Durumların tahmini, gelecek değerlerin tahmini için sadece bir adımdır. Buraya kadar olan süreçte ileriye doğru tahminlerin şartlı olasılıkları $\pi(\theta_n|y_{1:t})$ elde edilir. Diğer bir nokta da eğer filtreleme problemi çözümü için şartlı olasılık hesaplanıyorsa bu durumda ($n = t$), durum tahmini yapılıyorsa ($n > t$) ve son olarak eğer düzleştirme için şartlı olasılık hesaplanıyorsa ($n < t$) olarak değerlendirilir.

Gelecek durumların tahmini ise elde edilen değerlerden Bayesci yöntemler ile modele uygun örneklemeler ile ortalama ve varyans değerleri elde edilir. Bir adım ileriye yönelik tahmin için, bir başka deyişle, $Y_{1:t}$ verilerine dayanarak bir sonraki gözlemin $Y_{1:t+1}$ 'i tahmin etmek için, ilk önce durum vektörünün $\theta_{1:t}$ değeri tahmin edilir ve sonra bu tahmine dayanarak hesaplanır. Bir adım ilerideki durumun prediktif yoğunluğu $\pi(\theta_{t+1}|y_{1:t})$ ve θ_t filtreleme yoğunluğuna dayanır. Bundan bir adım ileri doğru tahmin yoğunluğu $\pi(y_{t+1}|y_{1:t})$ elde edilir. Gelecek değerlerin tahmini için durum vektörünün θ_{t+k} ($k>1$) ve gözlemlerin Y_{t+k} değerini bulmak gerekir. Durum tahmini problemi k adım ilerideki durum tahmini yoğunluğu $\pi(\theta_{t+k}|y_{1:t})$ hesaplanarak çözülür. Bu yoğunluğa bakarak t+k adım ilerideki gözlemler için öngörülen yoğunluk $\pi(y_{t+k}|y_{1:t})$ bulunabilir. Bir DLM için durumlar ve gözlemler için bir adım ilerideki tahmin dağılımları Kalman filtresi aracılığıyla, Kalman filtresinin bir yan ürünü olarak elde edilir.



Şekil 4.3. Dinamik Lineer Model ile Tahmin Aşamaları

4.3. Model Seçimi

DLM ile tahmin sürecinde model seçimi, tahmin adımlarının yüksek doğruluk yüzdesi ile gerçekleştirebilmek için büyük rol oynamaktadır. Bu kapsam da kurulacak modelin ele alınan zaman serisinin karakteristiklerini yansıtması gerekir.

Bir zaman serisi artan ya da azalan bir trende, sezonsal olarak artan ya da azalan etkilere sahip olabilir. Bunların yanında konjonktürel ve rassal etkiler zaman serilerinde bulunabilir. Tüm bu etkiler göz önüne alındığında kurulacak DLM modelinin bu etkileri de ayrıştırmalıdır.

DLM'lerin trend etkisini ve sezonsal etkileri ayrıştıran çeşitleri olduğu gibi kurucuya özel varyasyonları olabilir. Bu başlıkta trend etkisini, sezonsal etkiyi ve ARMA

süreçlerini işleten DLM modelleri anlatılmıştır, anlatılan modeller modelleyicinin elindeki temel yapı taşlarıdır ve veri setinin analizi için uygun bir biçimde bir araya getirilmesi gerekir. Bölüm sonunda ise tahmin için kullanılan model matrisleri ile birlikte açıklanmıştır.

4.3.1. Trend modeli

Bir zaman serisinin trendi, zaman içindeki pürüzsüz uzun vadeli gelişimini gösterir. Tren bileşeni n. dereceden polinomal süreç olarak modellenebilir ve zaman serisinde trend bileşeninin açık olarak görülebildiği DLM modellerinde yaygın olarak Polinomal DLM kullanılır [19].

Polinom derecesi “n” ne kadar büyük olursa trende olan yaklaşım o kadar yakın olur. En yaygın kullanılan polinom modelleri rastgele yürüyüş modelleri (n=1) ve doğrusal büyüme modelleridir(n=2).

“n.” dereceden polinomal model n boyutlu durum uzay modeli ile ifade edilir. “n.” dereceden polinomal süreç için DLM matrisleri;

$$F = (1, 0, \dots, 0) \quad (4.3)$$

$$G = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 1 \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix} \quad (4.4)$$

$$W = \text{diag}(W_1, \dots, W_n) \quad (4.5)$$

şeklinde tanımlanır. Bu modeli oluşturan denklemler aşağıdaki gibidir;

$$Y_t = \theta_{t,1} + v_t \quad (4.6)$$

$$\theta_{t,j} = \theta_{t-1,j} + \theta_{t-1,j+1} + w_{t,j} \quad (4.7)$$

$$\theta_{t,n} = \theta_{t-1,n} + w_{t,n} \quad (4.8)$$

Burada durum vektörünün ilk elemanı $\theta_{t,1}$ herhangi bir t anındaki serinin ortalamasını gösterir ve $j = 2, \dots, n$ için durum vektörünün j . elemanın sonraki zaman aralığı süresince $j-1$. elemanın değişimini gösterir.

4.3.2. Sezonsal model ve mevsimsellik gösterimi

Mevsimsellik etkisi gösteren bir serinin modellenmesi için iki farklı DLM yöntemi vardır. Birincisi sezonsal faktör modeli ve Fourier formda sezonsal model. Sezonsal modeller incelenirken trend bileşeninin bulunmadığı ve serinin ortalamasının 0 olmadığı (durağan olmadığı) varsayılmıştır.

Yıllık olarak sezonsal davranış sergileyen bir model için üç aylık verilerin olduğu bir durum incelendiğinde serileri ortalamanın mevsimsel sapmalarını tanıtarak tanımlayabiliriz. Her bir seri $i=1, \dots, 4$ için a_i ile ifade edilirsin. $t - 1$ birinci çeyrek, t 'nin ikinci çeyrek olarak varsayıldığı modelin tanım denklemleri;

$$Y_{t-1} = a_1 + v_{t-1} \quad (4.9)$$

$$Y_t = a_2 + v_t \quad (4.10)$$

Tanımlanan formüller için $t - 1$ zamanındaki durum $\theta_{t-1} = (a_1, a_4, a_3, a_2)'$ ve F_t matrisi her t anı için $F = (1, 0, 0, 0)$ olarak ifade edilir. Bu ifadeleri DLM formuna göre formüle edecek olursak DLM denklemi;

$$Y_t = F\theta_t + v_t \quad (4.11)$$

olur. Buradaki amaç durum denkleminin deęişimini ifade eden $\theta_t = (a_2, a_1, a_4, a_3)'$ permütasyon matrisini tanımlamaktır. Bu nedenle yılın ikinci çeyreęi için sistem denklemini için;

$$Y_t = F\theta_t + v_t = a_2 + v_t \quad (4.12)$$

eşitlięi sağlanır. Bu tür bir permütasyon ařaęıdaki G matrisinin kullanımını sayesinde elde edilebilir:

$$G = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (4.13)$$

Yukarıdaki G matrisinin kullanımını ile birlikte durum denklemini;

$$\theta_t = G\theta_{t-1} + w_t = (a_2, a_1, a_4, a_3)' + w_t \quad (4.14)$$

elde edilir.

Duraęan olan mevsimsel modellerde W_t genelde 0 olur ancak mevsimsel etkiler zamanla deęiřtięinden dolayı W_t sıfır deęildir ve dikkatlice tanımlanması gerekir. Mevsimsel katsayılar serinin belirli bir ortalama deęer etrafındaki deęişimlerini tanımlamayı amaçlamaktadır. Bu nedenle sezonsal katsayıların toplamı genellikle 0 olur.

$$\sum_j^s a_j = 0 \quad (4.15)$$

Sezonsal etkileri modellemek için alternatif bir yöntem ise katsayılardan birini sezonsal etki için normal deęer kabul edip kalan deęerleri periyodik deęişimler olarak kabul etmektir. Bu şekildeki bir varsayımda durum vektörü, F ve G matrisleri (s-1) boyutlu matrisler olarak tanımlanır. Sezonsal etkinin a_1 'den a_2 'ye geçmesi için durum

matrisinin $\theta_{t-1} = (a_1, a_4, a_3)'$ 'den $\theta_t = (a_2, a_1, a_4)'$ 'e dönüştürülmelidir. Sistem evrim hatasının olmadığı ve $\sum_{i=1}^4 a_i = 0$ statik bir model için dönüşüm işlemi;

$$G = \begin{bmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (4.16)$$

matrisi ile sağlanır.

Genel olarak, s periyoduna sahip bir sezonsal faktör modeli $(s - 1)$ boyutlu $F = (1, 0, \dots, 0)$ ve

$$G = \begin{bmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & & 1 & 0 \end{bmatrix} \quad (4.17)$$

durum uzay modeli olarak tanımlanır [19,26].

Bu durumda sezonsal bileşenlerin varyansı $W = \text{diag}(\sigma_w^2, 0, \dots, 0)$ olarak tanımlanan sistem evrim hatası olarak hesaplanır.

4.3.3. ARIMA süreçli dinamik lineer model

Herhangi bir ARIMA modeli DLM olarak tanımlanabilir. Daha doğrusu, ARIMA işlemi için verilen Y_t süreci aynı dağılıma sahip bir DLM olarak tanımlanabilir. ARMA(p,q) süreci Denklem 2.8.'de tanımlanmıştır. Durağan bir durum inceleneceği için μ değeri sıfır olarak kabul edilmiştir.

$$Y_t = \sum_{j=1}^r \varphi_j Y_{t-j} + \sum_{j=1}^{r-1} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (4.18)$$

ARMA modeli için genel DLM matrisleri $r = \max\{p, q + 1\}$, $\varphi_j = 0$ durumunda;

$$F = [1, 0, \dots, 0] \quad (4.19)$$

$$G = \begin{bmatrix} \varphi_1 & 1 & 0 & \dots & 0 \\ \varphi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \\ \varphi_{r-1} & 0 & \dots & 0 & 1 \\ \varphi_r & 0 & \dots & 0 & 0 \end{bmatrix} \quad (4.20)$$

şeklinde tanımlanır.

Bir ARMA modelini bir DLM olarak göstermenin başlıca iki avantajı vardır. Birincisi, DLM ile tanımlanan ARMA bileşeninin trend ve mevsimsel gibi diğer etki bileşenlerinin artık otokolerasyonunu açıklayabilmesidir. İkinci sebep ise ARMA modelinin olabilirlik fonksiyonu tahmininin DLM'nin özyinelemeli yapısıyla daha efektif olmasıdır [19].

4.4. Maksimum Olabilirlik Tahmin Yöntemi

Bir DLM tanımlanırken çoğu zaman matrislerde bilinmeyen parametrelerin olması olasıdır. Bu noktada bilinmeyen parametrelerin bulunabilmesi için kullanılan genel tahmin yöntemlerinden birinden faydalanılabilir. Bu çalışmada ise parametre tahmin yöntemlerinden olan Maksimum Olabilirlik Tahmin Yöntemi (Maximum Likelihood Estimation-MLE) kullanılmıştır.

Anakütleyi temsil eden olasılık yoğunluk fonksiyonu $p(x; \vartheta)$ ile gösterilsin ve elimizde dağılımı bilinmeyen ϑ parametresine bağlı Y_1, \dots, Y_n şeklinde n tane rastgele vektörümüz olsun. Bu gözlem değerlerinin olasılık yoğunluk fonksiyonu da $p(y_1, \dots, y_n; \vartheta)$ şeklinde tanımlayalım.

Maksimum olabilirlik tahmin yöntemi belli bir örneklem değerlerinin gerçekleşme olabilirliğini en yüksek yapan anakütle parametrelerini bulmaya çalışır.

MLE, sabit bir faktöre kadar, ϑ 'nin bir fonksiyonu olarak okunan gözlemlenen verilerin olasılık yoğunluğu olarak tanımlanmaktadır. Elimizde bir rassal örneklem olduğundan ve bunların çekildiği anakütlenin olasılık yoğunluk fonksiyonu bilindiği varsayıldığından, bağımsızlık özelliğinden hareketle DLM için olasılık yoğunluk fonksiyonu;

$$p(y_1, \dots, y_n; \vartheta) = p_1(x_1; \vartheta) \times p_2(x_2; \vartheta) \times \dots \times p_n(x_n; \vartheta) \quad (4.21)$$

$$p(y_1, \dots, y_n; \vartheta) = \prod_{t=1}^n p(y_t | y_{1:t-1}; \vartheta) \quad (4.22)$$

olarak yazılabilir. Maksimum olabilirlik tahmin yöntemi olabilirlik fonksiyonunu en yükseğe çıkararak tahmin etmeyi amaçlar. Anakütlenin dağılımı biliniyorsa, bu durumda Gaussian dağılıma sahip ortalama (f_t) ve varyans (Q_t) için logaritmik olasılık;

$$l(\vartheta) = -\frac{1}{2} \sum_{t=1}^n \log|Q_t| - \frac{1}{2} \sum_{t=1}^n (y_t - f_t)' Q_t (y_t - f_t) \quad (4.23)$$

şeklinde tanımlanır. Q_t ve f_t dolaylı olarak ϑ 'ye bağlıdır. Denklem 4.24 ile ϑ 'nin maksimum olasılığı sayısal olarak maksimize edilir. Burada maksimizasyon probleminin çözümünde kolaylık sağlaması için ortak olasılık fonksiyonunun logaritması alınabilir.

$$\hat{\vartheta} = \underset{\vartheta}{\operatorname{argmax}} l(\vartheta) \quad (4.24)$$

ϑ k adet bilinmeyen parametreden oluşuyorsa logolabilirlik fonksiyonunun bu parametrelere göre birinci türevleri sıfır (gerekli koşul), ikinci türev matrisi (Hessian-H) negatif belirli olmalıdır (yeterli koşul). Belirtilen denklemler ile H^{-1} matrisi DLM için maksimum olabilirlik fonksiyonunun varyansını ($\operatorname{Var}(\hat{\vartheta})$) hesaplanmasına imkân tanır [19].

4.5. Sakarya İli Hanehalkı Elektrik Talep Tahmini İçin DLM

Sakarya ili hanehalkı elektrik talebi için kullanılacak DLM açıklamadan önce tahmin için kullanılacak veri setinin tanımlanması gerekmektedir. Bu başlık altında ilk olarak veri setinin kaynağı ve biçimi anlatılmış ardından kullanılan DLM'nin kurucu matrislerinin yapıları açıklanmıştır.

4.5.1. Tahmin için kullanılan dinamik lineer model varyasyonları

Dinamik doğrusal modeller ile zaman serileri analizinde, serinin farklı özellikleri (trend, mevsimsellik vb.) tek bir birleşik modelde toplanabilir. Tez çalışmasında kullanılan Sakarya ili için hanehalkı elektrik tüketim verisinin de birleşik bir dinamik lineer modele ihtiyaç duyduğu görülmüştür. Çalışma kapsamında birkaç farklı varyasyonda DLM kullanılmış ve performans değerleri sunulmuştur. Çalışmada kullanılan DLM varyasyonlarının kurucu matrisleri aşağıda başlıklar halinde incelenmiştir. Kurucu matrislerin belirlenmesinden sonra gözlem ve sistem denklemlerinin normal dağılımlı kabul edilen V_t ve W_t hataları, maksimum olasılık fonksiyonu ile tahmin edilerek DLM kurulumu tamamlanmış olacaktır. Bu aşamaların gerçekleştirilmesi beşinci bölümde ele alınmıştır. Bu bölümde DLM kurulum parametreleri verilecektir.

a. Model A

Model A uygulamasında zaman serisinin trend ve sezonsal bileşenden oluştuğu düşünülmüştür. Model A diğer bir tanım ile sezonsal lineer büyüme modeli olarak da nitelendirilebilir. Model A veriye uygun olarak 3 sezona ayrılmış ve trend bileşeni eklenmiştir. Sezon bileşeninin $(s - 1)$ 2x2 boyutlu matris ile gösterilmiş. Lineer Büyüme modeli ve trend (2x2) boyut Bu durumda Model A için G_t ve F_t matrisleri aşağıdaki gibi olacaktır.

$$G_{4 \times 4} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (4.25)$$

$$F_{1 \times 4} = [1 \quad 0 \quad 1 \quad 0] \quad (4.26)$$

G_t ve F_t matrislerine bağılı olarak kurulan DLM'nin V_t ve W_t hataları maksimum olasılık fonksiyonu ile tahmin edilecektir. Tahmin edilecek V_t ve W_t matrislerinin yapısı aşağıdaki gibi olacaktır.

$$W_{4 \times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_{w1}^2 & 0 & 0 \\ 0 & 0 & \sigma_{w2}^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (4.27)$$

$$V_{1 \times 1} = \sum_{i=1}^n V_t^{(i)} \quad (4.28)$$

Burada V_t gözlem hatası birleşik model için toplam değeri ifade ederken W_t lineer büyüme trend bileşeni ve sezonsal etki bileşeni için σ_{w1}^2 ve σ_{w2}^2 şeklinde iki farklı hata parametresi ile ifade edilir.

Bütün bu parametrelere bağılı olarak θ_t durum matrisi;

$$\theta_{4 \times 1} = \begin{bmatrix} \mu_t \\ \beta_t \\ a_2 \\ a_1 \end{bmatrix} \quad (4.29)$$

olmakla birlikte seri hakkında herhangi bir ön bilgi olmadığından dolayı uygulamada başlangıç durumu tanımlanmamış DLM'nin varsayılan değerleri ile uygulama gerçekleştirilmiştir.

b. Model B

Model B için birleşik DLM'de zaman serisi rassal yürüyüş sürecini, sezonsal etkileri ve ARMA bileşenlerini içerdiği düşünülmüştür. Rassal yürüyüş sürecinden oluşan bu model durgun model olarak da geçmektedir. Model B'de rassal yürüyüş süreci birinci dereceden polinom matris olarak (1x1), sezonsal etki (s - 1) prensibi kapsamında (2x2) ve son olarak ARMA süreçleride uygulamadaki (p, d, q) paramadetleri için (3x3) boyutlu matris ile ifade edildiğinde G_t ve F_t matrisleri aşağıdaki gibi olacaktır.

$$G_{6 \times 6} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (4.30)$$

$$F_{1 \times 6} = [1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0] \quad (4.31)$$

G_t ve F_t matrislerine bağlı olarak kurulan DLM'nin V_t ve W_t hataları Model A maksimum olabilirlik fonksiyonu ile tahmin edilecektir. Tahmin edilecek V_t ve W_t matrislerinin yapısı aşağıdaki gibi olacaktır.

$$W_{6 \times 6} = \begin{bmatrix} \sigma_{w1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{w2}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \psi_1 & \psi_2 \\ 0 & 0 & 0 & \psi_1 & \psi_1^2 & \psi_1 \times \psi_2 \\ 0 & 0 & 0 & \psi_2 & \psi_1 \times \psi_2 & \psi_2^2 \end{bmatrix} \quad (4.32)$$

$$V_{1 \times 1} = \sum_{i=1}^n V_t^{(i)} \quad (4.33)$$

Burada V_t gözlem hatası birleşik model için toplam değeri ifade ederken W_t rassal yürüyüş bileşeni ve sezonsal etki bileşeni için σ_{w1}^2 ve σ_{w2}^2 ile ARMA süreçleri için

$R = [0 \dots 0 \ 1 \ \psi_1 \dots \psi_{r-2} \ \psi_{r-1}]$ olmak üzere $w_3 = RR'\sigma^2$ şeklinde üç farklı hata parametresi ile ifade edilir. w_3 hata terimi hesaplanırken uygulamada kullanılan ARIMA(p,d,q) parametrelerine göre $r = 3$ olacağından 3x3 boyutlu bir matris olarak bulunacaktır. R matrisi hata düzeltme katsayısı olarak bilinir ve $k \geq 1$ notasyon sayısı olmak üzere $R_t(k) = Var(\theta_{t+k}|y_{1:t})$ şeklinde tanımlanır.

Bütün bu parametrelere bağlı olarak θ_t durum matrisi;

$$\theta_{6 \times 1} = \begin{bmatrix} \mu_t \\ a_2 \\ a_1 \\ \theta_1[t] \\ \theta_2[t] \\ \theta_3[t] \end{bmatrix} \quad (4.34)$$

olarak tanımlanır. $\theta_1[t]$, $\theta_2[t]$ ve $\theta_3[t]$ ARIMA parametreleridir. μ_t rassal yürüyüş sürecinin, a_1 ve a_2 sezonsal faktör modelinin katsayılarından oluşurur.

c. Model C

Model C için birleşik DLM'de zaman serisi lineer büyüme sürecini, sezonsal etkileri ve ARMA bileşenlerini içerdiği düşünülmüştür. Model C'de lineer büyüme süreci ikinci dereceden polinom matris olarak (2x2), sezonsal etki $(s-1)$ prensibi kapsamında (2x2) ve son olarak ARMA süreçleride (p, d, q) paramadeleri için (3x3) boyutlu matris ile ifade edildiğinde G_t ve F_t matrisleri aşağıdaki gibi olacaktır.

$$G_{7 \times 7} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (4.35)$$

$$F_{1 \times 7} = [1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 0] \quad (4.36)$$

G_t ve F_t matrislerine bağılı olarak kurulan DLM'nin V_t ve W_t hataları Model A ve Model B'de olduğu gibi maksimum olabilirlik fonksiyonu ile tahmin edilecektir. Tahmin edilecek V_t ve W_t matrislerinin yapısı aşağıdaki gibi olacaktır.

$$W_{7 \times 7} = \begin{bmatrix} \sigma_{w1}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{w2}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{w3}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \psi_1 & \psi_2 \\ 0 & 0 & 0 & 0 & \psi_1 & \psi_1^2 & \psi_1 \times \psi_2 \\ 0 & 0 & 0 & 0 & \psi_2 & \psi_1 \times \psi_2 & \psi_2^2 \end{bmatrix} \quad (4.37)$$

$$V_{1 \times 1} = \sum_{i=1}^n V_t^{(i)} \quad (4.38)$$

Burada V_t gözlem hatası birleşik model için toplam değeri ifade ederken W_t lineer büyüme trend bileşeni için sistem hata terimi σ_{w1}^2 ve σ_{w2}^2 , sezon bileşeninin sistem hata terimi σ_{w3}^2 ve son olarak ARMA süreçleri için $R = [0 \dots 0 \ 1 \ \psi_1 \dots \psi_{r-2} \ \psi_{r-1}]$ olmak üzere $w_4 = RR' \sigma^2$ şeklinde dört farklı hata parametresi ile ifade edilir. w_3 hata terimi hesaplanırken uygulamada kullanılan ARIMA(p,d,q) parametrelerine göre $r = 3$ olacağından 3×3 boyutlu bir matris olarak bulunacaktır. R matrisi hata düzeltme katsayısı olarak bilinir ve $k \geq 1$ notasyon sayısı olmak üzere $R_t(k) = Var(\theta_{t+k} | y_{1:t})$ şeklinde tanımlanır.

Bütün bu parametrelere bağılı olarak θ_t durum matrisi;

$$\theta_{7 \times 1} = \begin{bmatrix} \mu_t \\ \beta_t \\ a_2 \\ a_1 \\ \theta_1[t] \\ \theta_2[t] \\ \theta_3[t] \end{bmatrix} \quad (4.39)$$

olarak tanımlanır. $\theta_1[t]$, $\theta_2[t]$ ve $\theta_3[t]$ ARIMA parametreleridir. μ_t rassal yürüyüş sürecinin, β_t lineer büyüme trendinin, a_1 ve a_2 sezonsal faktör modelinin katsayılarıdır. β_t parametresi ise lineer büyüme modelinin katsayısıdır.

BÖLÜM 5. GERÇEKLEME, BULGULAR VE DEĞERLENDİRME

5.1. R Platformu ve “dlm” Uygulama Paketi

R programı Yeni Zelanda Aucland Üniversitesinden R. Ihaka ve R. Gentleman tarafından ortaya çıkarılmış ve günümüzde R Geliştirme Çekirdek Ekibi tarafından geliştirilmektedir. Çoğunlukla araştırma ve öğretim için kullanılan temeli S diline dayanan bir yazılımdır [8].

R programı, birçok istatistiksel yöntemi bünyesinde barındıran, veri görmesi anlamında ileri derecede grafik çizim araçlarına sahip istatistiksel bir bilgisayar programıdır. Aynı zamanda özel alanlarda kullanıcılar tarafından oluşturulmuş zengin bir açık kaynak kütüphanesine sahiptir. Kullanıcılar mevcut fonksiyonları kullanarak ya da kendi yöntemlerini mevcut fonksiyonlar üzerinden geliştirerek kolaylıkla uygulayabilirler.

Dinamik Lineer Modeller ise R programında “dlm” ile isimlendirilen bir sınıf altında temsil edilir. “dlm” paketi Dinamik Lineer Modellerin Bayesci analizine odaklanır. Aynı zamanda bilinmeyen model parametreleri için maksimum olabilirlik fonksiyonunu, Kalman Filtreleme metodunu ve düzleştirme araçlarını bünyesinde barındırır. İlgili varyans matrislerinin durum vektörü tabanlı tekil değer ayrıştırması ve olasılık ölçümü için kullanılan Maksimum olabilirlik fonksiyonu diğer algoritmalara göre Kalman filtresi için sayısal kararlılığı artırır [29].

“dlm” paketi kapsamında özel işlevli fonksiyonlar bulunur. Bunlardan bazıları;

Tablo 5.1. Özel işlevler için tanımlanabilen "dlm" fonksiyonları

Fonksiyon İsmi	İşlev
dlmModSeas	Periyodik-Mevsimsel Faktörler için DLM
dlmModTrig	Periyodik-Fourier Trigonometrik DLM
dlmModPoly	n. dereceden polinomal DLM
dlmModARMA	ARMA sürecini temsil eden DLM nesnesi tanımlar
dlmModReg	Doğrusal Regresyon modelinin DLM gösterimini tanımlar
dlmSmooth	Kalman Filtresi için düzleştirme metodu
dlmForecast	DLM için tahmin fonksiyonu

“dlm” paketi sınıflarına bakıldığında birçok farklı yöntem için entegre bir ortam sağladığı görülmektedir. Özellikle zaman serisi analizinde en yaygın kullanılan model ve Box&Jenkins tarafından da çokça kullanılan otoregresif hareketli ortalama (ARMA) modeli sınıfı, “dlmModARMA”, tek değişkenli veya çok değişkenli durumlar için tanımlanabilir [19].

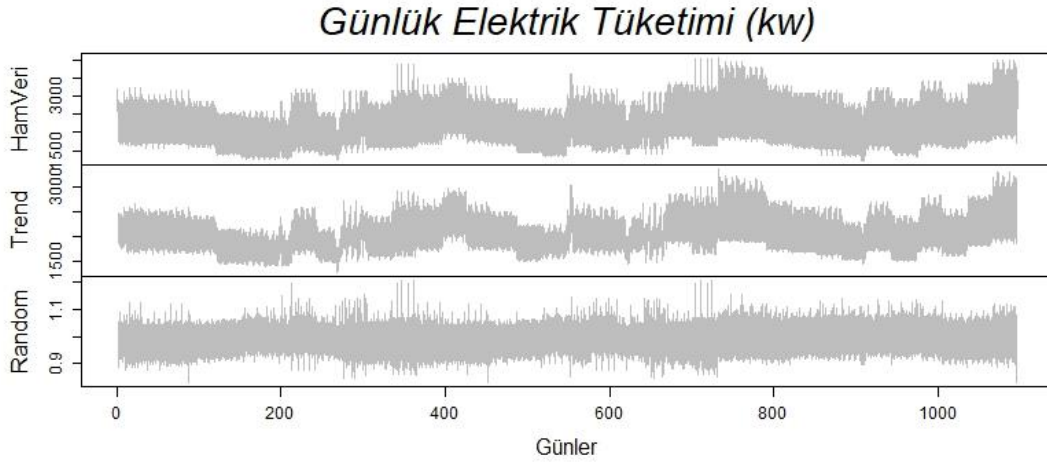
Ayrıca dlm paketi kapsamında filtreleme ve yumuşatma işlemi için “dlmFilter” ve “dlmSmooth” fonksiyonları vardır. Bu fonksiyonların haricinde tahmin işlemi gerçekleştirmek için “dlmForecast” fonksiyonunu bünyesinde barındırır. Çalışmada varyans ve sistem hatalarının tespiti için kullanılan bir diğer yöntem Gibbs örnekleme yöntemi, dlm paketinde “dlmGibbsDIG” fonksiyonu ile entegredir.

5.2. Modellerin Gerçeklemesi ve Çıktıları

Çalışmada ele alınan dinamik lineer modelin performansının ölçümü açısından oluşturulan zaman serisi farklı modeller ile uygulama gerçekleştirilmiştir. Modellerin grafiksel ve tahminsel performans değerleri bu başlık altında sunulmuştur. Aynı zamanda kıyaslama açısından ARIMA model uygulamaları da gerçekleştirilmiştir.

Uygulamalara geçmeden önce ele alınan zaman serisinin özelliklerinin belirlenmesi tutarlı sonuçlar elde etmek açısından gereklidir. Şekil 5.1.’de trend ve random etki bileşenine ayrılmış Sakarya iline ait mesken elektrik tüketim verisi gösterilmiştir. Burada sezon bileşeninin veri sıklığından dolayı görülebilirliği mümkün olmazken, trend bileşeninin zaman içerisindeki değişimi açık bir şekilde görülebilmektedir.

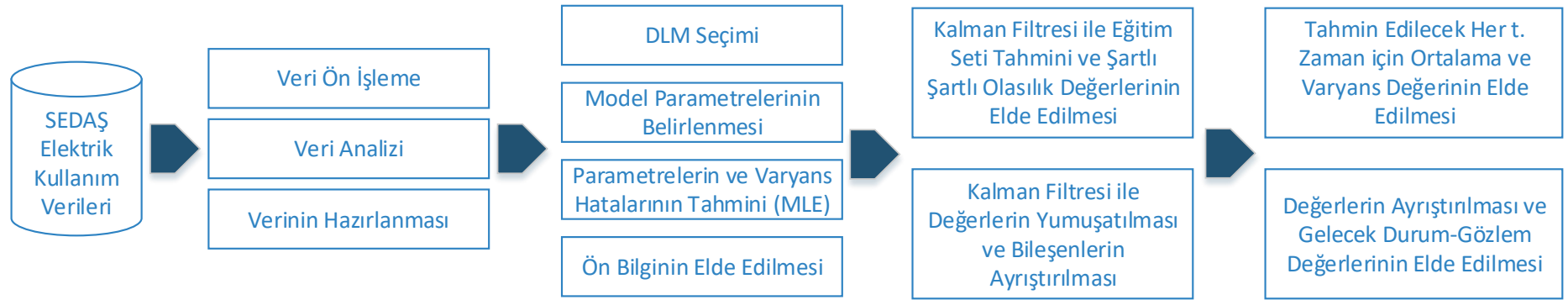
Random (rassal) etkiler ise zaman serisini etkileyen trend ya da sezon bileşenine ait olmayan etkilerdir.



Şekil 5.1. Trend ve Random Etki Bileşenine Ayrılmış Sakarya İli Mesken Elektrik Tüketimi

Zaman serisinin analiz ve tahminine geçildiğinde uygulamada kullanılan DLM modellerinin trend, sezonsal ve ARMA bileşenlerini içerdiği Bölüm 4.'te açıklanmıştı. Bu kapsamda R ortamında gerçekleştirilen modellerin uygulanmasındaki genel adımlar aşağıda maddeler halinde verilmiş, Şekil 5.2.'de ise akış şeklinde gösterilmiştir.

- Modele özgü parametrelere uygun DLM'nin R studio ortamında dlm kurucu fonksiyonu ile tanımlanması.
- Gözlem ve sistem denklemlerinde yer alan hata terimlerine ait varyans parametrelerinin ve bilinmeyen değerlerin, dlmMLE fonksiyonundan yararlanarak maksimum olabilirlik tahminlerinin elde edilmesi
- Elde edilen maksimum olabilirlik tahminlerinin modelde kullanılması ve dlmFilter fonksiyonu ile Kalman filtresi sonuçlarının ve dlmSmoothing fonksiyonu Kalman Düzleştirici (Kalman Smoothing) ile ayrıştırılmış verinin elde edilmesi.
- dlmForecast fonksiyonu ile Kalman Filtresi parametrelerine dayalı gelecek dönem tahmin ortalaması ve buna bağlı olarak tahmin değerlerinin belirlenmesi



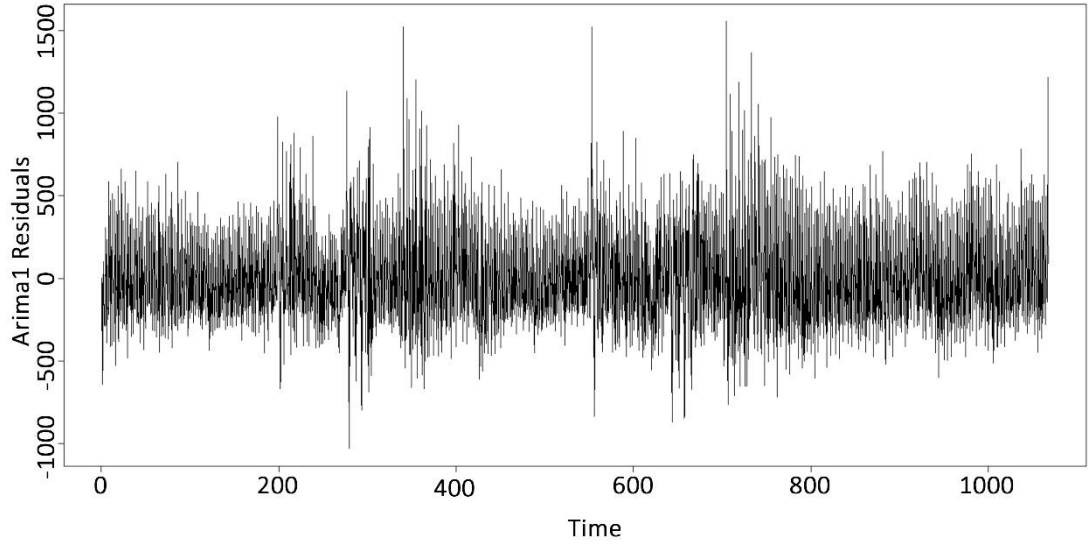
Şekil 5.2. DLM ile tahmin uygulamasında izlenen yol

5.2.1. ARIMA uygulamaları

Zaman serileri analizinde en çok kullanılan yöntemlerden birisi ARIMA modelleri ile tahmin yöntemidir. ARIMA modelleri durağan ya da çeşitli istatistiksel yöntemler ile durağan hale getirilen seriyi oluşturan veriler arasında doğrusal bir ilişkinin olduğunu varsayan ve bu doğrusal ilişkiyi modelleyebilir. Bu çalışmada DLM ile tahmin uygulamasının yanısıra ARIMA modellerinin bir türü olan ARIMAX (Reg-ARIMA) ve SARIMA (Seasonal ARIMA) modelleme yöntemi ile Sakarya ili Mesken Elektrik Tüketim verilerinin 2017 yılının Aralık ayı verileri tahmini gerçekleştirilmiştir.

ARIMAX yönteminde bağımlı değişken ARIMA yönteminde farklı olarak sadece geçmiş değerleri ile değil, açıklayıcı değişkenler ve geçmiş değerler ile birlikte tahmin edilir. Bu çalışmada Sakarya ili elektrik tüketim verileri bağımlı değişken, Sakarya ili için günlük sıcaklık değerleri ve haftanın günleri ise bağımsız değişken olarak kabul edilmiştir. SARIMA yöntemi ise ARIMA yönteminin sezonsal parametreleri barındıran varyasyonudur.

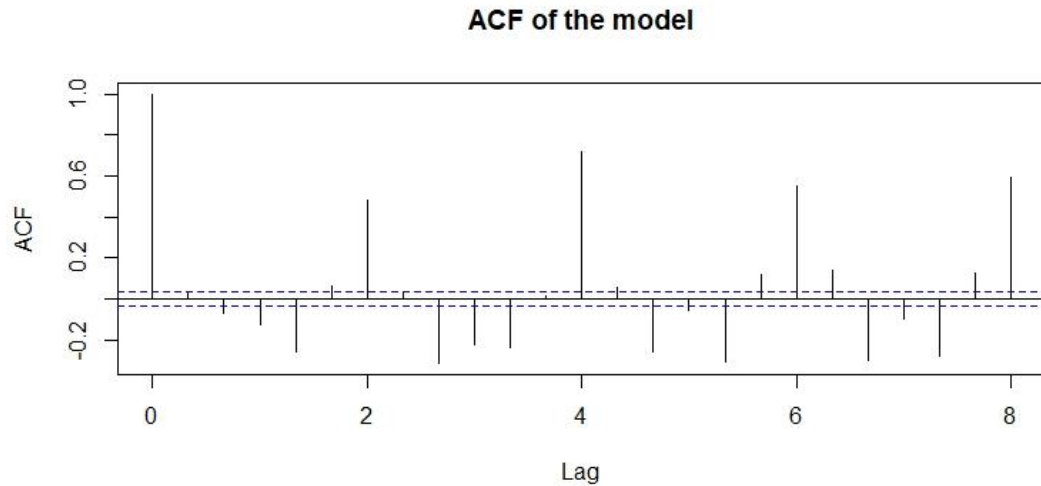
ARIMA yöntemleri ile tahmin işlemine geçmeden önce Sakarya ili için Mesken Elektrik tüketim verilerinin fark alma işlemine tabii tutularak durağan hale getirilmesi gerekir. Durağan hale getirilen veri AR ve MA parametrelerinin belirlenmesi gerekir. Bu noktada R ortamında “auto.arima” fonksiyonundan yararlanılmıştır. Bu fonksiyon AIC (Akaike Bilgi Kriteri) ya da BIC (Bayesci Bilgi Kriteri) değerine göre en iyi ARIMA modelini arar ve kurulumunu gerçekleştirir. Fonksiyon bir yandan arama kriteri ile parametreleri belirlerken aynı zaman da serinin durağanlığını elde etmeyi amaçlar. Bu aşamada durağanlık parametrelerinin kıyaslanması ve durağanlığı sağlayan modelin seçimi için bir durağanlık test yöntemi olan “kpss” testinden faydalanılmıştır. Uygulamanın sonucunda oluşan modele (ARIMA (2,1,2) (0,0,2) [3]) göre elde edilen seri Şekil 5.3.’de gösterilmiştir.



Şekil 5.3. Durağanlaştırılmış Sakarya İli Mesken Elektrik Tüketim Verisi

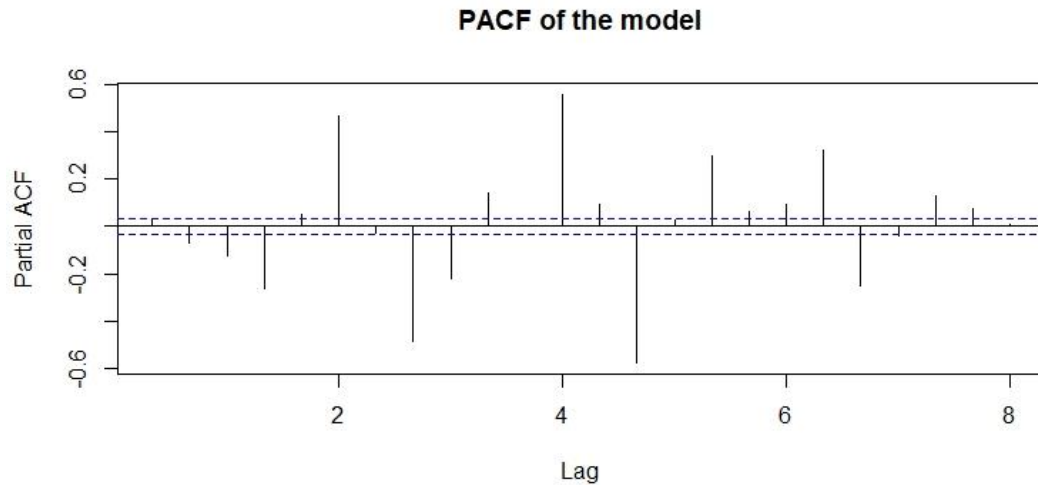
Serinin ACF ve PACF sonuçları ise Şekil 5.4. ve 5.5.'de sunulmuştur. Otokorelasyon fonksiyonu (ACF), zaman serisine ait bir değişkenin bir ya da daha fazla gecikmeli dönemi arasında korelasyonlu olması hali olarak tanımlanmaktadır ve zaman serilerinin durağanlığının belirlenmesinde kullanılır. Bu fonksiyonlar Box-Jenkins modelleme yönteminde AR ve MA süreçlerinin derecelerinin belirlenmesinde kullanılmaktadır.

ACF grafiğine bakıldığında zaman serisinin otokorelasyon katsayılarının hızlı bir şekilde sönerek sıfıra yaklaştığı ve mevsimsel serilerde olduğu gibi periyodik olarak sıçrama yapıp azaldığı farkedilebilir.



Şekil 5.4. Seçilen Modele göre Serinin ACF Testi

PACF grafiğine bakıldığında ise serinin ilk değerden itibaren güven sınırı içinde olduğu görülmektedir. Bu durumda ilk gecikmeden sonrası istatistiksel olarak anlamsız olarak kabul edilir.



Şekil 5.5. Seçilen Modele göre Serinin PACF Testi

Kurulan modelin Box-Ljung test sonuçları aşağıdaki gibidir.

Box-Ljung test

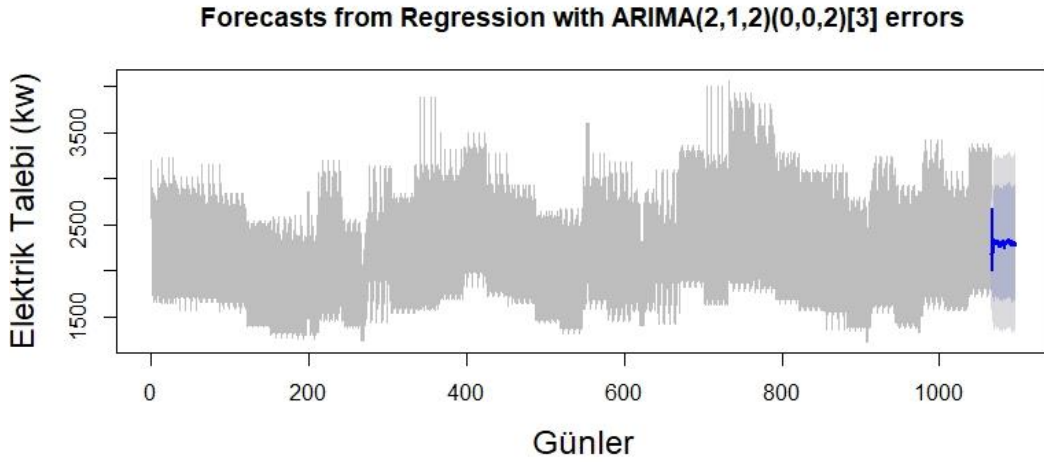
data: arima1\$residuals^2

X-squared = 770.51, df = 20, p-value < 2.2e-16

Modelin uygunluğuna yönelik analizlerin ardından elde edilen modele ait parametre değerleri Tablo 5.2.'de verilmiştir.

Tablo 5.2. ARIMAX Model Parametreleri

	%2,5	%97,5
AR1	-0.4819	-0.4044
AR2	-0.3767	-0.2893
MA1	-0.0381	-0.0097
MA2	-0.9392	-0.9115
SMA1	-0.3291	-0.2331
SMA2	0.7462	0.7842



Şekil 5.6. Sakarya İli Elektrik Tüketimi ve Hava Durumu Verileri ile Kurulan ARIMAX Modeline Göre Günlük Elektrik Talep Tahmini Sonucu (2017 Aralık)

Serinin durağanlaştırılması, en uygun modelin seçimi ve ARIMA parametrelerinin belirlenmesinden sonra Model A, Model B ve Model C' de olduğu gibi ARIMAX modelinde de 2017 yılının Aralık ayına dair günlük elektrik tüketim talebinin tahmini gerçekleştirilmiştir. Tahmin sonuçları Şekil 5.6.'te sunulmuştur.

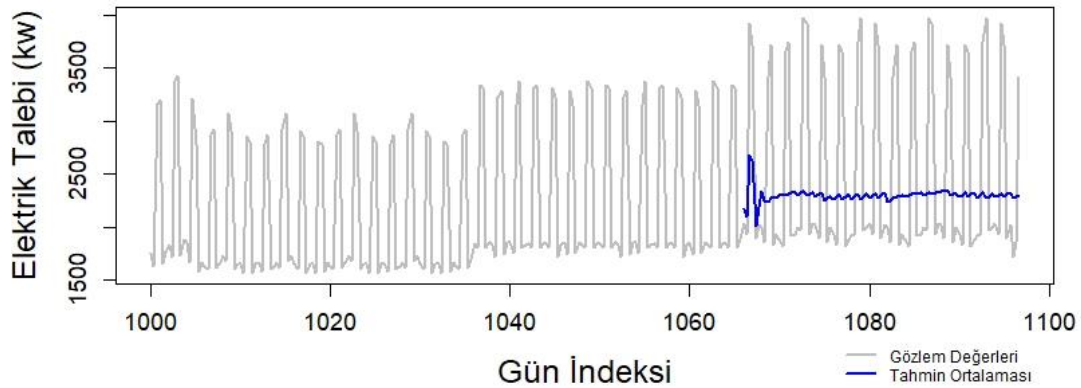
Yapılan tahmin sonucunda elde edilen günlük elektrik tüketim değerleri ile test veri seti kıyaslanmıştır. Kıyaslama sonucu farklı hata tespit yöntemleri ile Tablo 5.3.'de bulunmaktadır.

Tablo 5.3. Tutarlılık Testi-Hata Sonuçları (ARIMAX)

	ME	RMSE	MAE	MPE	MAPE	ACF1
Test Seti	230,2871	873,4183	709,0544	0,0390	25,0545	0,2156

Son olarak Şekil 5.7.'de serinin son 2017 yılı son 3 aylık verisi ve ARIMAX modeli ile yapılan 1 aylık (2017 Aralık) tahmininin ilerleyişi sunulmuştur. Dikkat edildiğinde tahmin değerlerinin, tahmin serisinin sonuna doğru, dağılım aralığı daralmıştır ve gözlem değerlerine yaklaşmamaktadır.

ARIMAX ile Günlük Elektrik Talep Tahmini



Şekil 5.7. 2017 Yılı Son 3 Aylık Verisi ve ARIMAX için Aralık 2017 Tahmin Değerleri

5.2.2. SARIMA uygulaması

SARIMA modeli uygulamasında ARIMAX modelinde olduğu gibi seri ilk önce durağanlaştırılmış ve zaman serisi Şekil 5.3.'de olduğu gibi elde edilmiştir. Elde edilen serinin ACF ve PACF grafikleri de ARIMAX uygulamasındaki Şekil 5.4. ve Şekil 5.5. ile aynıdır. Kullanılan fonksiyonda BIC ve KPSS testi sonuçlarına göre elde edilen en iyi model ARIMAX uygulamasında olduğu gibi (ARIMA (2,1,2) (0,0,2) [3]) olmuştur. SARIMA modelinin AR, MA ve Sigma parametreleri Tablo 5.4.'de sunulmuştur.

Tablo 5.4. SARIMA Model Parametreleri

	%2,5	%97,5
AR1	-0.4874	-0.4095
AR2	-0.3805	-0.2921
MA1	-0.0373	-0.0090
MA2	-0.9382	-0.9106
SMA1	-0.3243	-0.2239
SMA2	0.7480	0.7867

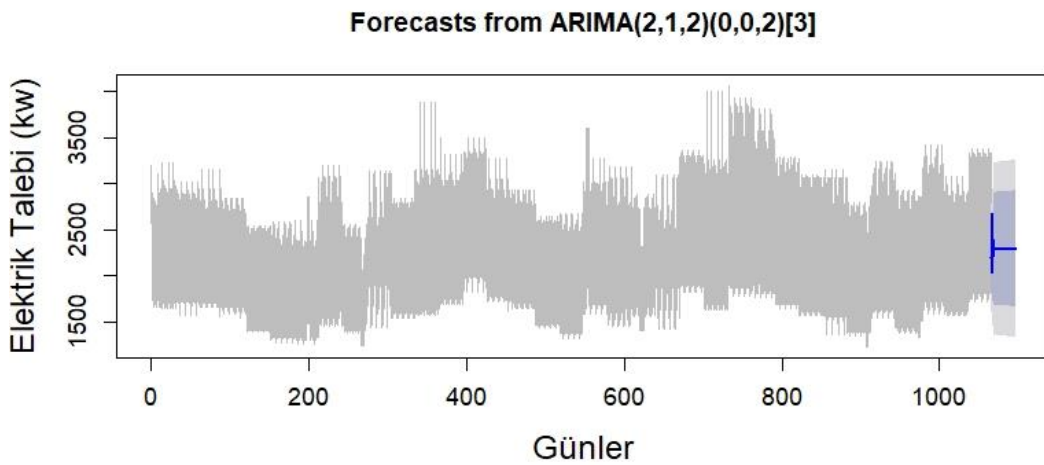
Kurulan model için Box-Ljung test sonuçları ve model seçimi için kullanılan BIC verisi ise aşağıdaki gibidir.

Box-Ljung test

data: arima1\$residuals^2
X-squared = 810.34, df = 20, p-value < 2.2e-16

sigma^2 estimated as 91162: log likelihood=-22793.15
AIC=45600.3 AICc=45600.33 BIC=45642.79

Serinin durağanlaştırılması ve model parametrelerinin belirlenmesinin ardından diğer modellerde olduğu gibi Sakarya ili mesken elektrik tüketim verileri için tahmin işlemi gerçekleştirilmiştir. Tahmin sonucu Şekil 5.8.'de gösterilmiştir.



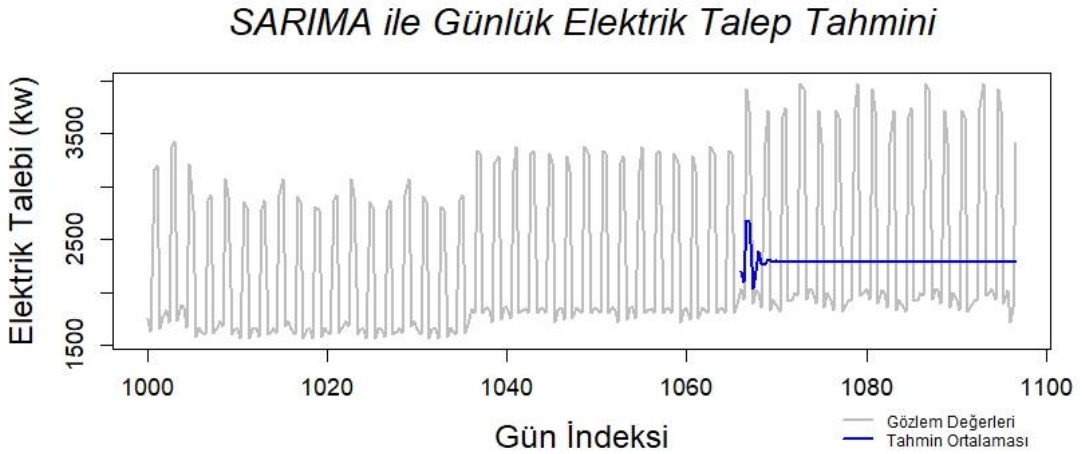
Şekil 5.8. SARIMA Modeline Göre Günlük Elektrik Talep Tahmini Sonucu (2017 Aralık)

2017 yılının Aralık ayı için yapılan tahminin test veri seti ile kıyaslama sonuçları Tablo 5.5.'de verilmiştir. Bu tabloya bakıldığında SARIMA uygulaması, hava durumu ve haftanın günleri ile yapılan ARIMAX uygulamasına göre yaklaşık %0,01 daha iyi bir performans göstermiştir.

Tablo 5.5. Tutarlılık Testi ve Hata Sonuçları (SARIMA)

	ME	RMSE	MAE	MPE	MAPE	ACF1
Test Seti	235,0637	875,8529	708,0960	0,2309	24,9573	0.2167

Şekil 5.9.'de serinin son 2017 yılı son 3 aylık verisi ve SARIMA modeli ile yapılan 1 aylık (2017 Aralık) tahmininin ilerleyişi sunulmuştur. Dikkat edildiğinde tahmin değerlerinin, tahmin serisinin sonuna yaklaştıkça dağılım aralığının daraldığı, ARIMAX modelinden farklı olarak bir süre sonra tahmin değerinin ilerleyişinin düz bir çizgi şeklinde olduğu farkedilmektedir.



Şekil 5.9. 2017 Yılı Son 3 Aylık Verisi ve SARIMA Aralık 2017 Tahmin Değerleri

5.2.3. Model A (Sezonsal faktör (SF) ve lineer büyüme (LB) modeli)

Daha önce Bölüm 4.'de açıklamaları ve kurulum matrislerini verilen Model A, R studio ortamında “dlm” paketi fonksiyonları ile gerçekleştirilmiştir. İlk olarak lineer büyüme modeli ve sezonsal faktör parametreleri ile DLM kurulmuştur. Kurulan DLM'nin R “dlm” paketi ile tanımlama işlemi “dlmModPoly(2) + dlmModSeas(3)” fonksiyonları ile DLM'nin toplamsallık özelliği sayesinde gerçekleştirilmiştir.

Kurulan dlm modelinin bilinmeyen parametreleri dlmMLE ile tahmin edildiğinde gözlem denklemi için V ve sistem denklemi için W hata matrisleri aşağıdaki gibi olmuştur.

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0,02 & 0 & 0 \\ 0 & 0 & 1,34 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

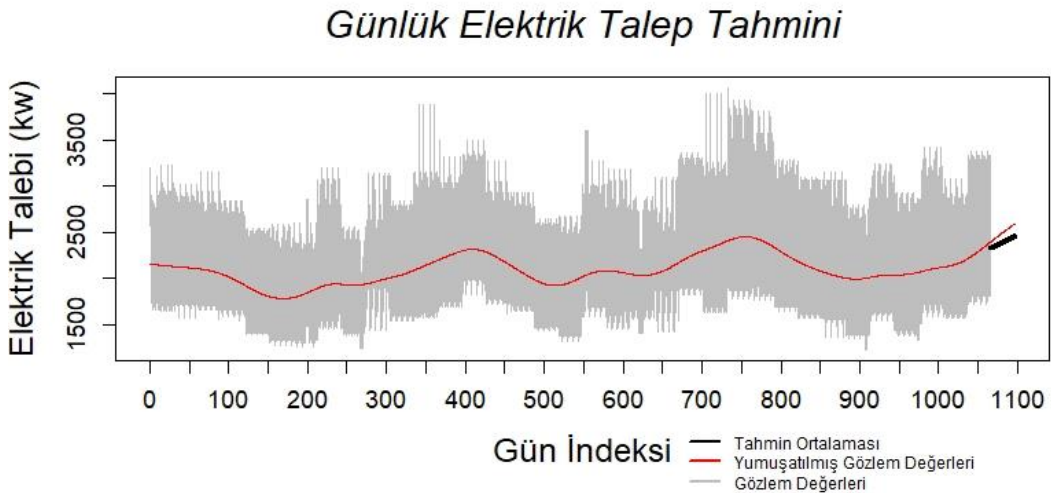
$$V = [266,717]$$

V ve W matrislerinin değerlerinin Hessian matrisine göre standart hata oranı Tablo 5.6.'da verilmiştir. Tabloya bakıldığında standart hata oranı düşük olduğundan elde edilen MLE değerlerinin gerçek değerler olduğu kabul edilebilir.

Tablo 5.6. Model A için Sistem ve Gözlem Matrislerinin MLE Standart Hata Oranı

Parametre	MLE Değeri	Standart Hata
V	0.0252255	0.0252255
W_1	0.5312138	0,5312138
W_2	1.34265606	0.8932793

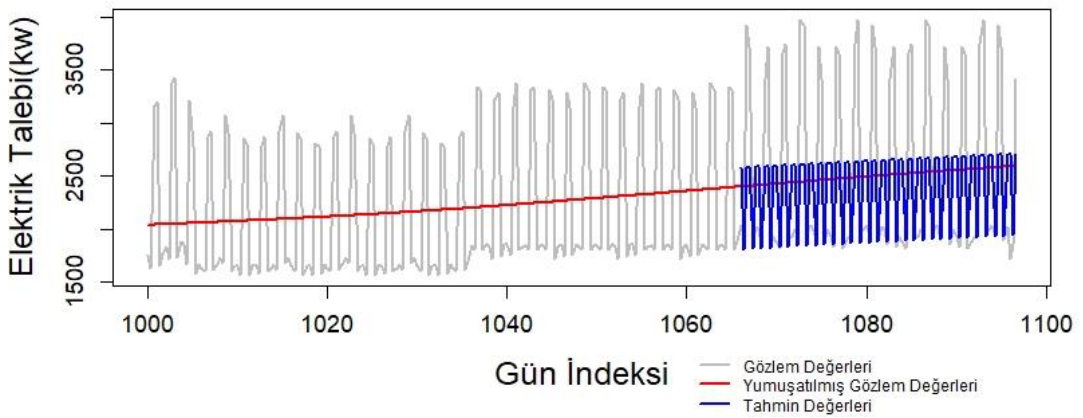
MLE ile bilinmeyen parametrelerin tahmininin ardından kurucu dlm fonksiyonunu temel alan Kalman Filtresi ile eğitim setinin tahmini yapılmış ve “dlmForecast” metodu ile Kalman Filtresi sonuçlarına göre 2017 yılının Aralık ayı için günlük elektrik talep tahmin sonuçları elde edilmiştir. Şekil 5.10.'a bakıldığında tahmin ortalaması değerinin Model A'nın kurulum şekline uygun olarak artan bir trende sahip olduğu görülmüştür. Ayrıca yumuşatılmış gözlem değerlerinin de zaman serisine uygun olarak trendi açıkça belirleyebildiği tespit edilmiştir.



Şekil 5.10. Lineer Büyüme ve Sezonsal Faktör Modeline Göre Günlük Elektrik Talep Tahmini

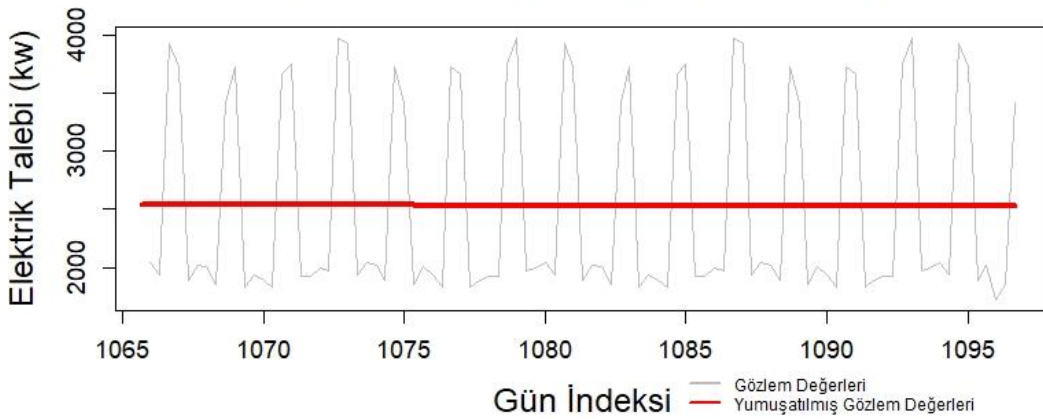
Şekil 5.11. ve Şekil 5.12.'de ise tahmin edilen Aralık ayının gerçek gözlem değerleri ve kurulan dlm modeline göre yumuşatılmış Kalman Smoothing gözlem verileri vardır. Grafikte görüldüğü üzere gerçek değerler son 3 aylık veri için artan bir değere sahipken 2017 Aralık ayı için yumuşatılmış değerler durgun bir süreci izlemiştir. Bu durumun sebebi kurulan modelin zaman serisi verisindeki etkileri yeterince kapsayamamasıdır.

Günlük Elektrik Talep Tahmini



Şekil 5.11. 2017 Yılı Son 3 Aylık Verisi ve Model A için Aralık 2017 Tahmin Değerleri

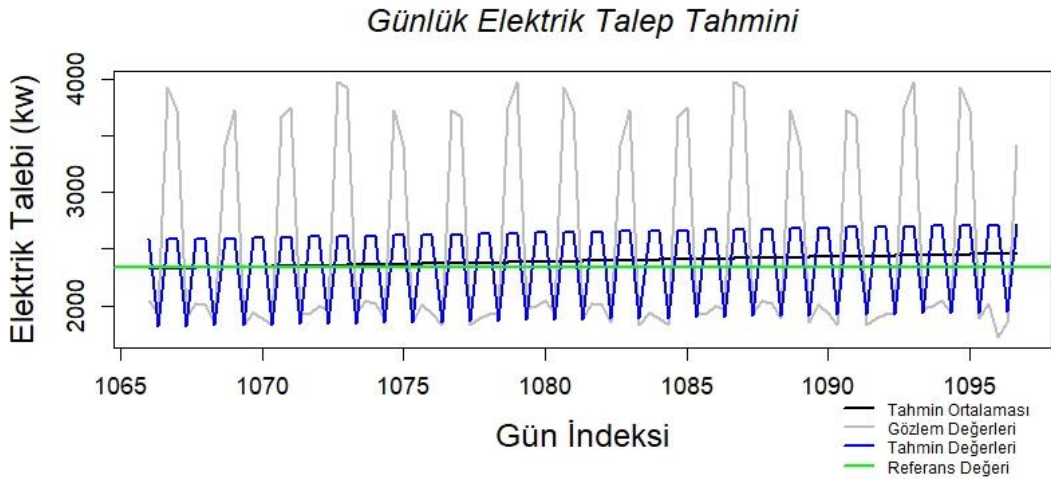
DLM ile Yumuşatılmış Elektrik Talebi



Şekil 5.12. 2017 Aralık Günlük Elektrik Talebi ve Model A DLM'ye Göre Yumuşatılmış Gözlem Değerleri

Şekil 5.13.'de ise Model A'ya göre günlük elektrik talep tahmini değerleri, tahmin ortalaması ve günlük gerçekleşen elektrik tüketimi verileri gösterilmiştir. Grafiğe

bakıldığında tahmin ortalamasının artan bir trende sahip olduğu ve tahmin değerlerinin gözlem değerlerinden uzak bir dağılıma sahip olduğu görülmüştür.



Şekil 5.13. Model A'ya Göre Günlük Elektrik Talep Tahmini Değerleri, Ortalaması ve Günlük Gerçekleşen Elektrik Tüketimi

Tablo 5.7.'de Model A ya göre yapılan DLM tahmininin test verisi ile tutarlılığı farklı hata tespit yöntemlerine göre kıyaslanmıştır.

Tablo 5.7. Tutarlılık ve Hata Sonuçları (Model A)

	ME	RMSE	MAE	R^2	MAPE	ACF1
Eğitim Seti	3.7856	531.732	435.514	0,273	20.266	0.490
Test Seti	139,284	748,777	605,782	0,30	22,136	0,7434

5.2.4. Model B (SF, Random Walk (RW) ve ARMA modeli)

Daha önce Bölüm 4.'te açıklamaları ve kurulum matrislerini verilen Model B, R studio ortamında “dlm” paketi fonksiyonları ile gerçekleştirilmiştir. İlk olarak rassal yürüyüş süreci, sezonsal faktör ve ARMA parametreleri ile DLM kurulmuştur. Kurulan DLM'nin R “dlm” paketi ile tanımlama işlemi “dlmModPoly(1) + dlmModSeas(3)+dlmModARMA()” fonksiyonları ile DLM'nin toplamsallık özelliği sayesinde gerçekleştirilmiştir.

Kurulan dlm modelinin bilinmeyen parametreleri Model A da olduğu gibi dlmMLE ile tahmin edildiğinde gözlem denklemi için V ve sistem denklemi için W hata matrisleri aşağıdaki gibi olmuştur.

$$W = \begin{bmatrix} 15579,72 & 0 & 0 & 0 & 0 & 0 \\ 0 & 60,6744 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,0696 & 0,1785 & -2,1809 \\ 0 & 0 & 0 & 0,1785 & 0,4577 & -5,5909 \\ 0 & 0 & 0 & -2,1809 & -5,5909 & 68,2948 \end{bmatrix}$$

$$V = [0,67417]$$

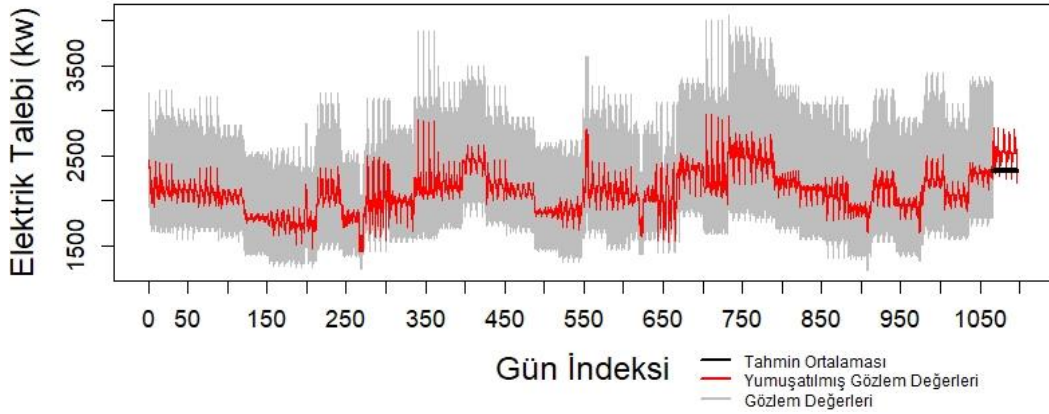
V ve W matrislerinin değerlerinin Hessian matrisine göre standart hata oranı Tablo 5.8.'de verilmiştir. Tabloya bakıldığında standart hata oranı düşük olduğundan elde edilen değerlerin gerçek değerler olduğu kabul edilmiştir.

Tablo 5.8. Model B için Sistem ve Gözlem Matrislerinin MLE Standart Hata Oranı

Parametre	MLE Değeri	Standart Hata
V	0,67417	4,4765
W_1	15579,72	0,0299
W_2	60,6744	0,2122

MLE ile bilinmeyen parametrelerin tahmininin ardından kurucu dlm fonksiyonunu temel alan Kalman Filtresi ile eğitim setinin tahmini yapılarak, *dlmForecast* metodu ile Kalman Filtresi sonuçlarına göre 2017 yılının Aralık ayı için günlük elektrik talep tahmin sonuçları elde edilmiştir. Şekil 5.14.'e bakıldığında tahmin ortalaması değerinin Model B'nin kurulum şekline uygun olarak durgun olduğu ve rassal yürüyüş sürecini yansıttığı görülmüştür. Ayrıca yumuşatılmış gözlem değerlerinin de zaman serisine uygun olarak hareket ettiği belirlenmiştir.

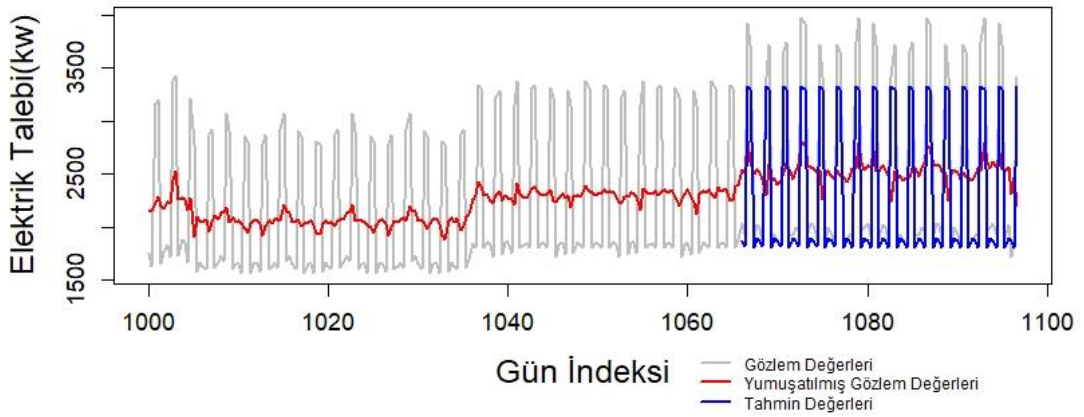
Günlük Elektrik Talep Tahmini



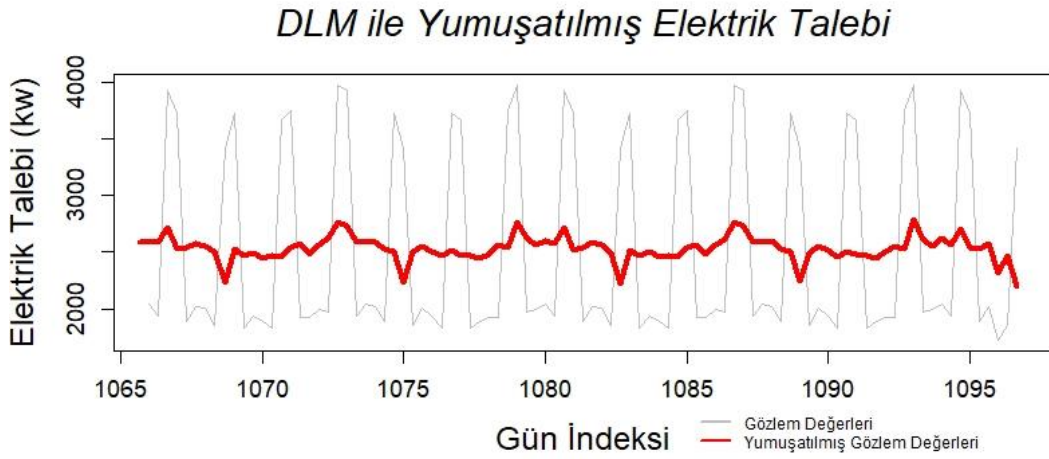
Şekil 5.14. Rassal Yürüyüş Süreci, Sezonluk Faktör ve ARMA Bileşenine Sahip Modele Göre Günlük Elektrik Talep Tahmini

Şekil 5.15. ve Şekil 5.16.'de ise tahmin edilen Aralık ayının gerçek gözlem değerleri ve kurulan dlm modeline göre yumuşatılmış Kalman Smoothing gözlem verileri vardır. Grafikte görüldüğü üzere gerçek değerler ve yumuşatılmış veriler uyum içindedir. Bunun yanında kırılma noktalarında DLM modeli iyi bir performans göstermiştir.

Günlük Elektrik Talep Tahmini

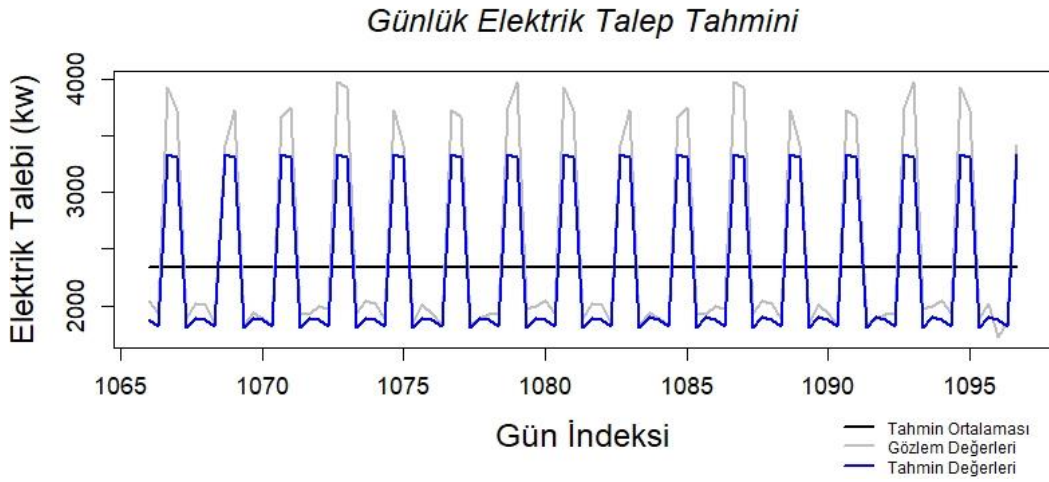


Şekil 5.15. 2017 Yılı Son 3 Aylık Verisi ve Model B için Aralık 2017 Tahmin Değerleri



Şekil 5.16. 2017 Aralık Günlük Elektrik Talebi ve Model B DLM'ye Göre Yumuşatılmış Gözlem Değerleri

Şekil 5.17.'de ise Model B'ye göre günlük elektrik talep tahmini değerleri, tahmin ortalaması ve günlük gerçekleşen elektrik tüketimi verileri gösterilmiştir. Grafiğe bakıldığında tahmin ortalamasının durgun bir rassal yürüyüş sürecini belirttiği ve trende sahip olmadığı, tahmin değerlerinin gözlem değerlerine yakın bir dağılıma sahip olduğu görülmüştür.



Şekil 5.17. Model B'ye Göre 2017 Aralık Günlük Elektrik Talep Tahmini, Ortalaması ve Gözlem Değerleri

Tablo 5.9.'da Model B' ye göre yapılan DLM tahmininin test verisi ile tutarlılığı farklı hata tespit yöntemlerine göre kıyaslanmıştır.

Tablo 5.9. Tutarlılık ve Hata Sonuçları (Model B)

	ME	RMSE	MAE	R ²	MAPE	ACF1
Eğitim Seti	2.8025	158.7529	87.95521	0.9392	4.124753	-0.2084
Test Seti	193,4155	273,0124	196,9341	0.9829	6,6312	0,2513

5.2.5. Model C (SF, LB ve ARMA modeli)

Daha önce Bölüm 4.' Te açıklamaları ve kurulum matrislerini verilen Model C, R studio ortamında “dlm” paketi fonksiyonları ile gerçekleştirilmiştir. İlk olarak lineer müyüme süreci, sezonsal faktör ve ARMA parametreleri ile DLM kurulmuştur. Kurulan DLM'nin R “dlm” paketi ile tanımlama işlemi “dlmModPoly(2) + dlmModSeas(3)+dlmModARMA()” fonksiyonları ile DLM'nin toplamsallık özelliği sayesinde gerçekleştirilmiştir.

Kurulan dlm modelinin bilinmeyen parametreleri Model B de olduğu gibi dlmMLE ile tahmin edildiğinde gözlem denklemi için V ve sistem denklemi için W hata matrisleri aşağıdaki gibi olmuştur.

$$W = \begin{bmatrix} 11157,31 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 7,0820 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 503,4807 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,0215 & 1,4144 & -0,3534 \\ 0 & 0 & 0 & 0 & 1,4144 & 92,8093 & -23,1934 \\ 0 & 0 & 0 & 0 & -0,3534 & -23,1934 & 5,7961 \end{bmatrix}$$

$$V = [59,8143]$$

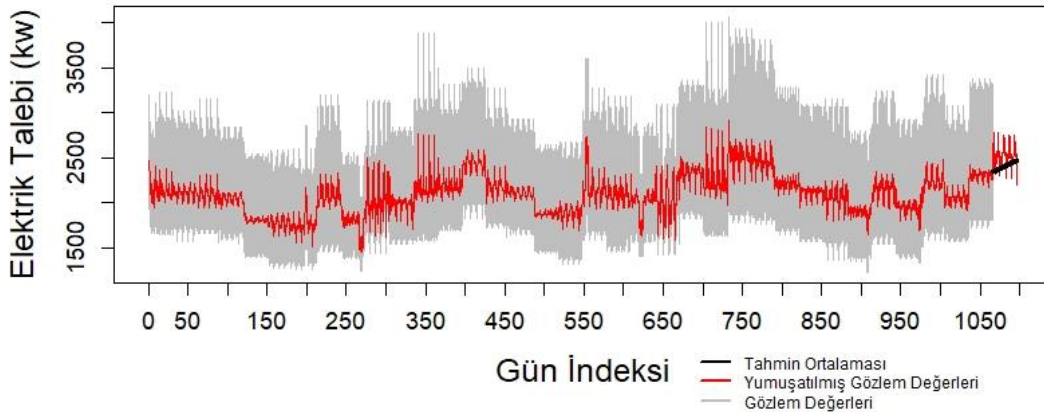
V ve W matrislerinin değerlerinin Hessian matrisine göre standart hata oranı Tablo 5.10.'da verilmiştir. Tabloya bakıldığında standart hata oranı düşük olduğundan elde edilen değerlerin gerçek değerler olduğu kabul edilmiştir.

Tablo 5.10. Model C için Sistem ve Gözlem Matrislerinin MLE Standart Hata Oranı

Parametre	MLE Değeri	Standart Hata
V	59,8143	0,4352
W_1	11157,31	0,1104
W_2	7,0820	0,3208
W_3	503,4807	0,4351

MLE ile bilinmeyen parametrelerin tahmininin ardından kurucu dlm fonksiyonunu temel alan Kalman Filtresi ile eğitim setinin tahmini yapılarak, “dlmForecast” metodu ile Kalman Filtresi sonuçlarına göre 2017 yılının Aralık ayı için günlük elektrik talep tahmin sonuçları elde edilmiştir. Şekil 5.18.’e bakıldığında tahmin ortalaması değerinin Model C’nin kurulum şekline uygun olarak artan bir trende sahip olduğu ve lineer büyüme özelliğini yansıttığı görülmüştür. Ayrıca yumuşatılmış gözlem değerlerinin de zaman serisine uygun olarak hareket ettiği belirlenmiştir.

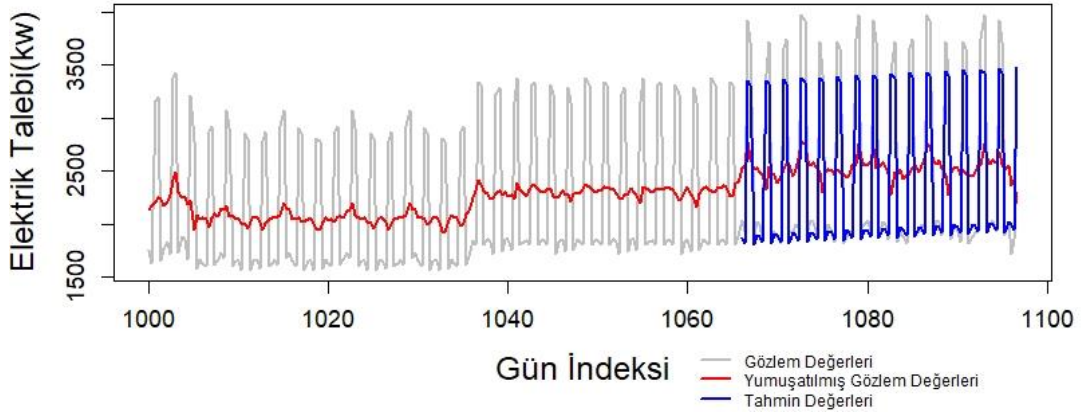
Günlük Elektrik Talep Tahmini



Şekil 5.18. Lineer Büyüme, Sezon Faktör ve ARMA Bileşenini İçeren Modele Göre Günlük Elektrik Talep Tahmini

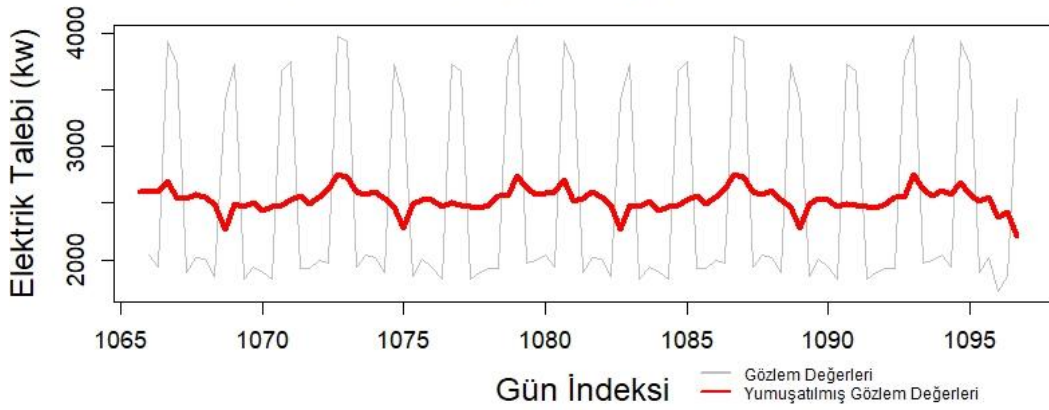
Şekil 5.19. ve Şekil 5.20.’de ise tahmin edilen Aralık ayının gerçek gözlem değerleri ve kurulan dlm modeline göre yumuşatılmış Kalman Smoothing gözlem verileri vardır. Grafikte görüldüğü üzere gerçek değerler ve yumuşatılmış veriler Model B de olduğu gibi uyum içindedir. Bunun yanında kırılma noktalarında DLM modeli iyi bir performans göstermiştir.

Günlük Elektrik Talep Tahmini



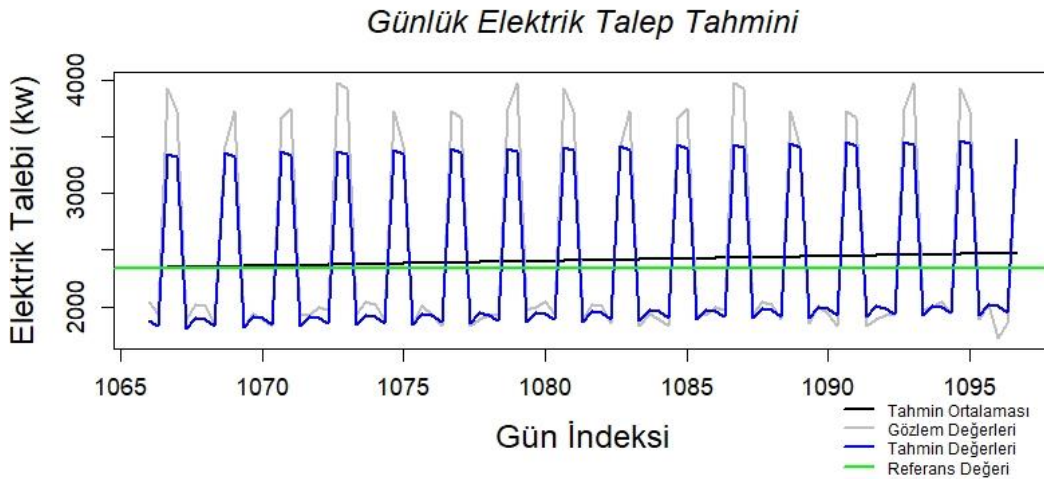
Şekil 5.19. 2017 Yılı Son 3 Aylık Verisi ve Model C için Aralık 2017 Tahmin Değerleri

DLM ile Yumuşatılmış Elektrik Talebi



Şekil 5.20. 2017 Aralık Günlük Elektrik Talebi ve Model C DLM'ye Göre Yumuşatılmış Gözlem Değerleri

Şekil 5.21.'de Model C'ye göre günlük elektrik talep tahmini değerleri, tahmin ortalaması ve günlük gerçekleşen elektrik tüketimi verileri gösterilmiştir. Grafiğe bakıldığında tahmin ortalamasının durgun bir linner büyüme sürecini yansıttığı ve artan bir trende sahip olduğu, tahmin değerlerinin gözlem değerlerine yakın bir dağılıma sahip olduğu görülmüştür.



Şekil 5.21. Model C'ye Göre 2017 Aralık Günlük Elektrik Talep Tahmini, Ortalaması ve Gözlem Değerleri

Tablo 5.11.'de Model C'ye göre yapılan DLM tahmininin test verisi ile tutarlılığı farklı hata tespit yöntemlerine göre kıyaslanmıştır.

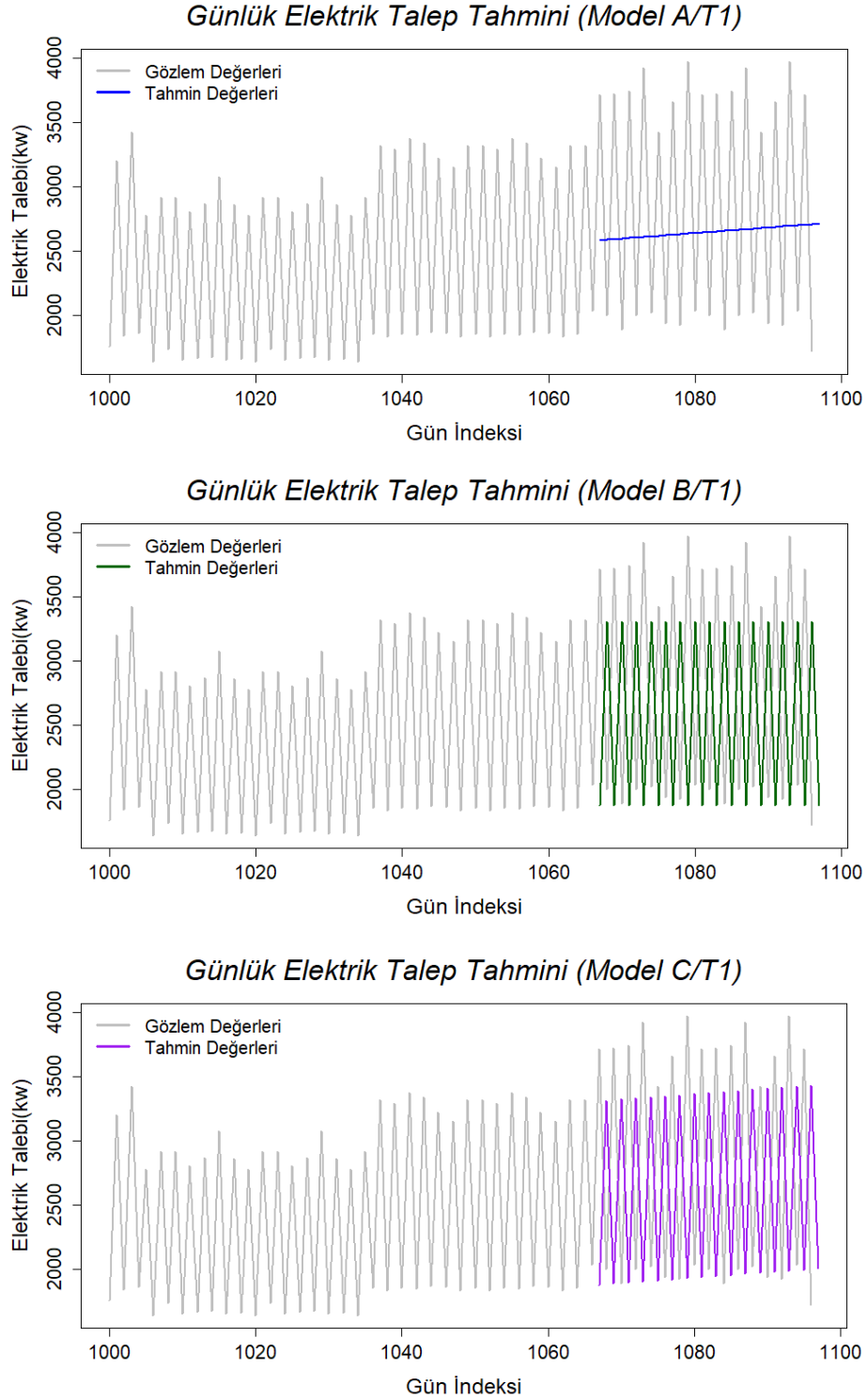
Tablo 5.11. Tutarlılık ve Hata Sonuçları (Model C)

	ME	RMSE	MAE	R ²	MAPE	ACF1
Eğitim Seti	0.7976	157.7638	92.556	0,940	4.3503	-0.28
Test Seti	124,3914	230,7329	154,637	0,979	5,0836	0,21

5.3. Bulgular ve Değerlendirme

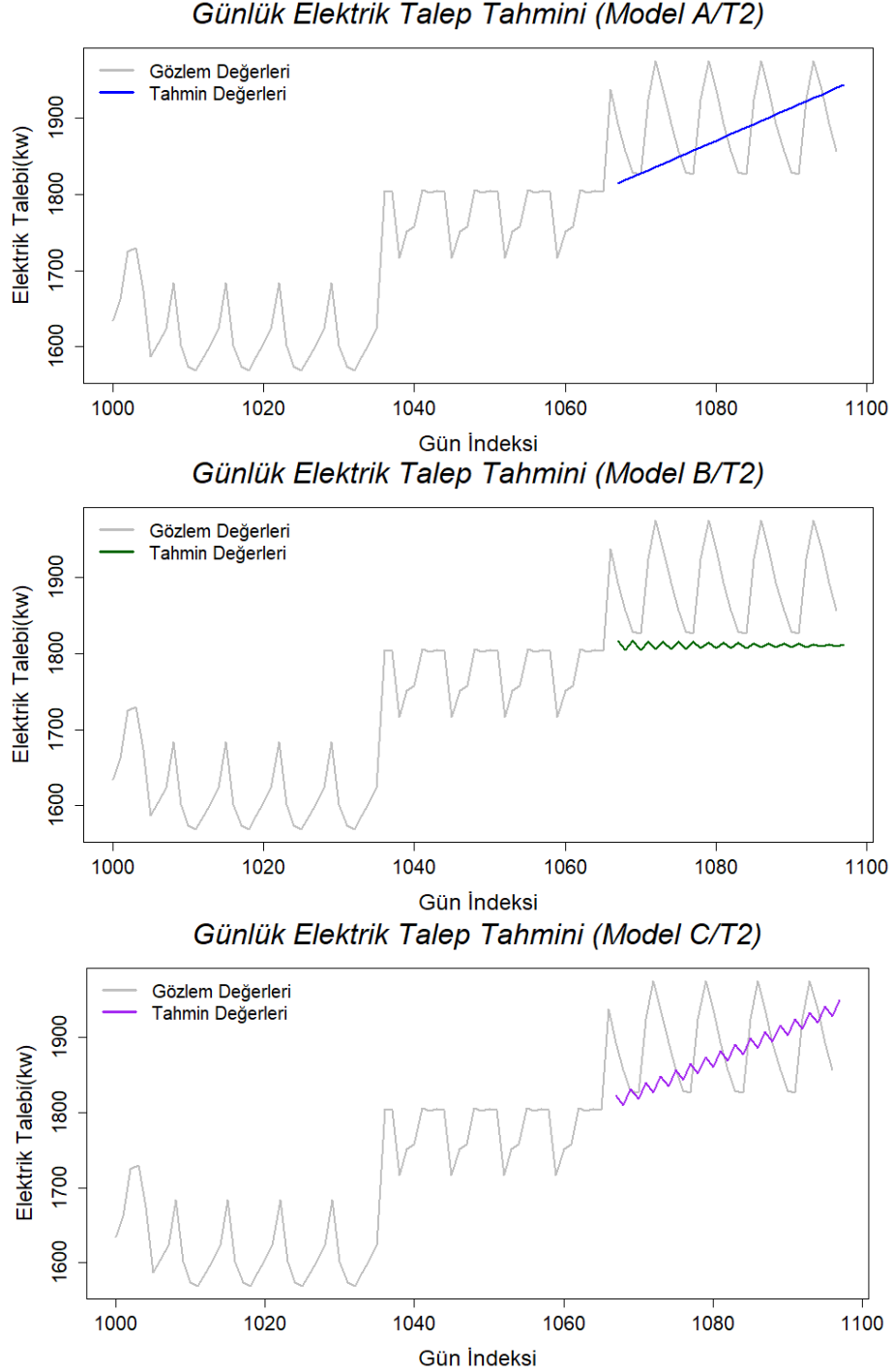
Gerçekleştirilen uygulamalara ve elde edilen sonuçlar sonuçlara bakıldığında Sakarya ili hanehalkı elektrik tüketim verileri ile yapılan tahminlerde DLM uygulamaları ARIMA yöntemine kıyasla daha başarılı olmuştur. ARIMAX modeli ile 2017 Aralık ayı için yapılan günlük üç sezondan oluşan toplamda 93 adet tahminin, gerçek veriler ile kıyaslandığında MAPE (Mean absolute percentage error-Ortalama Mutlak Yüzdellik Hata) ölçütüne göre hata oranı yüzde 25,0545 olmuş, Diğer bir ARIMA modeli uygulaması olan SARIMA modelinde ise aynı şartlar altında elde edilen MAPE hata oranı yüzde 24,957 şeklinde gerçekleşmiştir. Sadece bu iki uygulama arasında bir kıyas yapacak olursak aralarında ilişki olduğunu düşünülen hava durumu ve

haftanın günü bilgisinin ARIMAX modeli için elektrik tüketimi ile ilişkisi tahmin sonucuna ek bir katkıda bulunmamıştır.



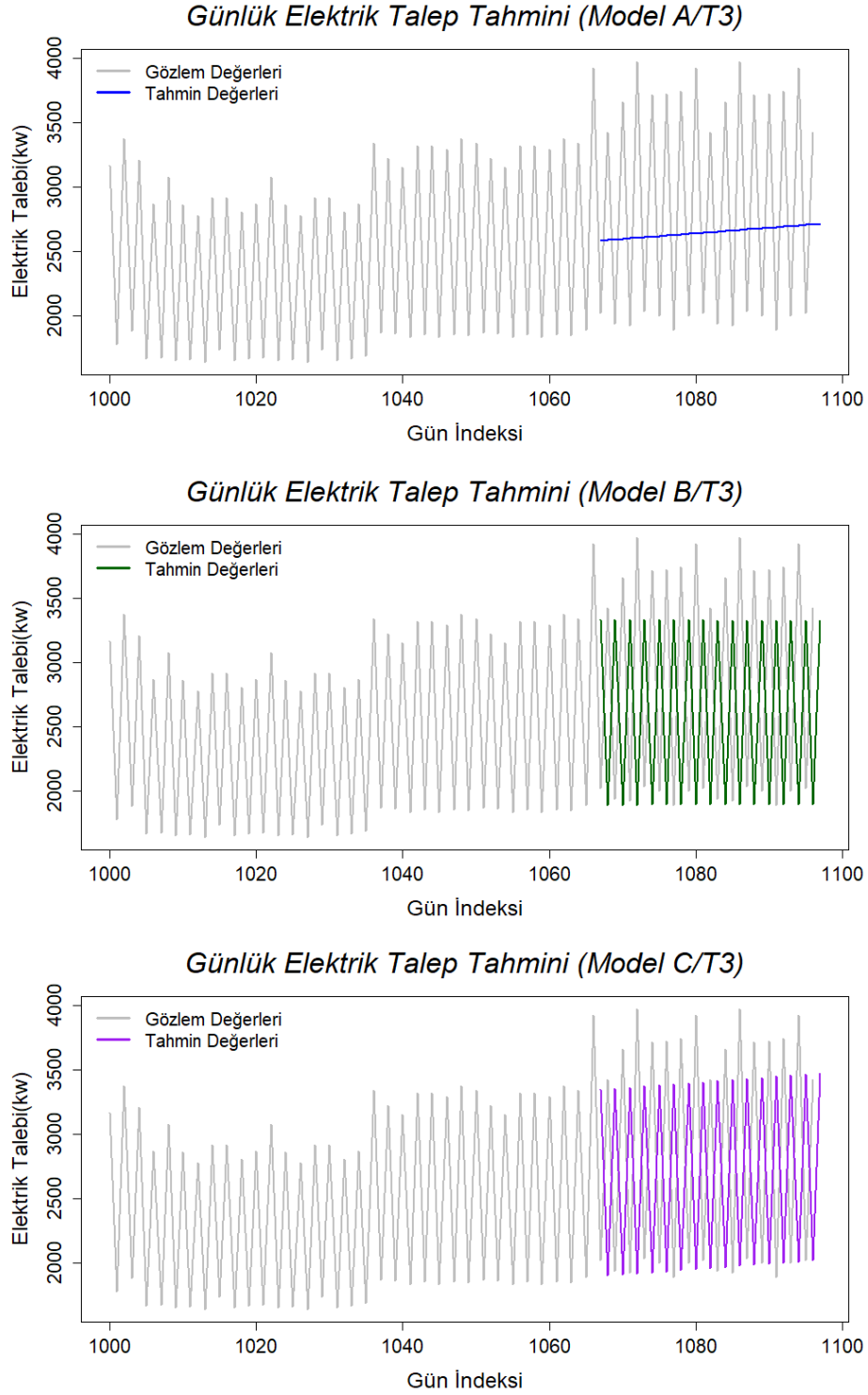
Şekil 5.22. T1 Zaman Aralığı için Model A, Model B ve Model C Aralık 2017 Tahmin Sonuçları

DLM modelleri ile yapılan tahminlerde ise ilk uygulama olan SF ve LB parametrelerini içeren Model A için 2017 yılı Aralık ayı için elde edilen 93 adet günlük bazdaki verinin, test verileri ile kıyaslandığında MAPE hata ölçüm yöntemine göre elde edilen hata oranı yüzde 22,1361 olmuştur.



Şekil 5.23. T2 Zaman Aralığı için Model A, Model B ve Model C Aralık 2017 Tahmin Sonuçları

İkinci DLM uygulaması olan SF, RW ve ARMA parametreleri ile kurulan Model B'de ise 2017 Aralık Ayı için gerçekleştirilen 93 adet günlük bazdaki veride, test verileri ile kıyaslandığında MAPE hata ölçüm yöntemine göre elde edilen hata oranı yüzde 6,6312 olmuştur.



Şekil 5.24. T3 Zaman Aralığı için Model A, Model B ve Model C Aralık 2017 Tahmin Sonuçları

Üçüncü ve son DLM uygulaması olan SF, LB ve ARMA parametrelerine sahip Model C’de, 2017 yılı Aralık ayı için elde edilen 93 adet günlük bazdaki veri, test verileri ile karşılaştırıldığında MAPE hata ölçüm yöntemine göre elde edilen hata oranı yüzde 5,0836 olmuştur.

Şekil 5.22’de T1, Şekil 5.23.’de T2 ve Şekil 5.24.’de T3 zaman dilimi için Model A, Model B ve Model C uygulamalarından elde edilen değerler grafik şeklinde sunulmuştur. Bahsedilen grafikler incelendiğinde Model A’nın T1, T2 ve T3 değerlerini belirli bir trend yönüne bağlı olarak artan bir tahmin işlemi gerçekleştirmiştir. Bunun aksine Model B ve Model C için T1,T2 ve T3 değerleri incelendiğinde ise ARMA süreçlerinin DLM şeklinde modellenmesinin sezonsallık ve trend etkileri ile beraber tahmin işlemi gerçek değerlere oldukça yakınsadığı görülmektedir. Ayrıca Model C, T2 değerlerinin tahmini konusunda gözlem değerlerine yaklaşmakta Model B’den daha etkin bir performans göstermiştir.

Elde edilen sonuçlara bakıldığında DLM ile modellenen Sakarya iline ait hanehalkı elektrik tüketimi zaman serisinin ARIMA modellerine nazaran daha iyi bir sonuç ürettiği gösterilmiştir. Bunun yanında farklı DLM modelleri arasında da süreci daha fazla kapsayan Model B (SF, RW ve ARMA) ile Model C (SF, LB ve ARMA) Model A’dan (SF ve LB) çok daha iyi sonuç üretmiştir. Bu durum ilgili bölümlerde elde edilen tahmin grafikleri Şekil 5.17. ve Şekil 5.21.’de görülmektedir. Gerçeklenen tüm uygulamaların çeşitleri ve hata oranları özet şeklinde Tablo 5.12.’de verilmiştir.

Tablo 5.12. Gerçeklenen Uygulamaların Hata Oranları Karşılaştırması

Model Adı\ Hata Ölçütü	MAPE %	MASE	RMSE	R ²
SARIMA	24,957	0,8717	875,8529	-
ARIMAX	25,0545	0,8729	873,4183	-
Model A	22,1361	0,4758	748,7778	0,273
Model B	6,6312	0,2837	273,0124	0,982
Model C	5,0836	0,3144	230,7329	0,979

BÖLÜM 6. SONUÇ VE ÖNERİLER

Tahmin, geçmişte gerçekleşmiş olayların analiz edilerek gelecek dönemlerde meydana gelebilecek olayların sonuçlarını önceden, doğru ya da doğruya yakın bir şekilde görebilme olarak tanımlanır. Genel olarak uzun, orta ve kısa vadeli olarak nitelendirilen tahminler stratejik ve kritik kararların verilmesinde kullanılmaktadır. Birçok özel şirket, kurum ya da kuruluşlar, kritik devlet birimleri; ekonomik, siyasi, üretimsel, finansal ve işletmeye yönelik konularda kararlar almadan önce konu ile ilgili analizleri, raporları ve tahmin sonuçlarını göz önüne alarak kararlarına bir yön çizmektedir.

Tahmin işlemi birçok alanda olduğu gibi enerji sektöründe de yoğun olarak kullanılmaktadır. Günümüzde enerji kullanıcılarının tamamına yakını elektrik enerjisinden faydalanmakta ve bu durum elektrik enerjisinin önemini arttırmaktadır. Elektrik diğer tüm emtia kaynaklarından farklı olarak depolanamaz ve üretildiği anda tüketilmesi gereken bir yapıya sahiptir. Ayrıca elektrik enerjisinin, kullanıcıların belirli bir an için değil istedikleri zaman erişim imkânı sağlanması gerekirken sürekli arz edilmesi ve talebi anında karşılaması beklenmektedir. Bu durum elektrik enerjisi sistemlerinin sürekliliğini sağlamak zorunda olan kurumlar için dikkate alınmadığında yüksek maliyetler getirecek ucu açık bir noktadır. Elektrik sağlayıcıları bu tip durumlarda fazla elektrik üretimi ve altyapının yeterli düzeyde tutulmasını sağlamak için bölgesel olarak araştırmalar gerçekleştirirler. Örneğin TEİAŞ Türkiye genelinde bölgesel elektrik arz ve talepleri ile ilgili 10 yıllık talep tahminleri gerçekleştirerek elektrik dağıtım şirketleri ile paylaşmaktadır.

Elektrik enerjisinin sürekliliğini sağlamak ve gereksiz maliyetlerden kurtulmak için yapılan tahminler, genellikle zaman serileri üzerinden gerçekleştirilmektedir. Zaman serisi olarak elektrik verileri günlük, aylık ve yıllık bazda sezonlara ayrılarak işleme tabii tutulur. Ancak zaman serileri ile tahmin işlemlerinde zaman serilerinin gerçekte

stokastik bir süreç olması, rassal etkiler ve durağanlık birçok tahmin uygulama yönteminde karşılaşılan bir problemdir ve tahmin işlemini güçleştirir. Bunun sebebi gerçek hayatta elde edilen verilerin birçoğu farklı etkileri bünyesinde barındırmasıdır ve bu durum zaman serilerinin modellenmesi konusunda araştırmacılara zorluk yaşatmaktadır. Elektrik tüketim verileri de bünyesinde rassal, sezonsal, trend ve konjonktürel etkileri barındırmaktadır.

Durum uzay modellerinin Bayesci bir yaklaşımı olan Dinamik Lineer Modeller, zaman serilerinin modellenmesi konusunda karşılaşılan ve yukarıda bahsedilen problemleri çözecek güçlü bir tahmin yöntemi olmuştur. Bayeci yaklaşım sayesinde Dinamik Lineer Modeller, dış etkenlere bağlı oluşan bilgiyi analiz sürecine dahil edebilir ve bunlarla ilişkili belirsizlikler olasılık dağılımları kullanılarak elde edilen zaman serisi bilgisine eklenebilir hale getirilir. Bu durum sürecin yeni olasılıklara açık olması anlamına gelmektedir. Bayesci yaklaşım sayesinde Dinamik Lineer Modeller, zaman serisi sürecinin Kalman Filtreleme yöntemi ile tahmini ve artıkların elde edilmesiyle ön bilginin oluşmasını sağlayarak tahmin için olasılık dağılımlarını oluşturur.

Bu çalışmada, zaman serisi yöntemlerinden ARIMA modelleri ve Dinamik Lineer Modeller matematiksel ifadeleri ile birlikte sunulmuş, çalışma prensipleri açıklanmıştır. SEDAŞ aracılığıyla elde edilen Sakarya iline ait hanehalkı elektrik tüketim verileri temel alınarak, iki farklı ARIMA (ARIMAX ve SARIMA) uygulaması ve üç farklı tipte Dinamik Lineer Model ile tahmin işlemi gerçekleştirilmiş böylece elektrik tüketimi konusunda talep tahmini gerçekleştiren ilgili alan araştırmacılarına Dinamik Lineer Modellerinde güçlü ve pratik bir tahmin yöntemi olduğuyla ilgili bir başvuru kaynağı sağlanmıştır.

R platformunda “dlm” paketi fonksiyonlarının kullanıldığı bu çalışmada, karmaşık bir zaman serisinin bilinmeyen parametreleri ile birlikte modellenmesi ve tahmin uygulaması yapılmıştır. İçerdiği fonksiyonlarla farklı süreçlerin karma bir şekilde modellenmesine olanak tanıyan “dlm” paketi, zaman serilerinin analizi için güçlü bir ortam sağlamaktadır. Bu şekilde bünyesinde farklı etkileri barındıran zaman serilerinin DLM ile modellenmesi, rassal etkilere ve kırılmalara duyarlı tahmin işlemini

kolaylaştırır. Özellikle ARIMA modellerinde, ani kırılmalara verilen tepkinin hata oranının yüksekliği ve karmaşık zaman serilerinin modellenmesi konusunda yaşanan zorluk, durağanlığı gerektirmeyen ve daha başarılı tahmin gerçekleştirebilecek bir uygulama yapma konusunda motivasyon sağlamıştır.

Yapılan uygulamada, sezonlara ayrılmış zaman serisi, lineer büyüme, rassal yürüyüş süreci ve ARMA modelini içeren dinamik lineer modeller şeklinde modellenerek, 1 aylık veri (2017 Aralık) için tahmin işlemi gerçekleştirilmiştir. Tahmin işlemini gerçekleştirirken bilinmeyen parametreler için maksimum olabilirlik fonksiyonunu; sürecin gözlem parametrelerini (V ve W), olasılık dağılımlarını ve artık değerleri elde etmek için Kalman filtresi kullanılmıştır.

Ele alınan üç farklı modelde sırasıyla, sezonsal faktörleri içeren lineer büyüme modeli, sezonsal faktörleri içeren rassal yürüyüş süreci ve ARMA modeli son olarak da sezonsal faktörleri içeren lineer büyüme ve ARMA modeli işlenmiştir. Her bir model için bilinmeyen parametreler tahmin edilerek belirli bir akış içerisinde tahmin işlemi adım adım ilerlenmiş ve modeller başarı ile uygulanmıştır. Ayrıca DLM modellerinin zaman serisi analizi ve tahminindeki etkisini görmek amacı ile iki farklı ARIMA (ARIMAX ve SARIMA) uygulaması yapılmıştır. ARIMA uygulamalarından ilki, zaman serisi verilerinin elde edildiği her gün için ilişkili hava durumu verileri ve haftanın gün bilgisi ile regresyon işlemini gerçekleştiren ARIMAX uygulaması, ikincisi ise zaman serisini sezonsal olarak ele alan SARIMA uygulamasıdır.

Uygulama sonucunda elde edilen tutarlılık ve hata oranları beşinci bölümde sunulmuştur. Bu sonuçlara göre dinamik lineer modeller ARIMA uygulamalarından daha iyi bir performans göstermiştir. Tablo 5.12.'de görüldüğü gibi Model C şeklinde ele alınan zaman serisi; sezonsal faktör, lineer büyüme ve ARMA süreçlerinin DLM çerçevesi altında birleştirilerek modellenmesi ile en düşük hata oranı ile tahmin işleminin gerçekleştirilmesini sağlamıştır.

DLM ile tahmin işlemi her ne kadar yüksek oranda doğruluk ile gerçekleştirilse de zaman serilerinin modellenmesi konusunda karşılaşılan karmaşıklık model kurucuya

zorluklar çıkarmaktadır. Eksik bir şekilde kurulan model zaman serisinin etkilerini tam olarak bünyesinde barındıramayacağından dolayı da tahmin hata oranı yüksek olacaktır. Model A bu duruma örnek olarak gösterilebilir. ARMA süreçlerinin de DLM'ye eklenmesi ile birlikte Model B ve C'de tahmin doğruluğu büyük oranda artmıştır.

Bu tez çalışmasında, zaman serilerinin dinamik lineer modeller ile modellenmesi konusunda edinilen deneyimler kullanılarak, elektrik tüketim verileri için çok değişkenli tahmin modellerinin geliştirilmesi ve genel bir tahmin modeli oluşturulması gibi çalışmaların yapılması planlanmaktadır. Ayrıca bu çalışmada kullanılan teorik temel ve yapılan uygulamaların ekonomi, iklim, üretim ve diğer birçok alanda araştırmacılara benzer uygulamalar yapma imkânı sağlayacağı düşünülmektedir.

KAYNAKÇA

- [1] Karagöz K., Öngörü ve Zaman Serileri Analizinde Bayesyen Yaklaşım, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, 2004.
- [2] Dawn E. Holmes, Lakhmi C. Jain, Data Mining: Foundations And Intelligent Paradigms, 978-3-642-23240-4, 2012.
- [3] Sevüktekin M. , Nargeleçekenler M. , Ekonometrik Zaman Serileri Analizi, 3. Baskı Nobel Yayınları, ISBN 978-975-591-755-9 , 2010.
- [4] Duru Ö. , Zaman Serileri Analizinde ARIMA Modelleri Ve Bir Uygulama, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, 2007.
- [5] Özek T. , Zaman Serisi Modelleri Üzerine Bir Simülasyon Çalışması, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Yüksek Lisans Tezi, 2010.
- [6] Sevüktekin, M. , Çınar, M. , Ekonometrik Zaman Serileri Analizi, 5. Baskı, Dora Yayınları, ISBN 978-605-4485-22-2 , 2017.
- [7] ATALAY A. , TORTUM A. , GÖKDAĞ M., Türkiye’de 1977-2006 Yılları Arasında Meydana Gelen Aylık Trafik Kazalarının Zamansal Analizi Pamukkale Üniversitesi, Mühendislik Bilimleri Dergisi, Cilt 18, Sayı 3, Sayfa 221-229, 2012.
- [8] Gürkan, H. Y., Markov Zinciri Monte Carlo Yönteminin Dinamik Doğrusal Modellere Uygulanması, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Bölümü, Yüksek Lisans Tezi, 2013.
- [9] Şeker S., Zaman Serileri Analizi, İstanbul Medeniyet Üniversitesi, YBS Ansiklopedi Dergisi, Cilt 2, Sayı 4, 2015.
- [10] Oğurlu, H., Matematiksel modelleme kullanarak Türkiye'nin uzun dönem elektrik yük tahmini, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 2011.

- [11] Taşcı T. , Veri Birleştirmeye Dayalı Parçacık Filtreleme ile Gerçek Zamanlı Hareket İzleme, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü, Doktora Tezi, 2014.
- [12] Bircan, H., & Karagöz, Y. (2003). Box-Jenkins Modelleri ile Aylık Döviz Kuru Tahmini Üzerine Bir Uygulama.
- [13] Benli, Y. K., Yıldız, A. (2014). Altın Fiyatının Zaman Serisi Yöntemleri ve Yapay Sinir Ağları ile Öngörüsü. Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 42(2), 213-224.
- [14] SÄRKKÄ S. , Bayesian filtering and smoothing, Cambridge University Press, 978-1-107-61928-9, 2013.
- [15] Altuntaş M. , İstatistiksel Model Seçiminde Bayesci Yaklaşımlar ve Bayes Faktörü, Sinop Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Yüksek Lisans Tezi, 2011.
- [16] LATIF Abdullatif M. , A Kalman Filter Approach To Multisite Precipitation Modeling İn Meteorology, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 1999.
- [17] Kalman, R. E., A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, 82(1), 35-45, 1960.
- [18] Çayıroğlu, İ., Kalman Filtresi ve Programlama, Fen ve Teknoloji Bilgi Paylaşımı, 2012.
- [19] Petris G., Petrone S., Campagnoli P. , Dynamic linear models. In Dynamic Linear Models with R. , Springer, 978-0-387-77237-0, 2009.
- [20] Kurt, B. (2007). Japon bıldırcınlarında (*Coturnix coturnix Japonica*) bazı yumurta kalite özelliklerine ait genetik parametre tahminlerinin Gibbs örnekleme yöntemi ile elde edilmesi (Doctoral dissertation, Adnan Menderes Üniversitesi).
- [21] Cengiz, M. A., Terzi, E., Şenel, T., & Murat, N. (2012). Lojistik Regresyonda Parametre Tahmininde Bayesci Bir Yaklaşım. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 12(1), 15-22.
- [22] ÖZBEK, L. , ÖZTÜRK, F. Lineer Olmayan Kesikli-Zaman Durum-Uzay Modelleri ve İlerletilmiş Kalman Filtresi Üzerine Bir Çalışma. VI. Ulusal Ekonometri ve İstatistik Sempozyumu, Mayıs 2003, Gazi Üniversitesi, Ankara, 2003.

- [23] J. J. F.; KOOPMAN, S. J. An Introduction to State Space Time Series Analysis. Oxford: OUP Oxford, 2007. 978-0-19-922887-4.
- [24] West M, Harrison J. , Bayesian Forecasting and Dynamic Models-Second Edition, ISBN 0-387-94725-6, 1997.
- [25] Kirchgässner, G., & Wolters, J. (2007). Introduction to modern time series analysis. Springer Science & Business Media.
- [26] Piccoli, P. P. (2015). Identification of a dynamic linear model for the American GDP (Bachelor's thesis, Università Ca'Foscari Venezia).
- [27] Davis A. R. , Brockwell J. P. , Introduction to Time Series And Forecasting Second Edition, 2002, ISBN 0-387-95351-5.
- [28] Kalman, R. E., Busy, R. S., New Results in Linear Filtering and Prediction Theory, Trans. ASME, Journal of Basic Engineering, Series D, 83, 95-108, 1960.
- [29] Petris, G. (2009). dlm: an R package for Bayesian analysis of Dynamic Linear Models. University of Arkansas.
- [30] ÜNSAL G. M., ZAMAN SERİLERİ ANALİZİNDE ARTIK TÜRLERİ VE UYGULAMALARI, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Bölümü, Yüksek Lisans Tezi, 2010.
- [31] Uyanık Ü. , Zaman Serilerinde Yapısal Kırılma ve Vergi Affi Üzerine Bir Uygulama, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Programı, Yüksek Lisans Tezi, 2014.
- [32] Demirci M. , BAYES TEOREMİ VE TEOREMİN İŞLETME BÖLÜMÜNDE UYGULAMALARI , The Journal of Academic Social Science Studies, Sayı: 43 , Sayfa: 439-462, 2016.
- [33] GUO, Wensheng. Dynamic state-space models. Journal of Time Series Analysis, 2003, 24.2: 149-158.
- [34] Steven Durlauf, L. Blume, Macroeconometrics and Time Series Analysis, pp.35-45, 2010, 978-230-23885-5.
- [35] Aoki, M., State Space Modeling of Time Series, Springer-Verlag, 1990.
- [36] Boltürk, E. Elektrik Talebi Tahmininde Kullanılan Yöntemlerin Karşılaştırılması (Doctoral dissertation, Fen Bilimleri Enstitüsü).

- [37] J.J. Allaire, RStudio: Integrated Development Environment for R , The R User Conference 2011, University of Warwick.

ÖZGEÇMİŞ

Emre Doruk, 17.10.1993'de Gümüşhane'de doğdu. İlkokul ve ortaokul eğitimini Gümüşhane'de tamamladı. 2011 yılında Gümüşhane Fen Lisesi'nden mezun oldu. 2011 yılında başladığı Sakarya Üniversitesi Bilişim Sistemleri Mühendisliği Bölümü'nü 2015 yılında bitirdi. Aynı zamanda 2012 yılında başladığı Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümünde Çift Anadal programını 2016 yılında bitirdi. 2015 yılında Sakarya Üniversitesi Bilgisayar ve Bilişim Sistemleri Mühendisliği anabilim dalında yüksek lisans eğitimine başladı. 2017 yılında Rize Recep Tayyip Erdoğan Üniversitesi'nde Araştırma Görevlisi olarak çalışmaya başladı akabinde yüksek lisans eğitimine Sakarya Üniversitesi Bilgisayar ve Bilişim Sistemleri Mühendisliği Anabilim dalında devam etti. Halen Sakarya Üniversitesi Bilişim Sistemleri Mühendisliği Bölümü'nde Araştırma Görevlisi olarak görev yapmaktadır.