

T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Parametrik Olmayan MIDAS Regresyon

DOKTORA TEZİ
Mesut ÖZDİNÇ

İstatistik Ana Bilim Dalı
İstatistik Programı

Tez Danışmanı: Prof.Dr.Gülay BAŞARIR

Aralık 2022

Sevgilim eşim Selcen'e ve kızım Zeynep'e,



PARAMETRİK OLMAYAN MIDAS REGRESYON

ÖZET

Olgular arasındaki zaman ve frekans farklılıkları, bunların modellenmesine yönelik arayışları ilginç kılar. MIDAS(Mixed Data Sampling) regresyon yöntemleri, yüksek frekanslı bağımsız değişkenlerin düşük frekanslı bağımlı değişkenle ilişkisi sağlamada önemli katkılar içeren bir yaklaşım ortaya koymuştur. Bu çalışmanın amacı, parametrik olmayan MIDAS regresyon yöntemlerini incelemek ve gelişmeye açık konuları ele almaktır. Bu amaç doğrultusunda, izleyen bölümlerde şu konular incelenmiştir.

Birinci bölümde, MIDAS regresyon modelinin tanımı, işlevi ve kullanım alanları tanıtılmıştır. Yüksek frekanslı değişkenlerin dönüştürülmesinde yaygın kullanılan, ortalama, toplama ya da son değer alma gibi yöntemlerin sakıncalarına ve MIDAS regresyon yönteminin çözüm önerilerine değinilmiştir. Literatürde yer alan MIDAS regresyon türlerinden ve kullanım alanlarından bahsedilmiştir.

İkinci bölüm, ağırlıklı MIDAS regresyon öncesi karma frekanslı diğer yöntemler üzerine literatür özeti niteliğindedir. MIDAS regresyon öncesi öneriler yöntemler, Zaman Toplulaştırma ve İterasyon Yöntemi, Köprü Denklemi Yöntemi ve Durum Uzayı Modeli alt başlıkları altında incelenmiştir. Bunlardan, Zaman Toplulaştırma ve İterasyon, regresyon analizinde bağımlı ve bağımsız değişkenin gözlem sayısını eşitlemede en yaygın kullanılan yöntemdir. Bu yöntem basit ve kullanışlı olması nedeniyle sıkça tercih edilir. Ancak, hem yüksek frekanslı serinin gecikmeli değerlerini, hem de değişkenliğini gereği gibi yansıtamaz. Bu nedenle de neden-sonuç ilişkilerinin bozulmasına neden olabilir. Öte yandan, köprü denklemi yöntemi, yüksek frekanslı değişkenleri düşük frekanslı değişkene bağlayan doğrusal regresyon modelleri ile iterasyon yapar. Son olarak, Durum Uzayı Modeli, düşük frekanslı değişkenin eksik gözlemlerle yüksek frekanslı bir değişken olarak modellenmesine dayanır. Kalman filtresi kullanılarak eksik gözlemler elde edilir. MIDAS regresyon modeli ise yüksek frekanslı serinin bir ağırlıklandırma işleviyle dönüştürülmesi esasına dayanır.

Üçüncü bölüm, parametrik MIDAS regresyon modeli ve uzantıları konusunda bir literatür özeti niteliğindedir. MIDAS regresyon yöntemi bir zaman serisi analizi olması nedeniyle dinamik regresyon modelinin özellikleri taşır. Bu kapsamda, gecikmesi dağıtılmış (DL) ve otoregresif gecikmesi dağıtılmış modellerin işleyişi ele alınmıştır. Ardından, MIDAS regresyon modeli, temel kavramları ve ağırlıklandırma işlevinin nitelikleri incelenmiştir. Ayrıca, yüksek frekanslı serinin gecikmeli değerleri ile frekansını gösteren veri matrislerinin işleyişi açıklanmıştır. Parametrik ağırlıklandırma işlevleri olan, Almon Polinomu, Üssel Almon, Beta ve Adımsal ağırlıklandırma işlevlerinin yapısı cebirsel ve görsel olarak aktarılmıştır. Üssel Almon işlevinin dayandığı Almon polinomunun, gecikme uzunluğunun az sayıda parametreye indirgenmesine ilişkin cebirsel ispatlarına yer verilmiştir. MIDAS regresyon modelinin, diğer uzantı ve türleri hakkında genel bilgilere yer verilmiştir. Son olarak

parametrik yöntemde kullanılan bilgi ölçütlerinin, 3n uygun gecikme uzunluğunun elde edilmesine dair tartışmaları sunulmuştur.

Dördüncü bölümde, parametrik olmayan regresyon yöntemleri üzerine literatür incelemesi ve MIDAS regresyon yöntemlerine uyarlanması aktarılmıştır. Splayn regresyon türleri ve çekirdek regresyon modelleri ele alınmıştır. Ardından, uygulamada yer alan Düzleştirilmiş MIDAS regresyon modelinin işleyişi açıklanmıştır. Çekirdek regresyon yöntemleri hakkında genel bilgi verilmiş, MIDAS regresyon yönteminde potansiyel kullanılabilirliği tartışılmıştır.

Beşinci bölümde, benzetim deneyi ile parametrik ve parametrik olmayan MIDAS regresyon yöntemleri sınanmıştır. Benzetim deneyinde, literatürde bu amaçla geliştirilmiş özel ağırlıklandırma işlevleri rassal veri üretimi için kullanılmıştır. Yüksek frekanslı değişken için, üssel azalan, tümsek, doğrusal azalan ve çevrimsel olmak üzere dört farklı ağırlıklandırma işlevi yürütülmüştür. Rassal veri setleri, artan-azalan, kısa-uzun dönem gecikme uzunluğu, varyans ölçeklendirilmesi yoluyla, yöntemler stres testine tabii tutularak sonuçları tartışılmıştır.

Altıncı bölümde, güncel bir uygulama olarak, Türkiye’de kaydedilen COVID-19 günlük vakaları düşük frekanslı bağımlı değişken ve yüksek frekanslı bağımsız değişken niteliğindeki saatlik paylaşılmış (yüksek akışkan nitelikli) COVID-19 içerikli Twitter (tweet ve retweet) mesajlarının hacmi üzerinden, MIDAS regresyon yöntemleri yürütülmüştür. 20 Ocak 2020 ve 31 Aralık 2021 tarihleri arasında, açık veri kaynaklarından hem bağımlı, hem de bağımsız değişkenin verileri elde edilmiştir. İki yıla yakın süre içerisinde, özellikle yüksek frekanslı Twitter mesajları hacim olarak büyük veri niteliğindedir. Elde edilen küresel çapta paylaşılmış, 1 milyar 960 bin Twitter mesajı 3.4 terabyte büyüklüğündedir. Bundan analize dahil edilen Türkçe 18.7 milyon Twitter mesajı bulunmaktadır.

Yedinci bölümde, MIDAS regresyon modelleri hakkında bulgular, sonuçlar ve ileriye dönük çalışma konuları tartışılmıştır. Parametrik olmayan MIDAS regresyon modeli ile elde edilen sonuçlara göre, yüksek frekanslı saatlik sosyal medya paylaşım verilerinden, bir süre önceden vaka tahmini imkanı olduğu görülmüştür. Özellikle, pandeminin başlangıç evresinde gelişen belirsiz ve olağandışı duruma intibaksızlık nedeniyle, vaka sayıları verilerinde var olan gürültü faktörüne rağmen, yüksek frekanslı değişken yardımıyla yapılan bağımlı değişken tahmini oldukça açıklayıcı olmuştur. Pandeminin ilerleyen aşamalarında, vaka sayıları ve model tahmini arasında daha da uyumlu bir seyir gözlenmiştir. Sonuç olarak, gürültü ve eksik verinin ve bilginin söz konusu olduğu durumlarda, model bir erken uyarı sistemi olarak çalışma becerisine sahiptir.

Düzleştirilmiş MIDAS regresyon modeli, cezalandırılmış splayn regresyon modeli baz alınarak kurulmuştur. Parametrik olmayan çekirdek regresyon modelleri Nadaraya-Watson, Priestly-Choa ve Gasser-Müller modellerinde MIDAS regresyon için uyarlanabileceği düşünülmektedir. Ancak bu uyarılama için, bir dizi teorik ve uygulamalı sorunlarla karşılaşmıştır. Parametrik olmayan MIDAS regresyon modelinde, hesaplama yükü parametrik yöntemle göre çok daha fazladır. Çekirdek regresyon modellerinin, yüksek frekanslı serinin gecikmeli değerlerini de içerecek biçimde genişletildiğinde,

hesaplama yükü çok daha fazla olmaktadır. İleri çalışmalarda, hesaplama yükünün kaldırılabilmesi için paralel hesaplama modellerinin ve uygun donanım altyapısının gereği ortadadır. Ayrıca, çekirdek regresyon modellerini MIDAS regresyona uyarlarken, yüksek frekanslı değişkenin iki aşamalı filtreden geçirilmesi öngörülmesine rağmen, aşamalar arasında geçişte, tutarsızlıklar ve bilgi ölçütüne uyarlama zorlukları halen çözüm beklemektedir.

Parametrik olmayan MIDAS regresyon, özellikle akışkan büyük veri setleri için artan bir öneme sahiptir. Akışkan veri sistemleriyle uyumlu çalışan, parametrik olmayan modeller, karar alıcılar için kritik olabilecek bilgiler üretme potansiyeline sahiptir. Şimdinin kestirimi, ekonomiden risk yönetimine, salgın hastalıklarda erken uyarı sistemlerinden kriz yönetime kadar geniş bir yelpazede, parametrik olmayan MIDAS regresyona kullanım alanı sağlamaktadır. Şimdinin kestirimi, zaman zaman hayati olabilecek durumlara ilişkin karar destek mekanizmalarına ilgili bilgileri aktarabilir.

NONPARAMETRIC MIDAS REGRESSION

ABSTRACT

In the course of life, individuals often intuitively try to make sense of a given situation and its future with the information that they can observe in their proximity. On the other hand, such observations usually constitute high-frequency variables. While observations are of high-frequency nature as a variable, the phenomenon that is tried to be defined is often a low-frequency dependent variable. The time and frequency differences between the variables make modeling them an attractive quest. The vast increase in the size of observable data has led to an emerging need for non-parametric regression methods. Developing non-parametric MIDAS regression models enables better comprehension of numerous different phenomena in their current and future states with promising contributions to their solutions. Nowcasting has the potential to inform decision-making mechanisms in life or death matters from time to time.

This study aims to examine MIDAS regression methods that allow for the analysis of high-frequency variables and to discuss the areas of improvement relevant to its non-parametric types. In this context, developing a non-parametric MIDAS model and testing kernel regression methods are also aimed. The performance of the established model is also tested through a simulation experiment. Objectives also include applying the developed non-parametric MIDAS regression to real big data to test it with a timely and relevant topic. Thus, the developed thesis involves both theoretical and practical contributions.

An analytical analysis of parametric and non-parametric MIDAS regression methods is presented together with a comparison of results obtained from the simulation experiment designed. Then, the methods are further tested with real-life event-based data for a discussion of the arrived conclusions.

First, the definition and use areas of the MIDAS regression model are introduced, and related literature is summarized. Methods proposed prior to MIDAS are examined in sub-headings of Temporal Aggregation and Iteration Method, Bridge Equation Method, and State Space Model. Of these, Temporal Aggregation and Iteration stand out as the most widely adopted, equating the number of observations of the dependent and independent variables in regression analysis. This method is often preferred because it is simple and convenient. However, it cannot adequately reflect the delayed values of the high-frequency series and the variability simultaneously. Thus, it may lead to the deterioration of the cause-effect relationships. In addition, the bridge equation method iterates with linear regression models that connect high-frequency variables to low-frequency variables. Finally, the State Space Model is based on modeling the low-frequency variable as a high-frequency variable with missing observations. Missing observations are obtained using the Kalman filter. The MIDAS regression model, on the other hand, is based on transforming the high-frequency series with a weighting function.

MIDAS regression method, as a time series analysis, has the characteristics of the dynamic regression model, as explained in the second part. In this context, the operation of distributed lag (DL) and

autoregressive distributed lag models are discussed. Third, the MIDAS regression model, its basic concepts, and the attributes of the weighting function were examined. In addition, the operation of the data matrices showing the lagged values and frequency of the high-frequency series is explained. The structure of the Almon Polynomial, Exponential Almon, Beta, and Stepwise weighting functions, which are the parametric weighting functions, are explained algebraically and visually. Algebraic proofs of the Almon polynomial, on which the exponential Almon function is based that reduce the lag length to a small number of parameters are given. General information about other extensions and types of the MIDAS regression models is given. Finally, discussions on optimizing the lag length of the information measures used in the parametric method are presented.

Fourth, a literature review on non-parametric regression methods and their adaptation to the MIDAS regression methods is presented. Spline regression types and kernel regression models are discussed. Then, the operation of the SLS MIDAS regression model in the application is explained. General information about kernel regression methods is given, with its potential usability in the MIDAS regression method is discussed.

Fifth, parametric and non-parametric MIDAS regression methods were tested with a simulation experiment. In the simulation experiment, special weighting functions developed for this purpose in the literature were used for random data generation. For the high-frequency variable, four different weighting functions, including exponential declining, hump-shaped, linear declining, and cyclical, were carried out. Random data sets, increasing-decreasing, short-long-term lag length, variance scaling, and methods were subjected to stress tests, and the results were discussed.

Sixth, for a timely and relevant application, MIDAS regression methods were implemented to the volume of hourly shared (high stream) Twitter (tweet and retweet) messages with COVID-19 content consisting of the high-frequency independent variable and the low-frequency dependent variable of daily COVID-19 cases recorded in Turkey. Datasets obtained from open data sources for independent and dependent variables covered the period of January 20, 2020, and December 31, 2021. Almost two years of data, especially for high-frequency Twitter messages, led to massive big data in volume. The obtained global data on 1 billion 960 thousand Twitter messages shared constitutes 3.4 Tera Bytes in size. Of these, 18.7 million Twitter messages posted in the language of Turkish were used in the model.

According to the results obtained with the non-parametric MIDAS regression model application, a possibility of case prediction from the high-frequency hourly social media sharing data was revealed in the seventh (final) part. In particular, the dependent variable estimation with the help of a high-frequency variable was quite explanatory, despite the noise factor present in the case numbers data, due to the lack of adaptation to the uncertain and extraordinary situation that had emerged in the initial phases of the pandemic. In the later stages of the pandemic, an even more congruent trend was observed between the number of cases and the model prediction. As a result, the model has the proven ability to serve as an early warning system in situations where noise and missing data and information are involved.

The SLS MIDAS regression model was built based on the penalized spline regression model. It is believed that non-parametric kernel regression models Nadaraya-Watson, Priestly-Choa, and Gasser-Müller models can be adapted for MIDAS. However, several theoretical and practical problems encountered hindered this adaptation. The computational load is much higher in the non-parametric MIDAS regression model when compared to the one in the parametric method. When the kernel regression models are extended to include the lagged values of the high-frequency series, the computational load becomes even much higher. Future studies require relevant parallel computing models and appropriate hardware infrastructures to tackle the inherent computational challenges. In addition, while adapting the kernel regression models to MIDAS, although the high-frequency variable is expected to be filtered in two stages, inconsistencies and difficulties in adapting the information criterion remain unsolved.

Non-parametric MIDAS regression is of increasing importance, especially for stream big data sets. Non-parametric models that work in harmony with stream data systems have the potential to generate information that may be critical for decision-makers. The nowcasting provides a wide range of possible uses of non-parametric MIDAS regression ranging from economics, risk management, early warning systems in pandemics, and crisis management.

The SLS MIDAS regression model was built based on the penalized spline regression model. It is believed that non-parametric kernel regression models Nadaraya-Watson, Priestly-Choa, and Gasser-Müller models can be adapted for MIDAS. However, several theoretical and practical problems encountered hindered this adaptation. The computational load is much higher in the non-parametric MIDAS regression model when compared to the one in the parametric method. When the kernel regression models are extended to include the lagged values of the high-frequency series, the computational load becomes even much higher. Future studies require relevant parallel computing models and appropriate hardware infrastructures to tackle the inherent computational challenges. In addition, while adapting the kernel regression models to MIDAS, although the high-frequency variable is expected to be filtered in two stages, inconsistencies and difficulties in adapting the information criterion remain unsolved.

ÖNSÖZ

Doktora tezi çalışmam, uzun, yoğun emek gerektiren bir yarıyla yorucu, diğer yandan keyifli bir öğrenme sürecinin ürünü. Görece ileri bir yaşta başladığım doktora serüvenimde, doğrudan ve dolaylı o kadar emek ve sabır var ki...

Bu tezin oluşumu sürecinde, etkili yönlendirmeleriyle yolumu kaybetmeden ilerlememi sağlayan, yapıcı eleştirileriyle beni yönlendiren ve bana inanan kıymetli tez danışmanım Prof.Dr. Gülay BAŞARIR'a teşekkür ederim.

Program boyunca regresyon teorisinin derinliklerinde bana yol gösteren, basamakları tırmanmakta destek olan, zaman ayıran ve emek veren Prof.Dr. Aydın ERAR, benim gelişimde özel bir yere sahip. Teori ve uygulama arasında bağları öğrenmemde desteği yanında, sıkça görüştüğüm ve rehberliğinden çok yararlandığım Prof.Dr.Eylem DENİZ'e de çok teşekkür ederim. Zaman serilerini kavramamdaki katkısı için Prof.Dr. Barış Aşıkil'e ve model kurulumu ve kodlanmasında her zaman rehberliğinden yararlandığım Dr. Elif Özge ÖZDAMAR'a teşekkür ederim. İstatistik Bölümünde aldığım derslerde, katıldığım seminerlerde, kişisel sohbetlerimizde gelişimime katkıda bulunan değerli hocalarıma emekleri ve arkadaşlarıma destekleri için çok teşekkür ediyorum.

İstatistik doktora programına başvurmak için kaygılarımı giderip, beni yapabileceğime inandıran ve teşvik eden, lisans yıllarından sınıf arkadaşım Prof.Dr Ünal ÖZDEN'in bu süreçte benim için özel bir yeri var. Gençlik yıllarımızda birlikte akademiye başlayıp, yoluna devam eden Prof.Dr. Serhat YANIK her zaman dostum olarak, zorlandığımda her zaman yanımda oldu. Pandemi döneminde eve kapanmışken, eski meslektaşım ve sevgili dostum, ortak araştırma partnerim Prof.Dr. Kerem ŞENEL'in desteğinin de önemi büyük.

Tez konusunu arayışında iken konuya ilk ilgi duymamı sağlayan, İsveç'te tanıştığım Prof.Dr.Pär SJÖLANDER'e ve Prof.Dr. Kristofer MÅNSSON'a ilgi çekici sohbetleri için teşekkür ederim. Prof.Dr.Nihat KASAP ile aynı ekipte ortak akademik çalışmalar yapma olanağına sahip olduğum ve tezim boyunca desteğini aldığım için ayrıca şanslıyım.

Tez sürecinin meşakkatli iniş çıkışlarında, aynı zamanda İsveç'te zorlu göçmenlik maceramızı birlikte yaşadığım, yol arkadaşım sevgili eşim, aynı zamanda bilim insanı Selcen'in büyük desteği benim için özellikle çok değerli. Kızım Zeynep, her zaman kendine özgü tarzıyla hayatıma ilk girdiği günden itibaren babalık yolculuğumda beni sürekli evrimleştiriyor. Zeynep ile aynı dönemde MSGSÜ'de öğrenci olmamız hayatta az rastlanır güzel bir tesadüf. Tez sürecinde ömrü yettiğince beni hiç yalnız bırakmayan, kedimiz Tıpır'ın duygusal desteği çok değerliydi ve aramızdan ayrılışının ardından boşluğunu çok hissettim.

Aralık 2022
Mesut ÖZDİNÇ

İçindekiler

Özet	viii
Abstract	xii
Önsöz	xvi
Kısaltmalar	xxiii
Semboller	xxv
Tablo Listesi	xxvii
Şekil Listesi	xxix
1 Giriş	1
2 Karma Frekanslı Analiz Yöntemleri	5
2.1 Karma Frekanslı Yöntemlere Genel Bakış	5
2.1.1 Zamansal Toplulaştırma ve İterasyon Yöntemi	6
2.1.2 Köprü Denklemleri Yöntemi	7
2.1.3 Durum Uzayı Modeli	8
2.1.4 MIDAS Regresyon Modelleri	9
2.1.5 Parametrik Olmayan MIDAS Regresyon	10
2.1.6 Modellerin Karşılaştırılması	10
3 MIDAS Regresyon Modelleri	13
3.1 Zaman Serisi Regresyonu	13
3.1.1 Klasik Regresyon Modelinin Varsayımları	13
3.1.2 Dinamik Regresyon Modelinin Türleri	14
3.1.2.1 Gecikmesi Dağıtılmış (DL) Model	15
3.1.2.2 Otoregresif Gecikmesi Dağıtılmış Model (ADL)	16
3.2 MIDAS Regresyon Modelinin Yapısı	17
3.2.1 MIDAS Regresyon Modelinin Temel Kavramları	18
3.2.2 MIDAS Gecikmeli Veri Matrisleri	19
3.3 MIDAS Ağırlıklandırma İşlevleri	21

3.3.1	Almon Gecikme Polinomu Ağırlıklandırması	21
3.3.2	Üssel Almon Gecikme Polinomu Ağırlıklandırması	24
3.3.3	Beta İşlevi	27
3.3.4	Adımsal Ağırlıklandırma	28
3.3.5	Diğer Ağırlıklandırma İşlevleri	28
3.4	MIDAS Regresyon Modellerindeki Yöntemsel Farklılıklar ve Sınıflandırma	29
3.4.1	Temel MIDAS Regresyon Modeli	29
3.4.2	Otoregresif Gecikmesi Dağıtılmış MIDAS (ADL-MIDAS)	30
3.4.3	Kısıtsız MIDAS Regresyon Modeli	31
3.4.4	Faktör MIDAS	32
3.4.5	Adımlama Yöntemi	33
3.4.6	MIDAS Modellerinin Diğer Uzantıları	33
3.5	Model Seçimi ve Bilgi Ölçütleri	34
4	Parametrik Olmayan MIDAS Regresyon	35
4.1	Temel kavramlar	35
4.1.1	Parametrik yaklaşım	35
4.1.2	Parametrik olmayan yaklaşım	36
4.1.3	Yarı-Parametrik yaklaşım	36
4.2	Splayn regresyon modelleri	37
4.2.1	Temel Splayn Regresyon Modelleri	37
4.2.1.1	Doğrusal Splayn Regresyon	37
4.2.1.2	Karesel Splayn Regresyon	39
4.2.1.3	Kübik ya da Doğal Splayn Regresyon	40
4.2.2	Düzleştirilmiş Splayn Regresyon Modelleri	41
4.2.2.1	En uygun bant genişliği λ değeri	45
4.2.2.2	Düğüm Noktası Sayısı	46
4.3	Düzleştirilmiş MIDAS regresyon modeli	47
4.4	Çekirdek regresyon yöntemleri	49
4.4.1	Nadaraya-Watson Çekirdek kestiricisi	52
4.4.2	Priestley-Chao Çekirdek kestiricisi	53
4.4.3	Gasser-Müller Çekirdek kestiricisi	53
4.5	MIDAS çekirdek regresyon modeli üzerine tartışma	54
5	Benzetim Deneyi	57
5.1	Deney tasarımı	57
5.2	Rassal Veri Üretim Süreci	58
5.3	Ağırlıklandırma İşlevleri	59
5.3.1	Üssel azalan ağırlıklandırma işlevi	59
5.3.2	Tümsek azalan ağırlıklandırma işlevi	60
5.3.3	Doğrusal azalan ağırlıklandırma işlevi	62
5.3.4	Çevrimsel ağırlıklandırma işlevi	63
5.4	Yöntemlerin Uygulanması	64
5.5	Benzetim deneyi sonuçları	64
5.5.1	Deney Sürecinde karşılaşılan zorluklar ve sınırlandırmalar	65
5.5.2	Deney sonuçlarının karşılaştırılması	65
5.5.2.1	Örneklem hacmindeki değişimin etkisi	65

5.5.2.2	Gecikme uzunluęu deęişiminin etkisi	67
5.5.2.3	Deęişkenlik ölçek faktörü (α) deęişiminin etkisi	70
5.5.2.4	Göreceli performans incelemesi	72
5.6	Sonuç	74
6	Uygulama	75
6.1	Veri yapısı ve işlem süreçleri	75
6.2	Model seçimi	77
6.3	Parametrik MIDAS-NLS modeli sonuçları	77
6.4	Parametrik olmayan MIDAS-SLS modeli sonuçları	80
6.5	Sonuçlar üzerine deęerlendirme	81
7	Sonuç ve Tartışma	83
A	Kodlamalar	87
A.1	Parametrik Aęırlıklandırma İşlevleri	87
A.1.1	Almon Polinomu İşlevi	87
A.1.2	Üssel Almon İşlevi	87
A.1.3	Beta İşlevi	87
A.1.4	Adımlama İşlevi	88
A.2	Rassal veri üretim süreci	88
A.2.1	Rassal Veri Seti Üretimi	88
A.2.2	Veri Üretimi İçin Aęırlıklandırma İşlevi Modelleri	89
A.2.2.1	Üssel Azalan Aęırlıklandırma İşlevi	89
A.2.2.2	Tümsek Aęırlıklandırma İşlevi	89
A.2.2.3	Doęrusal Azalan Aęırlıklandırma İşlevi	89
A.2.2.4	Çevrimsel Aęırlıklandırma İşlevi	90
	Kaynakça	91

Kısaltmalar

AIC	: Akaike Information Criteria (Akaiki Bilgi Ölçütü)
AICc	: Akaike Information Criteria corrected (Düzeltilmiş Akaiki Bilgi Ölçütü)
AR	: Autoregressive (Oto regresif)
ARDL	: Autoregressive Distributed Lag (Oto regresif Gecikmesi Dağıtılmış)
BIC	: Bayesian Information Criteria (Bayesçi Bilgi Ölçütü)
B-Splines	: Basis Splines (Temel Splayn)
CV	: Cross Validation (Çapraz Geçerlilik)
DL	: Distributed Lag (Gecikmesi Dağıtılmış)
EKK	: En Küçük Kareler (Least Square Error)
EM	: Expectation Maximum (En Yüksek Beklenti)
GCV	: Genarilized Cross Validation (Genelleştirilmiş Çapraz Geçerlilik)
HKO	: Hata Kareler Ortalaması (Mean Squared Error)
HKT	: Hata Kareler Toplamı (Sum of Squared Error)
IDD	: Independent and Identically Distributed (Bağımsız Özdeş Dağılım)
MA	: Moving and Average (Hareketli Ortalama)
MIDAS	: Mix Data Sampling (Karma Frekanslı Veriler)
ADL-MIDAS	: Autoregressive Distributed Lag - MIDAS (Oto regresif Gecikmesi Dağıtılmış MIDAS)
DL-MIDAS	: Distributed Lag - MIDAS (Gecikmesi Dağıtılmış MIDAS)

- MIDAS-GARCH** : **MIDAS - Generalized AutoRegressive Conditional Heteroskedasticity**
(MIDAS - Genelleştirilmiş Otoresif Koşullu Değişken Varyans - MIDAS)
- MIDAS-NLS** : **MIDAS Nonlinear Least Square** (Doğrusal Olamayan En Küçük Kareler - MIDAS)
- ö.d.i** : **öteki değerler için** (Otherwise)
- P-Splines** : **Penalized Splines** (Cezalandırılmış Splayn)
- SLS** : **Smoothing Least Square** (Düzleştirilmiş EKK)
- UMIDAS** : **Unrestricted MIDAS** (Kısıtsız MIDAS)



Semboller

Σ	: Toplam
y^L	: Düşük frekanslı bağımlı değişken
x^H	: Yüksek frekanslı bağımsız değişken
x^L	: Düşük frekanslı bağımsız değişken
m	: Yüksek frekanslı serinin yinelenme sayısı
\mathbf{X}	: Bağımsız değişken matrisi
ε	: Hata terimi
β	: Regresyon katsayıları vektörü
\mathbf{H}	: Şapka matrisi
exp	: Üssel
Γ	: Gamma dağılımı
θ	: Theta parametresi, MIDAS ağırlık işlevi parametresi
λ	: Lambda parametresi, bant genişliği ya da lagrange çarpanı
κ	: Düğüm noktaları vektörü
w_i	: Ağırlık değerleri
$t - i$: Gecikme uzunluğu
α	: Değişkenlik ölçek parametresi
M	: Aylık frekans
Q	: Üç aylık frekans

η : Durum uzayı model hata terimi
 R : Durum uzayı model hata terimi
 E : Durum uzayı model hata terimi
 $t + h$: t+h dönem tahmini
 u_h : Zaman serisi hata terimi
 k : Gecikme uzunluğu
 p : Polinom derecesi
 F : Değişkenlerin ortak faktörleri
 Φ : Faktör yükleme matrisi

Tablo Listesi

3.1	ARDL(1,1) için çeşitli model kombinasyonları	15
3.2	Doğrusal azalan gecikme ağırlıkları	16
4.1	Çekirdek İşlevleri ve Özellikleri	52
5.1	MIDAS-SLS ve MIDAS-NLS göreceli performansları	73
6.1	Model değişkenleri	77

Şekil Listesi

3.1	Almon ağırlıklandırma polinomu	22
3.2	Üssel Almon gecikme polinomu	26
3.3	Beta Gecikme İşlevi	27
3.4	Adımsal Ağırlıklandırma İşlevi	28
4.1	Doğrusal splayn regresyon modeli	38
4.2	Karesel Splayn Model	39
4.3	Kübik Splayn Model	41
4.4	Düzleştirme parametresi λ 'nın değişiminin splayn regresyon modeline etkisi	44
4.5	Çekirdek işlevlerinin karşılaştırmalı görünümü	51
5.1	Üssel Azalan Ağırlıklandırma işlevi	60
5.2	Tümsek azalan ağırlıklandırma işlevi	61
5.3	Doğrusal Azalan Ağırlıklandırma işlevi	62
5.4	Çevrimsel ağırlıklandırma işlevi	63
5.5	MIDAS-NLS modelinin örnek büyüklüğüne göre performans değişimi	66
5.6	MIDAS-SLS modelinin örnek büyüklüğüne göre performans değişimi	67
5.7	MIDAS-NLS modelinin gecikmesi uzunluğuna göre performans değişimi	68
5.8	MIDAS-SLS modelinin gecikmesi uzunluğuna göre performans değişimi	69
5.9	MIDAS-NLS ve MIDAS-SLS modellerinin gecikmesi uzunluğuna göre performans karşılaştırması	70
5.10	MIDAS-NLS modelinin değişkenlik ölçek faktörüne göre performans değişimi	71
5.11	MIDAS-SLS modelinin değişkenlik ölçek faktörüne göre performans değişimi	72
6.1	Birikimli günlük hasta / vaka sayıları (kişi)	76
6.2	MIDAS-NLS modeli ağırlık işlevi kestirimi	78
6.3	MIDAS-NLS modeli vaka sayıları(y)ve kestirimi(\hat{y})	79
6.4	MIDAS-SLS modeli ağırlık işlevi kestirimi	80
6.5	MIDAS-SLS modeli vaka sayıları(y) ve kestirimi(\hat{y})	81

Bölüm 1

Giriş

Farklı frekans sıklıklarına sahip zaman serisi değişkenleri arasında model kurma yöntemlerinde önemli gelişmeler ve önermeler ortaya çıkmıştır. Regresyon modellerinde, bağımlı değişkenin ve bağımsız değişkenlerin eşit boyutlu olması zorunludur. Regresyon analizi için, yüksek frekanslı değişkenlerin gözlem sayısı ile düşük frekanslı bağımlı değişkenin gözlem sayısının eşitlenmesi için dönüşüm işlemlerine ihtiyaç vardır.

Farklı zaman aralıklarından oluşan veri setleriyle bir regresyon modeli oluştururken, dengeli bir veri seti elde etmek için en basit ve yaygın yaklaşım, yüksek frekanslı verilerin toplamını, ortalamasını ya da son değerini alarak düşük frekanslı veri ile aynı zaman dilimine getirmektir. Böylece regresyon analizi için, bağımlı ve bağımsız değişkenin zaman aralığı düşük frekanslı seriyle kolayca eşleştirilmiş olur. Bu yöntem basit ve kullanışlı olmakla birlikte, zaman aralıklarının toplamının, ortalamasının ya da son değerinin alınması gibi işlemler, yüksek frekanslı değişkenlerin içerdiği bilgilerde önemli kayıplara yol açar.

Yüksek frekanslı bağımsız değişkenlerin dönüştürülmesinde, değişkenin gecikmeli değerleri ve dönem içi değişkenliğinin modele daha etkin yansıtıldığı, nedensellik ilişkilerinin daha belirgin ortaya konduğu yöntemler geliştirilmiştir.

Yüksek frekanslı serinin ortalama, toplam ya da son değerini alma uygulamaları, zamansal toplulaştırma ve iterasyon yöntemi olarak adlandırılmaktadır. Uygulamada uzun süredir kullanılan, ileri teknikler olarak yer edinmiş olan, Köprü Denklemi ve Durum Uzayı Modeli (Kalman Filtresi) yöntemleri ikinci bölümde karşılaştırmalı olarak incelenmiştir.

MIDAS (Mixed Data Sampling) regresyon modeli, yüksek frekanslı serilerdeki dönem içi değişkenliği ve gecikmeli değerlerini kapsayan bir ağırlıklandırma işlevinin, bilgi kayıplarını önleyebileceği, böylece değişkenler arasındaki nedensellik ilişkilerini daha doğru yansıtılabilecek bir model kurma iddiasıyla ortaya çıkmıştır. Yöntem ilk olarak [Ghysels, Santa-Clara ve Valkanov](#) tarafından 2002 yılında tartışmaya açılmıştır. MIDAS regresyon modeli, yüksek frekanslı bağımsız değişkenin bir ağırlıklandırma işlevi yardımıyla, düşük frekanslı bağımlı değişkenle eşleştirilmesi esasına dayanır. Ağırlıklandırma işlevi, az sayıda girdi parametresiyle sınırlanarak ağırlık katsayıları üretir. Böylece yüksek frekanslı serinin sınırlanmasıyla, indirgenmiş formda bir regresyon modeli kurmak mümkün hale gelir.

MIDAS regresyon modeli, ilk olarak finansal veriler üzerinde [Ghysels, Santa-Clara ve Valkanov \(2004\)](#) tarafından yayınlanmasının ardından geniş ilgi görmüştür. Makroekonomik veri setleri üzerine uygulamanın [M. P. Clements ve Galvão \(2009\)](#) tarafından yayınlanmasının ardından, geleceğe yönelik tahmin yanında, ekonomide şimdinin tahmini (nowcasting) üzerine çok geniş bir ilginin doğmasına yol açmıştır. MIDAS regresyon yöntem üzerine, uygulama **R** programlama dilindeki *midasr* ([Ghysels, Kvedaras ve Zemlys, 2016](#)) paketi ve Matlab uygulaması ([Sinko, Sockin ve Ghysels, 2010](#)) yayınlanmıştır. Küresel ölçekte, merkez bankaları, ekonomi yönetiminden ve istatistik tahmin yayınlayan pek çok kurum yöntemin uygulamasına ilgi göstermiştir. İlgili kurumlarda uygulamacılar tarafından yoğun kullanılan Eviews zaman serisi yazılım paketin 11. versiyonunda itibaren yönteme yer verilmiştir. Yazılım araçlarında yer verilmesinin ardından, yöntemi uygulayan pek çok ampirik çalışma yayınlanmıştır.

MIDAS yöntemi ampirik çalışmalarda çok yoğun kullanılmasına rağmen, yöntemin geliştirilmesine yönelik sınırlı sayıda çalışma ortaya çıkmıştır. Doğrusal olmayan MIDAS regresyon modeli (MIDAS-NLS)'in asimtotik özellikleri [Andreou, Ghysels ve Kourtellos \(2010\)](#) tarafından tartışılmıştır. Ayrıca, [Andreou, Ghysels ve Kourtellos \(2013\)](#) tarafından makro ekonomik verilerin günlük finansal verilerle tahmin edilmesi üzerine, klasik yöntemlerle karşılaştırmalı bir analiz yapılmıştır. Söz konusu çalışmada, bine yakın yüksek frekanslı finansal veri gruplanarak, makro ekonomik değişkenlerle ilişkileri MIDAS-NLS ve diğer yerleşik yöntemlerle hesaplanmış, MIDAS-NLS'in görece başarımın yüksek olduğu gösterilmiştir. Devam eden çalışmalarda Gecikmesi Dağıtılmış MIDAS (MIDAS-DL) ve Otoregresif MIDAS (MIDAS-ADL), MIDAS-GARCH gibi, faktör-MIDAS üzerine, özellikle finansal getiriler, risk yönetimi ve ekonomik veriler üzerinden yöntemsel ve uygulamalı geliştirmeler yapılmıştır.

MIDAS regresyon modelinde, yaygın olarak Üssel Almon ve Beta ağırlıklandırma işlevleri kullanılmaktadır. Bu yaygın yöntemlerin yanısıra Almon Polinomu ve Adımlama ağırlıklandırma işlevleri de kullanılmaktadır. Yaygın olarak kullanılan bu işlevlerinin yanısıra farklı matematiksel ağırlıklandırma işlevleri önerilse de kullanımları sınırlıdır. Söz konusu ağırlıklandırma işlevlerinin şekli önceden bilindiği için parametrik yöntemlerdir.

Parametrik MIDAS yöntemlerinin, yüksek frekanslı değişkende uzun vadeli etkinin kısa vadede yüksek olması ve artan azalan değerler alabilen çevrimsel yapıda form içeren verilerde, parametre

kestirimlerinin zayıfladığı ve tahmin gücünün azaldığı iddia edilmiştir. Bu bağlamda, parametrik olmayan yöntemler önerilmiştir. Düzleştirilmiş MIDAS regresyon yöntemi ilk olarak [Roling \(2014\)](#) ve [Breitung ve Roling \(2015\)](#) tarafından, cezanlandırılmış splayn regresyon yöntemleri bazında geliştirilerek, parametrik MIDAS-NLS ve parametrik olmayan yöntemleri karşılaştırılmıştır.

Tez, MIDAS regresyon yöntemlerini karşılaştırmalı olarak inceleyerek, parametrik olmayan MIDAS regresyon yöntemlerinde gelişmeye açık alanları ele almayı amaçlamıştır. Bu amaç doğrultusunda, konu yedi bölümde ele alınmıştır. İkinci bölüm, MIDAS öncesi geliştirilmiş olan Zamansal Toplulaştırma ve İterasyon, Köprü Denklemi ve Durum Uzayı Modeli yöntemleri üzerine bir literatür özeti ve MIDAS yöntemlerinin genel bir karşılaştırması yer almaktadır. Üçüncü bölümde parametrik MIDAS regresyon modeli detaylı olarak incelenmiştir. Parametrik ağırlıklandırma işlevleri olan Almon Polinomu, Üssel Almon, Beta ve Adımlandırma detaylı incelenmiştir. Ayrıca yüksek frekanslı serinin veri dizilimi üzerine algoritmalar ve veri dizilim matrislerinin işleyişi anlatılmıştır. Dördüncü bölümde, parametrik olmayan regresyon yöntemleri ve MIDAS regresyon yöntemine uyarlanması tartışılmıştır. Beşinci bölümde, benzetim deneyi ile parametrik ve parametrik olmayan MIDAS yöntemlerinin performansı karşılaştırılmıştır. Altıncı bölümde, MIDAS-NLS ve parametrik olmayan MIDAS-SLS yöntemleri gerçek veri üzerinde çalıştırılmıştır. Türkiye’de COVID-19 konusu üzerine sosyal medya tartışmalarının yoğunluğu ile raporlanan günlük bulaşma verileri arasında ilişkisi MIDAS regresyon modelleriyle analiz edilmiştir. Yedinci bölüm sonuç bölümü olup, bulgular, mümkün ileri çalışmalar ve sınırlandırmalar tartışılmıştır. Yöntemlerin işleyişin anlaşılmasında önem arz eden kodlar -Python programlama dilinde- [ekler](#) bölümünde yer almaktadır.

Bölüm 2

Karma Frekanslı Analiz Yöntemleri

Bu bölümde, karma frekanslı analiz yöntemleri ele alınmıştır. Literatürde yer alan karma frekanslı yöntemlere genel bir bakış sunularak, sıklıkla kullanılan yöntemlerden, Zamansal Toplulaştırma ve İterasyon (Temporal Aggregation and Iteration), Köprü Denklemi (Bridge Equation) ve Durum Uzayı (State Space) modellerine değinilmiştir. Yaygın kullanılan yöntemlere yer verilmesinin ardından, tezin ana konusu olan parametrik ve parametrik olmayan MIDAS regresyon yöntemlerinin tanıtımı yapılarak, karma frekanslı yöntemler karşılaştırılmıştır. Karma frekanslı analiz yöntemlerinin gruplandırılmasında ağırlıklı [Foroni ve Marcellino \(2013\)](#) tarafından önerilen yaklaşım temel alınmıştır.

2.1 Karma Frekanslı Yöntemlere Genel Bakış

Karma frekanslı analiz, bağımlı değişken ile bağımsız değişken serilerinin farklı zaman dilimlerine sahip olması durumunu ifade eder. Regresyon modelleri için değişkenler aynı frekansta olmalıdır, farklı frekanslı serilerde bunu sağlamak için yüksek frekanslı seriler en düşük frekansa toplulaştırılır ya da düşük frekanslı veriler mevcut en yüksek frekansa interpolate edilir.

Veri setlerinin farklı frekanslarda olması sıklıkla karşılaşılan bir durumdur. Örneğin iktisadi istatistiklerde Gayri Safi Yurt İçi Hasıla (GSYİH) üç aylık frekansla yayınlanan zaman serisi iken, Sanayi Üretim Endeksi ve Tüketici Fiyat Endeksleri aylık; diğer bazı veri türleri örneğin borsa indeksleri, kur ve faiz oranları günlük, seanslık, saatlik, dakikalık frekanslarda olabilmektedir.

İlerleyen alt bölümlerde literatürde kullanılan yöntemler genel hatlarıyla özetlenmiştir.

2.1.1 Zamansal Toplulaştırma ve İterasyon Yöntemi

Zamansal toplulaştırma yöntemi, uygulamanın basit ve kullanışlı olması nedeniyle ampirik araştırmalarda en yaygın başvuru olan yöntemdir.

Zamansal toplulaştırma, yüksek frekanslı serinin ortalama, toplama ya da son dönem değerini alma yoluyla, düşük frekanslı değişkenin zaman aralığına getirilmesi işlemidir. Aşağıda denklem (2.1)'de x_t^L bağımsız değişkeni, ağırlıklı ortalama alınarak düşük frekanslı seri ile eşleştirilmiştir.

$$x_t^L = \sum_{i=0}^{m-1} w_i x_{t-i/m}^H \quad (2.1)$$

Yukarıda w_i ağırlıklık değerlerini, L düşük frekans ve H yüksek frekans temsil ederken, m ifadesi düşük frekanslı değişkenin bir zaman dilimine karşılık gelen yüksek frekanslı değişkenin yinelenme sayısını göstermektedir. Yüksek frekanslı bağımlı değişkenin bir dönem içindeki yinelenme sıklığı indisi $i = 0, \dots, m - 1$ aralığında değişir. Örneğin üç ayda bir yayınlanan GSYİH verisi ve aylık Sanayi Üretim Endeksi arasında bir model kurulmaya çalışıldığında $m = 3$ değerini alır. Değişken vektörlerinin satır sayısı, düşük frekanslı serinin gözlem sayısı kadar olup, zaman indisi $t = 1, \dots, T$ aralığındadır. Ağırlıklar toplamı cebirsel olarak $\sum w_i = 1$ bire eşit olmalıdır. Basit aritmetik ortalama için ağırlıklar eşittir. Üssel ağırlıklandırmada cari dönem t 'den geriye doğru uzaklaştıkça ağırlık değerleri azalır.

$$y_t^L = \beta_0 + \beta x_t^L + \epsilon_t^L \quad (2.2)$$

Yüksek frekanslı bağımsız değişkenin ağırlıklandırılarak düşük frekanslı serinin frekansına eşleştirilmesinin ardından, regresyon modeli kurulabilir. Denklem (2.2)'de frekansların eşleştirildiği regresyon denklemi görülmektedir.

$$\hat{y}_{t+1}^L = \hat{\beta}_0 + \hat{\beta} x_{t+1}^L \quad (2.3)$$

Regresyon modelinin parametre kestirimleri yapıldıktan sonra denklem (2.3)'de gösterildiği üzere, kurulan modelle $t + 1$ ile gösterilen ileri dönem tahmini yapılabilir.

İterasyon yöntemi, düşük frekanslı seride gözlenmeyen aralıkların yüksek frekanslı serideki zaman aralıklarına uygun olarak, eksik verinin kestirimini yapılmasıdır. Toplulaştırma yönteminin tersine bir işleyişe sahiptir.

$$y_t^H = \beta_0 + \beta x_t^H + \epsilon_t^H \quad (2.4)$$

Yukarıda denklem (2.4)'de düşük frekanslı bağımlı değişken y_t^L iterasyonla frekansları y_t^H dönüştürülmüş ve yüksek frekanslı bağımsız değişken X_t^H ile eşleştirilmiştir. İterasyon yöntemi, toplulaştırma yöntemi kadar yaygın bir kullanıma sahip değildir.

2.1.2 Köprü Denklemleri Yöntemi

Köprü denklemleri yöntemi, özellikle ekonomik verilerde mevcut açıklanmış yüksek frekanslı verileri kullanarak şimdiki durumun tahmini (nowcasting) için kullanılan yöntemlerden biridir.

Köprü denklemi yöntemi yüksek frekanslı değişkenleri düşük frekanslı değişkene bağlayan doğrusal regresyon modelleri ile iterasyon yapılması olarak tanımlanabilir. Yüksek frekanslı değişkenler kullanılarak düşük frekanslı değişkenin erken tahminleri hesaplanır. Köprü yönteminin uygulanabilmesi için yüksek frekanslı açıklayıcı değişken setinin projeksiyon dönemi boyunca bilinmesi gerekir. Böylece düşük frekanslı değişkeninin yalnızca mevcut dönemi hesaplanabilir. Dolayısıyla yöntemin ileriye tahmin ufku oldukça kısadır.

Örneğin üç aylık ve aylık serilerden oluşan bir veri setiyle çalışıldığında prosedür aşamaları şöyle olur: (i) aylık göstergeler, genellikle tek değişkenli zaman serisi modeli kullanılarak, üç aylık dönemin geri kalanı için tahmin edilir. Bir çeyrek dönem içindeki, üç ayrı aylık değerlerin oluşmasının ardından, düşük frekanslı çeyrek döneme gelen değerleri elde etmek için aylık değerler toplanır. (ii) Toplanan değerler köprü denkleminde açıklayıcı değişken olarak kullanılır.

$$y_t = \beta_0 + \beta_i(L)x_t^Q + \varepsilon_t \quad (2.5)$$

Burada y_t düşük frekanslı bağımlı değişkeni örneğin çeyrek dönemlik GSYİH, x_t^Q ise aynı dönem aralığındaki bağımsız değişkeni, örneğin aylık Sanayi Üretim Endeksini gösterir. Sabit terim β_0 ve x_t^Q yüksek frekanslı açıklayıcı değişkenin düşük frekanslı değişkeninin zaman aralığına toplulaştırılmasını ifade eder. Üç aylık gecikme polinomu $\beta(L)$ p 'inci derecen tanımlı olarak $\beta(L) = \sum_{i=0}^p \beta_{i+1}L^i$ ifadesini temsil ederken, $Lx_t^Q = X_{t-1}^Q$ ifadesine eşittir. Gecikme polinomunun oluşturulurken, gecikme döneminin uzunluğu bilgi ölçütleri yardımıyla belirlenir (Schumacher, 2016).

Aylık serinin üç aylık döneme yansıtılması şöyle çalışır:

$$x_t^Q = x_t^M + x_{t-1/3}^M + x_{t-2/3}^M \quad (2.6)$$

Yukarıda denklem (2.6)'de M aylık frekansı gösterirken, Q çeyrek dönemlik değeri gösterir. Henüz aylık seri $X_{t-2/3}$ döneminde ve diğer dönemler henüz bilinmiyorsa $x_{t-1/3}^M$ ve x_t^M dönemlerinin tahmini yapılır. Elde edilen tahmin değerleri denklem (2.6)'deki yerlerine konduğunda köprü denklemi tamamlanmış olur.

Şimdinin tahmini bağlamında, geniş değişkenler kümesinden oluşan veri setleri için faktör analizi ve köprü denklemi yaklaşımını birleştiren "faktörle köprü oluşturma" yöntemi geliştirilmiştir. Faktörlerle köprü denklemi oluşturularak, finansal ve ekonomik göstergelerin tahmini üzerine uygulama [Giannone, Reichlin ve Small \(2008\)](#) tarafından ortaya koyulmuştur. Köprü Denklemi ve diğer karma frekanslı analiz yöntemleri arasındaki farklar kuramsal ve ampirik uygulamalı olarak [Schumacher \(2016\)](#) makalesinde karşılaştırılmıştır. Zaman toplulaştırma yöntemleri üzerine tartışmalar geriye doğru [Chow ve Lin \(1971\)](#) uzanmaktadır. Çok sayıda farklı frekansa sahip tahmin edicinin, temel bileşenler analiziyle az sayıda değişkenlerle indirgenmesine yönelik çalışma [Stock ve Watson \(2002\)](#) tarafından ele alınmıştır.

2.1.3 Durum Uzayı Modeli

Durum uzayı modeli, düşük frekanslı değişkenin eksik gözlemlerle yüksek frekanslı bir değişken olarak kabul edilmesi esasına dayanır. Kalman Filtresi kullanılarak eksik gözlemler tahmin edilerek daha pürüzsüz değerler elde edilir. Durum uzayı modelleri bir sistem yaklaşımıdır ve Kalman filtresinin kullanımı sayesinde eksik yüksek frekanslı verilerin tahminine olanak sağlar. Öte yandan, bu yaklaşımın kullanımının bazı dezavantajları da vardır. Hesaplama açısından karmaşıktır, karmaşıklık seviyesi modele dahil edilen değişkenler arttıkça hızla yükselir. Bu nedenle çoğu zaman yalnızca küçük ölçekli model kurulumlarında kullanılabilir ([Forni ve Marcellino, 2013](#)).

Durum uzayı modelinde Kalman filtresi hem gözlemlenebilir, hem de gözlemlenemeyen değişkenlerin tahminlerini sağlar. Kalman filtresi, parametrelerin en çok olabirlik (maximum likelihood) tahmin edicilerini türetmek için kullanılır. Kalman Filtresi, ilk defa [Kalman](#) tarafından 1960 yılında önerilmesinin ardından havacılık ve uzay çalışmalarından, nakil vasıtalarının kontrolüne, navigasyon gibi pek çok teknolojik alanda kullanılmaktadır.

Durum uzayı modeli ve Kalman filtresinin işleyişi için kapsamlı açıklamalar [Hamilton \(2020\)](#) ve [Ghysels ve Marcellino \(2018\)](#) yer almaktadır.

Doğrusal Durum Uzayı modelinin genel formu şöyledir:

$$\beta_t = \beta_{t-1} + \eta \quad (2.7)$$

$$y_t = X_t \beta_t + \varepsilon_t, \quad (2.8)$$

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \stackrel{\text{i.i.d.}}{\sim} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & H \end{bmatrix} \quad (2.9)$$

$$E(\eta, \eta') = \begin{cases} Q & t = \tau \text{ için} \\ 0 & \text{ö.d.i} \end{cases} \quad (2.10)$$

$$E(\varepsilon, \varepsilon') = \begin{cases} R & t = \tau \text{ için} \\ 0 & \text{ö.d.i} \end{cases} \quad (2.11)$$

Yukarıda denklem (2.7)'de β_t gözlemlenemeyen durum değişkeni olarak adlandırılır ve β_t 'nin tahmin değerinin hesaplanmasının yapıldığı ilk aşamadır. İkinci aşamada denklem (2.7)'de elde edilen β_t değeri denklem (2.8)'de yerine konularak gözlemlenebilir değişken vektörü y_t hesaplanır. Yukarıdaki denklem (2.8)'deki ifade gözlemlenebilir durum değişkeninin doğrusal bir işlevidir. "Hata terimleri η_t ve ε_t sembolleriyle, ortalaması sıfır, bileşik normal dağılıma sahip, birbiriyle ilişkili olmayan hata terimlerini gösterir" (Ghysels ve Marcellino, 2018, s.421).

MIDAS Regresyon ile Kalman filtresi modelleri arasındaki ilişki ve tahmin performansı Bai, Ghysels ve Wright (2013) tarafından tartışılmıştır. Söz konusu makalede Durum Uzayı Modeli ve MIDAS regresyon modeli genelde benzer tahminler vermekle birlikte, Kalman Filtresi yönteminin biraz daha performanslı tahmin değerleri ürettiği tespiti yapılmıştır. Öte yandan, Kalman Filtresi karmaşık yapıda bir denklem sistemi gerektirirken, klasik parametrik MIDAS regresyon modeli tek bir denklem içermektedir.

2.1.4 MIDAS Regresyon Modelleri

Karma Frekanslı Veri Örnekleme anlamına gelen ve kısaltma olarak MIDAS (Mixed Data Sampling) olarak adlandırılan regresyonu yöntemi ilk olarak Ghysels, Santa-Clara ve Valkanov 2002 yılındaki öncü çalışmanın ardından, 2004 ve 2007 yıllarında geliştirilerek kapsamlı yayınlarla ortaya konulmuştur. Söz konusu çalışmalarda, farklı frekanslı değişkenlerin modellenmesinde bilgi kaybının önlenmesi, parametre kestirimlerinde yüksek frekanslı serideki dönem içi değişkenlikleri ve gecikmeli etkilerinin dahil edilmesiyle etkin ileri dönem (forecasting) ve şimdiki dönem tahmininin (nowcasting) yapılabilmesinin yöntemleri ele alınmıştır.

MIDAS regresyonu ilk olarak Ghysels, Santa-Clara ve Valkanov tarafından 2004 yılında, hisse senetleri piyasasındaki ortalama, koşullu varyans ile getiriler arasındaki ilişkilerin modellenmesi için uygulanmıştır. İzleyen dönemde, makroekonomik veri setleri üzerine uygulanmasının (M. P. Clements ve Galvão, 2009) ardından, özellikle merkez bankaları başta olmak üzere finansal kurumlar ve ekonomi araştırmacıları tarafından yoğun olarak kullanılmaya başlanmıştır.

MIDAS regresyon analizinde kullanılan başlıca yazılım paketleri şunlardır: Programlama dili R 'da "midasr" modülü (Ghysels, Kvedaras ve Zemlys, 2016), Matlab için yazılmış (Sinko, Sockin ve Ghysels, 2010) ve (zaman serisi analizde uygulamalı araştırmalarda yoğun olarak kullanılan Eviews programının 11. versiyonundan itibaren yer almaktadır.

Literatüre görece yeni girmesine bağlı olarak, Türkçe'de kimi ampirik çalışmalarda MIDAS regresyon için "Karma frekanslı veriler" ve "Farklı frekanslı veriler" terimleri kullanılmasına rağmen, bu terimler yukarıda incelediğimiz diğer yöntemlerle bir ayrışmayı ifade etmemektedir. Bu nedenle, mevcut uluslararası yayınlardaki kullanımına paralel olarak MIDAS regresyonu teriminin tez boyunca kullanılması tercih edilmiştir.

MIDAS regresyon modeli, karma frekanslarda örneklenen verilerle, düşük frekanslı değişkenin doğrudan tahminini sağlayan, gecikmesi dağıtılmış polinomlara dayanan az sayıda parametre ile yalınlaştırılmış indirgenmiş form denklemlerinden oluşur (Ghysels ve Marcellino, 2018).

$$y_{t+h}^L = \beta_0 + \beta_1 C(L^{1/m}; \theta) x_t^H + \varepsilon_{t+h}^L \quad (2.12)$$

Gecikmesi dağıtılmış yüksek frekanslı değişkenin az sayıda parametre ile ağırlık işlevi aracılığı ile yalınlaştırıldığı denklem (2.12)'de, ağırlıklandırma polinomu $C(L^{1/m}; \theta_h) = \sum_{j=0}^{j_{max}-1} c(j; \theta) L^{j/m}$ ifade edilmiştir. Uygulamada yalınlaştırma fonksiyonu olarak yaygın olarak Üssel Almon ve Beta işlevleri kullanılmakla birlikte, takip eden bölümde incelendiği üzere yalınlaştırmaya elverişli farklı matematiksel işlevler de yer almaktadır.

2.1.5 Parametrik Olmayan MIDAS Regresyon

MIDAS regresyon ağırlıklandırma işlevlerinde Üssel Almon ve Beta işlevleri parametrik yöntemlerdir. Hangi ağırlıklandırma işlevinin kullanılacağı araştırmacının tercihinin göre belirlenir. Gerçek veri üzerinde yapılan analizlerde parametrik ağırlıklandırma işlevlerinin yalınlaştırma özelliği yetersiz kalabilir. Parametrik işlevlerinin yetersiz kaldığı durumlar iki başlık altında toplanabilir: (i) Bağımsız değişkenin uzun vadeli etkisinin, kısa vadeli etkiye göre daha küçük olduğu durumlar, (ii) Bağımsız değişkenin bağımlı değişkeni her gecikme aralığında aynı etkilemediği, yani belirli aralıklarda pozitif gecikme katsayılarına karşılık, bazı aralıklarda negatif katsayı olabileceği durumlarda, parametrik ağırlıklandırma işlevleri yetersiz kalmaktadır (Breitung ve Roling, 2015).

Parametrik olmayan yöntem uygulamasında, Breitung ve Roling (2015) parametrik bir ağırlıklandırma işlevi yerine, cezalandırılmış en küçük kareler (penalized least-square) yöntemiyle gecikme uzunluklarını belirli derecede yalınlaştırarak daha iyi bir performans elde edebileceğini iddia etmişlerdir. Bunun için parametrik olmayan cezalandırılmış ya da düzleştirilmiş regresyon modeli kullanılmıştır. Parametrik olmayan yöntemlerle ilgili kuramsal çerçeve dördüncü bölümde etraflıca incelenmektedir.

2.1.6 Modellerin Karşılaştırılması

Zamansal toplulaştırma yöntemi uygulanmasının pratik ve kolay olması nedeniyle en yaygın kullanıma sahiptir. Ancak toplulaştırma yöntemi, yüksek frekanslı serinin içerdiği bilginin kaybına ve modelin nedensellik ilişkilerinde bozulmaya yol açmaktadır. Düşük frekans ile yüksek frekans arasındaki zaman aralığının ve yinelenme sıklığının artması bu kayıpların etkisinin daha fazla olmasına yol açar.

Köprü denklemi, toplulaştırma yöntemine göre yüksek frekanslı verinin modele doğrudan yansımısını sağlamakla birlikte, klasik modelin tahmin ufku kısıtlıdır. Bayeşçi yöntem ile tahmin yeteneđi ve tahmin vadesinin uzatıldıđı geliřtirmeler yapılmıřtır.

Durum-Uzayı modeli oldukça geliřmiř bir yapı olup, tahmin yeteneđini de gerek veriler üzerinde en yüksek performansa sahip olduđu ifade edilmiřtir. Kalman filtresine dayanan yöntem karmařık ve ok katmanlı yapıda olduđundan uygulamada ciddi zorluklar ortaya ıkarmaktadır. Durum Uzayı Modelleri ile MIDAS regresyon arasında karřılařtırma [Bai, Ghysels ve Wright \(2013\)](#) tarafından yapılmıřtır. Buna göre MIDAS regresyonun, Durum-Uzayı modellerine ok yakın performans gstermesi ve uygulamasının grece kolay olması finansal ve ekonomik uygulamalarda sađladıđı tahmin performansı nedeniyle hızla kullanımının yaygınlařtıđı ifade edilmiřtir([Ghysels, Santa-Clara ve Valkanov, 2004](#)) .



Bölüm 3

MIDAS Regresyon Modelleri

Karma frekanslı verilerle regresyon modeli oluşturulurken, literatürde yaygın olan yöntemlerin genel bir özeti [ikinci bölümde](#) incelenmiştir. Bu bölümde parametrik MIDAS regresyon modellerinin kuramsal işleyişi ve türleri ele alınmıştır.

MIDAS regresyon, düşük frekanslı bağımlı değişken ile yüksek frekanslı bağımsız değişkenin ilişkilendirebilmesi için, dönüştürme sürecine gecikme uzunluğu özelliklerinin de dahil edildiği, gecikme katsayılarının az sayıda parametreye indirildiği, bir zaman serisi yöntemidir. Karma frekanslı değişkenlerle model kurarken, yüksek frekanslı serilerin toplulaştırılmasının bilgi kaybına yol açması önemli bir sorundur. Özellikle yaygın olarak uygulanan ortalamalar ya da toplamlar şeklinde yapılan dönüştürmelerde, neden-sonuç ilişkilerini etkileyen önemli bilgi kayıpları meydana gelebilir. MIDAS regresyon modeli, önceki bölümde tartışılan diğer yöntemlere göre, bir yandan yüksek frekanslı serileri dönüştürürken, hem bilgi kayıplarının azaltma, diğer yandan da modeli yalınlaştırarak, daha etkin parametre kestirimi yapabildiği iddiasındadır.

3.1 Zaman Serisi Regresyonu

MIDAS regresyon modelleri bir zaman serisi analiz yöntemidir. Zaman serisi analizleri, klasik regresyona ait bazı varsayımların esnetilmesini gerektirir. İlerleyen alt bölümlerde zaman serileri regresyonunun uygulanmasına dair kuramsal yapıya ilişkin literatürde yer alan tartışmalara yer verilmiş ve MIDAS regresyon ile bağları tartışılmıştır.

3.1.1 Klasik Regresyon Modelinin Varsayımları

Doğrusal regresyon modelinin doğru bir ilişki tanımlayabilmesi için çok sayıda kısıtlayıcı varsayıma sahiptir. Doğrusal regresyon denkleminin genelleştirilmiş matris notasyonu ile gösterimi denklem (3.1) yer almaktadır.

$$y = X\beta + \varepsilon \quad (3.1)$$

Doğrusal regresyon modelinin parametreleri kestirimleri (3.2) şöyledir:

$$\beta = (X'X)^{-1}X'y \quad (3.2)$$

Doğrusal regresyon modelinin doğru bir kestirim yapabilmesi için aşağıdaki varsayımları yerine getirmesi gerekir (Baltagi, 2011, s.50-54):

Varsayım 1. Hata terimi ε 'nin beklenen değeri $E(\varepsilon) = 0$ sıfırdır.

Varsayım 2. Hata terimlerinin varyansı $var(\varepsilon_i) = \sigma^2$, matris notasyonu ile $E(\varepsilon\varepsilon') = \sigma^2I$ sabit olmalıdır.

Varsayım 3. Hata terimleri arasında çapraz ilişki $i \neq j$ şartıyla $cov(\varepsilon_i, \varepsilon_j) = 0$ yoktur. Hata terimlerinin kovaryansı $cov(\varepsilon_i, \varepsilon_j) = \sigma^2I$ sabit varyans terimi ile birim matrisin çarpımına eşittir.

Varsayım 4. Hata terimi $\varepsilon = N(0, \sigma^2I)$ normal dağılıma uygundur.

Varsayım 5. Bağımsız değişken X stokastik değildir ya da zayıf formda durağan olup, hata terimi vektörü ε ile ilişkili değildir.

Gauss-Markov koşulları olarak da adlandırılan varsayımlar, matematiksel olarak tam bir tutarlılık içindedir. Bazı varsayımlar ihlal edildiğinde, doğru sonuçlar üreten bir model oluşturmak için, dönüşümler yapılarak varsayımlar ile uyumlu hale getirilmeye çalışılır. Zaman serisi regresyon analizinde ise yukarıdaki varsayımların bir ölçüde esnetilmesi gerekir.

3.1.2 Dinamik Regresyon Modelinin Türleri

Dinamik regresyon modelinin en genelleştirilmiş şekli otoregresif gecikmesi dağıtılmış (ARDL) modelle gösterebiliriz. ARDL(p,q) biçiminde otoregresif bir model tanımlandığında, p parametresi bağımlı değişken y 'nin gecikme değerleri ve q parametresi de x bağımsız değişkeninin gecikme uzunluğunu gösterir. Denklem (3.3)'de ARDL(1,1) şeklinde bir otoregresif model dinamik regresyon tanımlanmıştır.

$$y_t = \alpha_1 y_{t-1} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2) \quad (3.3)$$

Eğer denklem (3.3)'de $\alpha_1 = \beta_2 = 0$ durumu varsa aşağıdaki ifade ortaya çıkar:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (3.4)$$

Denklem (3.4) *statik regresyon modeli* olarak tanımlanmaktadır. Aşağıda ARDL(1,1) modellerinin parametre değerlerine göre, yaygın kullanılan model kombinasyonları ve literatürdeki adlandırmalarının bir kısmına yer verilmiştir (Tablo 3.1). Farklı kombinasyonların da yer aldığı geniş kapsamlı bir özet [Ghysels ve Marcellino \(2018, s.119-143\)](#) yer almaktadır.

Tablo 3.1: ARDL(1,1) için çeşitli model kombinasyonları

Varsayım	Denklem	Model
$\alpha_1 = \beta_2 = 0$	$y_t = \beta_0 + \beta_1 x_t$	Statik regresyon
$\beta_1 = \beta_2 = 0$	$y_t = \alpha_1 y_{t-1} + \varepsilon_t$	AR(1)
$\beta_i = 0, i = 0, 1, 2; \alpha_1 = 1$	$y_t = y_{t-1} + \varepsilon_t$	Rassal yürüyüş
$\alpha_1 = 1, \beta_1 = -\beta_2$	$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \varepsilon_t$	Fark Modeli
$\alpha_1 = \beta_1 = 0$	$y_t = \beta_0 + \beta_2 x_{t-1} + \varepsilon_t$	Öncü gösterge
$\alpha_1 = 0$	$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t$	Gecikmesi dağıtılmış(DL)
$\beta_2 = 0$	$y_t = \beta_0 + \beta_1 x_t + \alpha_1 y_{t-1} + \varepsilon$	Kısmi düzeltme

Zaman serileri geniş kapsamlı bir alan oluşturmaktadır. Sonraki bölümlerde MIDAS regresyon için temel teşkil eden gecikmesi dağıtılmış ve otoregresif gecikmesi dağıtılmış modeller incelenmiştir.

3.1.2.1 Gecikmesi Dağıtılmış (DL) Model

Gecikmesi dağıtılmış model (Distributed Lag Model), yalnızca bağımlı değişken x 'in cari dönem x_t ve gecikmiş x_{t-s} değerleriyle bağıntı içinde olduğu, ancak bağımlı değişken y_t 'nin kendi gecikmiş değerleriyle bağıntı içinde olmadığı durumu tanımlar ([Davidson ve MacKinnon, 1999, s.565](#)). Gecikmesi dağıtılmış modeller x 'in y üzerindeki etkisinin zamana bağlı etkisinin var olduğu dinamik modellerdir. Tek değişkenli gecikmesi dağıtılmış modeli şöyle gösterebiliriz:

$$y_t = \alpha + \sum_{s=1}^q \beta_s x_{t-s} + \varepsilon_t \quad (3.5)$$

Yukarıda denklem (3.5)'de α sabit terimi, y_t bağımlı değişkeni, bağımsız değişkenin cari dönem değerini x_t , gecikmeli uzunluğu q ve durağan hata terimi ya da beyaz gürültü (white noise) terimi ε_t sembolleri temsil eder. Gecikmesi dağıtılmış modelde, bağımsız değişken x vektörünün ilk satırından gecikme genişliği q kadar gözlem kaybı anlamına gelir. Ayrıca, gecikme uzunluğu kadar parametre kestirimi β katsayısı kadar serbestlik derecesi gerektireceğinden veri kaybı artar ([Baltagi, 2011, s.131](#)). Denklem (3.5)'deki β_s katsayıları gecikme ağırlıklarını ve toplamlarını gösteren $\sum_{s=1}^q \beta_s x_{t-s}$ ifadesi de toplam gecikme dağılımını ifade eder. Gecikme dağılımının uzunluğu q pratikte zaman uzadıkça ağırlığın sıfıra yaklaşması nedeniyle sonlu bir değere sahiptir.

$$\frac{\partial y_t}{\partial x_{t-s}} = \beta_s \quad (3.6)$$

Basit doğrusal model için ([Parker, 2021, s. 35-46](#)), denklem (3.6)'de gösterildiği gibi ağırlık katsayısı β_s , bağımlı değişken y_t ile bağımsız değişken x_{t-s} birinci dereceden türevine eşit olur. Gecikme sayısının

sonlu olduğu durumda, denklem (3.7)'de gösterilen durağan bir seride, gecikme uzunluğu q ile gecikme ağırlıkları katsayıları β_s arasındaki ilişki tanımlanabilir. Her bir gecikme ağırlığı β_s kendisinden önceki ağırlık β_{s-1} den daha küçük olur.

$$\beta_s = \frac{q+1-s}{q+1}\beta_0 \quad s = 1, 2, \dots, q \quad (3.7)$$

Gecikme ağırlıkları $\beta_1, \beta_2, \dots, \beta_q$ biçiminde zaman içinde doğrusal olarak azalan oranlarda ise, cari dönem t 'deki parametre katsayısı β_0 ile ilişkileri tablo 3.2'de gösterildiği gibi olur.

Tablo 3.2: Doğrusal azalan gecikme ağırlıkları

Gecikme (s)	Ağırlık Katsayısı (β_s)
0	β_0
1	$q/(q+1)\beta_0$
2	$(q-1)/(q+1)\beta_0$
...	...
$q-1$	$2/(q+1)\beta_0$
q	$1/(q+1)\beta_0$
$q+1$	0

Bu durumda doğrusal azalan ağırlıkların varlığı durumunda, eğer gecikme uzunluğu q biliniyorsa her gecikme katsayısı β_s için denklem (3.6) ifadesi denklem (3.5)'de yerine konulduğunda aşağıdaki ifadeye ulaşılır:

$$y_t = \alpha + \beta_0 \left[\sum_{s=1}^q \frac{q+1-s}{q+1} \beta_0 x_{t-s} \right] + \varepsilon_t \quad (3.8)$$

Yukarıdaki denklem, doğrusal azalan ağırlıklarının gecikme uzunluğu ile ilişkisini gösteren kısıtlanmış bir durağan gecikmesi dağıtılmış modelin açık yazılmış halidir. Gecikme uzunluğu ile gecikme katsayıları arasındaki ilişki doğrusal biçimlerin dışında, farklı işlevler; örneğin polinom ya da karesel biçimlerde de yazılabilir.

3.1.2.2 Otoregresif Gecikmesi Dağıtılmış Model (ADL)

Otoregresif gecikmesi dağıtılmış model (Autoregressive Distributed Lag), bağımlı değişken cari dönemi y_t 'nin, açıklayıcı değişken X_t ve X_{t-s} gecikmeli değerlerinin yanı sıra, kendi gecikmeli değerleri, örneğin önceki değeri olan y_{t-1} de açıklayıcı değişken olarak modele dahil edilmesi durumudur. Denklem (3.9)'de bağımlı değişkenin kendinden bir önceki dönemle ve açıklayıcı değişkenin yer aldığı bir otoregresif gecikmesi dağıtılmış model yer almaktadır. Bağımlı değişkenin gecikmeli değerlerinin açıklayıcı değişken seti içinde yer almasının tahmin etme gücünü artırdığı durumlarda modele dahil edilir. Özellikle makroekonomi ve finansal verilerde sıklıkla başvurulan modellerdir.

$$y_t = \alpha + \lambda y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \quad (3.9)$$

Bağımlı değişken y 'nin gecikme sayısının p ve bağımsız değişken x 'in gecikmelerinin q ile gösterildiği durumda otoregresif gecikmesi dağıtılmış model ADL(p,q) olarak gösterilir. Denklem (3.9)'de ADL(1,1) modelinde, y_t cari dönem bağımlı değişkeni, α sabit terimi, y_{t-1} bağımlı değişkenin önceki dönem değerini ve λ katsayı kestirimini gösterirken, x_t ve x_{t-1} değişkenlerinin katsayı kestirimleri de sırasıyla β_0 ve β_1 olarak gösterilmiştir. Ayrıca, ε_t hata terimini ya da beyaz gürültü terimini ifade etmektedir.

Otoregresif modellerde, zamana bağlı olarak değer alan değişkenler gecikme işlemcisi L ile ifade edilebilir. Gecikme işlemcisi $L^n y_t = y_{t-n}$ eşitliğiyle ifade edilirse ve denklem (3.9)'de yerleştirildiğinde,

$$y_t = \alpha + \lambda L y_t + \beta_0 x_t + \beta_1 L x_t + \varepsilon_t \quad (3.10)$$

$$(1 - \lambda L) y_t = \alpha + (\beta_0 + \beta_1 L) x_t + \varepsilon_t \quad (3.11)$$

denklem (3.10) elde edilir ve gerekli işlemler yapılırsa denklem (3.11) ifadesine ulaşılır. Bağımsız değişken x 'in gecikmeli değerleri için gecikme işlemcisi çalıştırıldığında, birinci dereceden polinom için β_1 ifadesi ile $\lambda \beta_0$ eşit olur. Bu işlemlerin ardından denklem:

$$(1 - \lambda L) y_t = \alpha + \beta_0 (1 - \lambda L) x_t + \varepsilon_t \quad (3.12)$$

halini alır. Bağımsız değişken y_t için düzenlenerek yeniden yazıldığında:

$$y_t = \frac{\alpha}{(1 - \lambda L)} + \beta_0 x_t + \frac{\varepsilon_t}{(1 - \lambda L)} \quad (3.13)$$

$$y_t = \alpha (1 - \lambda L)^{-1} + \beta_0 x_t + \tilde{\varepsilon}_t \quad (3.14)$$

Denklem (3.14) ifadesine ulaşılır. Denklem (3.13) yer alan hata terimi, $\tilde{\varepsilon}_t = \frac{\varepsilon_t}{(1 - \lambda L)}$ ifadesine eşittir. Bu ifade gecikmesi dağıtılmış otoregresif modeller için çok önemli bir yalınlaştırma sağlar, ilk olarak [Hendry ve Mizon tarafından 1978](#) ortaya konmuştur, otoregresif gecikmesi dağıtılmış modelinin işleyişi ve ADL-MIDAS modeline geçiş açısından çok önemli bir yere sahiptir.

3.2 MIDAS Regresyon Modelinin Yapısı

MIDAS regresyon modelleri yalnızca gelecek dönem tahminin (*forecasting*) değil, aynı zamanda şimdiki dönemin tahmini (*nowcasting*) için de kullanılmaktadır. Şimdinin tahmini, düşük frekanslı

değişkene ait henüz yayınlanmamış verinin, yüksek frekanslı değişkenlerle tahmin edilerek, güncel durumun tahmin edilmesini ifade eder.

3.2.1 MIDAS Regresyon Modelinin Temel Kavramları

MIDAS regresyon analizi farklı frekanslardaki değişkenlerle model oluşturulmasını sağlayan, indirgenmiş biçim (reduced form) zaman serisi analiz yöntemi olarak tanımlanabilir. MIDAS regresyonu ilk olarak, [Ghysels, Santa-Clara ve Valkanov](#) tarafından 2002 yılında ilk taslağı tartışmaya açılarak, [2004](#), [2006](#) yüksek frekanslı örneklenmiş değişkenlerin modelde doğrudan yer aldığı, tek değişkenli zaman serisi regresyonu olarak önerilmiştir.

$$y_t = \alpha + \sum_{i=1}^p \beta_i x_{(t-q/m)} + \varepsilon_t \quad (3.15)$$

MIDAS regresyon modelleri, farklı frekanslarda örneklenen zaman serisi verilerini içerir. MIDAS regresyon, *gecikmesi dağıtılmış* yüksek frekanslı değişkenlerin, koşullu beklenen değerlerini ifade eder. Tek bağımsız değişkenli MIDAS regresyon modelinin cebirsel ifadesi denklem (3.15)'de yer almaktadır. Yukarıda denklem (3.15)'de y_t bağımlı değişkenin cari dönem t 'deki değerlerini, α sabit katsayı kestirimini gösterir. Yüksek frekanslı değişken x 'nin t dönem içinde yinelenme sıklığı m , gecikme uzunluğu q , gecikmeli x değerlerinin katsayı kestirimleri de β ile gösterilmiştir. MIDAS regresyonun ayırt edici özelliği q gecikme uzunluğuna karşı, p sayıda indirgenmiş parametre ile β tahminin yapılmasına olanak sağlamasıdır. Başka bir ifadeyle, gecikme uzunluğu $p \leq q$ ilişkisi vardır, gecikmeler daha az sayıda parametre ile ifade edilebilmektedir.

Gecikme uzunluğu katsayılarının kestiriminin doğrudan yapılması, düşük frekanslı değişken ile yüksek frekanslı değişkenin periyotlarının arasının açılması, diğer bir ifadeyle, sıklık katsayısının (m) artması, tahmini yapılacak katsayı adedini hızla yükseltir. Örneğin, borsa endeksinin günlük kapanış değerleri ile GSYİH gibi üç aylık bir değişken arasında bir regresyon modeli kurarsak, 66 iş günü yinelenme sıklığı gözlem setinde yer almalıdır. Bu durum yönetilmesi zor ve modelin anlamlı bir örutü oluşturmasını güçleştiren, çoğalma (proliferation) sorununa yol açmaktadır.

MIDAS Regresyon, farklı frekanslarda örneklendirilmiş değişkenlerin, yüksek frekanslı örneklemelerin düşük frekansa dönüşümüyle sıkı biçimde parametrelendirilmiş, indirgenmiş biçim modelleridir. Aşağıda denklem (3.16) 'de doğrusal MIDAS Regresyon modeli yer almaktadır ([Ghysels vd., 2004](#)).

$$y_t^L = \beta_0 + \beta_1 C(L^{1/m}; \theta) x_t^{(m)} + \varepsilon_t^{(m)} \quad (3.16)$$

Yukarıda L işareti düşük frekans ve m işareti yüksek frekanslı değişkenin bir periyot içindeki yinelenme sıklığını ifade ederken, y_t^L ifadesi düşük frekanslı bağımlı değişkenin t periyodundaki değerini gösterir. Örneğin, düşük frekanslı değişken Y 'nin üç aylık ve yüksek frekanslı değişken X 'in

aylık olduğu durumda yinelenme sıklığı $m = 3$ olur. Bu denklem setinde MIDAS regresyonun ayırt edici özelliğini $C(L^{1/m})$ işlevi sağlar. $L^{j/m}$ işlemcisi X_t 'nin j/m periyodu içindeki gecikmeli değerlerini üretir. β_1 parametresi kestirimi $C(L^{1/m})$ polinomu ile ağırlandırılmış olarak modelde yer alır.

$$C(L^{1/m}) = \sum_{j=0}^{jmax} C(j)L^{j/m} \quad (3.17)$$

$$L^{j/m}X_t^m = X_{t-j/m}^m \quad (3.18)$$

$$E(\epsilon_t|y_{t-1}, \dots, y_{t-p}) \quad (3.19)$$

MIDAS regresyon modeli, bir sonraki dönemin (y_{t+1}) tahmini olarak yazılırsa:

$$y_{t+1} = \beta_0 + \beta_1 C(L^{1/m}; \theta_1) x_t + \epsilon_{t+1}^L \quad (3.20)$$

Yukarıdaki denklem (3.20) ifadesi, ağırlıklandırma işlevi ile yüksek frekanslı değişken x_t gecikmeli değerleri de içerilmek üzere, bağımlı değişken y 'nin, $t+1$ dönemindeki tahmin değerini vermektedir. MIDAS regresyonun, yinelenme sıklığı ve gecikmeli değerleri ile ilişkileri açık matris notasyonu yoluyla gösterimi bir sonraki alt bölümde açıklanmıştır.

3.2.2 MIDAS Gecikmeli Veri Matrisleri

MIDAS regresyon analizi için, ilk olarak yüksek frekanslı değişkenin gecikmeli değerleriyle dizildiği bir matris formuna ihtiyaç vardır (Ghysels vd., 2016). Örneğin, düşük frekanslı bağımlı değişken y_t^L 'nin üç aylık, bağımsız değişken x_t^H 'in aylık olduğu durum ele alındığında, yinelenme sıklığı $m = 3$ olur. Veri setlerini y_t^L ve x_t^H için açık matris notasyonu ile aşağıda yer almaktadır.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & x_3 & x_2 & x_1 \\ 1 & x_6 & x_5 & x_4 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{3n} & x_{3n-1} & x_{3n-2} \end{bmatrix}_{n \times 4} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_3 \end{bmatrix}_{4 \times 1} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n \times 1} \quad (3.21)$$

Yukarıda denklem (3.21)'de yer alan X matrisi aylık verilerden oluşan serinin veri dizilimi göstermektedir. Veri üç aylık dönem içinde kalmakta, daha öncesine ait gecikme uzunluğunu içermediği varsayılmaktadır.

Yukarıda denklem (3.21)'de, eğer gecikme uzunluğu p 'nin değerini, yüksek frekanslı değişkenin yinelenme sıklığı m değerinden daha uzun alınır, veri matrisinin sütun sayısı artar. Bu artışa paralel olarak gecikme katsayısı vektörü β 'nin satır sayısı da artar. Aşağıda yer alan matriste $m = 3$ olmasına

karşın, gecikme uzunluğu $p = 5$ olarak tanımlanmıştır. Bu durumda veri dizilimi ve parametre vektörünün şekli şöyle olur:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{nx1} = \begin{bmatrix} 1 & x_5 & \cdots & x_1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{3n} & \cdots & x_{3n-5} \end{bmatrix}_{nx6} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_5 \end{bmatrix}_{6x1} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{nx1} \quad (3.22)$$

Yukarıda bağımlı değişken $y = [y_1, \dots, y_n]$ aralığında $n * 1$ bir sütun vektörüdür. Buna karşılık yüksek frekanslı bağımsız değişkenin yinelenme sıklığı $m = 3$ olmasına rağmen, gecikme parametresi sayısı $p = 5$ olduğundan X matrisi ($n * 6$) boyundadır. Gecikme katsayıları β_1, \dots, β_5 ve sabit terim β_0 'dan oluşan parametre matrisi de ($6 * 1$) boyutunda olmaktadır. Hata terimi ε 'da ($n * 1$) boyunda olduğu görülmektedir. Burada X matrisinin şekli ve boyutu belirleyici öneme sahiptir. Bu şekliyle MIDAS regresyon modeli, gecikmesi dağıtılmış (DL) modeldir.

Otoregresif MIDAS Regresyonun, bağımlı değişkenin bir dönem öncesi ile bağıntı içinde olduğu modelinin açık matris gösterimi de (3.23)'de yer almaktadır.

$$\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix}_{nx1} = \begin{bmatrix} y_1 \\ \vdots \\ y_{n-1} \end{bmatrix}_{nx1} \alpha_1 + \begin{bmatrix} 1 & x_5 & \cdots & x_1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{3n} & \cdots & x_{3n-5} \end{bmatrix}_{nx6} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_5 \end{bmatrix}_{6x1} + \begin{bmatrix} \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{nx1} \quad (3.23)$$

Modelde bağımlı değişken y vektörünün, y_{t-1} değerleri açıklayıcı değişken olarak yer almaktadır. Bağımlı değişken y 'nin parametre kestirimi α_1 ile gösterilmiştir.

Bağımlı değişkenin gecikme uzunluğu için p genelleştirme yapılırsa, modelde veri dizilimi aşağıdaki şekli almaktadır:

$$\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{l-1} & \cdots & y_{l-k} \\ \vdots & \ddots & \vdots \\ y_{n-1} & \cdots & y_{n-k} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} + \sum_{i=0}^p X^{(i)} \begin{bmatrix} \beta_0^{(i)} \\ \vdots \\ \beta_p^{(i)} \end{bmatrix} + \begin{bmatrix} \varepsilon_p \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (3.24)$$

Yukarıda denklem (3.24)'de, k gecikme uzunluğunda, bağımlı değişkenin otoregresif gecikmesi dağıtılmış matrisi ve bir skaler yerine α katsayılar vektörü yer almaktadır. Bağımsız değişken X matrisi için gösterim, y gecikmeli değerleriyle eşleşme ihtiyacı nedeniyle daha karmaşık bir dizilime neden olarak, aşağıdaki şekle dönüşmektedir. Denklem (3.24)'de yer alan $\sum_{i=0}^p X^{(i)}$ ifadesinin açık matris yazılımı şöyledir:

$$X^{(i)} = \begin{bmatrix} x_{um_i}^i & x_{um_i-p}^i & \cdots & x_{um_i-p}^i \\ x_{(u+1)m_i}^i & x_{(u+1)m_i-p}^i & \cdots & x_{(u+1)m_i-p}^i \\ \vdots & \vdots & \ddots & \vdots \\ x_{1m_i}^i & x_{1m_i-p}^i & \cdots & x_{1m_i-p}^i \\ \vdots & \vdots & \ddots & \vdots \\ x_{(n-1)m_i}^i & x_{(n-1)m_i-p}^i & \cdots & x_{(n-1)m_i-p}^i \\ x_{nm_i}^i & x_{nm_i-p}^i & \cdots & x_{nm_i-p}^i \end{bmatrix} \quad (3.25)$$

Denklem (3.25)'de yer alan matris $u > k$ koşulu geçerli iken, u ise $um_i - p$ koşulunu sağlayan en küçük doğal sayıdır. Veri dizilimi yapılması bağımlı ve bağımsız değişkenlerin dönemlerinin hizalanmasını sağlar. Bu işlemlerin ardından, MIDAS regresyon modeli parametrelerinin kestirimi yapılabilir. Veri dizilimine ilişkin detaylar Ghysels vd. (2016)'de yer almaktadır.

3.3 MIDAS Ağırlıklandırma İşlevleri

Gecikmesi dağıtılmış zaman serisi regresyon modeline, yüksek frekanslı değişkene ait gecikmeli değişkenler dahil edildiğinde, kestirilecek parametre sayılarında anlamlı sonuçlar elde etmeyi ve yorumlamayı güçleştirecek çok büyük artışlar meydana gelir. Yüksek frekanslı seride önemli bilgi kaybına yol açmadan, katsayı kestirimi yapılması, az sayıda parametre kullanarak modelin yalınlaştırılması ağırlıklandırma işlevleri aracılığıyla gerçekleştirilir.

Örneğin bölüm 6'da erken uyarı tahmin modeli uygulamasında, açıklanan COVID-19 vaka sayısı günlük iken, bu konuya ait sosyal medya platformu Twitter üzerinde paylaşılan saatlik mesaj sayısı ile ilişkilendirildiğinde, gecikmesiz durumda bir güne ait 24 sütun için parametre kestirimi yapılması gerekir. Gecikme sayısı p 'nin dönem içi yenilenme sayısı m 'den daha büyük olması nedeniyle, modelde ise kestirimi yapılacak parametre sayısı çok daha fazla artar.

Aşağıdaki alt başlıkta, Almon polinomu, Üssel Almon, Beta ve Adımlama ağırlıklandırma işlevleri görsel olarak incelenmiştir. Bu ağırlıklandırma işlevlerinin görselleştirilmesinde, açık kaynak R programlama dilinde yazılmış olan midasr paketinin kodlarından yararlanılmıştır. Python programlama dilinde yeniden yazılan kodlar, ek A'de paylaşılmıştır.

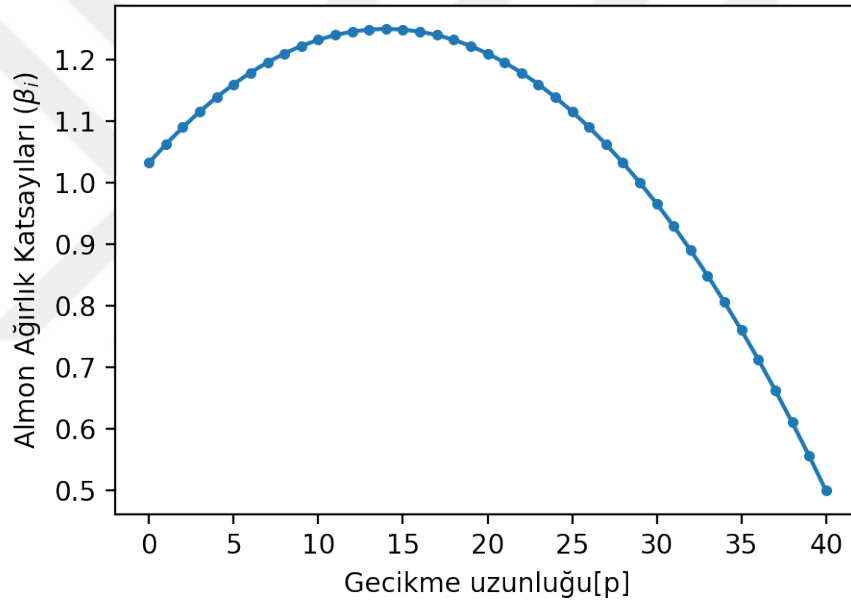
3.3.1 Almon Gecikme Polinomu Ağırlıklandırması

Almon gecikme polinomu, ilk olarak Almon tarafından 1965 yılında, imalat sanayinde geçmiş ve mevcut yatırım ödeneklerinin yatırım harcamasına dönüşme tahmini için, gecikmelerin az sayıda parametreyle ifade edilmesi amacıyla kullanılmıştır. Farklı sektörlerle ait verilere ilişkin örüntüler üzerinden, gecikme sayısı kadar parametrenin kestirimini yol açtığı karmaşa yerine, yalınlaştırma polinomu kullanılarak

daha az sayıda parametre ile model kurulmuştur. Tek değişkenli gecikmesi dağıtılmış(DL) regresyon modeli aşağıda (3.26)'de yer almaktadır.

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_p X_{t-p} + u_t \quad t = 1, 2, \dots, T \quad (3.26)$$

Yukarıda gecikme katsayıları β_i , gecikme uzunluğu $[1, 2, \dots, p]$ olurken, düşük frekanslı seri $t = [1, \dots, T]$ aralığındadır. Aşağıda denklem (3.27)'de β_i katsayılarının ikinci dereceden bir polinom yoluyla yalınlaştırılması gösterilmiştir. Polinom derecesi konusunda teorik olarak bir kısıtlama olmamakla birlikte, uygulamada dördüncü dereceyi aşan polinomlar tercih edilmez. Yalınlaştırma için Almon ağırlıklandırma prosedürünü işletildiğinde, polinom derecesi k 'nin gecikme uzunluğu p 'den küçük olduğu, yani $k < p$ koşulunun sağlandığı varsayılır (Judge, Hill, Griffiths, Lütkepohl ve Lee, 1988).



Şekil 3.1: Almon ağırlıklandırma polinomu

Yukarıda şekil 3.1'de ikinci dereceden bir Almon Gecikme Polinomununun, gecikme uzunlukları ile Almon gecikme katsayıları arasındaki ilişki görülmektedir.

$$\beta_i = \gamma_0 + \gamma_1 i + \gamma_2 i^2 + \dots + a_k i^k \quad i = 0, 1, 2, \dots, p \quad (3.27)$$

Bu denklemde, polinom ikinci dereceyle sınırlandırıldığında, kestirilecek parametre sayısı üç olup, indis değerleri i sıfırdan başlayarak p değerine kadar artırıldığında, (3.28) ve (3.31) denklemleri arasındaki ifadeler elde edilir.

$$\beta_0 = \gamma_0 \quad (3.28)$$

$$\beta_1 = \gamma_0 + \gamma_1 + \gamma_2 \quad (3.29)$$

$$\beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2 \quad (3.30)$$

⋮

$$\beta_p = \gamma_0 + p\gamma_1 + p^2\gamma_2 \quad (3.31)$$

Yukarıdaki denklem (3.31) ifadesi, regresyon denkleminin yer aldığı denklem (3.26)'de yerine konduğunda, denklem (3.32) ifadesi elde edilir. Denklem (3.32)'deki $\sum_{i=0}^p X_{t-i}$ ifadesi polinomdaki γ_i değerleriyle çarpıldığında (3.33) denklemi elde edilir. Denklem (3.33)'de toplam ifadeleri Z_0, Z_1, Z_2 olarak kısaltılırsa, denklem (3.34) ifadesine ulaşılır.

$$y_t = \alpha + \sum_{i=0}^p (\gamma_0 + \gamma_1 i + \gamma_2 i^2) X_{t-i} + u_t \quad (3.32)$$

$$= \alpha + \underbrace{\gamma_0 \sum_{i=0}^p X_{t-i}}_{Z_0} + \underbrace{\gamma_1 \sum_{i=0}^p i X_{t-i}}_{Z_1} + \underbrace{\gamma_2 \sum_{i=0}^p i^2 X_{t-i}}_{Z_2} + u_t \quad (3.33)$$

$$y_t = \alpha + \gamma_1 (Z_1 + Z_0) + \gamma_2 (Z_2 - Z_0) + u_t \quad (3.34)$$

Yukarıdaki denklem setinde, (3.33)'deki $Z_0 = \sum_{i=0}^p X_{t-i}$, $Z_1 = \sum_{i=0}^p i X_{t-i}$ ve $Z_2 = \sum_{i=0}^p i^2 X_{t-i}$ dönüşümleri yapılarak, denklem 3.34'deki sadeleştirilmiş ifadeye ulaşılır (Baltagi, 2011).

İkinci dereceden Almon gecikme işlevini, p gecikme uzunluğu (3.26) ve k dereceden polinom (3.27) ifadesinde yeniden düzenlendiğinde:

$$\begin{aligned} y_t &= \gamma_0 x_t \\ &+ (\gamma_0 + \gamma_1 + \gamma_2 + \dots + \gamma_k) x_{t-1} \\ &+ (\gamma_0 + 2\gamma_1 + 4\gamma_2 + \dots + 2^k \gamma_k) x_{t-2} \\ &\quad \vdots \\ &+ (\gamma_0 + n\gamma_1 + n^2\gamma_2 + \dots + n^k \gamma_k) x_{t-p} \\ &+ u_t \end{aligned}$$

Yukarıda yer alan γ_i polinom katsayıları ve k derece ile parantez içinde ifade edilirken, x ifadesi de x_t ile x_{t-p} aralığında düzenlenmiştir. Bu düzenlemenin ardından, x değerleri çarpım yoluyla parantez içine alarak ve γ_i değerleri ile yeniden düzenlendiğinde aşağıdaki ifadeye ulaşılır:

$$\begin{aligned}
 y_t &= \gamma_0(x_t + x_{t-1} + x_{t-2} + \dots + x_{t-p}) \\
 &+ \gamma_1(x_t + x_{t-1} + x_{t-2} + \dots + x_{t-p}) \\
 &+ \gamma_2(x_t + 2x_{t-1} + 4x_{t-2} + \dots + 2^k x_{t-p}) \\
 &\vdots \\
 &+ \gamma_k(x_t + x_{t-1} + 2^k x_{t-2} + \dots + n^k x_{t-p}) \\
 &+ u_t
 \end{aligned}$$

Denklem (3.27) ifadesi matris formunda ifade edildiğinde (3.35)'e ulaşılır:

$$\beta = H\gamma \quad (3.35)$$

Bu denklemde $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_p)$ olurken, H matrisi şu şekli alır:

$$H = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 2^2 & \dots & 2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s & s^2 & \dots & s^k \end{bmatrix} \quad (3.36)$$

Denklem (3.35)'deki, β vektörü genel regresyon denkleminde $y = X\beta + \varepsilon$ yerine konulduğunda:

$$y = XH\gamma + u = Z\gamma + \varepsilon \quad (3.37)$$

Yukarıda (3.37) ifadesine en küçük kareler yöntemi uygulandığında, $\hat{\gamma}$ parametre kestirimi $\hat{\gamma} = (Z'Z)^{-1}Z'y$ olur. Hata terimleri normal dağılıma sahiptir. $N(\gamma, \sigma^2(Z'Z))^{-1}$ ve β katsayılarının kestirimi $\hat{\beta} = H\hat{\gamma}$ denkleminin çözümüyle elde edilebilir (Judge vd., 1988, s.350-369).

3.3.2 Üssel Almon Gecikme Polinomu Ağırlıklandırması

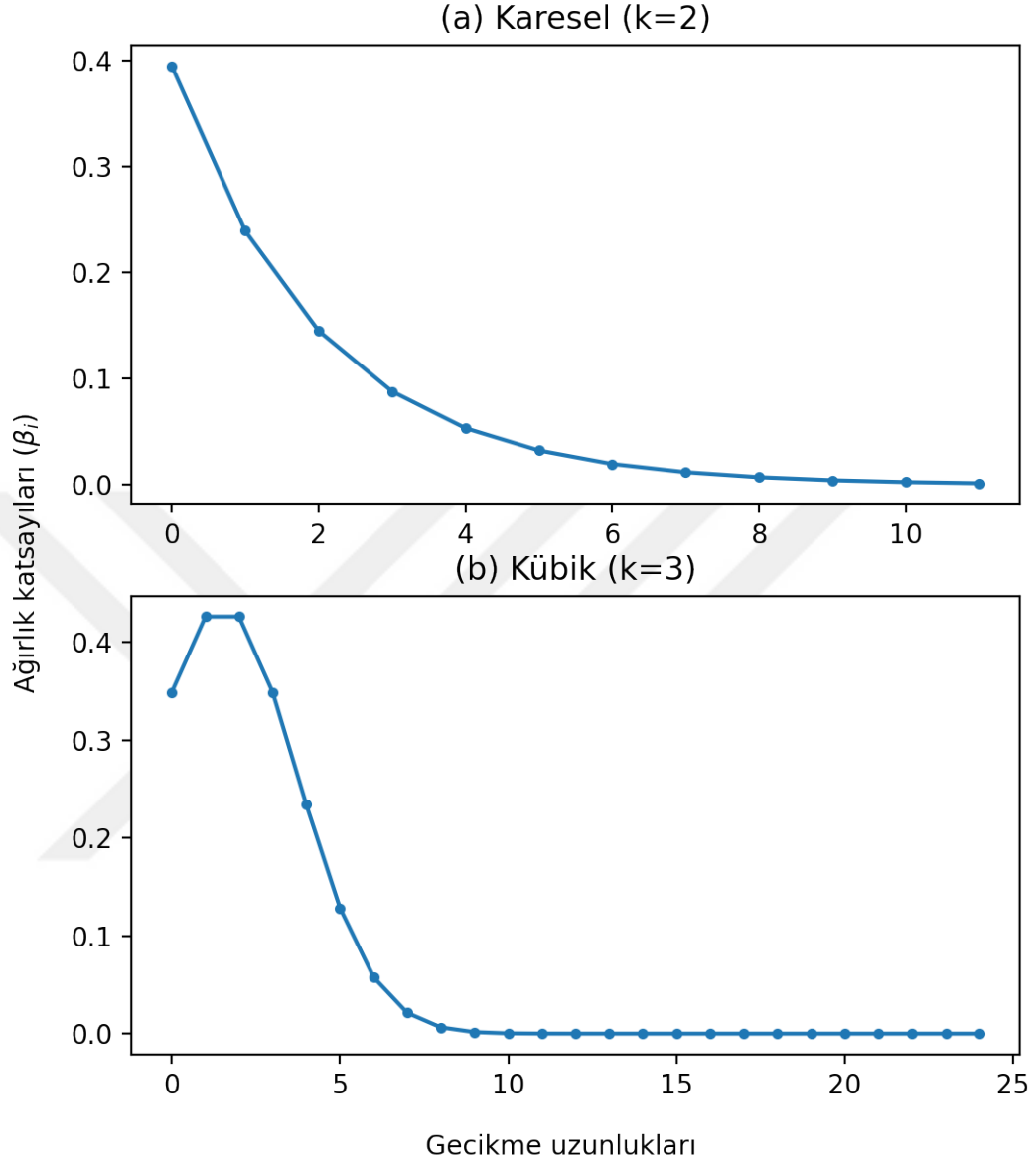
Üssel Almon gecikme işlevi, uygulamada en çok kullanılan ağırlıklandırma yöntemidir. Üssel Almon ağırlıklandırma işlevinin cebirsel gösterimi denklem (3.38)'de yer almaktadır. Parametre sayısı, polinom

derecesi k ile kısıtlanmıştır. İşlevin çıktısı, gecikme boyutunun uzunluğu p ve polinomun derecesi k tarafından belirlenir.

$$\beta(p; \theta) = \frac{\exp(\theta_1 i + \theta_2 i^2 + \dots + \theta_k i^k)}{\sum_{i=0}^p \exp(\theta_1 i + \theta_2 i^2 + \dots + \theta_k i^k)} \quad (3.38)$$

Bu denklemde, \exp ifadesi doğal logaritma tabanı $e = 2,7172..$ gösterirken, β ağırlık katsayılarını, $i=[0,1,\dots,p]$ olmak üzere p gecikme uzunluğunu, k polinom derecesini ve $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ boyutlu kısıtlanmış parametre vektörünü ifade eder.

Üssel Almon gecikme polinomu'nun matematiksel ifadesi beklenen en büyük değer algoritmasına (Expectation Maximum -EM- Algorithm) dayanır. EM algoritması, en çok olabilirlik kestirimine dayanan bir iterasyon yöntemidir. EM algoritması, eksik veri probleminin giderilmesine yönelik olarak, en çok olabilirlik yöntemi ile çözüm bulmaya çalışan [Dempster, Laird ve Rubin \(1977\)](#) tarafından önerilmiştir.



Şekil 3.2: Üssel Almon gecikme polinomu

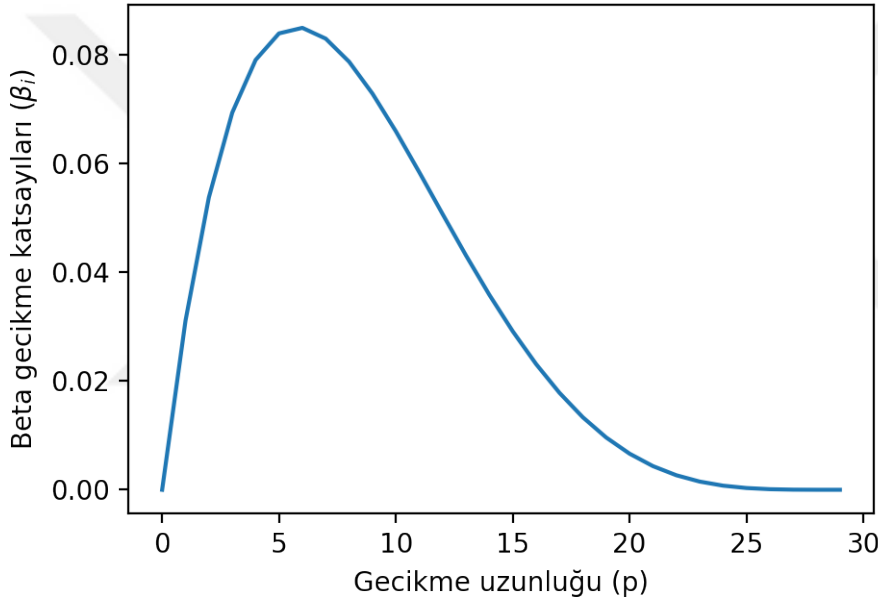
Şekil 3.2’de üssel Almon gecikme polinomunun karesel ve kübik derecelerinin formları yer almaktadır. Ağırlıklandırma işlevinin temel parametreleri, gecikme uzunluğu (p) ve polinomun derecesi (k) değerleridir. Polinom derecesinde kübik, yani üçüncü dereceden daha yüksek değerler genelde tercih edilmez. Modele yüksek frekanslı serinin özelliklerinin yansıtılması gecikme uzunluğu p tarafından sağlanır ve en uygun değeri aranır.

3.3.3 Beta İşlevi

Beta işlevi, θ_j parametrelerinin değerlerine bağlı olarak oldukça esnek bir ağırlıklandırma yapısı sunar. Aşağıda görüldüğü üzere, p gecikme sayısı, θ_1 ve θ_2 parametreleri ile sınırlandırılmış bir yalınlaştırma yöntemi olup, beta işlevini temel alır.

$$C(p; \theta_1; \theta_2) = \frac{f(\frac{i}{p}; \theta_1; \theta_2)}{\sum_{i=0}^p f(\frac{i}{p}; \theta_1; \theta_2)}, \quad f(p, \theta) = \frac{\Gamma(p + \theta)}{\Gamma(p + 1)\Gamma(\theta)} \quad (3.39)$$

Burada, i gecikme indisini gösterirken p en büyük gecikme uzunluğunu ifade eder. Beta ağırlıklandırma işlevi tek parametre içerir, $\theta_1 = 2$ iken, $\theta_2 > 5$ olarak tanımlanmıştır.



Şekil 3.3: Beta Gecikme İşlevi

Yukarıda şekil 3.3’de beta ağırlıklandırma işlevinde, gecikme uzunlukları ile beta gecikme katsayıları arasındaki ilişki görülmektedir. Buna göre, beta ağırlıklandırma işlevinin kırılma noktası 25. gecikme civarında yatay hale yakınsamaktadır.

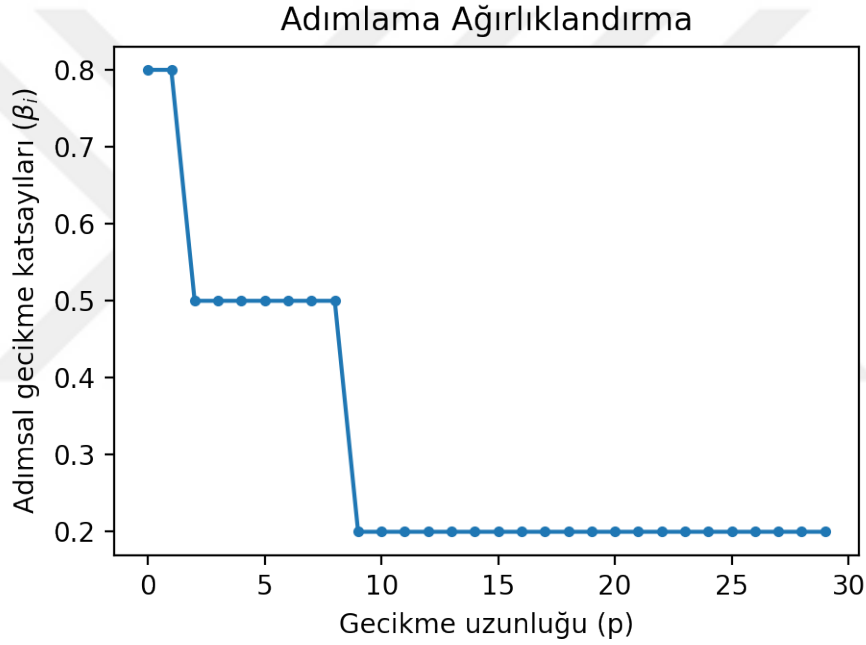
Beta ağırlıklandırma işlevinin MIDAS regresyonda kullanılması ilk olarak [Ghysels, Sinko ve Valkanov \(2007\)](#) tarafından önerilmiştir. Beta ağırlıklandırma işleviyle ilgili genel bilgiler ve diğer ağırlıklandırma işlevlerine yönelik kapsamlı tartışmalar [Ghysels, Valkanov ve Rubia \(2009\)](#)’da ele alınmıştır.

3.3.4 Adımsal Ağırlıklandırma

MIDAS regresyonun adımsal işlev ile ağırlıklandırılması görece basitleştirilmiş parametrik bir formdur. Adımsal ağırlıklandırma, doğrusal olmayan ve çok değişkenli modellerde önemli kolaylıklar sağlar. Adımsal ağırlıklandırmanın M adımla çalıştırıldığı model şöyledir:

$$y_t = \beta_0 + \sum_{i=1}^M \beta_i X_t(K_i, m) + \varepsilon_t \quad (3.40)$$

Burada m yüksek frekanslı değişkenin yinelenme frekansını gösterirken, $X_t(K_i, m) = \sum_{j=1}^K x_{t-j/m}^m$ ve $K_1 < \dots < K_M$ değerleri X^m 'in kısmi toplamlarını ifade eder. Adımlama sayısı M ile gösterilir, değeri arttıkça MIDAS regresyonunun yalınlaştırma seviyesi azalır (Ghysels vd., 2007).



Şekil 3.4: Adımsal Ağırlıklandırma İşlevi

Şekil 3.4'de adımsal ağırlıklandırma işlevinin, gecikme uzunluğuna göre katsayı değişimi ve kırılma noktaları görülmektedir. Adımsal ağırlıklandırma işlevinin şekli merdivene benzer, kırılma noktaları ağırlık katsayılarının değiştiği gecikme aralıklarını gösterir. Gecikme uzunluğu arttıkça, ağırlık katsayıları azalır ya da sıfır olur.

3.3.5 Diğer Ağırlıklandırma İşlevleri

Ağırlıklandırma işlemlerine uygun farklı işlevler de vardır. Yukarıda incelenen, Üssel Almon, beta ve adımlama işlevleri en çok kullanılan yöntemlerdir. Almon gecikme polinomu kuramsal arka planın anlaşılmasında önemli bir role sahiptir. Uygulamada yaygın kullanılmamasına rağmen, Gompertz,

Log-Cauchy ve Nakagami işlevleri de ağırlıklandırma için kullanılabilir. MIDAS regresyon için R programlama dilinde yazılmış "*midasr*" paketinde, Gompertz, Log-Cauchy ve Nakagami ağırlıklandırma işlevleri için, sırasıyla, "gompertzp", "lcauchyp" ve "nakagamip" fonksiyonlarına yer verilmiştir (Ghysels, Kvedaras ve Zemlys, 2016). Ayrıca, yine uygulamada sık kullanılmayan, doğrusal, hiperbolik ve geometrik yapıda ağırlıklandırma yöntemleri de önerilmiştir (Ghysels, Valkanov ve Rubia, 2009). Zaman serisi analizlerinde yaygın kullanılan Eviews yazılımında, MIDAS modülünde sadece Üssel Almon, Beta, Almon polinomu ve Adımlama yöntemlerine yer verilmiştir.

Benzetim deneyinde(bölüm 6), üssel azalan (exponential decline), tümsek (hump-shaped), doğrusal azalan ve çevrimsel (cyclical) ağırlıklandırma işlevleri, rassal yüksek frekanslı serinin dönüştürülmesinde kullanılmıştır. Üssel azalan ağırlıklandırma işlevi, Üssel Almon işlevinin ikinci dereceyle sınırlandırıldığı özel bir formudur. Tümsek işlevi, yine Üssel Almon gecikme polinomunun ikinci dereceden özel bir hali olup, θ_2 parametresi negatif değer alır. Doğrusal azalan ağırlıklandırma işlevi, gecikme uzunluğu arttıkça, doğrusal bir eğimle azalır. Çevrimsel ağırlıklandırma işlevi, trigonometrik sinüs işlevini baz alarak, artan ve azalan, negatif ve pozitif değerler alan bir yapıdadır. Benzetim deneyinde kullanılan ağırlıklandırma işlevleri, benzetim deneyinde stres testi yapılmasına olanak sağlayan formlar oluşturmak için kullanılmıştır.

3.4 MIDAS Regresyon Modellerindeki Yöntemsel Farklılıklar ve Sınıflandırma

MIDAS regresyon modellerinin, kullanıldığı alanlar genişledikçe, yeni yöntemsel geliştirmeler yapılmıştır. Literatürde MIDAS regresyon yöntemini temel alan, farklı değişken ve veri setlerinin analizleri için geliştirilen yöntemler, aşağıda genel olarak sınıflandırılmıştır.

3.4.1 Temel MIDAS Regresyon Modeli

Temel MIDAS regresyon modeli, farklı frekanslarda örneklem içerene, az sayıda parametreyle yalınlaştırılmış, indirgenmiş form regresyon modeli olarak tanımlanır. MIDAS regresyon, gecikmesi dağıtılmış yüksek frekanslı değişkenlerin, koşullu beklenen değerlerini ifade eder. Yüksek frekanslı seri ve gecikmesi dağıtılmış değerlerinin yol açtığı parametre çoğalması sorununun, ağırlıklandırma işlevleri aracılığıyla yalınlaştırılarak oluşturulan regresyon modelleri olarak tanımlanır (Ghysels ve Marcellino, 2018, s.453-502).

$$y_t^L = \beta_0 + \beta_1 C(L^{1/m}; \theta_h) x_t^H + \varepsilon_t^L \quad (3.41)$$

Yukarıdaki ifade, basit gecikmesi dağıtılmış MIDAS regresyonu ifade eder, aynı zamanda DL-MIDAS regresyon modeli olarak da adlandırılır.

$$y_t^L = \beta_0 + \beta_1 x_t^L + \varepsilon_t \quad (3.42)$$

Gecikmesi dağıtılmış regresyon modelinde, düşük frekans L ile yüksek frekans Y arasındaki fark açıldıkça ve gecikme uzunluğu artıkça, parametre çoğalması sorunu yaşanır. MIDAS regresyon ağırlıklandırma işlevi $C(L^{1/m})$ aracılığıyla model yalınlaştırılarak ve az sayıda parametre ile kestirim yapılır. Yüksek esneklik avantajlarıyla Üssel Almon ve Beta ağırlıklandırma işlevleri, uygulamalarda iki ya da üç parametre ile ağırlıkların hesaplanmasına imkan verdiklerinden yaygın olarak kullanılırlar.

MIDAS Regresyon, farklı frekanslarda örneklendirilmiş değişkenlerin, yüksek frekanslı örneklemelerin düşük frekansa dönüşümüyle sıkı biçimde parametrelendirilmiş, indirgenmiş form modelleridir (Ghysels ve Qian, 2019). Denklem (3.43)'de doğrusal linear MIDAS regresyon modeli yer almaktadır.

$$Y_t^L = \beta_0 + \beta_1 C(L) X_t + \varepsilon_t \quad (3.43)$$

$$Y_t^L = \beta_0 + \beta_1 C(L^{1/m}) X_t^m + \varepsilon_t^m \quad (3.44)$$

$$Y_t^L = \beta_0 + \beta_1 C(L^{1/m}; \theta) X_t^m + \varepsilon_t^m \quad (3.45)$$

Denklem (3.43), (3.44) ve (3.45)'te L işareti düşük frekans ve m işareti yüksek frekanslı değişkenin bir periyot içindeki yinelenme sıklığını ifade ederken, Y_t^L ifadesi düşük frekanslı bağımlı değişkenin t periyodundaki değerlerini gösterir. Yukarıdaki denklemleri $t + h$ dönem sonrası tahmini olarak yazıldığında:

$$Y_{t+h}^L = \beta_0 + \beta_1 C(L^{1/m}; \theta) X_t^H + \varepsilon_{t+1}^H \quad (3.46)$$

Yüksek frekanslı bağımsız değişken X 'in dönüşümü ağırlıklandırma işleviyle yapılır. Ağırlıklandırma işlevi, polinom derecesi ve gecikme uzunluğu girdileri ile ağırlık katsayıları kestirimi yapar. Örneğin, Üssel Almon ağırlıklandırma işleviyle dönüşüm yapıldığında, ikinci dereceden bir polinom seçilirse iki katsayı, eğer kübik polinom seçilirse üç katsayı kestirimi yapılır. Polinom derecesi ve gecikme uzunluğunun optimizasyonu problemi, kısıt altında doğrusal olmayan regresyon ile çözülebilir. Kısıtlanmış doğrusal olmayan regresyonda iteratif yöntemlerle en uygun çözüm aranır.

3.4.2 Oto-regresif Gecikmesi Dağıtılmış MIDAS (ADL-MIDAS)

MIDAS regresyon modeline oto-regresif terim eklenerek makroekonomik tahminlerde kullanılması için geliştirilen, oto-regresif (AR-MIDAS) ve oto-regresif gecikmesi dağıtılmış MIDAS (ADL-MIDAS) modellerinin incelenmesi, M. Clements ve Galvão tarafından 2008 yılında ele alınmıştır. ADL-MIDAS modeli, Stock ve Watson (2003) yılında ABD ekonomisi üzerine oto-regresif (ADL) temelli

oluşturdukları öncü göstergeler modeli ile [Ghysels vd. \(2004\)](#) temel MIDAS regresyon modeli önermelerinin, [M. Clements ve Galvão \(2008\)](#) tarafından bir araya getirilmesiyle yeniden tanımlanmıştır.

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/3}, \theta) x_{t-1}^{(3)} + \varepsilon_t \quad (3.47)$$

Denklem (3.47), aylık öncü göstergelerden x_{t-1} ile üç aylık bağımlı değişkenin y_t ADL-MIDAS modelini ifade etmektedir. Denklemde, bağımlı değişken y_t kendi gecikmeli değeri y_{t-1} ile ilişki içindedir, $B(L^{1/3}, \theta)$ MIDAS ağırlık polinomunu ifade eder, yinelenme sıklığı $m = 3$ olduğu için gecikme işlemcisi $L^{1/3}$ ifadesiyle gösterilmiştir.

ADL modelinde denklem (3.14)'de yapılan sadeleştirme ile (3.48) ifadesi yeniden yazılarak aşağıdaki denklem elde edilir.

$$y_t = \beta_0(1 - \lambda)^{-1} + \beta_1(1 - \lambda)^{-1} B(L^{1/3}, \theta) x_{t-1}^{(3)} + \tilde{\varepsilon}_t \quad (3.48)$$

Denklem (3.48)'de, önceki denklem (3.14)'de de çözümü yeralan $\tilde{\varepsilon}_t = (1 - \lambda L)^{-1}$ eşitliği geçerlidir. Bir adım öteye giderek, bağımlı değişken y_t ile bağımsız değişken x arasında, denklem (3.14)'te olduğu gibi ortak gecikme faktörü olduğunu varsayıldığında, ADL-MIDAS modeli şöyle olur:

$$y_t^* = \beta_0 + \lambda y_{t-1} + \beta_1(1 - \lambda)^{-1} B(L^{1/3}, \theta) x_{t-1}^{(3)} + \tilde{\varepsilon}_t \quad (3.49)$$

Burada bir dönem öncesi yerine, bağımlı değişken için d ve bağımsız değişken için h gecikme değerleri biçiminde yazılırsa:

$$y_t^* = \lambda y_{t-d} + \beta_1 + \beta_2 B(L^{1/3}, \theta) (1 - \lambda^d)^{-1} x_{t-h}^{(3)} + \tilde{\varepsilon}_t \quad (3.50)$$

Denklem (3.50) ifadesi, doğrusal olmayan en küçük kareler yönetimi $\hat{\lambda}$ ve $\hat{\theta}$ değerleri iteratif yöntemlerle, hata kareler toplamını minimize edecek şekilde çözülebilir.

Yukarıda işleyişi açıklanan otoregresif MIDAS regresyonu, [M. Clements ve Galvão \(2008\)](#) tarafından AR-MIDAS olarak tanımlanmış; daha sonra çoklu öncü göstergeler modelleri olarak geliştirilerek, [M. P. Clements ve Galvão \(2009\)](#) tarafından M-MIDAS olarak adlandırmışlardır.

3.4.3 Kısıtsız MIDAS Regresyon Modeli

Kısıtsız MIDAS regresyon modeli (UMIDAS, Unrestricted MIDAS), ağırlıklandırma işlevi olmaksızın yüksek frekanslı bağımsız değişkenler ile bağımlı değişken arasında parametre kestirimi yapılmasına

imkan sağlar. Özellikle MIDAS regresyon analizinin yoğun olarak kullanıldığı makroekonomik verilerde, genellikle düşük ve yüksek frekanslı seri arasındaki fark küçüktür. Parametre çoğalmasının olmadığı durumlarda, gecikmesi dağıtılmış model kullanmak gerekli olmayabilir. Kısıtsız MIDAS regresyon modeli, UMIDAS olarak adlandırılmıştır. UMIDAS ve MIDAS regresyon sonuçları, Monte Carlo benzetimi ve kısa vadeli makroekonomik değişkenler üzerinde performans testi karşılaştırması, [Froni, Marcellino ve Schumacher \(2015\)](#) tarafından yapılmıştır. Genelde yüksek ve düşük frekans arasındaki uzunluğun küçük olduğu ve kısa vadeli durumlarda, UMIDAS iyi performans gösterirken, frekans sıklığı artıktıkça beklendiği üzere ağırlıklandırma işleviyle kısıtlanmış MIDAS regresyon modeli daha iyi performans sergilemiştir.

$$Y_t^L = \beta_0 + \sum_{p=1}^P \beta_p X_{t-p}^H + \varepsilon_t \quad (3.51)$$

Burada, X bağımsız değişken matrisini, y düşük frekanslı bağımlı değişkeni, m yinelenme frekansını, gecikme dönemi sayısı $p = 1, 2, \dots, P$ ve β_i gecikme katsayılarını gösterir. Regresyon katsayıları β_i , en küçük kareler (EKK) yöntemiyle hesaplanır. Katsayılar, bağımsız değişken matrisi X 'in gecikmeli değerlerine ait parametre kestirimleridir. Aşağıda denklem (3.51)'de, ağırlıklandırma işlevinin olmadığı, kısıtsız MIDAS modelinin genel ifadesi yer almaktadır.

Kısıtsız MIDAS (UMIDAS) regresyon modelinin sadece bir dönem gecikmeli basitleştirilmiş şekli aşağıdaki denklemde yer almaktadır:

$$y_t = \delta_1(L)x_{t-1} + \varepsilon_t \quad (3.52)$$

Temel MIDAS regresyonda ağırlık işlevi $C(L, \theta)$ iken, denklem (3.52)'deki kısıtsız MIDAS formülünde (L) olarak yer almaktadır. MIDAS regresyon modelinde ağırlık katsayıları kestirimi doğrusal olmayan regresyon (NLS) ile yapılırken, kısıtsız MIDAS (UMIDAS) yönteminde, kısa vade ve frekans sıklığının az olduğu durumda doğrusal regresyon modeliyle, en küçük kareler (EKK) yöntemi kullanılarak kestirim yapılabilir ([Froni, Marcellino ve Schumacher, 2015](#)). R programlama dilinde "midasr" paketinde de kısıtsız MIDAS yönteminde parametre kestirimi EKK ile yapılmaktadır ([Ghysels vd., 2016](#)).

3.4.4 Faktör MIDAS

Şimdinin tahmininde, bir kaç göstergeden oluşan değişken setlerinde MIDAS regresyon kullanılırken, çok sayıda değişken ve gözlem içeren geniş veri setlerinde Faktör MIDAS (Factor MIDAS) yöntemi kullanılmaktadır. Şimdinin tahmini modellerinin işletilmesinde, açıklanan verilerin zamanlamasındaki farklılıklar ve düzensizlikler nedeniyle eksik veriler ve dengesiz veri setlerinin oluşması, açıklayıcı değişkenlerin periyotlarındaki farklılıklar, önemli sorunlara ve model belirsizliğine yol açar. Bu sorunlara karşı etkin işleyen bir model yapısı [Kuzin, Marcellino ve Schumacher \(2013\)](#) tarafından önerilen Faktör MIDAS, geniş veri setleri için şimdinin tahminine olanak sağlar. Faktör-MIDAS

modelinde, faktör kestirim yöntemiyle faktör sayısı indirgenerek, gecikme uzunluğu ve gösterge seçimi yapılır.

$$y_{t+1} = \beta_+ \beta' F_t + \eta_{t+1} \quad (3.53)$$

$$x_t = \Phi F_t + \varepsilon_t \quad (3.54)$$

Burada, y bağımlı hedef değişken, F_t analize dahil edilen tüm değişkenlerin ortak faktörleri, β' faktör yükleri vektörünü gösterir. Ayrıca F_t , yükleme matrisi Φ tarafından belirlenir. Geniş ve dengesiz veri setlerinde değişken seçiminde bilgi kriterleri ve değişken indirgeme yöntemleri konusunu [Kapetanios, Marcellino ve Papailias \(2014\)](#) tarafından tartışılmıştır.

3.4.5 Adımlama Yöntemi

MIDAS regresyon analizinde, gecikme uzunlukları ağırlıklandırma işlevleriyle doğrusal olmayan kısıtlara tabii tutulduğundan, doğrusal olmayan kestirim yöntemlerinin kullanılması gerekmektedir. Adımlama Yöntemi (Step Method) görece daha basitleştirilmiş bir parametre kestirim yöntemi olarak ortaya çıkmıştır. Adımlama yöntemi ilk olarak [Forsberg ve Ghysels \(2007\)](#) tarafından ifade edilirken, [Ghysels, Sinko ve Valkanov \(2007\)](#) tarafından da ele alınmıştır.

$$y_t = \beta_0 + \sum_{i=0}^M \beta_i X_t(K_i, m) + \varepsilon_t \quad (3.55)$$

Burada, $X_t(K, m) \equiv \sum_{j=1}^K x_{t-j/m}^{(m)}$ denkliğine sahip ve yüksek frekanslı seri $x^{(m)}$ 'in kısmi toplamlarını ifade ederken, $K_1 < \dots < K_M$ ilişkisi geçerlidir. Yüksek frekanslı değişken $x_t^{(m)}$ 'in tüm kısmi etkileri $\sum_{i=1}^M \beta_i$ ile ölçülür. Kısmi etkiler $x_t^{(m)}$ 'in, $K_1 < j$, K_2 değeri $\sum_{i=2}^M \beta_i$ tarafından belirlenir. Regresyon modelindeki adım sayısı arttıkça, yalınlaşmadan uzaklaşılar. Yalınlaşmadan uzaklaşma ise, MIDAS regresyon yaklaşımının amacıyla örtüşmez ([Ghysels vd., 2007](#), s.61-62).

3.4.6 MIDAS Modellerinin Diğer Uzantıları

MIDAS regresyon modelleri yalınlaştırma yaklaşımı içerirler. Bu, zaman serisi analizlerinde kullanılan yöntemlerle bütünleştirilmeleri için araştırmacıların ilgisini çekmiştir. Bütünleştirmeye yönelik pek çok yöntem önerilmiştir, halen geliştirmeler devam etmektedir. Genel başlıklar altında bu yeni geliştirilen yöntemlere değinilmiştir. Örneğin, çok değişkenli MIDAS regresyon, parametrik olmayan MIDAS, asimetrik doğrusal olmayan ve yarı-parametrik MIDAS, düzleştirilmiş geçişli (Smooth Transition) MIDAS, Markov Yönlendirmeli (Markov Switching) MIDAS, Quantile MIDAS regresyon, MIDAS-VAR ve volatile tahminine yönelik olarak da MIDAS-GARCH. MIDAS regresyondaki

gelişme alanlarına ve yöntemlere ilişkin kapsamlı bir özete (Ghysels ve Marcellino, 2018, s.466-476) çalışmasında yer verilmiştir.

3.5 Model Seçimi ve Bilgi Ölçütleri

Gecikmesi dağıtılmış (DL) ve gecikmesi dağıtılmış otoregresif modeller (ADL), bağımlı ve bağımsız değişkenler zamana bağlı olarak değişim gösterdiğinden dinamik modeller olarak tanımlanır. Zaman serisi regresyon modelleri AR, MA, ARMA ve ARIMA modellerinde gecikme genişliği seçimi kritik öneme sahiptir. Söz konusu yöntemlerde en uygun gecikme uzunluğunu belirlemek için, Akaiki Bilgi Ölçütü (AIC), Düzeltilmiş Akaiki Bilgi Ölçütü (AICc), Bayeşçi Bilgi Ölçütü(BIC) ve Hannan-Quinn(HQ) Ölçütü yaygın olarak kullanılmaktadır.

MIDAS regresyonda modelinde, ağırlıklandırma işlevindeki gecikme sayısı parametre sayısını etkilemediğinden, bilgi ölçütlerindeki ceza işlevlerinin etkisi güçlü olmamaktadır. MIDAS regresyonda gecikme uzunluğunun (p) büyük seçilmesinin tek cezası başlangıç aşamasında daha fazla gözlem satırı kaybedilmesidir. Bunun nedeni, gecikme uzunluğunun yüksek değerlerinde ağırlık katsayılarının sıfıra yakınsamasıdır. Gecikme uzunluğu p 'i küçük seçmek, ağırlıklandırma işlevinin şeklini etkileyeceğinden daha sorunludur (Ghysels ve Marcellino, 2018, s.462).

Bölüm 4

Parametrik Olmayan MIDAS

Regresyon

Geleneksel MIDAS regresyon modelinde, çoğunlukla Üssel Almon ya da Beta gecikme ağırlıklandırma işlevleri kullanılmaktadır. Üçüncü bölümde incelenen klasik MIDAS regresyon modelleri, yüksek frekanslı verinin düşük frekanslı veri ile eşleşmesini sağlayan ağırlıklandırma işlevinin şeklinin önceden bilinmesi nedeniyle parametrik bir yaklaşımdır. Önsel olarak bir dağılım modelinin var olduğu varsayımı ortadan kaldırıldığında, parametrik olmayan analiz yöntemleri gerekli olur. Bu bölümde, parametrik olmayan MIDAS regresyon modelinin, parametrik MIDAS regresyon modellerine göre üstünlükleri ve zayıflıklarını ele alınmıştır. Parametrik olmayan yöntemlerden cezalandırılmış splayn regresyon yönteminin MIDAS regresyona uygulanmasına yönelik literatürde yer alan araştırmalar tartışılmıştır. Son olarak, parametrik olmayan yöntemlere katkı sağlamak üzere çekirdek regresyon yöntemlerinin, MIDAS regresyona yöntemime uyarlanabilirliği tartışılmıştır.

4.1 Temel kavramlar

Regresyon analizi için üç farklı beklenen koşullu olasılık modeli oluşturulabilir: i) Parametrik Regresyon Modeli ii) Parametrik Olmayan Regresyon Modeli, iii) Yarı Parametrik Regresyon Modeli. Bu üç temel gruba ilişkin kavramlar ve özellikler aşağıdaki alt başlıklarda genel olarak incelenmiştir.

4.1.1 Parametrik yaklaşım

Parametrik regresyon modellerinde anakütle dağılımı bilinmektedir. Örneklem üzerinden hesaplamalarla, merkezi limit teorimi ve normal dağılımdan yararlanarak parametre kestirimleri elde

edilmektedir. Bağımlı(Y) ve bağımsız değişkenler(X) arasındaki ilişkinin modeli doğrusal, üssel, yüksek dereceli polinom şeklinde olabilir.

Parametrik Regresyon Analizinde, Genelleştirilmiş Doğrusal Modelin temel varsayımlarını matematiksel olarak şöyle gösterilebilir (Ruppert, Wand ve Carroll, 2003a, s. 15-55):

$$Y_i = x_i' \beta + \varepsilon_i \quad i = 1, \dots, N \quad (4.1)$$

Parametrik regresyon analizinin, doğru bir ilişki tanımlayabilmesi için çok sayıda kısıtlayıcı varsayıma uyulması gerekir. Doğrusal regresyon modelinin temel varsayımları önceki bölümde ele alınmıştı.

4.1.2 Parametrik olmayan yaklaşım

Parametrik olmayan yöntemlerin, parametrik yaklaşımdan en ayırtıcı özelliği, örneklemin alındığı anakütlenin şekli için bir bilginin olmaması ya da sabit bir dağılımının bulunmamasıdır. Parametrik olmayan yaklaşımda, düzeltirici işlevi bilinmeyen olarak alınmakta ve ilişkinin şekli veri kümesi tarafından belirlenmektedir. Aşağıda denklem (4.2)'de $m(X_i)$ bilinmeyen işlevinin yani regresyon denkleminin genel şeklini ifade eder. Burada yer alan hata terimi ε_i 'lerin bağımsız ve sıfır ortalamaya sahip olmaları dışında hiçbir kısıtlayıcı varsayım bulunmamaktadır (Başarır, 2001, s. 275).

$$Y_i = m(X_i) + \varepsilon_i \quad (4.2)$$

Veri analizinde ilk adım, verilerin ham haliyle saçılım grafiğinin incelenmesidir. Değişkenler arasındaki ilişkinin şekli parametrik doğrusal olmayan regresyon yöntemleri kullanılarak yapıldığında, yüksek dereceli polinom ile parametre kestirimi yapılabilse bile, saçılım grafiklerinde gözlenen düzensiz tepe, vadi ve iniş-çıkışlar nedeniyle, belirgin bir işlevsel biçime uydurmak her zaman mümkün olmayabilir. Durağan olmayan düzensiz örüntülerin belirgin olduğu durumlarda, parametrik olmayan regresyon yöntemleriyle model kurmak, daha verimli sonuçlar elde edilmesi sağlayabilir.

4.1.3 Yarı-Parametrik yaklaşım

Yarı-Parametrik (semi-parametric) yaklaşım, modelde yer alan bazı açıklayıcı değişkenlerin bağımlı değişken ile bağıntısı parametrik bir dağılıma uygun iken, diğer değişkenlerin parametrik olmayan yöntemlerde kestirimin yapıldığı karma modellerdir. Yarı-Parametrik yaklaşımda $m(\bullet)$ düzeltirici işlevinin bazı parametreleri kestirilebilir iken, diğerlerinin şekli bilinmemekte ancak veri üzerinden belirlenebilmektedir.

$$Y_i = x_i' \beta + m(z_i) + \varepsilon_i \quad i = 1, \dots, N \quad (4.3)$$

Yarı-Parametrik yaklaşım, bağımlı değişken ve bağımsız değişkenler arasındaki ilişkinin modelinin ortaya konmasında, bazı değişkenler için parametrik, diğer bazı değişkenleri de parametrik olmayan yöntemlerle incelenmesine ilişkin ifade denklem (4.3)'de yer almaktadır (Hardle, Müller, Sperlich ve Werwatz, 2012). Yarı-parametrik yaklaşım karma bir model olması, modelin parametrik yöntemlerin uygulanabilmesinin yanı sıra, parametrik olmayan yöntemlerin esnek yapısını birlikte kullanabilmesi nedeniyle yaygın kullanım alanı bulmaktadır.

4.2 Splayn regresyon modelleri

Splayn regresyon modelleri, parametrik olmayan regresyon yöntemleri arasında yer alır. Literatürde, splayn model kavramı çatısı altında pek çok yöntem ve tanımlama olmasının yanında, temel splayn modelleri (B-splines) ve düzleştirilmiş ya da cezalandırılmış splayn regresyon modelleri (P-Spline) olarak iki ana gruba ayırmak yaygındır.

4.2.1 Temel Splayn Regresyon Modelleri

Splayn Regresyon Yöntemi, parametrik olmayan bir regresyon yöntemidir. Bir veri setine ait değişkenlerin saçılım grafiği görünümü, farklı derinlik ve yüksekliklerde vadi ve tepeler ya da çok sayıda dönüş ve kırılma noktasının olduğu durumlarda verimli sonuçlar üreten yöntemlerdir. Splayn regresyonda model ve parametre kestirimleri düzleştirme parametresi, düğüm sayısı ve düğümlerin konumları üzerinde oluşur. Doğrusal, karesel ve kübik splayn işlevleri temel splayn modelleri (B-splines) olarak adlandırılır.

Aşağıdaki alt başlıklarda, temel splayn regresyon modellerinin açıklanmasında, literatürde yaygın örnek olarak kullanılan LIDAR veri setinden yararlanılmıştır. LIDAR veri seti, atmosferdeki kimyasal bileşenlerin incelenmesi için, lazer ışınlarının yansımalarının ölçülmesine dair değişkenlerin yer aldığı bir veri setidir. Söz konusu veri, 221 gözlemden oluşmakta, bağımsız değişken ışık seviyesi ve bağımlı değişken yansıma düzeyinin logaritması biçimindedir (Ruppert vd., 2003a).

4.2.1.1 Doğrusal Splayn Regresyon

Doğrusal splayn regresyon modelinde, saçılım grafiğinde gözlenen belirgin kırılma noktalarının tespit edilmesi ve bu kırılma noktaları arasında, hata kareler ortalamasının minimum yapılarak, doğru

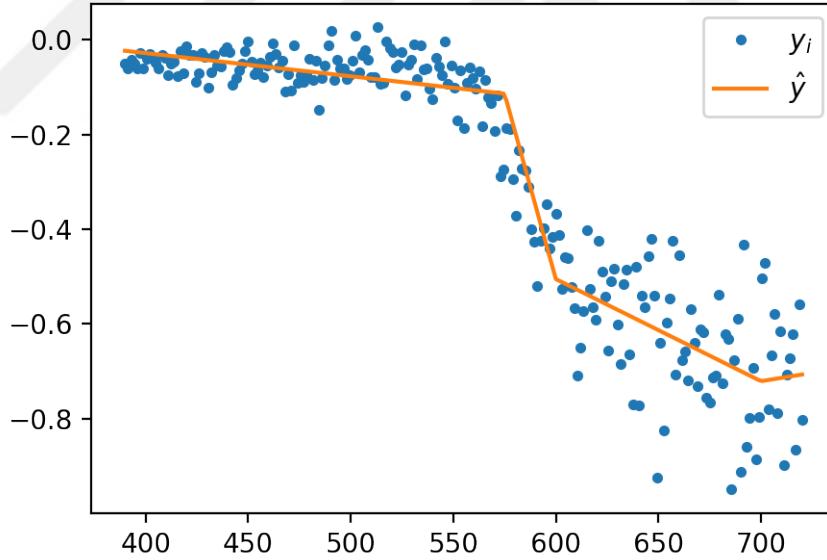
işlevinin çizilmesi esasına dayanır. Burada dikkat çekici hususlar, kaç adet kırılma noktasının seçileceği ve kırılma noktalarına ait bağımsız değişken x_i değerlerinin nasıl tespit edileceği sorularıdır.

Doğrusal splayn regresyon modelinde, farklı düğüm noktaları için cebirsel ifade şöyledir:

$$f(x) = \beta_0 + \beta_1 X_i + \sum_{k=1}^{\kappa} b_k (x - \kappa_k) \quad (4.4)$$

$$(x - \kappa_k) \begin{cases} (x - \kappa_k) & \text{için } \geq 0 \\ 0 & \text{ö.d.i} \end{cases}$$

Burada κ_k değeri kırılma noktaları, k indisi kırılma noktasını ifade ederken κ ifadesi kırılma noktalarının sayısını ifade eder. Denklemde β katsayıları ile gösterilen ilk bölümü klasik EKK ile aynı iken, ek bölümünde yer alan $\sum_{k=1}^{\kappa} b_k (x - \kappa_k)$ ifadesi kırılma noktalarını ve yön değişikliklerinin konumunun bulunmasını sağlamaktadır. Şekil 4.1'de saçılım grafiği üzerine, kırılma noktalarının dışsal olarak belirlendiği splayn regresyon modeli çizgisi görülmektedir.



Şekil 4.1: Doğrusal splayn regresyon modeli

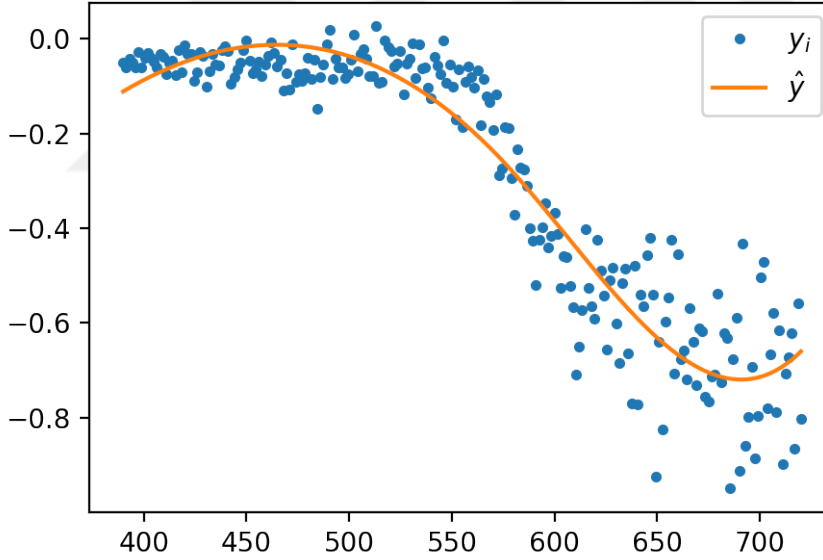
Splayn regresyon modellerinin uygulamasında, klasik E.K.K. modelindeki işlem süreçlerine benzer şekilde, bağımsız değişken matrisi oluşturulur. Aşağıdaki matris setinde, doğrusal splayn regresyon modelinin tanımlarına uygun olarak oluşturulmuştur.

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & (x_1 - \kappa_k) \\ \vdots & \vdots & \vdots \\ 1 & x_n & (x_n - \kappa_k) \end{bmatrix} \quad (4.5)$$

Burada, ilişkinin kırıldığı ya da yön değiştirdiği noktalar olan κ_k değerleri, \mathbf{x} vektöründen çıkarılarak, yeni diziler elde edilmiş şekli verilmektedir. κ bir vektör seti olarak, kırılma noktalarını gösterir.

4.2.1.2 Karesel Splayn Regresyon

Doğrusal splayn regresyon modelinde, doğrusallığın doğal yapısı gereği keskin dönüşler vardır. Bu keskin dönüşlerden kaçınmak için, $(x - \kappa_k)$ doğrusal formu ile modele x^2 terimi dahil edilebilir. Bu durumda süreklilik arz eden birinci türev elde edilebilir, birinci türevinin alınabilir olması analitik çözümleri kolaylaştırır. Ayrıca splayn regresyon eğrisinin dönüşleri, vadi ve tepeler daha yumuşak hale gelir.



Şekil 4.2: Karesel Splayn Model

Yukarıda şekil 4.2'de görülen örnek veri setinin karesel splayn modeli ile uyumu gözlenmektedir. Eğer şekil 4.1'deki doğrusal splayn regresyon ile karşılaştırılırsa, dönüşlerin daha yumuşak ve düzgün hale geldiği, sivri uçların olmadığı gözlenmektedir.

$$f(x) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \sum_{k=1}^K b_k(x - \kappa_k) \quad (4.6)$$

Aşağıda kırılma noktaları için, gerekli koşullar görülmektedir.

$$(x - \kappa_k) \begin{cases} (x - \kappa_k) & \text{için } \geq 0 \\ 0 & \text{ö.d.i} \end{cases}$$

Bu denklemde, doğrusal splayn regresyondan farklı olarak $\beta_2 X_i^2$ ifadesinin yer aldığı gözlenmektedir. Böylece örnek veri seti için, regresyon modeli için ikinci dereceden bir polinom oluşturulmuş olur. İkinci dereceden bir ifadenin denklemde yer alması, birinci türevinin alınabilir olmasını ve bu sayede daha yumuşak kırılma noktalarının oluşmasını mümkün hale gelir. Aşağıda örnek veri setine ait bağımsız değişken matrisinin formu (4.7)'de yer almaktadır.

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & (x_1 - 575)_+ & \dots & (x_1 - 600) \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & (x_n - 575)_+ & \dots & (x_n - 600) \end{bmatrix} \quad (4.7)$$

Matriste üçüncü sütün üzerinde x_i^2 ifadesi dikkat çekicidir ve doğrusal splayn regresyon modeline ait bağımsız değişken matrisinden farklıdır. Bu ifadenin eklenmesi şekil 4.2'de gözlendiği üzere dönüş noktalarının daha yumuşak geçişli olmasına olanak sağlamaktadır.

4.2.1.3 Kübik ya da Doğal Splayn Regresyon

Kübik splayn regresyon modeli, bağımsız değişkene matrisine X_i^3 değişkeni eklenerek, üçüncü dereceden bir polinom oluşturulması sağlanır. Kübik splayn model, doğal splayn regresyon modeli (Natural Spline Regression Model) olarak da adlandırılır. Üçüncü dereceden bir polinom, hem birinci ve hem de ikinci türevin alınmasına imkan verdiği için, optimum noktalarının belirlenebilmesi için gerekli ve yeter koşulların analitik olarak incelenmesine olanak sağlar.

$$f(x) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \sum_{k=1}^K b_k (x - \kappa_k) \quad (4.8)$$

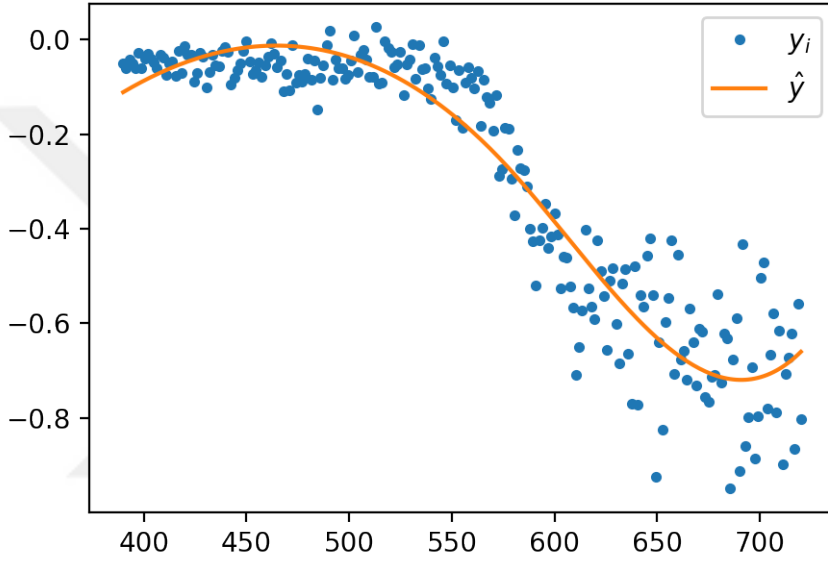
Bu denklemde, öncekilerden farklı olarak $\beta_3 X_i^3$ ifadesi eklenmiştir. Kırılma noktaları konumları ve kırılma noktası adedi, önceki doğrusal ve karesel modeller ile aynıdır. Karesel modelden farklı olarak grafiğin sağ tarafının üçüncü dereceden olmasının etkisi ile model eğrisi \hat{y} yukarı eğimi daha belirgindir.

$$(x - \kappa_k) \begin{cases} (x - \kappa_k) & \text{için } \geq 0 \\ 0 & \text{ö.d.i} \end{cases}$$

Bağımsız değişkenler matrisi X 'in dördüncü sütununa x_i^3 ifadesi eklenmiştir.

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & (x_1 - 575)_+ & \dots & (x_1 - 600) \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & (x_n - 575)_+ & \dots & (x_n - 600) \end{bmatrix} \quad (4.9)$$

Şekil 4.3'de yumuşatılmış dönüşlere sahip bir splayn regresyon eğrisi elde edilmiştir. Görüldüğü üzere, Şekil 4.1 ve 4.2'e göre dönüş noktaları daha da yumuşak hale gelerek, sağ uçtaki yukarı doğru eğim daha da belirgin hale gelmiştir.



Şekil 4.3: Kübik Splayn Model

4.2.2 Düzleştirilmiş Splayn Regresyon Modelleri

Splayn regresyon modelinde, düğüm sayısı κ arttıkça, uyumluluk düzeyi artma eğilimindedir. Bir yandan uyum artarken, diğer yandan regresyon çizgisi daha pürüzlü hale gelir. Bu durumda modelin soyutlama ve genellik düzeyini ifade etme yeteneği azalır. Düğüm noktası sayısının en uygun olması için, pürüzlük ve karmaşıklığın etkisini azaltmak için cezalandırılmış ya da diğer bir adlandırmayla düzleştirilmiş splayn regresyon modelleri geliştirilmiştir. Modelin pürüzlü olması durumu aşırı uyum sorunu (over fitting) olarak tanımlanır. Tersine, düğüm sayısının az olması modelin tahmin performansını azaltır, dikkate değer değişime neden olan faktörün ya da aralığın modelde yer almaması durumu, düşük uyum (under fitting) olarak tanımlanır. Düğüm noktası sayısının en uygun olması için, aşırı dalgalanmanın ve karmaşıklığın cezalandırıldığı düzleştirilmiş splayn regresyon modeli önerilmektedir.

Düzleştirilmiş splayn regresyon modellerinin, temel splayn regresyon modellerinden farkı, aşırı belirlenmiş modeli oluşumunu engellemek için cezalandırma ifadesinin kısıt olarak denkleme dahil edilmesidir. Doğrusal splayn modelde, kısıt olmadığında çok sayıda farklı eğri uydurmak mümkündür, fakat hangisinin en uygun değer olduğuna dair bir kestirim bulunmadığından, düğüm noktaları araştırmacı tarafından seçilir. En uygun model için, performans kriterleri değerlerinin hedef değışken olarak alındığı, kısıt altında en uygun çözüm aranır.

$$EKK(\lambda(g)) = \sum_{i=1}^n (y_i - g(x_i))^2 + \int (g''(x))^2 dx \quad (4.10)$$

Denklem (4.10)'de görülen, düzleştirilmiş splayn regresyon modelinde, model $g(x)$ in ikinci türevi $\int (g''(x))^2 dx$ cezalandırma işlevi olarak kullanılır.

$$\hat{\mathbf{y}} = \mathbf{X}\beta \quad (4.11)$$

Burada genel regresyon denklemi ifadesi β parametre vektörünün en uygun değerleri için $(\mathbf{y} - \mathbf{X}\beta)^2$ en küçük yapacak uygun değerlerin bulunması gerekir. $\beta = [\beta_0, \beta_1, \beta_1 1, \dots, \beta_1 \kappa]'$ olur, $\beta_{1\kappa}$ ifadesinde κ indisi ile düğüm noktasını göstermektedir. Düğüm sayısının aşırı ya da az olması durumu, karmaşıklığın cezalandırıldığı, kısıt altında bir optimizasyon problemi olarak ele alınır. Kısıt altında optimizasyon probleminin çözümü için,

$$\max |\beta_{1\kappa}| < C \quad (4.12)$$

$$\sum |\beta_{1\kappa}| < C \quad (4.13)$$

$$\sum \beta_{1\kappa}^2 < C \quad (4.14)$$

koşullarından birinin sağlanması durumuna bakılmalıdır. Burada en hassas ve dikkat gerektiren nokta, bilenen bir kısıt değeri C seçimi olup, saçılım noktalarına düzgün bir eğri uydurulmasında belirleyicidir. Yukarıda verilen üç kısıt ifadesinden denklem (4.14) diğer iki kısıta göre hesaplanması görece daha kolaydır. Eğer düğüm sayısı κ 'ya göre $(\kappa + 2) \times (\kappa + 2)$ boyutunda bir \mathbf{D} matrisi tanımlanırsa,

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times \kappa} \\ \mathbf{0}_{\kappa \times 2} & \mathbf{I}_{\kappa \times \kappa} \end{bmatrix} \quad (4.15)$$

kısıt altında en uygun değer bulmak için aşağıdaki denklem Lagrange çarpanları yöntemiyle çözülebilir:

$$\min \|y - X\beta\|^2 \quad \text{kısıt} \quad \beta' D \beta \quad (4.16)$$

Bu amaç ve kısıt denklemine göre, Lagrange İşlevi matris notasyonu ile şöyle yazılabilir:

$$\|y - X\beta\|^2 + \lambda^2 \beta' D \beta \quad \lambda \geq 0 \quad (4.17)$$

Lagrange denklemi β için çözüldüğünde:

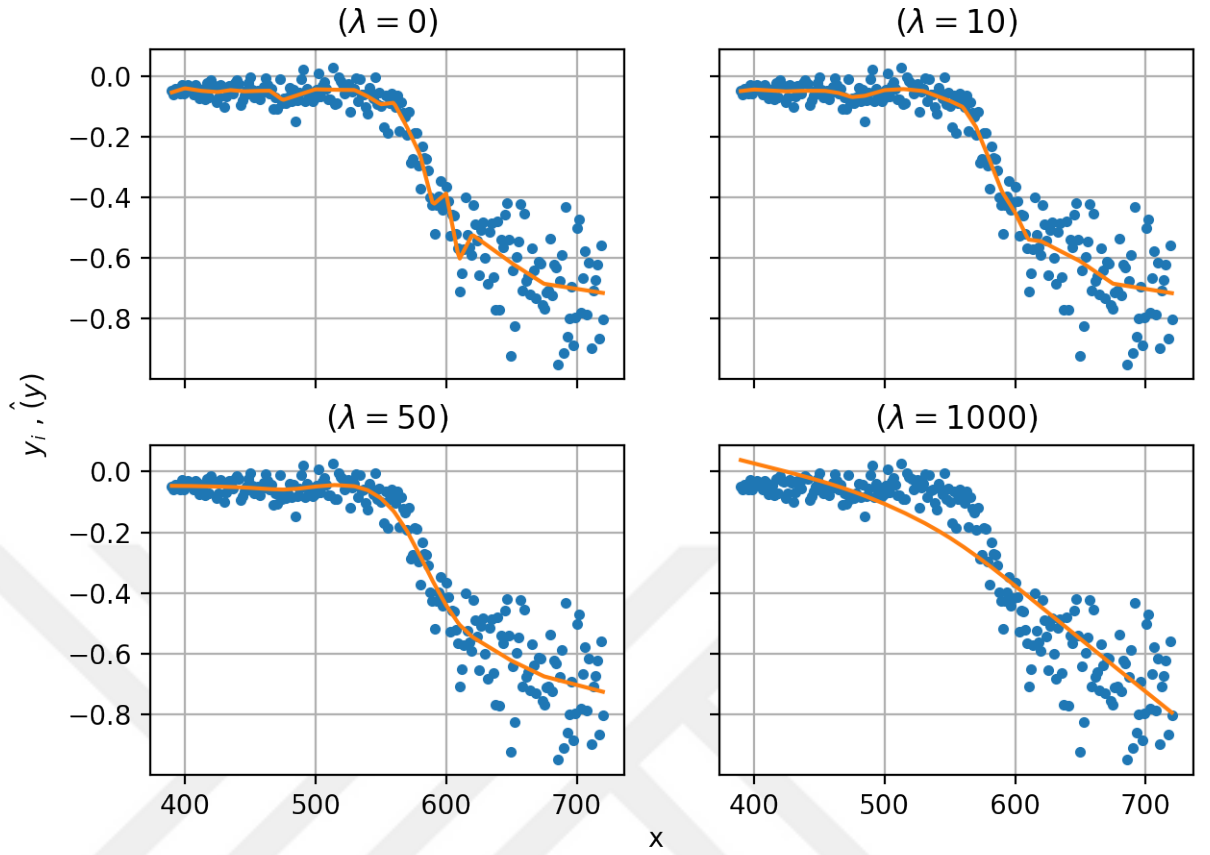
$$\beta_\lambda = (\mathbf{X}'\mathbf{X} + \lambda^2\mathbf{D})^{-1}\mathbf{X}'\mathbf{y} \quad (4.18)$$

Burada $\lambda^2\beta'D\beta$ ifadesi, uydurulan eğrinin aşırı dalgalanmasını cezalandırması nedeniyle, pürüzlülük cezası olarak adlandırılır. λ değeri aynı zamanda düzleştirme seviyesini belirlediğinden düzleştirme parametresi ya da bant genişliğini gösterir.

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda^2\mathbf{D})^{-1}\mathbf{X}'\mathbf{y} \quad (4.19)$$

Cezalandırılmış kestirim, seçilmiş düğüm noktalarını karşılayan splayn işlevinin katsayılarını sıfıra doğru baskılar. Ancak bazı düğüm noktaları sıfır olurken, bazılarının da daraltılmadan kalmasını sağlar. Denklem (4.19) aynı zamanda bir Ridge Regresyon türüdür (Ruppert vd., 2003a).

Düzleştirme parametresi λ regresyon modelinin diğer belirleyici bileşenidir. Aşağıda Şekil (4.4)'de, LIDAR veri setinin λ 'nın farklı değerleri için, 24 düğüm noktalı doğrusal splayn regresyon eğrisinin durumu incelenmektedir (Ruppert, Wand ve Carroll, 2003b).



Şekil 4.4: Düzleştirme parametresi λ 'nın değişiminin splayn regresyon modeline etkisi

Düzleştirme parametresinin $\lambda = 0$ olduğu durumda düğüm noktalarının pürüzlülük seviyesinin fazla olduğu, orta seviye değerler olan $\lambda = 10$ ve $\lambda = 50$ pek değişkenlik göstermediği, çok yüksek değerlerde $\lambda = 1000$ aşırı düzleştiğini gözlenebilir.

Cezalandırılmış splayn regresyon modelinin genel formunu elde etmek için aşağıdaki ifade yazılabilir:

$$\hat{\mathbf{y}} = \mathbf{S}_\lambda \mathbf{y} \quad (4.20)$$

$$\mathbf{S}_\lambda = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda^2 \mathbf{D})^{-1} \mathbf{X}' \quad (4.21)$$

Splayn Regresyon ile en küçük kareler temel özellikler açısından farklı olmakla birlikte, denklem (4.20), H şapka matris formuna benzer. Denklem (4.21) yer alan p 'inci derecen polinom değerini gösterir ve \mathbf{S}_λ ifadesi **düzleştirme matrisi** olarak adlandırılır.

4.2.2.1 En uygun bant genişliği λ değeri

Splayn regresyonda en uygun model seçimi yapmak için, üç temel belirleyici parametre vardır. Bunlar *düzgünleştirme parametresi*, *düzleştirme derecesi* ve veri üzerinden seçilecek olan *düğümünün noktalarıdır*. En uygun seçim, bu üç bileşen üzerinden aranır.

En uygun bant genişliği (λ) değerinin bulunması için, λ_j değerlerinden oluşan bir vektörün, her bir elemanı için ayrı bir performans değeri hesaplanır. Model seçiminde, Hata Kareler Toplamı (HKT), Çapraz Geçerlilik(CV) ve Genelleştirilmiş Çapraz Geçerlilik (GCV) yöntemleri kullanılabilir. Bu bilgi ölçütü değerlerini, minimum yapan λ değeri en uygun değerdir. Parametrik olmayan regresyon literatüründe, Mallows'un C_p ölçütü, Akaiki Bilgi Ölçütü (AIC) ve Düzeltilmiş Akaiki Bilgi Ölçütü (AICc) performans ölçütleri de kullanılmaktadır.

a) Hata Kareler Ortalaması:

Bir regresyon modeli için uyumun iyiliği, regresyon eğrisi ile saçılım grafiğindeki veri eşleşme noktaları arasındaki fark, Hata Kareler Toplamı olarak hesaplanır. Aşağıda yer alan denklem setinde, HKT'yi minimize edecek, optimal bant genişliği λ aranır.

$$HKT = \sum_{i=1}^n (y_i - \hat{y})^2 = \|y - \hat{y}\|^2 \quad (4.22)$$

$$HKT(\lambda) = \sum_{i=1}^n (y_i - \hat{f}(x_i, \lambda))^2 \quad (4.23)$$

Burada düzleştirme parametresi λ değişken olduğu için, $HKT(\lambda)$ tanımlanmıştır. Splayn regresyon modeli için HKT'nı matris notasyonu ile ifade edilirse,

$$HKT = y'(I - S)y \quad (4.24)$$

olur. Burada yer alan S matrisi splayn regresyon düzleştirme matrisi olarak adlandırılır.

b) Çapraz Geçerlilik

Düzleştirme parametresi λ 'nın farklı değerleri için, splayn modelin kestirimi işlevi $\hat{f}(x; \lambda)$ olur. Çapraz Geçerlilik(CV), HKT için yazılan denklem (4.22)'teki ifadenin, her bir gözlem satırı için ayrı $(x_i; y_i)$ hesaplama dışı bırakılarak, kalan veri setiyle kestirim yapılmasıdır. Kalan verilerle yapılan kestirim denklemine, çıkarılan satırların bağımsız değişkenleri girdi olarak verilir, elde edilen \hat{y}_i değeri hesaplanarak, y_i değeriyle farklı alınarak, bütün veri seti toplam değer alınarak Çapraz Geçerlilik(CV) değeri elde edilir.

Çapraz geçerlilik değerleri her bir λ değeri için ayrı ayrı hesaplanır, bu nedenle veri ve test edilecek λ vektörünün büyüklüğüne bağlı olarak ağır bir hesaplama yükü ortaya çıkar. Hesaplama yükü gözlem sayısı ve test edilecek λ sayısı kadar, n gözlem sayısı için, n^2 adet CV değeri hesaplamak gerekir.

$$CV(\lambda) = \sum_{i=1}^n (y_i - \hat{f}_{-i}(x_i, \lambda))^2 \quad (4.25)$$

Denklem (4.25)'de $\lambda \geq 0$ olmak üzere, $CV(\lambda)$ değerini minimum yapacak λ değeri için splayn işlevi aranır.

Çapraz Geçerlilik değeri CV farklı λ değerleri için minimum yapacak değer arayışının getirdiği hesaplama yükünü azaltarak, parametrik olmayan regresyon modelleri için geliştirilmiş hızlı hesaplama yapabilen algoritmalar geliştirilmiştir (Hutchinson ve de Hoog, 1985).

$$CV(\lambda) = \sum_{i=1}^n \left(\frac{y_i - \hat{f}(x_i; \lambda)}{1 - S_{\lambda, ii}} \right)^2 \quad (4.26)$$

$$= \sum_{i=1}^n \left(\frac{[(\mathbf{I} - \mathbf{S}_\lambda)\mathbf{y}]_i}{1 - S_{\lambda, ii}} \right)^2 \quad (4.27)$$

$$= \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - S_{\lambda, ii}} \right)^2 \quad (4.28)$$

Denklem (4.28) çapraz geçerlilik CV değerini yalnızca sıradan artıklar ve düzleştirme matrisi S' in köşegen elamanları $S_{\lambda, ii}$ ile hesaplanmasını mümkün kılar (Ruppert vd., 2003a, s. 116-117).

c) Genelleştirilmiş Çapraz Geçerlilik

Genelleştirilmiş Çapraz Geçerlilik (GCV) cebirsel ifadesinde, yukarıdaki denklem (4.28)'deki S_λ düzleştirici matrisinin köşegen değerleri yerine, matrisinin izinin ortalaması $\frac{1}{n}tr(S_\lambda)$ kullanılmıştır.

$$GCV(\lambda) = \sum_{i=1}^n \left(\frac{[(\mathbf{I} - \mathbf{S}_\lambda)\mathbf{y}]_i}{1 - n^{-1}S_{\lambda, ii}} \right)^2 \quad (4.29)$$

$$= \sum_{i=1}^n \frac{AKT_\lambda}{[1 - n^{-1}S_{\lambda, ii}]^2} \quad (4.30)$$

Burada genelleştirilmiş çapraz geçerlilik(GVC), HKT'nin $tr(S_\lambda)$ ortalama değeriyle ilişkilendirilmiş ifadesini göstermektedir.

4.2.2.2 Düğüm Noktası Sayısı

Düğüm sayısını belirlemek için genel olarak iki algoritma kullanılmaktadır.

a) Miyop Algoritma:

Miyop Algoritma, düğüm sayısı κ için deneme değerlerinden oluşan bir vektörün sıralı sayısal değerleri üzerinden test yapılır. Düğüm sayısı x_i tekrarlı olmayan değerleri n_{tekil} iken, $n_{tekil-p-1}$ den küçük olması dışında bir kısıtlama yoktur.

Algoritma işleyişi için, belirli düğüm sayısı aralıkları belirlenir. Ardından her bir sınıf aralığına cezalandırılmış splayn regresyon uygulanır. Örneğin düğüm sayısı 5 ve 10 arasında ise, her durum için $GCV(\lambda)$ değerini minimize eden λ değeri hesaplanır. Eğer düğüm sayısı $\kappa = 10$ 'daki GCV değeri,

$\kappa = 5$ 'teki düğüm sayısından büyükse, düğüm sayısı κ 'nın yükseltilebilir. κ düğüm sayısı 5 ile 10 arasında, her durumda GCV minimum yapacak değer aranır. Aynı biçimde $\kappa = 10$ ve $\kappa = 20$ için de arama yapılır, yine minimum GCV aranır. Bu şekilde κ vektörü içindeki değerler sınırlanır. Algoritma hangi κ değerinde duracağını önceden öngöremediği için "Miyop Algoritma" olarak adlandırılmıştır (Ruppert vd., 2003a, s. 127). Miyop algoritmanın avantajı görece daha az hesaplama yoluyla sonuca ulaşmasıdır.

b) Tam Arama Algoritması:

Tam arama algoritmasında, GCV'yi minimize eden tüm λ , bütün mümkün dizide yer alan κ değerleri incelenir. Uygun GCV'yi minimum yapan düğüm sayısı κ değeri bulunur. Bütün mümkün durum kümesi hesaplamaya dahil edildiğinden, ağır bir işlem yükü ortaya çıkar.

4.3 Düzleştirilmiş MIDAS regresyon modeli

Parametrik olmayan MIDAS regresyon modeli, Breitung ve Roling tarafından 2015 yılında incelenmiştir. Breitung ve Roling tarafından, gecikme dağılımını düzleştirilmiş hale getirmek için, parametrik olmayan kestirim yöntemi olarak, cezalandırılmış EKK yönteminin uyarlaması olan, kübik spline regresyon kullanılmıştır. Düzleştirilmiş En Küçük Kareler (Smoothed Least Squares - SLS), Monte Carlo benzetim denemeleriyle ve ampirik verilerle, bazı durumlarda parametrik olmayan yaklaşımın daha güvenilir sonuçlar ürettiği ve geleneksel yöntemlere göre esnek bir yaklaşım ortaya koyduğu iddia edilmiştir.

Splayn regresyon yöntemlerinde, parametre kestirimlerini elde etmek ve en uygun bant genişliği λ değerini bulmak, böylece aşırı belirlenmiş (overfitting) ya da yetersiz model (underfitting) oluşumunu meydan vermemek için, sayısal analize dayalı ağır hesaplama yükü oluşturan yöntemler kullanılmaktadır. Bu hesaplama yükünün azaltılması, özellikle büyük veri analizlerinde daha önemli hale gelmektedir. Düzleştirilmiş EKK kareler yönteminin, bant genişliği parametresi λ 'nın özel değerlerinden yararlanmak suretiyle, hesaplama yükününün azaltılmasına yönelik öneriler içeren ve modelleri karşılaştıran çalışma Toker, Özbay ve Månsson tarafından 2022 yayınlanmıştır. Düzeltilmiş Akaike Bilgi Ölçütü (AICc) değerini en küçük yapan bant genişliği (λ) değeri ve düğüm sayısı (κ) sayısal analiz yöntemleriyle en uygun değer kümesine ulaşılmaya çalışılmıştır.

Parametrik olmayan MIDAS regresyon modelinin, düzleştirilmiş EKK olarak adlandırılan Breitung ve Roling (2015) makalelerinde, MIDAS regresyon modeli basit EKK regresyonun kısıtlanmış bir formu olduğunu belirtmişlerdir. Devamında, parametrik olmayan MIDAS regresyon modelinin önceden şekli bilinmeyen regresyon işlevinin, kısıtlanmış cezalandırılmış EKK hedef işlevinin, minimum hata kareler toplamı (HKT) olarak göstermişlerdir. Bu tanımlamaya ait hedef işlevin matematiksel formu şöyledir:

$$S(\alpha, \theta) = \sum_{t=1}^T (y_{t+h} - \alpha_0 - \beta^T x_{t,\bullet})^2 + \lambda \sum_{j=2}^p (\nabla^2 \beta_j) \quad (4.31)$$

Bu denklemde eşitliğin sol tarafı hata kareler toplamını, α ve θ parametresiyle belirlendiğini ifade eder, hedef işlevi EKK uygun olarak minimize edilmelidir. Eşitliği sol tarafındaki ilk terim, düşük frekanslı bağımlı değişken serisinin değerleri y_{t+h} ile formu bilinmeyen regresyon modeli $\alpha_0 + \beta^T x_{t,\bullet}$ temsil eder. Eşitliğin solundaki ikinci terim $\lambda \sum_{j=2}^p (\nabla^2 \beta_j)$ cezalandırma ve kısıt terimidir. Denklemdeki λ aynı zamanda parametrik olmayan regresyon modelinin bant genişliği parametresidir.

Burada yer alan cezalandırma ifadesi $\lambda \sum_{j=2}^p (\nabla^2 \beta_j)$, şekli belirsiz düzleştirme işlevinin parametre katsayıları olan B_j , bu denklemde görüldüğü üzere ikinci farklarının mutlak değerlerini ifade eder. Bu yukarıda splayn regresyon modelinde, doğal splayn regresyon modelinin formuyla uyumludur.

$$\nabla^2 \beta_j = \beta_j - 2\beta_{j-1} + \beta_{j-2} \quad j = 2, \dots, p \quad (4.32)$$

Denklem (4.32)'de j indis değerleri aşağıda gösterildiği gibi yerine koyarak hesaplandığında:

$$\begin{aligned} j=2 & \Rightarrow = \beta_2 - 2\beta_{2-1} + \beta_{2-2} = \beta_2 - 2\beta_1 + \beta_0 \\ j=3 & \Rightarrow = \beta_3 - 2\beta_{3-1} + \beta_{3-2} = \beta_3 - 2\beta_2 + \beta_1 \\ & \vdots \\ j=p & \Rightarrow = \beta_p - 2\beta_{p-1} + \beta_{p-2} = \beta_p - 2\beta_{p-1} + \beta_{p-2} \end{aligned} \quad (4.33)$$

Eğer elde edilen doğrusal denklem sisteminde yer alan katsayıları ve çarpanları düzenlersek ikinci farklar $\nabla^2 \beta_j$ için matris formu elde edilir. Bu düzleştirme işlevinde kullanılacak katsayılar matrisidir.

$$\nabla^2 \beta_j = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & . & . & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad (4.34)$$

Yukarıdaki yer alan fark denkleminin polinom katsayıları \mathbf{D} matrisi olarak adlandırılır (denklem (4.35)), $(p-1)$ satır sayısına ve $(p+1)$ sütün sayısına sahiptir. D üç köşegenli (tridiagonal matrix) bir matris formundadır.

$$\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & . & 0 \\ 0 & 1 & -2 & 1 & . & 0 \\ . & . & . & . & . & . \\ 0 & . & . & 1 & -2 & 1 \end{bmatrix} (p-1) \times (p+1) \quad (4.35)$$

Hata kareler toplamını minimum yapan işlevinin, bant genişliği parametresi λ için çözümlerse, $\tilde{\beta}$ katsayıları vektörünün kestirimi şöyle ifade edilebilir:

$$\tilde{\beta}_\lambda = \arg \min \{ (y - X\beta)^T (y - X\beta) + \lambda \beta^T D^T D \beta \} \quad (4.36)$$

$$\tilde{\beta}_\lambda = (X^T X + \lambda D^T D)^{-1} X^T y \quad (4.37)$$

MIDAS regresyon modelinin temelini oluşturan, yüksek frekans bağımsız değişkenin gecikmeli değerlerinin, düşük frekanslı bağımlı değişken ile eşleştirilmesi süreci parametrik olmayan modelde de işletilir. Model seçimi için, Düzeltilmiş Akaiki Bilgi Ölçütü (AICc)'nün, λ parametresiyle belirlenen, düzeltilmiş versiyonu kullanılır.

$$\min AICc(\lambda) = \log([y - \hat{y}]^T [y - \hat{y}]) + \frac{2(\kappa_\lambda + 1)}{T - \kappa_\lambda - 2} \quad (4.38)$$

Düzleştirilmiş en küçük kareler yoluyla splayn regresyon modeli, en düşük düzeltilmiş AICc değerini sağlayan bant genişliği parametresi λ ve gecikme uzunluğu düğüm sayısı κ_λ tarafından temsil edilir, bu iki parametre yardımıyla model seçimi yapılır, ardından model gecikme katsayıları $\tilde{\beta}_\lambda$ vektörü oluşturulur.

4.4 Çekirdek regresyon yöntemleri

Parametrik olmayan tek değişkenli çekirdek kestirim modelinin bilinen en eski ve sık kullanılan modeli histogramlardır. Histogramları oluşturan çubuklar, bir başlangıç noktasından, belirli bir bant genişliğinde verilerin dağılım sıklığını gösterir. Veri kümesi üzerinde x_0 başlangıç noktası ve h aralık genişliği verildiğinde, m pozitif ve negatif bir tamsayı olmak üzere $[x_0 + mh, x_0 + (m + 1)h]$ aralıklarına karşılık gelen değerlerin sıklığı, histogram işlevini tanımlar.

$$\hat{f}(x) = \frac{1}{nh} B \quad (4.39)$$

Bu denklemde B ifadesi bir histogram çubuğundaki X_i sayısını gösterir. Histogram kutucuğunun genişliği (bin width) ya da *bant genişliği* olasılık dağılım işlevi $\hat{f}(x)$ şeklini belirler. Bant genişliği h arttıkça düzeltirici etki artar, tersine azaldığında olasılık değeri $\hat{f}(x)$ 'da değişkenlik artar (Silverman, 1986, chap. 2).

Çekirdek Yoğunluk Kestiricisi, kuyruklarda sıçramalar yapmayan, daha düzeltilmiş bir kestirim için çekirdek işlevi olarak adlandırılan kesiksiz ağırlıklandırma işlevleri kullanılır. Çekirdek Yoğunluk

Kestiricileri, yukarıda histogramlar için gösterilen kesikli olasılık dağılımından farklı olarak, düzleştirilmesi ve sürekli bir olasılık yoğun işlevi haline getirilmesidir. Eğer bant genişliği h çok küçük bir sayı ise ve f işlevi x için sürekli ise:

$$f(x) \approx \frac{1}{2h} P\{x-h < X < x+h\} \quad (4.40)$$

$$P\{x-h < X < x+h\} \approx \frac{1}{n} \sum_{i=1}^n I\{x-h < X_i < x+h\} \quad (4.41)$$

$$\hat{f}(x) = \frac{1}{2nh} \sum_{j=1}^n I\{x-h < X_j < x+h\} \quad (4.42)$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{I\{|X_i - x| \leq h\}}{2h} \quad (4.43)$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \mathbf{W}\left(\frac{x - X_i}{h}\right) \quad (4.44)$$

Gösterge İşlev \mathbf{I} 'nin tanım aralığı şöyledir:

$$w(x) \begin{cases} \frac{1}{2} & \text{için } |x| < 1 \\ 0 & \text{ö.d.i} \end{cases}$$

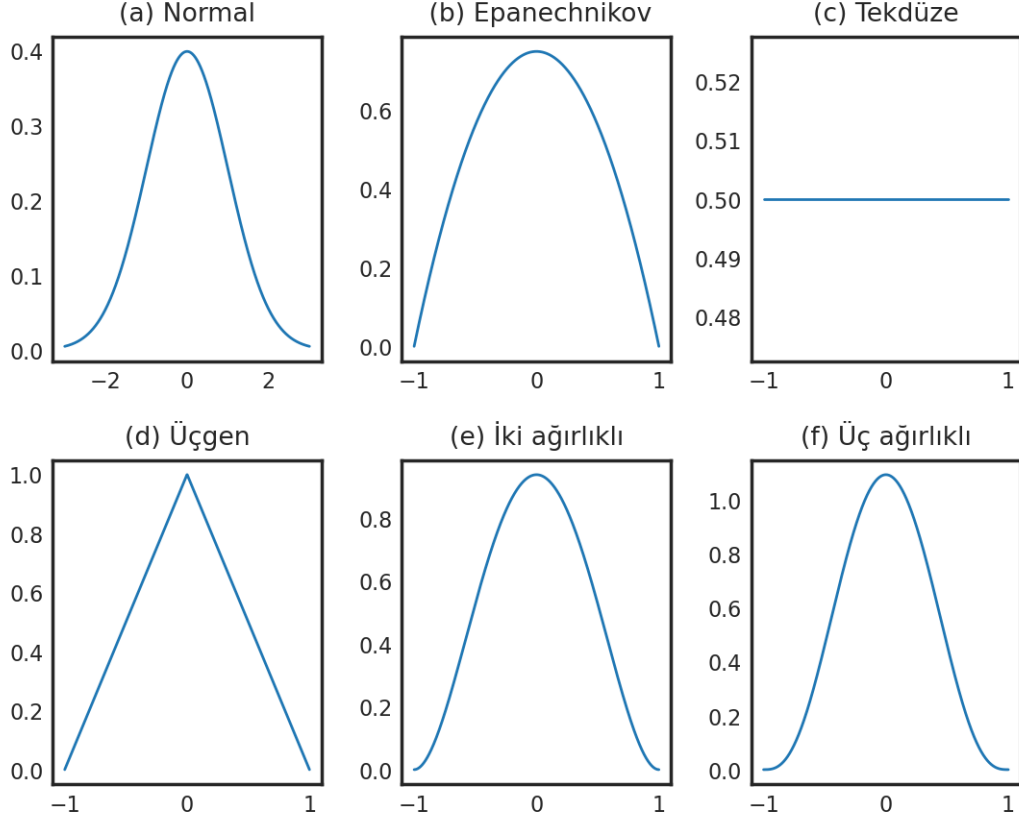
Denklem (4.44)'deki \mathbf{W} ifadesi çekirdek olarak adlandırılır ve bu denkliği sağlayan ifade şudur:

$$\mathbf{W}(x) = \frac{1}{2} I\{|x| \leq h\} \quad (4.45)$$

$$W_h(x) = \frac{1}{h} \mathbf{W}\left(\frac{x - X_i}{h}\right) \quad (4.46)$$

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (4.47)$$

Buarada h bant genişliği aynı zamanda düzleştirme parametresi olarak adlandırılmaktadır.



Şekil 4.5: Çekirdek işlevlerinin karşılaştırmalı görünümü

Şekil 4.5'de çekirdek işlevleri, Normal, Epanechnikov, Tekdüze (Uniform) ve Üçgen (Triangular), İki ağırlıklı (Biweight) ve Üç ağırlıklı (Triweight) formların karşılaştırmalı görselleri sunulmuştur (Seabold ve Perktold, 2010). Çekirdek işlevi özellikleri şöyledir :

$$K(u) \geq 0 \quad (4.48)$$

$$\int_{-\infty}^{\infty} K(u) du = 1 \quad (4.49)$$

$$\int_{-\infty}^{\infty} K^2(u) du < \infty \quad (4.50)$$

$$K(u) = K(-u) \quad (4.51)$$

Denklem (4.48) ve (4.49) olasılık yoğun işlevinin genel özelliklerini göstermekte, denklem (4.50) işlevin ikinci momentini ve (4.51) simetrik olma özelliklerini tanımlamaktadır.

Yukarıda tablo 4.1 çeşitli çekirdek yoğunluk işlevleri ve bu işlevlere ilişkin beklenen değer ve varyans değerlerine yer vermektedir. Uygulamada görece etkinliğinin yüksekliği nedeniyle, Epanechnikov çekirdek işlevi yaygın olarak kullanılmaktadır.

Tablo 4.1: Çekirdek İşlevleri ve Özellikleri

Çekirdek	Denklem	R(k)	σ_k^2
Normal	$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})$	$1/(2\sqrt{\pi})$	1
Epanechnikov	$K(u) = \frac{3}{4}(1-u^2)1(x \leq 1)$	3/5	1/5
Uniform	$K(u) = \frac{1}{2}1(e \leq 1)$	1/2	1/3
Üçgen	$K(u) = (1- x)1(x \leq 1)$	2/3	1/6
İki ağırlıklı	$K(u) = \frac{15}{16}(1-u^2)^2 1(x \leq 1)$	5/7	1/7
Üç ağırlıklı	$K(u) = \frac{35}{32}(1-u^2)^3 1(x \leq 1)$	350/429	1/9

4.4.1 Nadaraya-Watson Çekirdek kestiricisi

Nadaraya-Watson çekirdek kestiricisi, bağımlı değişkenin ağırlıklı ortalaması olarak tanımlanabilir. Regresyonun en genel ifadesi, bağımlı ve bağımsız arasında koşullu olasılık dağılımı olarak tanımlanabilir. Buna göre regresyonun genel formu aşağıda gibi gösterilebilir.

$$m(\cdot) = E(Y|X = x) = \int y f(y|x) dy = \int y \frac{f(x,y)}{f_x(x)} dy \quad (4.52)$$

Burada koşullu olasılık dağılımı, paydaki bileşik olasılık yoğunluk işlevi ve paydadaki bağımsız değişkenin marjinal olasılık işlevlerinin oranını ifade eder. Bu ifadeye, bileşik olasılık dağılımın yerine, çekirdek işlevinin kullanılması durumunda, regresyon için çekirdek kestiricisi elde edilmiş olur.

$$\hat{f}(x,y) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K\left(\frac{x-X_i}{h_1}\right) \frac{1}{h_2} K\left(\frac{y-Y_i}{h_2}\right) \quad (4.53)$$

Paydaki bileşik olasılık yoğunluk işlevi yazıldığında, koşullu olasılık ifadesine uygun olarak yazıldığında:

$$\int y \hat{f}(x,y) dy = \int y \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K\left(\frac{x-X_i}{h_1}\right) \frac{1}{h_2} K\left(\frac{y-Y_i}{h_2}\right) dy \quad (4.54)$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K\left(\frac{x-X_i}{h_1}\right) \int \frac{y}{h_2} K\left(\frac{y-Y_i}{h_2}\right) dy \quad (4.55)$$

İntegral ifadesinde $(\frac{y-Y_i}{h_2}) = t$ dönüşümü yapılırsa:

$$\int y \hat{f}(x,y) dy = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{h_1} K\left(\frac{x-X_i}{h_1}\right) \quad (4.56)$$

ifadesi elde edilir. Koşullu olasılık yoğunluk işlevi ifadesinin paydasını oluşturan, marjinal olasılık yoğunluk işlevi de denkleme eklendiğinde:

$$\hat{f}_x = \frac{1}{nh_1} \sum_{i=1}^n K\left(\frac{x-X_i}{h_1}\right) \quad (4.57)$$

Gerekli sadeleştirmeler yapıldıktan sonra Nadaraya-Watson kestiricisi,

$$\hat{m}(x) = \frac{\sum Y_i K\left(\frac{x-X_i}{h}\right)}{\sum K\left(\frac{x-X_i}{h}\right)} \quad (4.58)$$

denklemini elde edilir.

Bu ifade bağımlı değişkenin ağırlıkları şeklinde yazılırsa:

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^n W_i Y_i \quad (4.59)$$

$$W_i(x) = \frac{h^{-1} K\left(\frac{x-X_i}{h}\right)}{\frac{1}{nh} \sum K\left(\frac{x-X_i}{h}\right)} \quad (4.60)$$

$W_i(x)$ ifadesi X_i değerlerine bağlı olan ağırlıkları gösterir.

4.4.2 Priestley-Chao Çekirdek kestiricisi

Priestley ve Chao, faz spektrumu $\phi(w)$ kestiriminde, n sayıdaki gözlemin türevi $\phi'(w)$ ve kesikli ve eşit aralıklı frekans kümesi w_1, \dots, w_n ve frekans bant genişliğinin sonlu ve ϕ işlevinin $(0,1)$ aralığında uygun değer aralığı için yöntem önermişlerdir (Priestley ve Chao, 1972).

$$\hat{f}_{h,n}(x) = \sum_{i=1}^n y_i (x_i - x_{i-1}) k^{-1} W_0\left(\frac{x-X_i}{h}\right) \quad (4.61)$$

Priestley ve Chao tarafından verilen çekirdek kestiricisi ise sabit tasarım için önerilmiştir. Birçok çalışmada bağımsız değişken olan X , $[a,b]$ aralığında eşit aralıklı alınmaktadır. $[a,b]=[0,1]$ ve $X_i=i/n$ biçiminde alınması, genelliği bozmadan kolaylık sağlamaktadır. Bu tür veriler genellikle zamana bağlı verilerdir.

4.4.3 Gasser-Müller Çekirdek kestiricisi

Gasser-Müller çekirdek kestiricisi, periyodik modellere yönelik olarak, Priestly-Chao yaklaşımını geliştirmiştir. Eşit olmayan zaman aralıklarının dahil edilmesi ve yüksek dereceli çekirdek işlevlerinin,

küçük dereceden simetrik ve pozitif çekirdek işlevleri tarafından tanımlanmasına ilişkin sorunlara kuramsal çözümler önermişlerdir (Gasser ve Müller, 1979).

$$\hat{m} = \frac{1}{h} \sum_{i=1}^n \int_{s_{i-1}}^{s_i} Y_i K\left(\frac{x-u}{h}\right) du \quad (4.62)$$

Denklem (4.62)'de ağırlıklar s_0, s_1, \dots, s_n ve $s_0 < x_1 < s_1 < x_2 < \dots < x_n < s_n$ şeklindeki diziye göre oluşur.

Çekirdek fonksiyonu K 'nın v defa farkının alınabilir olması koşulu sağlanmalıdır (Gasser ve Müller, 1984):

$$K \subset [-\tau, +\tau] \quad (4.63)$$

$$\int_{-\tau}^{+\tau} K(x) dx = 1 \quad (4.64)$$

$$K^j(\tau) = K^j(-\tau) = 0 \quad (j = 0, \dots, v-1) \quad (4.65)$$

$$(4.66)$$

4.5 MIDAS çekirdek regresyon modeli üzerine tartışma

MIDAS regresyon modellerinde, yüksek frekanslı bağımsız değişken serisinin, düşük frekanslı bağımlı değişken serinin gözlem sayısına denkleştirilmesi, ağırlıklandırma işleviyle yapılır. Ağırlıklandırma işlevi hem parametrik, hem de parametrik olmayan yöntemlerde anahtar niteliğindedir.

Çekirdek regresyon yöntemleri, bant genişliği ve çekirdek düzleştirici parametrelerini baz alarak kestirim yapar. Literatürde, ağırlıkla Nadaraya-Watson yöntemi yaygın olarak kullanılır. Geliştirilmiş farklı versiyon çekirdek regresyon yöntemleri Priestley-Choa ve Gasser-Müller çekirdek regresyon kestirimlerine ilişkin bilgiler önceki başlıkta ele alınmıştır. MIDAS regresyon modelinde çekirdek kestiricilerden nasıl yararlanılabileceği sorusu ilgi çekicidir.

Yüksek frekanslı bağımsız değişkenin çekirdek yoğunluk işlevleri ve uygun bant genişliğinde düşük frekanslı bağımlı değişkenin gözlem sayısı ile eşleştirilmesi, iki aşamalı olarak yapılabilir. Birinci aşamada, $X_{t,p}$ boyutunda bir bağımsız değişken setinin, satır boyunca çekirdek $K(X_i - x)/b$ yardımıyla dönüştürülmesiyle her X_t satırı için bir değer oluşturulmaktadır. İkinci aşamada, elde edilen y_t ve x_t değişken seti üzerinden çekirdek regresyon yöntemi uygulanmaktadır. Ancak, gecikme değerlerindeki örüntünün ortaya çıkarılması için, her iki boyutun birlikte çözümüne ihtiyaç vardır.

Çekirdek regresyon yöntemlerinde model seçimi için, bilgi ölçütlerinin yeniden düzenlenmesine ihtiyaç vardır (Zhang, 2007). Ayrıca, tek değişkenli çekirdek regresyonda bile önemli bir hesaplama yükü vardır. Çekirdek regresyonun zaman serilerinde uygulanması da önemli zorluklar içermektedir. Modele gecikmeli değerlerin dahil edilmesi, hesaplama yükünü devasa boyuta taşımaktadır. Bu

nedenle ekirdek MIDAS regresyon modeline ynelik denemelerde, teorik tutarlılık sorunları yanında, hesaplama ykleri de nemli bir sorun oluřturmaktadır.



Bölüm 5

Benzetim Deneyi

MIDAS Regresyon modellerinin kuramsal işleyişi, parametrik ve parametrik olmayan yöntemler olarak iki ana başlık altında üçüncü ve dördüncü bölümlerde incelenmişti. Bu bölümde parametrik ve parametrik olmayan yöntemler üzerine Monte Carlo benzetim deneyi tasarlanarak, yöntemler farklı yapılarıdaki rassal veri setleriyle teste tabii tutulmuştur. Yöntemlerin, veri setleri içine gömülmüş doğru model seçim performansı hesaplanmış ve göreceli başarı oranları karşılaştırılmıştır. Farklı örneklem hacmi, gecikme uzunlukları, değişkenlik ölçek faktörü gibi değer aralıklarında yöntemlerin performansları ölçülmüş, üstünlükleri ve zayıflıkları mukayese edilmiştir.

5.1 Deney tasarımı

Monte Karlo deneyi, MIDAS regresyon modellerinin karşılaştırmasına yönelik olarak tasarlanmıştır. Rassal veri üretimi için dört farklı ağırlıklandırma işlevi yüksek frekanslı serinin dönüştürülmesinde kullanılmış, değerlere hata terimi de eklenerek gürültü faktörü yaratılmıştır. Bağımsız değişken için oluşturulan rassal veri setlerinin, ardından tanımlanmış model parametreleri ile bağımsız değişken vektörü oluşturulmuştur. Benzetim modeli için, 500 adet rassal veri üretilmiştir. Yüksek frekanslı bağımsız değişken için, veri üretimde otoregresif AR(1) modeli baz alınarak veri üretimi yapılmıştır. Rassal veri seti için, (i) Üssel azalan (ii) Tümsük azalan (iii) Doğrusal azalan ve (iv) Periyodik artan-azalan olmak dört farklı ağırlıklandırma işlevi kullanılmıştır. Azalan ağırlıklandırma işlevleri kullanılmasının nedeni, yakın tarihli gecikme değerinin etkisinin daha yüksek, geriye doğru gidildikçe ise azalacağı varsayımına dayandırılmıştır. Ağırlıklandırma işlevlerinin matematiksel formu ve işleyişleri ki alt başlıklarda detaylı incelenmiştir. İkinci aşamada, Parametrik MIDAS Üssel Almon Modeli bazlı doğrusal olmayan regresyon (MIDAS-NLS) ile rassal veriler üzerinde çalıştırılmıştır. Üçüncü aşamada, parametrik olmayan MIDAS regresyon modellerinden Düzleştirilmiş MIDAS Regresyon Modeli (MIDAS-SLS) ile hesaplama yapılmıştır. Dördüncü aşamada, kapsama alınan tüm modellerin, en uygun Hata Kareler Ortalaması (HKO) ve Düzeltilmiş Akaike Bilgi Ölçütüne (AICc) göre model

seçimi yaptırılmıştır. Beşinci aşamada, doğru model seçimi için performans karşılaştırma tablosu oluşturulmuştur. Sonuç tablosu üzerinden üstün ve zayıf yönler, yöntemlerin sınırlılıkları tartışılmıştır.

Deney tasarımı için veri üretim prosedürü, yapay veri setinin şeklini belirleyen ağırlık işlev türleri ve parametreleri için [Andreou vd. \(2010\)](#), [Roling \(2014\)](#), [Breitung ve Roling \(2015\)](#) ve [Toker vd. \(2022\)](#) çalışmalarına paralel değerler kullanılmıştır. Böylece literatürde yer alan benzetim modellerindeki yöntemler arasında karşılaştırılabilirlik sağlanmaya çalışılmıştır.

5.2 Rassal Veri Üretim Süreci

Veri üretimi sürecinde üç farklı örneklem genişliği $T = (100, 200, 400)$ olmak üzere, standart normal dağılım bazlı rassal veri üreticisi kullanılmıştır. İlk olarak, yüksek frekanslı bağımsız değişken x_t için, otoregresif seri AR(1) formunda rassal veri üretilmiştir.

$$x_{t,p} = c + \phi x_{t-1} + \varepsilon_{p,t} \quad (5.1a)$$

$$\varepsilon_{p,t} \stackrel{idd}{\sim} \mathcal{N}(0, 1) \quad (5.1b)$$

Rassal veri üretim sürecinde yukarıdaki denklem (5.1)'de otoregresyon katsayısı $\phi = 0.9$ ve $c = 0.5$ olarak alınmıştır. AR(1) rassal veri üretimi için Python yazılım dilinde "statsmodels" ([Seabold ve Perktold, 2010](#)) paketinin zaman serileri alt modülü "tsa" kullanılmıştır.

Yukarıdaki denklemde (5.1b)'de hata terimi $\varepsilon_{p,t}$ standart normal $N(0, 1)$ dağılıma sahiptir. Aynı denklemde yer alan $x_{t,p}$ ifadesinde t indisi düşük frekanslı serinin gözlem satır sayısını ve p indisi yüksek frekanslı serinin gecikme değerlerinin indisini gösterir. Yüksek frekanslı bağımsız değişken $x_{t,p}$ 'in, gecikmeli değeri x_{t-1} ile ilişkisi, AR(1) modelinin işleyişi gereği, ϕ katsayısı üzerinden kurulur.

Düşük frekanslı bağımlı değişken y 'nin gözlem sayısı T , yüksek frekanslı bağımsız değişken x_t yinelenme sıklığı m ve yüksek frekanslı bağımsız değişken serisindeki eleman sayısı, $T * m$ değerine eşittir. Rassal veri üretim sürecinin ardından, MIDAS regresyon için veri matrislerinin şekli ve x_t 'nin gecikmeli değerlerinin konumlandırılması üçüncü bölümde (3.2.2) ele alınmıştır.

Düşük frekanslı bağımlı değişken y ile yüksek frekanslı ve otoregresif gecikmesi dağıtılmış (ADL) bağımsız değişken matrisi X arasında ilişkinin şekli aşağıda denklem (5.2a)'de görülmektedir.

$$y_{t+h} = \sum_{p=0}^{P-1} \beta_p X_{p,t} + u_{t+h} \quad (5.2a)$$

$$\beta_p = \alpha \omega_p(\theta) \quad (5.2b)$$

$$u_{u+h} \stackrel{idd}{\sim} N(0, 1) \quad (5.2c)$$

Burada denklem (5.2a)'de, y_{t+h} ifadesi düşük frekanslı bağımlı değişkenin, $t = 1, \dots, T$ zaman indisi değerleri gösterirken, h değeri örneklem dışı tahmin periyodu değerini ifade eden bir tam sayıdır. Bağımsız değişken matrisi $X_{p,t}$ ile ifade edilir iken, p indisi $p = 0, \dots, P - 1$ kadar gecikme uzunluğuna ifade eder. Bağımsız değişken matrisinin önünde yer alan β_p vektörü, ağırlıklandırma işlevi tarafından kestirimi yapılan gecikme katsayıları vektörüdür. Bağımsız değişken X matrisi $(t * p)$ boyutunda iken, gecikme katsayıları vektörü β_p ($p * 1$) boyutundadır.

Gecikme katsayıları β_p vektörü, ω_p ağırlıklandırma işlevi tarafından kestirimi yapılmaktadır (5.2b). Ağırlıklandırma işlevinin, yalınlaştırma parametreleri θ katsayıları vektörü tarafından belirlenmektedir. Yalınlaştırma katsayıları θ genelde iki ya da üç elamandan oluşan bir küme olarak ele alınmaktadır. Denklem (5.2b)'de yer alan α ölçeklendirme parametresidir, $\alpha \in (0.2, 0.3)$ olarak kabul edilen ve değeri azaldıkça ağırlıklandırma işlevinin yalınlaştırıcı etkisi azaltan bir etkiye sahiptir. Rassal sayı üretim sürecinde, α değeri azaldıkça, yüksek frekanslı serinin değişkenliği ve yayılımı artış göstermektedir. Böylece model parametrelerinin kestirimi yapay olarak daha da zorlaştırılarak, model seçimi başarımı zorlu bir teste tabi tutulmaktadır.

Benzetim deneyi üssel azalan, tümsek, doğrusal azalan ve çevrimsel ağırlıklandırma işlevleri her biri için 500 adet veri seti, ayrı ayrı olmak üzere 100, 200 ve 400 gözlem sayısında olmak üzere üretilmiştir. Üretilen bu veriler, virgül ile ayrılmış (.csv) dosya olarak kaydedilmiştir.

5.3 Ağırlıklandırma İşlevleri

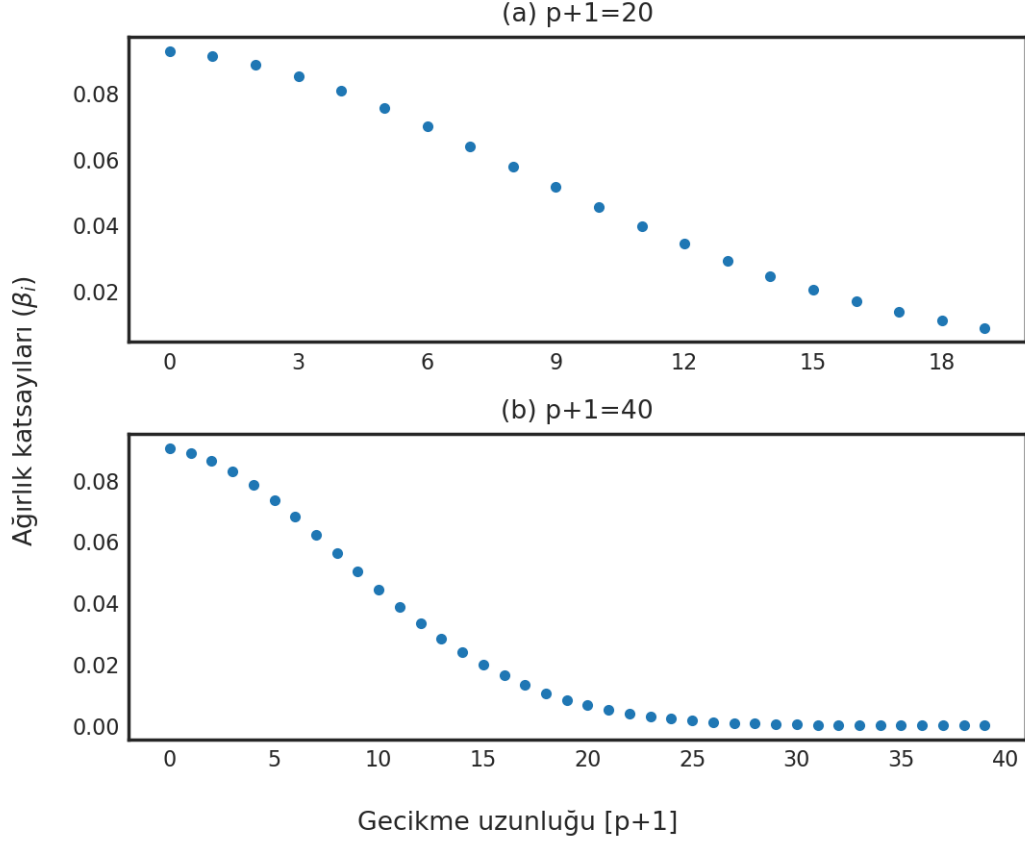
Ağırlıklandırma işlevleri yüksek frekanslı bağımsız değişkenin gecikmeli değerlerini, düşük frekanslı bağımlı değişkenin frekansına dönüştürürler. Yukarıda denklem seti (5.2)'de yer alan modelde, ağırlıklandırma işlevi $\omega_p(\theta)$ ile gösterilmiştir. Ağırlıklandırma işlevi ω_p , yalınlaştırma parametresi θ tarafından şekillendirilmektedir. Benzetim deneyinde, rassal veri seti üretimi için dört farklı ağırlıklandırma işlevi ω_p kullanılmıştır. Aşağıdaki alt başlıklarda ağırlıklandırma işlevlerinin özellikleri ele alınmaktadır.

5.3.1 Üssel azalan ağırlıklandırma işlevi

Üssel ağırlıklandırma işlevi, Üssel Almon ağırlıklandırma polinomunun parametre değerlerinin sınırlandırıldığı özel bir şeklini ifade eder. İşlevin matematiksel formu denklem (5.3)'de gösterilmiştir. Polinomun derecesi karesel değer ile sınırlandırılmıştır.

$$\omega(\theta) = \frac{\exp(\theta_1 j + \theta_2 j^2)}{\sum_{i=0}^p \exp(\theta_1 i + \theta_2 i^2)} \quad (5.3)$$

Yukarıda θ_1 ve θ_2 değerleri, ağırlıklandırma işlevi katsayılarını, j indisi ise $j = (0, \dots, P)$ kadar gecikme uzunluğu değerlerini ifade eder. Bu ifade sonucunda θ parametre değerleri tarafından sınırlandırılan, gecikme sayısı uzunluğu kadar ağırlıklandırma katsayısı elde edilir.



Şekil 5.1: Üssel Azalan Ağırlıklandırma işlevi

Yukarıda şekil 5.1'de gecikme uzunluğu p ile ağırlık katsayıları β_i arasındaki ilişki görülmektedir. Gecikme uzunluğu p ile işlev çıktıları olan gecikme katsayıları β_i 'le arasında, en yakın gecikme değerlerinde yüksek iken, gecikme uzunluğu arttıkça katsayıların değeri azalmaktadır. Gecikme uzunluklarının görece uzun olduğu $p + 1 = 40$ durumda, işlevin çıktısı olan beta katsayılarının sıfıra yakınsaması belirgin biçimde gözlenmektedir.

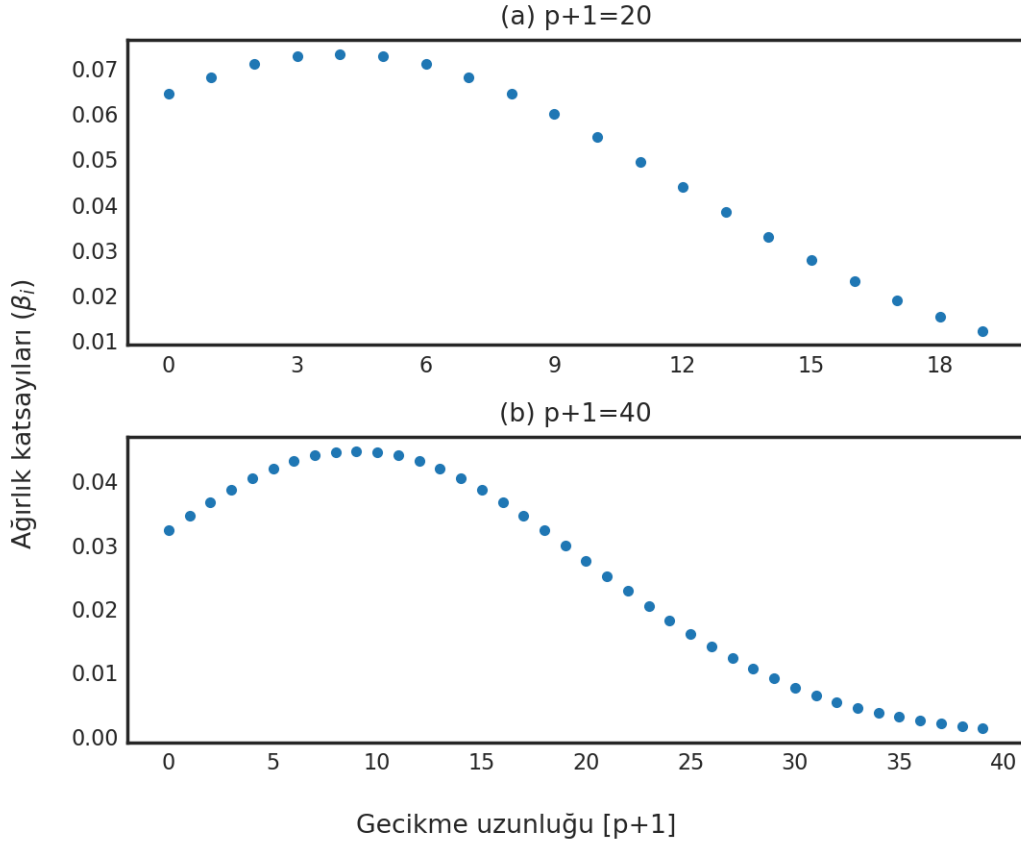
Benzetim deneyinde, üssel ağırlıklandırma işlevi için $\theta = (7 * 10^{-4}, -6 * 10^{-3})$ olurken, gecikme uzunlukları sırasıyla $p + 1 = (20, 40)$ olarak modelde yer almıştır. Ağırlık katsayıları toplamı bire eşit olacak şekilde, normalize edilmiştir.

5.3.2 Tümsek azalan ağırlıklandırma işlevi

Azalan Tümsek ağırlıklandırma işlevi de, Almon gecikme işlevinin özel bir halidir. Almon polinomunun ikinci derece ve eğim parametresinin negatif işaretli olması ayırt edici özelliğidir. Tümsek azalan işlevinin cebirsel olarak gösterimi şöyledir:

$$\omega(\theta) = \frac{\exp(\theta_1 j - \theta_2 j^2)}{\sum_{i=0}^p \exp(\theta_1 j - \theta_2 j^2)} \quad (5.4)$$

Yukarıda işlevde yer alan yalınlaştırma parametreleri, bütün gecikme uzunlukları için $\theta_1 = 0.08$ olarak alınırken, θ_2 parametresi için, sırasıyla, 20 ve 40 gecikme uzunlukları için θ_2 ve $\theta_1/10$ ve $\theta_1/20$ değerleri kullanılmıştır. İşlevde yer alan j indis değeri (i, \dots, p) arasında gecikme uzunlukları değerlerini ifade eder. Denklemdaki $\omega(\theta)$ değerleri Tümsek azalan ağırlıklandırma işlevi katsayılar vektörünü gösterir.



Şekil 5.2: Tümsek azalan ağırlıklandırma işlevi

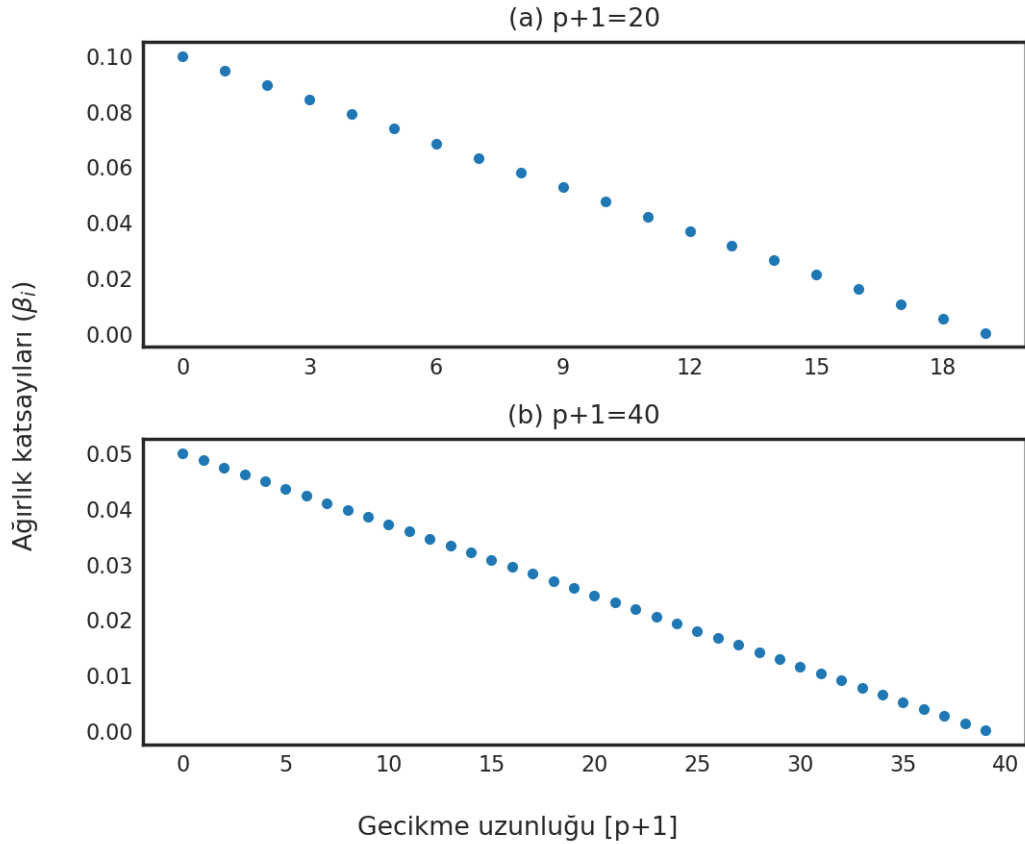
Almon polinomunun ikinci derece katsayısı θ_2 'nin negatif olması, şekil 5.2'de görüldüğü üzere, gecikme uzunluğu ile ağırlık katsayıları arasındaki ilişkinin sıfıra yakınsamasa hızının azalmasına neden olmaktadır. Gecikme uzunluğu ile ağırlık katsayıları arasında ilişki daha yumuşak bir azalış görünümü ortaya koymaktadır. Ağırlıklandırma işlevinin tanımı gereği, ağırlık katsayılarının toplamı bire eşit olmaktadır.

5.3.3 Doğrusal azalan ağırlıklandırma işlevi

Doğrusal azalan ağırlıklandırma, a_1 ve a_2 parametreleri tarafından belirlenen, aşağıya doğru eğimli bir işlevdir. Gecikme uzunluğu artıkça, ağırlık katsayılarının etkisi doğrusal olarak azalır (Roling, 2014). Breitung ve Roling (2015) tarafından yazılan makalede doğrusal ağırlıklandırma işlevi, azalan olarak nitelendirilmiş ise de söz konusu formül artan eğimli çıktı üretmektedir. Roling (2014) doktora tezinde, azalan doğrusal işlev olarak farklı yapıda bir ağırlıklandırma işlevine yer vermiştir, işlevin bu versiyonu, doğrusal azalan ağırlıklandırma işlevi olarak benzetim deneyine dahil edilmiştir. Toker, Özbay ve Månsson (2022) yayınladıkları makalede, (Breitung ve Roling, 2015) makalesindeki denklemi alarak azalan ifadesine yer vermeyerek, doğrusal ağırlıklandırma işlevi olarak ifade ederek etmişlerdir.

$$\omega_p = \frac{a_0 + a_1 p}{a_0 P + a_1 (P - 1) / 2} \quad (5.5)$$

Yukarıda doğrusal azalan ağırlık işlevinin formülü yer almaktadır. Doğrusal azalan ağırlıklandırma işlevinde $a_1 = 0.05$ ve $a_2 = -0.08$ olarak kullanılmıştır.



Şekil 5.3: Doğrusal Azalan Ağırlıklandırma işlevi

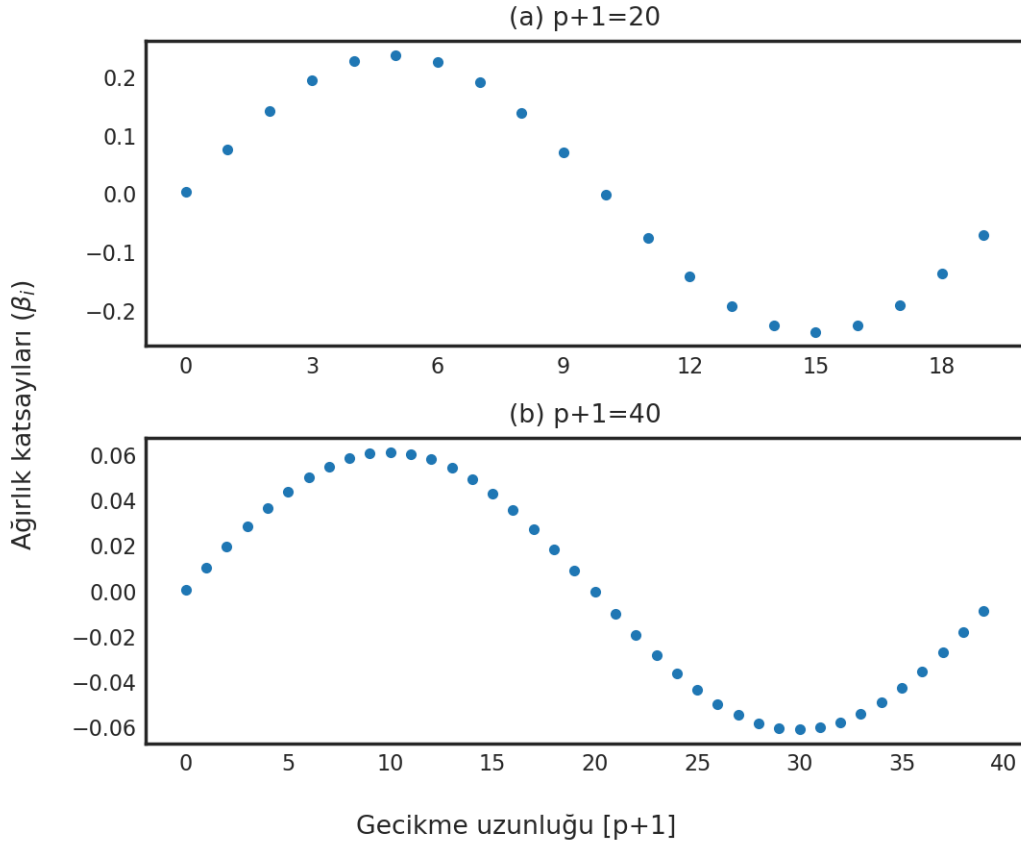
Yukarıda şekilde 5.3 görüldüğü üzere, ağırlık katsayıları β_i 'ler gecikme uzunlukları $p + 1$ artıkça azalmakta olduğundan, azalan doğrusal şekildedir. Benzetim deneyinde, $p + 1 = (20, 40)$ olarak çalıştırılmıştır.

5.3.4 Çevrimsel ağırlıklandırma işlevi

Çevrimsel (cyclical) ağırlıklandırma işlevi, c_1 ve c_2 parametreleri tarafından belirlenen, artan-azalan özelliklere sahip, sinüs eğrisi baz alınarak oluşturulmuştur. İşlev, gecikme uzunluğu $p + 1$ ve radyal çevrim için 2π değerleriyle ilişkilendirilmiştir.

$$\omega(c_1, c_2) = \frac{c_1}{p+1} \left[\sin \left(c_2 + \frac{j2\pi}{p} \right) \right] \quad (5.6)$$

Denklem 5.6'da, ağırlık katsayıları üretimi için $c_1 = (5, 2.5)$ değerlerinin gecikme uzunlukları $p + 1 = (20, 40)$ değerlerine sırasıyla paralel olarak değişirken, $c_2 = 0.01$ değeriyle deney boyunca sabit tutulmuştur.



Şekil 5.4: Çevrimsel ağırlıklandırma işlevi

Yukarıda şekil (5.4)'de $p + 1 = (20, 40)$ ağırlıklandırma katsayılarının değişimi gözlenmektedir. Çevrimsel ağırlıklandırma işlevi öncekilerden farklı olmak üzere, farklı gecikme değerlerinde artan ya da azalan değişim göstermektedir.

5.4 Yöntemlerin Uygulanması

Deney boyunca parametrik doğrusal olmayan MIDAS regresyon (MIDAS-NLS) ve parametrik olmayan Düzleştirilmiş MIDAS regresyon (MIDAS-SLS) modelleri kullanılmıştır. Adı geçen yöntemler için, Akaiki Bilgi Ölçütü düzeltilmiş versiyonu (AICc) ve Hata Kareler Ortalaması (HKO)yi minimum yapacak, yüksek frekanslı serinin gecikme uzunluğu bulunmuştur. Uygulama için, Python programlama dilinin bilimsel hesaplamalar için kullanılan "scipy" (Virtanen vd., 2020) modülünün, "optimize" alt modülü kullanılmıştır. Her iki yöntem içinde, en uygun değer arayışı için, sayısal çözümleme yöntemi "Nelder-Mead" kullanılmıştır.

MIDAS-NLS modeli için, yüksek frekanslı serinin gecikmeli değerlerinin dönüştürülmesi için üssel Almon ağırlıklandırma işlevi kullanılmıştır. Model yürütülürken, sayısal çözümleme yöntemiyle, AICc ve HKO bilgi ölçütlerinin en küçük değerini veren gecikme uzunluğu aranmaktadır. Modelde başlangıç değeri olarak, Üssel Almon İşlevinin yalınlaştırma parametreleri θ_1 ve θ_2 ve bir sabit θ_0 değeri verilmekte, her bir rassal veri seti için, gecikme uzunlukları değiştirilerek test edilmekte, Nelder-Mead simleks algoritması kullanılarak yalınlaştırma parametreleri yenilenecek iterasyon yapılmaktadır.

MIDAS-SLS modeli için, sayısal çözümleme yöntemi ile öncelikle bant genişliği parametresi λ hesaplanır. Her bir rassal veri seti için, $AICc(\lambda)$ ve $HKO(\lambda)$ bilgi ölçütü değerlerini minimize eden, gecikme uzunlukları test edilerek, en uygun λ değerlerine ulaşırlar.

Bu işlemlerin sonucunda, yöntemler ayrıştırılmış olarak, her bir veri seti grubu için gecikme uzunlukları, AICc ve HKO değerlerinden oluşan sonuç tablosu elde edilmektedir. AICc ve HKO değerleri minimum olan gecikme değerleri kaydedilerek gruplandırılır. Gecikme uzunluğunun frekans sıklığı tablosu elde edilir. Başarılı gecikme uzunluğu değeri ya da aralığı deney sayısına (500 adet) oranlanarak, yöntemin deney başarı oranı hesaplanır.

5.5 Benzetim deneyi sonuçları

Benzetim deneyinin sonuçları, MIDAS-NLS ve MIDAS-SLS olarak karşılaştırmalı ve göreceli başarı oranları olarak sunulmuştur. Ayrıca yöntemler, örneklem büyüklüğü $T = (100, 200, 400)$, bilgi ölçütü (AICc ve HKO), gecikme uzunluğu $p+1 = (20, 40)$ ve değişkenlik ölçeklendirme parametresi $\alpha = (0.2, 0.3)$ değerlerine göre başarı oranları incelenmiştir. Söz konusu kontrol parametreleri, test edilen yöntemleri farklı durumlardaki davranışı inceleyerek, stres testine tabii tutmaktadır.

5.5.1 Deney Sürecinde karşılaşılan zorluklar ve sınırlandırmalar

Rassal veri setleri üretiminde, yüksek frekanslı değişken için AR(1) modeli kullanılmıştır. MIDAS regresyon için veri matrisleri oluştururken, MIDAS-SLS için, gecikme uzunluğu $p + 1$ 'in yinelenme sıklığı m 'den büyük olması bazı belirsizlikler oluşturmaktadır. Bağımsız değişken $x_{t,p}$ için hata terimi $\varepsilon_{t,p}$ her bir gecikme uzunluğu için tanımlanır. Diğer bir ifadeyle, $p + 1 > m$ ise tek satırdaki hata terimi $\varepsilon_{t,p}$ ile yüksek frekanslı değişken $x_{t,p}$ kesintisiz AR(1) modeli yapısıyla çelişki oluşur. Ters durum için, $p + 1 < m$ için bu çelişki geçerli değildir. MIDAS-SLS için AR modelinde, $p + 1 > m$ bir belirsizlik sınırlandırması ortaya çıkar.

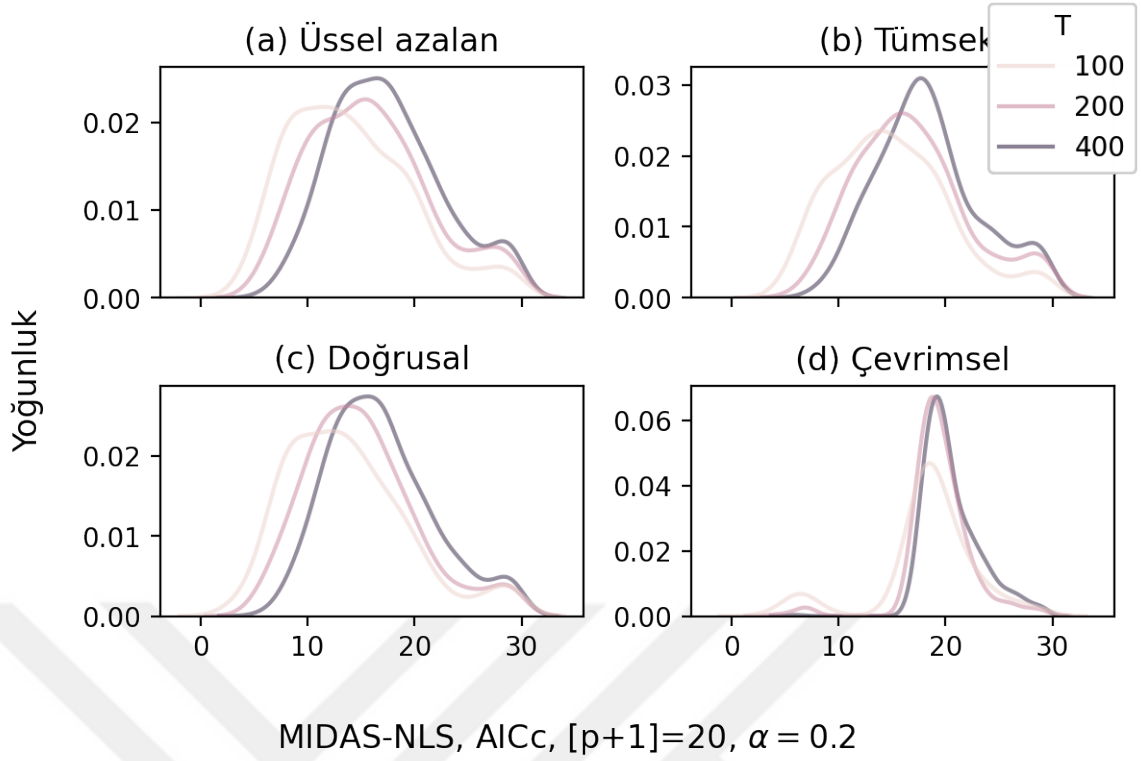
Deney kapsamı, teknik donanım sınırlamaları nedeniyle daraltılmak zorunda kalmıştır. Başlangıçta deney sayısı, 1000 adet planlanmasına ve denenmesine karşın, özellikle MIDAS-SLS modelinin çok yüksek bir donanım kapasitesi gerektirmesi nedeniyle 500 adete indirilmiştir. Sayısal hesaplama ve iterasyon işlemlerinin çok yüksek işlemci (CPU) kapasitesi gerektirmesi, destek olarak grafik işlem hızlandırıcısı (GPU) kullanılmasının rağmen planlanan kapasitede işlem yapılması mümkün olmamıştır. Yöntemlerin test edilmesinde 500 deney sayısının da önemli ölçüde yeterli olduğu görülmüştür.

5.5.2 Deney sonuçlarının karşılaştırılması

Deney sonuçları göre, parametrik MIDAS-NLS ve parametrik olmayan MIDAS-SLS yöntemlerinin karşılaştırılması yapılmıştır. Yöntemlerin, kontrol parametrelerindeki değişimlere duyarlılıkları izleyen alt başlıklarda değerlendirilmiştir.

5.5.2.1 Örneklem hacmindeki değişimin etkisi

Örnek büyüklüğü T artıkça beklendiği üzere MIDAS-NLS ve MIDAS-SLS modellerinin her ikisinin de performansı paralel artış göstermektedir. MIDAS-SLS modelinin, daha büyük örneklem hacimlerinde göreceli performansı daha yüksektir.

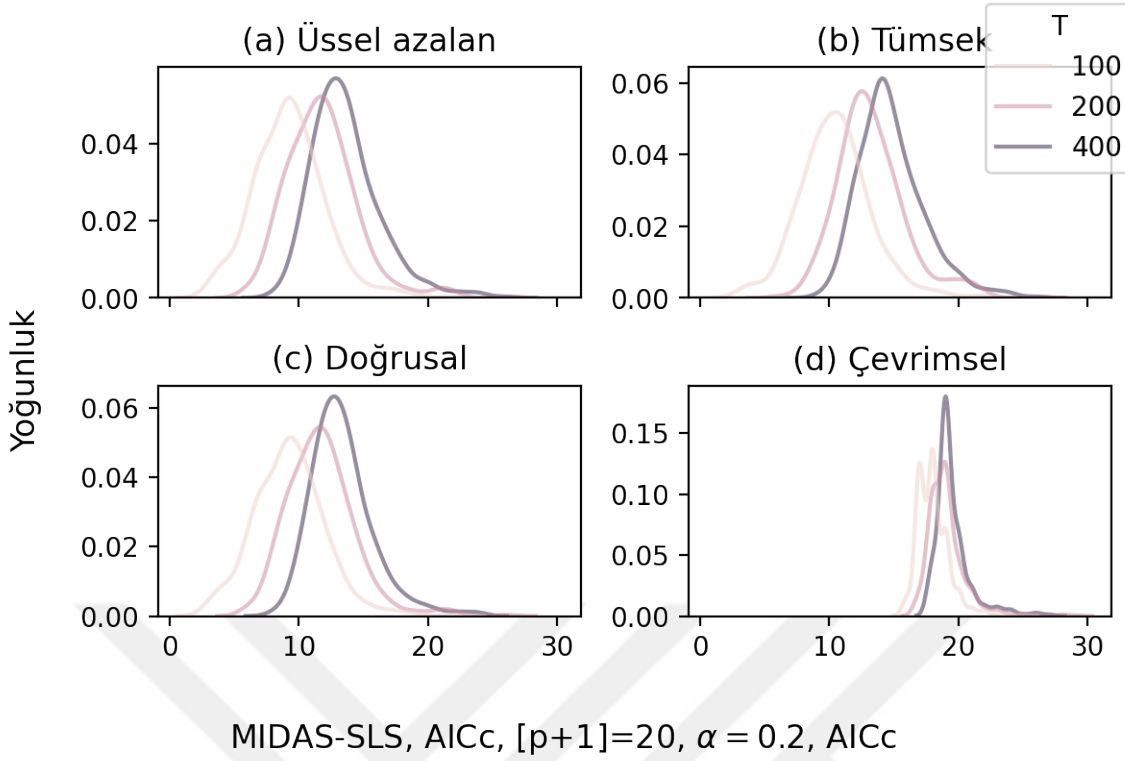


Şekil 5.5: MIDAS-NLS modelinin örnek büyüklüğüne göre performans değişimi

Gecikme uzunluğu arttıkça, MIDAS-SLS göreceli olarak daha yüksek performans göstermektedir. Yukarıda şekil (5.5)'te, MIDAS-NLS modelinin benzetim deneyi sonuçları, çekirdek yoğunluk çizimleri ile görsel olarak sunulmuştur. Doğru gecikme uzunluğu $(p+1)=20$ ve değişkenlik ölçek parametresi $\alpha = 0.2$ sabit kalırken, örneklem genişliği olarak gözlem sayısı $T=(100, 200, 400)$ olarak değişmektedir. Bağımlı değişken gözlem sayısı arttıkça, doğru modele, dört farklı ağırlık işlevinden üretilmiş rassal veri setlerinde de, yakınsamanın arttığı açıkça gözlenmektedir.

MIDAS-SLS modelinde de, yukarıda MIDAS-NLS modeli açıklamalarına paralel olarak, örnek hacmi arttıkça doğru gecikme uzunluğuna yakınsama eğilimi açıkça gözlenmektedir. MIDAS-SLS doğru modele yaklaşıırken, güven aralığının daha dar bir bantta oluştuğu gözlenmektedir.

Çevrimsel ağırlık işlevinden üretilmiş veri setinde, MIDAS-NLS güven aralığının $p + 1 = 20$ değerine oldukça dar bir bantta yakınsaması ilgi çekicidir (şekil 5.6). Diğer ağırlık işlevlerinden farklı olarak, artan-azalan ve negatif değerler alma özelliğini nedeniyle, parametrik olmayan MIDAS-SLS bu değişkenliği daha uygun bir örüntü uyarlaması beklentilere uygundur.

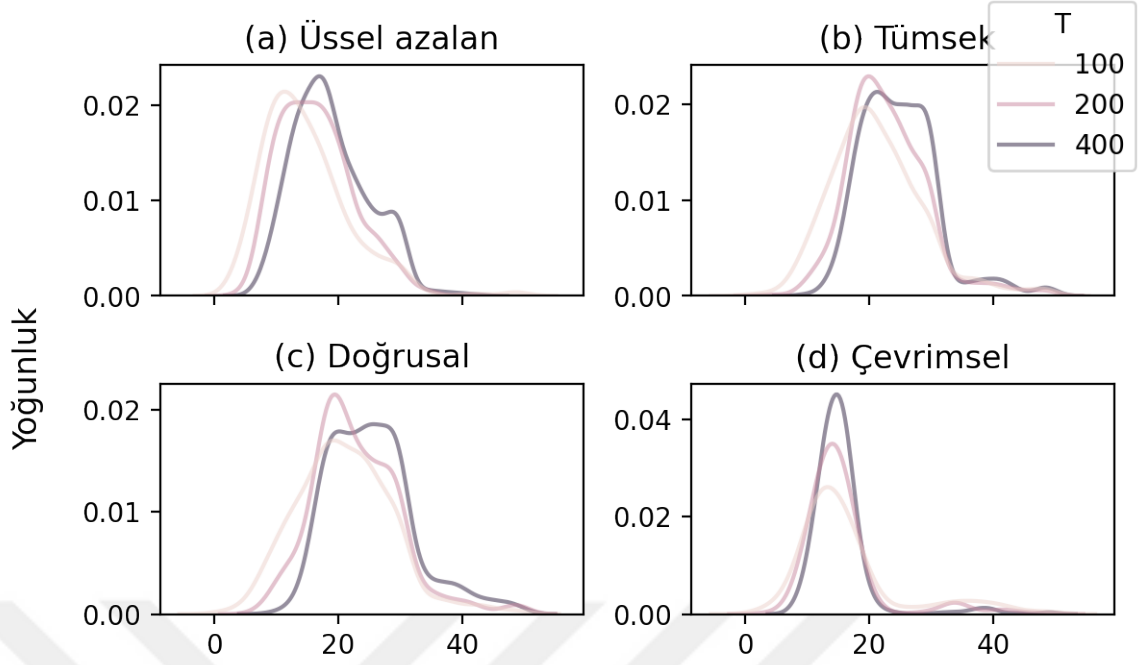


Şekil 5.6: MIDAS-SLS modelinin örnek büyüklüğüne göre performans değişimi

Çevrimsel ağırlık yapısında, MIDAS-SLS başarı seviyesi AIC bilgi ölçütünde MIDAS-NLS'ten dikkat çekici biçimde daha yüksektir. Ayrıca, değişkenlik ölçeklendirme katsayısı α artınca MIDAS-SLS'in başarı oranı artmaktadır. Buna karşılık, HKO ölçütü ile bakıldığında MIDAS-SLS'in performansı, MIDAS-NLS'in altında kalmaktadır.

5.5.2.2 Gecikme uzunluğu değişiminin etkisi

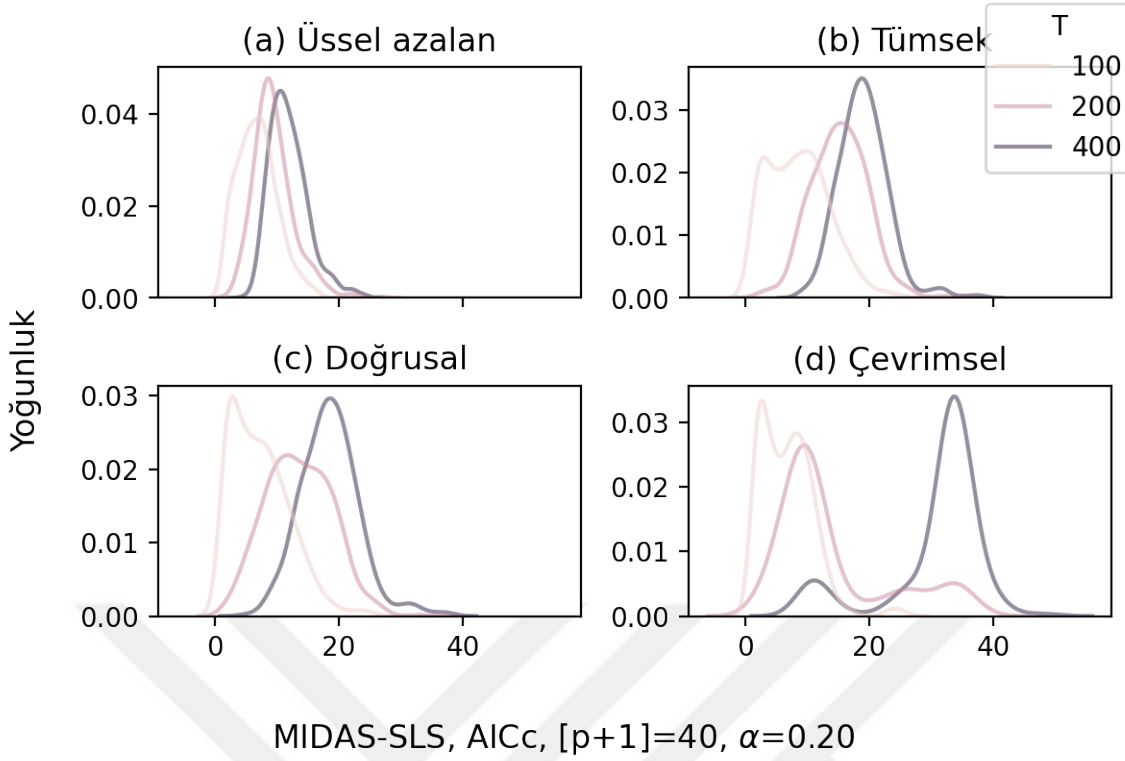
Yukarıdaki alt başlıkta, $p+1=20$ gecikme uzunluğunun doğru model olarak tanımlandığı durum incelenmişti. Koşullar aynı kalmak üzere, gecikme uzunluğu parametresi $p+1=40$ durumunda ilgi çekici farklılaşmalar ortaya çıkmaktadır.



MIDAS-NLS, AICc, $[p+1]=40$, $\alpha = 0.2$

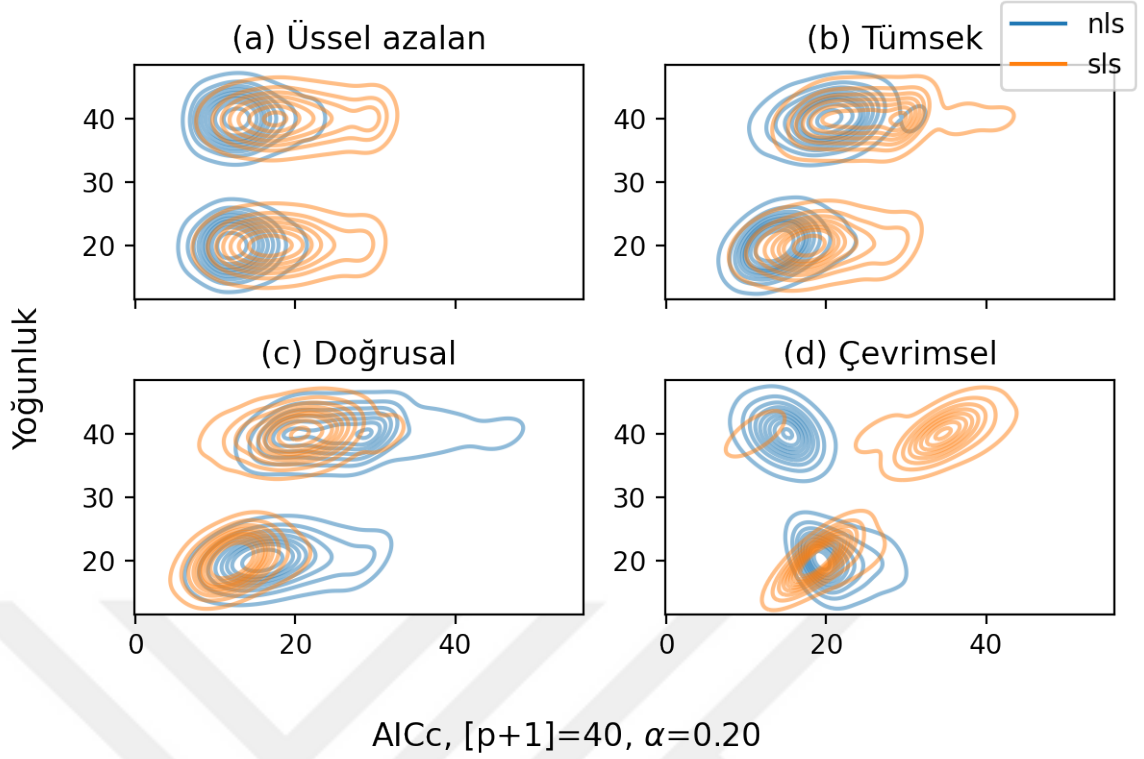
Şekil 5.7: MIDAS-NLS modelinin gecikmesi uzunluğuna göre performas değişimi

MIDAS-NLS modelinde, gecikme uzunluğu arttıkça, doğru modele yakınsama ve kestirim yeteneği, tüm ağırlıklandırma işlevlerinde azalmıştır. Rassal veri oluşturma sürecinde, yalınlaştırma parametrelerinin daha uzun gecikme uzunluklarını yansıtabilecek değerlerde seçilmesi daha iyi performansa yol açabilir. Parametrik MIDAS-NLS modelinin yapısal özellikleri, kısa gecikme değerlerin modele yansıtılmasına uygundur. Ayrıca, MIDAS-NLS modelinin doğrusal olmayan EKK'nin sayısal optimizasyon için, yüksek miktarda iterasyon gerektirmesinin doğru örüntüye ulaşmasını güçleştirdiği düşünülmektedir. Eğer MIDAS-NLS modelinde, birinci türevi alınabilir bir polinom yapısı kullanılırsa, Nelder-Mead simpleks algoritması yerine, BFGS (Broyden-Fletcher-Goldfarb-Shanno) algoritması kullanılabilir. Alternatif olarak, ikinci dereceden türevlenebilir bir işlev kullanılırsa, Hessian determinantlarının da optimizasyon sürecine dahil olduğu, Newton-CG (Newton-Conjugate-Gradient) algoritması, daha uygun olabilir. BFGS ya da Newton-CG algoritmaları daha hızlı çalışırlar.



Şekil 5.8: MIDAS-SLS modelinin gecikmesi uzunluğuna göre performans değişimi

MIDAS-NLS modelinin, gecikme uzunluğu artışından olumsuz etkilendiği görülmektedir. Bu durum, sadece çevrimsel ağırlıklandırma işlevinde büyük örneklem üzerinde çalıştırıldığında azalmaktadır. Üssel azalan ve Tümsek azalan ağırlıklandırma işlevlerinin, parametrik üssel Almon işlevinin özelleştirilmiş durumları olması ve kısa gecikme uzunluklarına uygun yapılar olmasının da etkisi olabilir. Doğrusal azalan ağırlıklandırma işlevinde, gecikme uzunluğu artıkça, ağırlık sayıları azalır. Diğer yandan, MIDAS-NLS, çevrimsel ağırlık işlevinde, yeterli örnek hacmi varsa doğru modele yakınsama yeteneğine sahiptir. Yukarıda şekil 5.7(d)'de, $p+1=40$ gecikme uzunluğunda ve $T=400$ örneklem büyüklüğünde, doğru modele belirgin bir biçimde yakınsamaktadır.

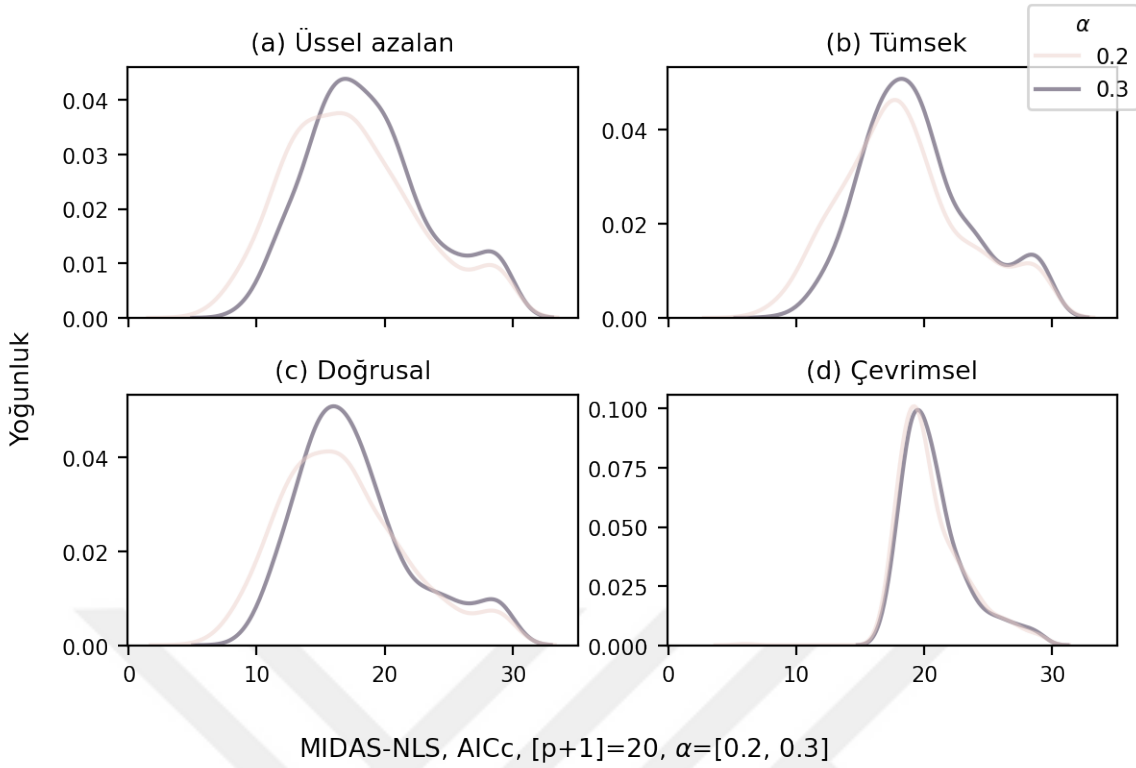


Şekil 5.9: MIDAS-NLS ve MIDAS-SLS modellerinin gecikmesi uzunluğuna göre performans karşılaştırması

Yukarıda kesit grafiğinde (şekil 5.9), MIDAS-NLS ve MIDAS-SLS, yöntemlerinin, $p+1=20$ ve $p+1=40$ gecikme uzunluklarının, doğru modele yakınsama eğilimleri karşılaştırmalı olarak gösterilmektedir. MIDAS-NLS çevrimsel ağırlıklandırma işlevinde gecikme uzunluğunu ayırmada belirgin olarak başarılı olduğu görülmektedir. Üssel azalan, tümsek ve doğrusal işlevlerde, parametrik ve parametrik olmayan yöntemler benzer bir ayırma örneğine sahiptir. Burada, MIDAS-NLS modeli gecikme uzunluğu artışı yansıtmakta yetersiz kalırken, MIDAS-SLS modeli gecikme uzunluğu artarken daha başarılı gözlenmektedir.

5.5.2.3 Değişkenlik ölçek faktörü (α) değişiminin etkisi

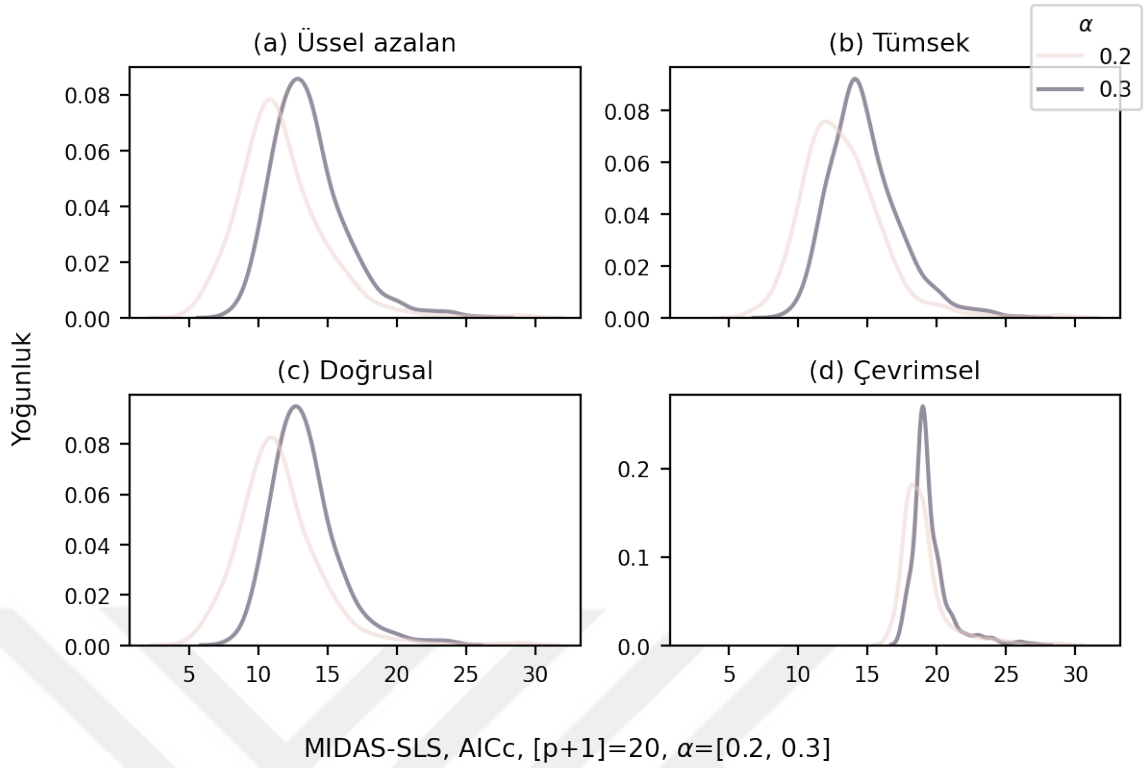
Değişkenlik ölçek faktörü α 'nın değeri azaldıkça, ağırlıklandırma katsayılarının etkisi zayıflar. Bu sebeple, yüksek frekanslı değişkenin değişkenliği artar. Değişkenliğin artması ve azalmasının MIDAS-NLS ve MIDAS-SLS modellerinin performansına etkisi, aşağıda görsel olarak sunulmuştur. Değişkenlik ölçek faktörü karşılaştırmasında, her iki modelinde yakın performans gösterdiği $p+1=20$ gecikme uzunluğu ve örneklem büyüklüğü $T=400$ baz alınmıştır.



Şekil 5.10: MIDAS-NLS modelinin değişkenlik ölçek faktörüne göre performans değişimi

Yukarıda (şekil 5.10), MIDAS-NLS modelinde, değişkenlik ölçek faktörü (α) azaldığında -yani değişkenlik artığında-, çevrimsel hariç olmak üzere, diğer ağırlıklandırma işlevinden üretilmiş veri setlerinde, performans azalışı görülmektedir. Çevrimsel ağırlıklandırma işlevinin artan-azalan karakterde olması nedeniyle, değişkenlik faktörünün etkisini önemli ölçüde zayıflattığı görülmektedir. MIDAS-NLS modelinde değişkenliğin artması, göreceli performansını azaltır.

MIDAS-SLS modeli de, aşağıda (şekil 5.11) görüldüğü üzere, farklı ağırlık işlevlerinden üretilmiş veri setlerinde, değişkenliğin artışı olumsuz etkilenmektedir. Değişkenlik azaldıkça doğru modele yakınsama belirgin olarak artış kaydetmektedir.



Şekil 5.11: MIDAS-SLS modelinin değişkenlik ölçek faktörüne göre performans değişimi

Değişkenlik artışına, parametrik MIDAS-NLS modelinin duyarlılığının daha zayıf olduğunu, buna karşılık parametrik olmayan MIDAS-SLS'nin duyarlılığının daha yüksek olduğu görülmüştür.

5.5.2.4 Göreceli performans incelemesi

Modelin başarısı, rassal veri setine gömülmüş olan doğru gecikme uzunluğundaki modeli belirleme yetkinliği olarak tanımlanmıştır. MIDAS-SLS ve MIDAS-NLS yöntemlerini toplu olarak karşılaştırabilmek için göreceli başarı oranları hesaplanmıştır. Göreceli performans oranı ise, MIDAS-SLS başarısının, MIDAS-NLS başarısına bölünmesiyle hesaplanmıştır. Göreceli performanslar aşağıda tablo 5.1'de yer almaktadır. Birden küçük değerler, MIDAS-SLS'in, MIDAS-NLS'e göre daha başarısız, birden büyük ise MIDAS-SLS daha başarılı olduğunu gösterir.

Çevrimsel işlevden üretilmiş rassal veri seti için, MIDAS-SLS modelinin, AICc bilgi ölçütünde, MIDAS-NLS göre belirgin bir başarı gösterdiği gözlenmektedir. Ancak hata kareler ortalaması (HKO) göre, MIDAS-SLS'in başarı oranı MIDAS-NLS'in altında kalmaktadır. Gecikme uzunluğu arttıkça, MIDAS-SLS başarı oranı artma eğilimindedir. Tabloda yer alan, bazı uç başarı değerlerin, deney sayısının artırılması durumunda, daha iyi ölçüm verecek orana yakınsamasını mümkün olabilir.

Doğrusal azalan ağırlık işlevinde, gecikme uzunluğunun $p+1=20$ olduğu veri seti için, MIDAS-NLS modelinde HKO'ya göre daha iyi performans gözlenmesine karşın, gecikme uzunluğu $p+1=40$ olduğu

Tablo 5.1: MIDAS-SLS ve MIDAS-NLS göreceli performansları

Ağırlık işlevi	T	p+1=20		p+1=40	
		$\alpha=0.2$	$\alpha=0.3$	$\alpha=0.2$	$\alpha=0.3$
Çevrimsel					
AICc	100	1.0706	1.2493	0.0000	0.0000
	200	1.1383	1.3075	1.0000	10.0000
	400	1.2958	1.2841	5.5556	148.0000
HKO	100	0.6637	0.6347	2.3143	5.5294
	200	0.5894	0.6160	6.5625	13.0000
	400	0.5514	0.6453	13.6000	57.6667
Doğrusal azalan					
AICc	100	0.0330	0.0579	0.0000	0.0000
	200	0.0614	0.1083	0.0000	0.0000
	400	0.1180	0.2131	0.0870	0.1818
HKO	100	0.8632	0.7360	3.2222	2.0769
	200	0.7179	0.7333	2.2400	1.3500
	400	0.5706	0.6648	2.0000	2.4333
Tümsek azalan					
AICc	100	0.0465	0.0753	0.0000	0.0000
	200	0.0943	0.2150	0.0000	0.0000
	400	0.2038	0.4050	0.1333	0.2000
HKO	100	0.7280	0.6887	3.5625	3.0556
	200	0.6711	0.5171	3.7857	2.7000
	400	0.5891	0.6260	2.7895	3.2857
Üssel azalan					
AICc	100	0.0275	0.0720	0.0000	0.0000
	200	0.0522	0.1257	0.0000	0.0000
	400	0.1607	0.2632	0.0000	0.0000
HKO	100	0.8600	0.7308	5.2000	4.3636
	200	0.6593	0.5967	4.4000	3.9167
	400	0.5965	0.5814	6.8571	4.4615

veri setinde, MIDAS-SLS modeli daha yüksek başarı göstermektedir. HKO göre, gecikme uzunluğu arttıkça MIDAS-SLS modelinin başarısı artma eğilimindedir. AICc bilgi ölçütü bu veri setinde, ayırt edici bir ölçüm farklılığı ortaya koyamadığı gözlenmektedir. AICc bilgi ölçütü, MIDAS-SLS modelinde, p+1=40 gecikme uzunluğu için başarılı model seçimi yapmadığı gözlenmiştir, deney sayısının daha yüksek olması durumunda bu değerlerde iyileşme gözlenebilir. Örneklem genişliğinin düşük olduğu T=(100,200) grubunda, gecikme uzunluğundaki artışın, hem serbestlik derecesini ve hem de zaman serisi gözlem sayısını azaltması nedeniyle kestirim yapma yeteneğini düşürdüğü gözlenmektedir.

Tümsek ağırlıklandırma işlevi veri seti için, HKO ölçütünde, uzun gecikme değerlerinde, MIDAS-SLS modelinin göreceli başarısının daha yüksek olduğu gözlenmektedir. AICc bilgi ölçütünün, T=400

örneklem büyüklüğü grubunda ayırt edici işlev göstermeye başlamaktadır. Bu grupta da, örneklemin görece az olduğu $T=(100,200)$ değerlerinde, AICc yeterli ayırt edici ölçüm yapamadığı görülmektedir. HKO ölçütü, MIDAS-SLS, $p+1=20$ veri grubunda daha az performans gösterirken, $p+1=40$ gecikme uzunluğunda, durumun belirgin olarak MIDAS-SLS lehine iyileştiği görülmektedir.

Üssel azalan veri setinde, HKO'nun her iki model için daha ayırt edici ölçüm yapmaktadır. Gecikme uzunluğu arttıkça MIDAS-SLS modelindeki başarı, MIDAS-NLS göre artmaktadır. Kısa gecikme uzunluklarında ise tersine MIDAS-NLS modelinin, MIDAS-SLS'e göre daha başarılıdır.

Göreceli başarı sonuçlarının genel değerlendirilmesinde, örneklem genişliğinin ve gecikme uzunluğunun, göreceli başarı oranı üzerinde çok etkili olduğu görülmektedir. MIDAS-SLS yöntemi daha büyük örneklem genişliğinde, uzun gecikme uzunluklarında iyi performans gösterebilir. MIDAS-NLS daha küçük örneklem genişliğinde ve kısa gecikme uzunluklarında daha başarılı sonuçlar gösterme eğilimindedir.

5.6 Sonuç

Benzetim deneyi sonuçlarına göre, parametrik olmayan MIDAS-SLS modeli, parametrik MIDAS-NLS modeline göre, yüksek frekanslı değişkenin gecikme uzunluğu arttıkça, model seçiminde daha başarılı olma eğiliminde olduğunu gözlenmektedir. Ancak büyük gecikme uzunluklarında, MIDAS-SLS modelinin başarılı kestirim yapabilmesi için daha büyük örneklem hacimlerine ihtiyaç vardır. Gecikme uzunluğunun kısa olduğu durumlarda, örneklem genişliği büyük olmasa bile, MIDAS-NLS başarılı sonuç üretme eğilimindedir.

MIDAS-SLS, parametrik olmayan bir yöntem olması nedeniyle, parametrik MIDAS-NLS yöntemine göre daha yüksek hesaplama kapasitesine sahip donanım ve daha fazla işlem zamanı gerektirmektedir. MIDAS-NLS görece daha az zaman ve basit donanım gerektirmektedir. MIDAS-NLS modelinde, birinci ya da ikinci dereceden türevi alınabilen parametrik ağırlıklandırma işlevleri tercih edildiğinde, daha kısa zamanda işlem yürütülmesi mümkündür.

Parametrik olmayan MIDAS-SLS modeli, artan-azalan ve negatif değerler alabilen çevrimsel veri yapısında, parametrik yönteme göre daha başarılı olduğu gözlenmektedir. Ayrıca, örneklem hacmi arttıkça her iki yönteminde başarı oranı artmaktadır. Parametrik olmayan yöntemin başarılı sonuç üretmesi için büyük örneklem hacmine ihtiyaç vardır. Değişkenlik azaldıkça, parametrik olmayan MIDAS-SLS başarısının da artma eğiliminde olduğu gözlenmektedir.

Benzetim deneyi, 500 tekrarlı olarak yapılmıştır. Deneyin daha yüksek tekrarlı olarak yapılmasının performans farklarını daha iyi ölçülebilir hale getireceği tahmin edilmektedir. Ancak tekrar sayısının artırılması, donanım ve işlem süresi kısıtları nedeniyle, denenmesine rağmen mümkün olmamıştır.

Bölüm 6

Uygulama

Korana Virüs salgını 2019 yılı sonunda başlayarak, kısa süre sonra 2020 yılı başında, Dünya Sağlık Örgütü (DTÖ) tarafından küresel ölçekte salgın anlamına gelen pandemi olarak ilan edilmiştir. Ülke yönetimleri, yaygın ve öldürücü nitelikteki COVID-19 pandemisine karşı önlem geliştirmekte zorlanmışlardır. Sadece az gelişmiş ülkeler değil, mücadele için kaynakları güçlü olan gelişmiş ülkeler de etkili ve önleyici önlemler geliştirmekte yetersiz kalmıştır. Pandeminin ilk dönemlerinde hastalığa karşı koruma sağlayacak aşuların hazırda bulunmaması, hızla artan hasta sayısı ile sağlık hizmetlerinin aşırı artan talep ve ihtiyacı karşılayamaması, COVID-19 hastalığından kaynaklı ölümlerin artmasına yol açmıştır. Sağlık sisteminin ayakta kalması için, hastalığın toplum geneline yayılmadan kısıtlayıcı önlemler alınması, erken uyarı sistemlerinin varlığını çok daha önemli hale gelmiştir. Bu konuda küresel çapta araştırmacıların ilgisini çekmiştir.

Solunum yoluyla bulaşan Korana grubu ve diğer değişim geçiren virüslerinin gelecekte de insanlığa karşı tehdit oluşturacağı öngörülmektedir. Bu nedenle, virüslerin henüz yayılmadan izole edilmesi hayati öneme sahiptir. Bu bağlamda, erken uyarı modellerinin geliştirilmesi büyük önem arz etmektedir.

MIDAS regresyon modelinin, yüksek frekanslı açıklayıcı değişkenler üzerinden yola çıkarak model kurması, erken uyarı sistemleri oluşturulmasında karar destekleyici bir araç ortaya koyabilir. Bu bölümde, MIDAS regresyon yöntemlerini gerçek veriler üzerinde çalıştırarak, model seçimi, parametre kestirimi ve tahmin performansı karşılaştırması yapılmıştır. Küresel ölçekteki COVID-19 vaka sayılarına ilişkin resmi istatistikler ve sosyal medya verileri kullanılarak, bir erken uyarı sisteminin geliştirilmesi ele alınmıştır.

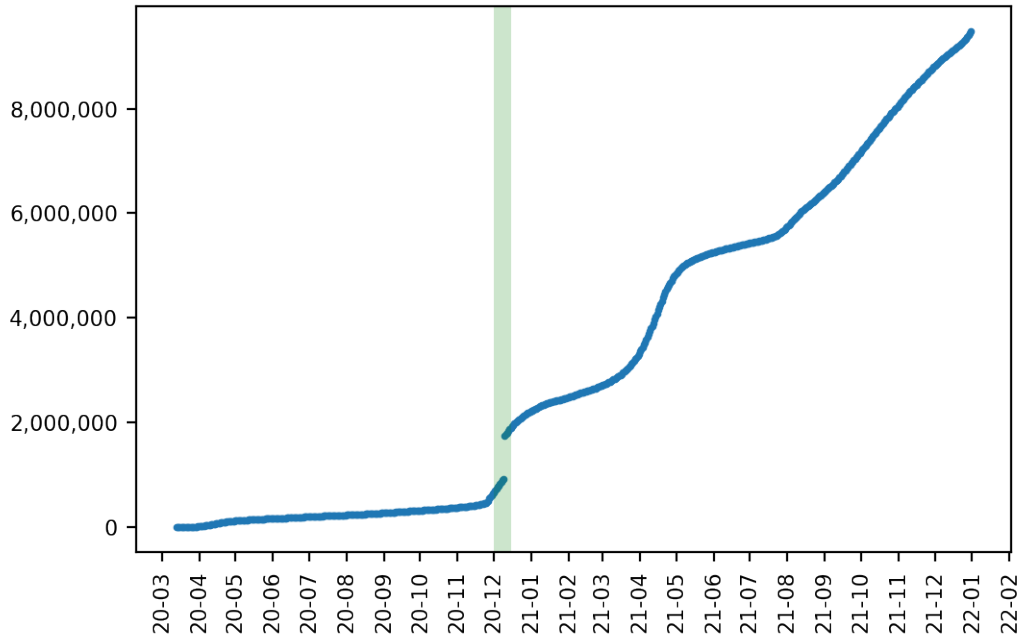
6.1 Veri yapısı ve işlem süreçleri

COVID-19 vakalarına ilişkin veriler için, John Hopkins Üniversitesi'nin küresel verileri derlediği veritabanı kullanılmıştır (Dong, Du ve Gardner, 2020). Veritabanının bulunduğu [proje sayfası](#), açık

kaynak proje paylaşım sitesi GitHub'da yer almaktadır. İlgili sayfada ülkelerdeki farklı veri kaynaklarına ilişkin bilgiler verilmiştir. Söz konusunu sayfada, Türkiye'ye ait verilerin Sağlık Bakanlığı ve Cumhurbaşkanlığı Dijital Dönüşüm Ofisi kaynaklarından düzenlendiği belirtilmektedir. Türkiye'nin COVID-19 verilerine ilişkin olarak içerik tanımlarında dönemler arasında zaman zaman önemli farklar yaratan değişiklikler yapılmıştır.

Erken uyarı sisteminin açıklayıcı değişkeni olarak, COVID-19 pandemisiyle ilgili sosyal medya paylaşımlarının hacim verisi kullanılmıştır. Sosyal medya platformu olan ve mini blog paylaşımları yapılan Twitter verileri kullanılmıştır. Küresel ölçekte bu verileri toplayan ve tweetID'leri paylaşarak erişilmesini sağlayan proje (Chen, Lerman ve Ferrara, 2020), GitHub üzerinde [proje sayfasında](#) paylaşımına açılmıştır. Açık paylaşımına sunulan Twitter mesajları, tweetID üzerinden, küresel ölçekteki tüm paylaşılan veriler indirilmiştir.

Veri işleme prosedürü aşamaları, ilk olarak indirilmiş CSV (Comma Separated Values) formatındaki metin veri dosyaları, saatlik parçalar halinde sadece tweetID bilgisi içermektedir. Tweet mesajları JSON (JavaScript Object Notation) veri yapısındadır. JSON veri yapısı, içi içe geçmiş (nested), yaklaşık 90 adet veri alanından oluşmaktadır. JSON formatındaki dosyalar indirilerek, MongoDB veri tabanına yüklenerek bir araya getirilmiştir. Twitter mesajlarının toplamı 1 milyar 960 bin Twitter mesajı 3.4 Tera Byte büyüklüğündedir. İkinci aşamada, Python programlama dilinin "pymongo" paketi kullanarak, analiz için ihtiyaç duyulan sorgular oluşturulmuştur. Türkçe dilinde paylaşılmış olan Twitter mesajları sorgulanarak bir veri seti oluşturulmuştur. Son aşamada, CSV formatında veri seti gerçek zaman aralıklı, günlük, saatlik, dakikalık ve saniyelik frekans aralıklarında, Türkçe paylaşılan veri setlerine dönüştürülmüştür. Bu veri seti 20 Ocak 2020 ve 31 Aralık 2021 tarihleri arasında paylaşılmış 18.7 milyon Twitter mesajlarından (tweet ve retweet) oluşmuştur.



Şekil 6.1: Birikimli günlük hasta / vaka sayıları (kişi)

Sağlık Bakanlığı tarafından açıklanan resmi COVID-19 istatistiklerde, birikimli bulaşma sayıları serisinde 'hasta sayısı' ve 'vaka sayısı' olmak üzere tanım değişikliği yapılmıştır. Tanımlardan 'vaka sayısı' COVID-19 testi pozitif olan her kişileri dahil ederken, 'hasta sayısı' sadece tıbbi desteğe ihtiyaç gösteren kişileri ifade etmektedir. Yukarıda şekilde(6.1) hasta ve vaka adetlerinin birikim günlük seyri gösterilmektedir. Hasta ve vaka sayısı tanım değişikliğinin yol açtığı farklılık açık renk bant çizgisi ile gösterilmiştir. İnceleme döneminden, birikimli bulaşma sayıları için, belirgin sıçrama noktası 10 Aralık 2020 tarihinde oluşmaktadır. Bu farklılık, analizde oluşturabileceği uç değer etkileri nedeniyle, 10 Aralık 2020 tarihten itibaren veriler 'vaka sayısı' nı, önceki tarihlerinde 'hasta sayısı' nı gösterdiği varsayılmıştır. Uç değer niteliğindeki 10 Aralık 2020 günü birikimli veri setinden çıkarılmıştır. Birikim veri setinin, birinci farkı alınarak, bağımlı değişken yeniden düzenlenmiştir. Böylece, birinci farkı alınan bağımlı değişken (y) serisi, günlük yeni eklenen vaka ya da hasta sayısı göstermektedir. Değişkenlere ilişkin bilgiler şöyledir:

Tablo 6.1: Model değişkenleri

Veri seti	Değişken	Sembol	Frekans	Tür	Açıklama
COVID-19	Bağımlı	y_t	(T,1)	Düşük	Günlük vaka (hasta) sayısı
Twitter	Bağımsız	$X_{t,m}$	(T, 24)	Yüksek	Saatlik tweet ve retweet sayısı
Twitter	Bağımsız	$X_{t,p}$	(T, p)	Yüksek	X'in gecikmeli değerleri

Türkiye COVID-19 vaka veri seti ile saatlik Türkçe Twitter mesaj sayıları verileri, Python programlama dilinde, *pandas*, *numpy* ve *scipy* paketleri kullanılarak, MIDAS veri matrislerinin gerektirdiği forma uygun yapılara dönüştürülerek, analize hazır hale getirilmiştir.

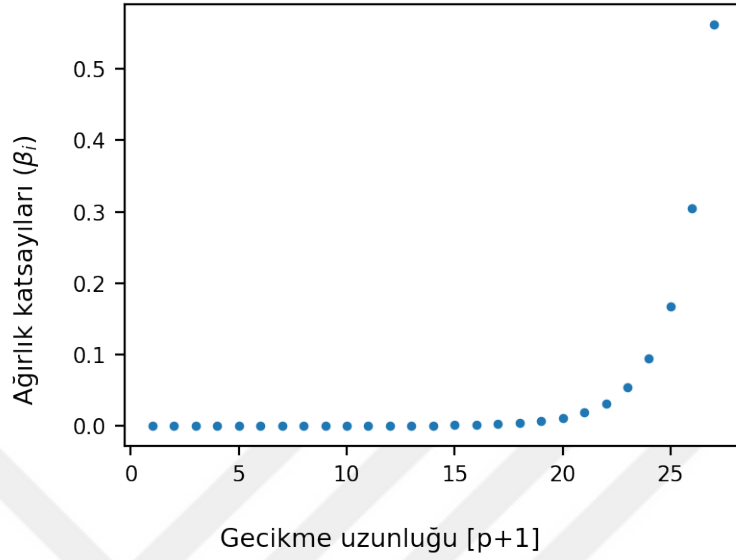
6.2 Model seçimi

MIDAS-NLS ve MIDAS-SLS modelleri, yukarıda tablo 6.1'de açıklanan veri setleri üzerinde çalıştırılmıştır. Yüksek frekanslı değişkenin gecikme yapısı örüntüsüne ait yalınlaştırma parametrelerinin kestirimleri yapılmıştır. Parametrik MIDAS-NLS modelinde gecikme uzunluğu ve ağırlık katsayıları; parametrik olmayan MIDAS-SLS modelinde bant genişliği, gecikme uzunluğu ve ağırlıklandırma katsayıları kestirimi yapılmıştır. Ardından, her iki modelin için tahmin değerleri (\hat{y}) hesaplanmıştır.

6.3 Parametrik MIDAS-NLS modeli sonuçları

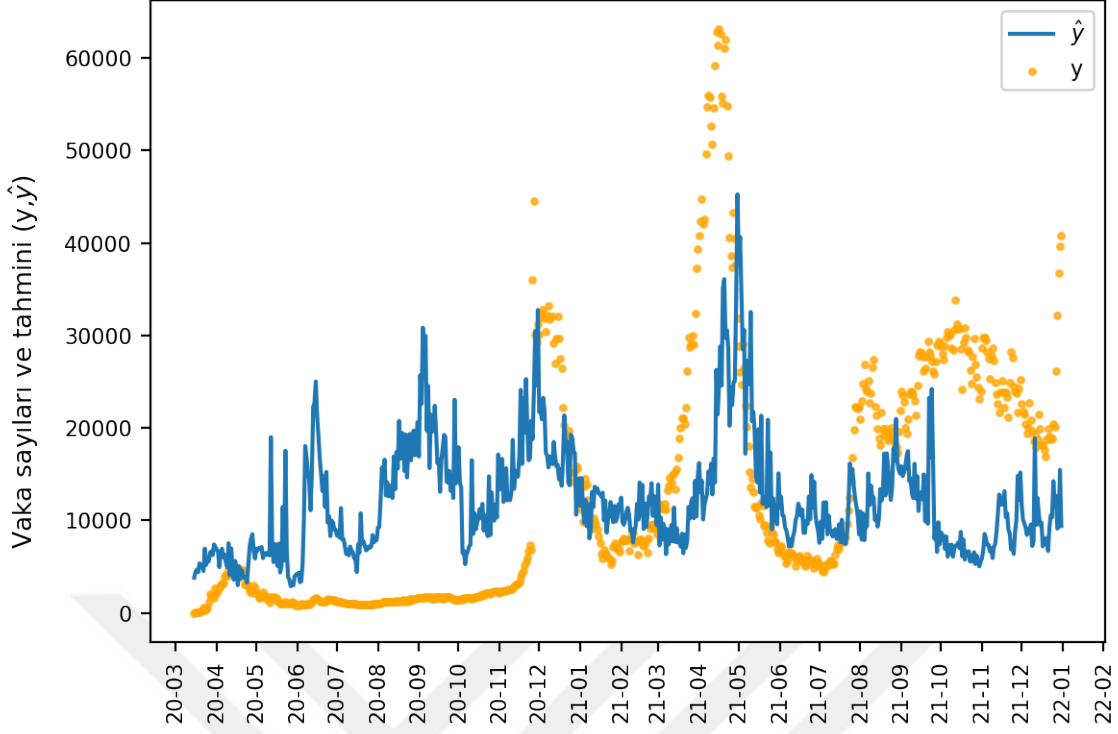
MIDAS-NLS modeli için, Üssel Almon ağırlıklandırma işlevi, sabit terim ile birlikte üç adet parametre tahmini gerektiren, ikinci derece polinom kullanılmıştır. Doğrusal olmayan regresyon modelinin parametre kestirimi için Nelder-Mead yöntemi kullanılmıştır. Elde edilen üç katsayıdan oluşan θ vektörü ile en uygun gecikme uzunluğu(p) üzerinden, ağırlık katsayıları vektörü (β_i) değerleri hesaplanmıştır.

Parametrik MIDAS-NLS modeli, en uygun gecikme uzunluđu kestirimini 27 saat olarak belirlemiřtir. Ařađıdaki řekil 6.2’de kestirimi yapılan parametrik ađırlık iřlevi katsayıları ile gecikme uzunluđu iliřkisi gzlenmektedir.



řekil 6.2: MIDAS-NLS modeli ađırlık iřlevi kestirimi

MIDAS-NLS modeli, vaka sayılarını 27 saat ncesinden itibaren yksek frekanslı serinin, COVID-19 vakalarının kestirimini yapabileceđi gzlenmektedir. Ařađıdaki grafikte vaka ya da hasta sayılarının aıklanan deđerleri y ile model kestirim deđerleri \hat{y} arasındaki iliřki grlmektedir.



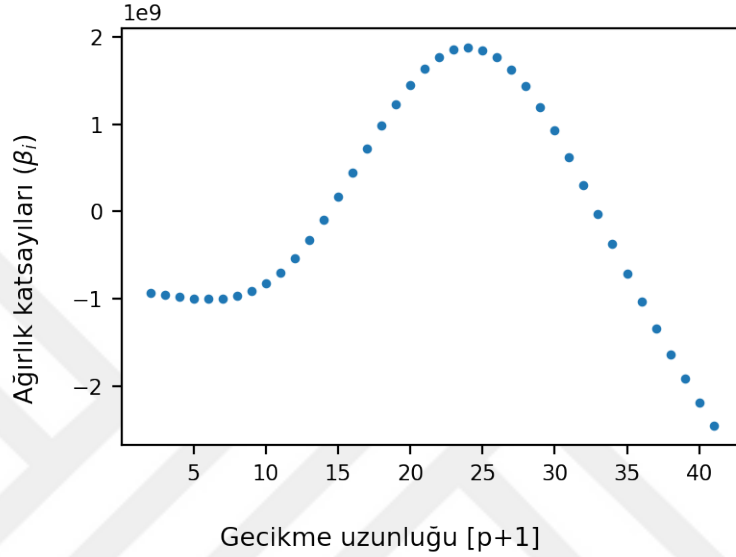
Şekil 6.3: MIDAS-NLS modeli vaka sayıları(y)ve kestirimi(\hat{y})

Yukarıdaki şekilde, bağımlı değişken y nokta, MIDAS-NLS modelinin tahmin değerleri \hat{y} kesintisiz çizgi ile gösterilmiştir. Şekil 6.3'te açıklanan ve tahmin edilen değerler arasındaki ilişki incelendiğinde, pandeminin başlangıç evresinde, toplumda COVID-19 konularının Twitter'da yoğun bir gündem oluşturduğu gözlenmektedir. Buradaki tahmin değerleri(\hat{y}), salgının yayılım hızına yönelik erken uyarı niteliğine sahiptir. Veri tanımında yapılan değişikliğin ardından, 10 Aralık 2020 tarihinden sonra, modelin ürettiği vaka sayısı tahminleri (\hat{y}) açıklanan vaka sayılarına yakınsamaktadır. Özellikle salgının erken dönemlerinde, MIDAS-NLS modelinin, karar alıcılara kamu sağlığını yönetmek için, önemli destek sağlayan bilgiler verebileceğine işaret etmektedir. Pandeminin ilerleyen dönemlerinde ise, MIDAS-NLS modelinin tahmin değerleri, açıklanan vaka sayılarının altında kalmıştır. Bu durumun, sosyal medyada olumlu haberlerin daha az, negatif haberlerin ise daha fazla gündem oluşturmasından kaynaklandığı düşünülmektedir.

Benzetim deneyi sonuçlarında tartışılan, parametrik MIDAS-NLS'in büyük gecikme uzunluklarında ve artan-azalan veri yapılarında örüntü oluşturma yeteneğinin zayıf kalması durumu, uygulama sonuçlarında da kendini göstermektedir.

6.4 Parametrik olmayan MIDAS-SLS modeli sonuçları

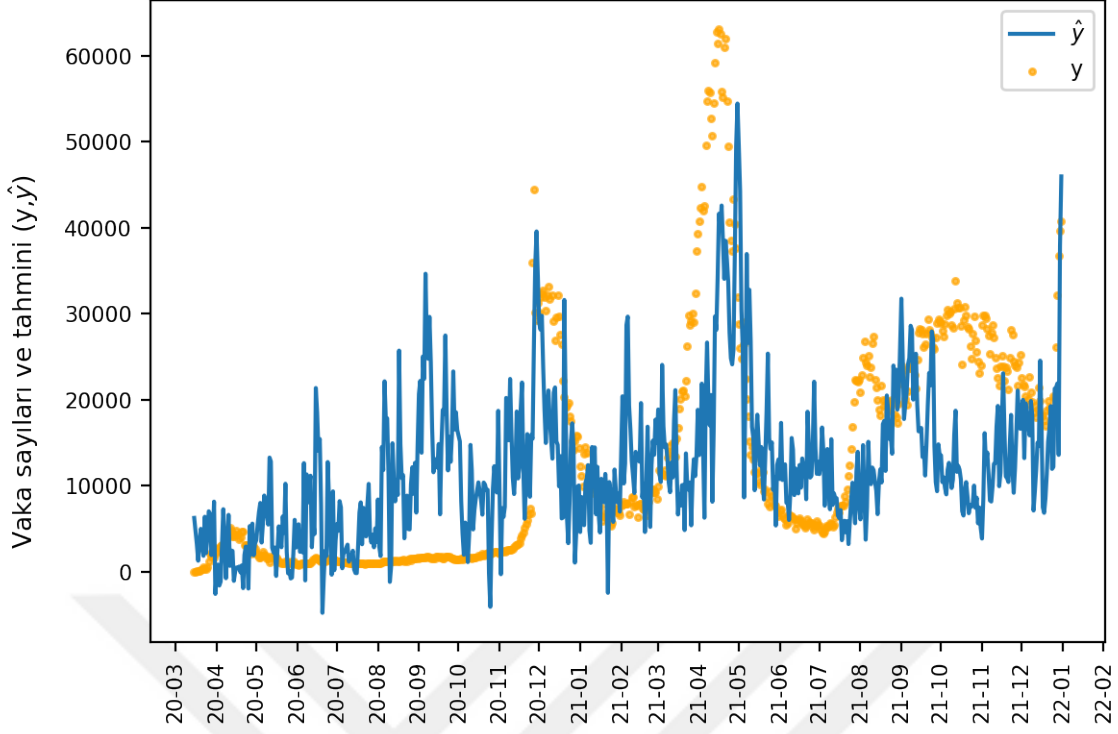
MIDAS-SLS modeli, cezalandırılmış splayn regresyon modelini baz almaktadır. Öncelikle bilgi ölçütü AICc değerini minimum yapan en uygun bant genişliği (λ) değeri ve gecikme uzunluğu parametre kestirimi yapılmıştır. En uygun parametre kestirimi için, Nelder-Mear sayısal çözümleme yöntemi kullanılmıştır. En uygun bant genişliği (λ) ve gecikme uzunluğu elde edilmesinin ardından, ağırlık katsayıları vektörü β_i hesaplanmıştır.



Şekil 6.4: MIDAS-SLS modeli ağırlık işlevi kestirimi

Modelin, bilgi ölçütü AICc= 22.88 ve en uygun bant genişliği parametresi $\lambda = 0.004618$ 'yü olarak hesaplanmıştır. MIDAS-SLS modelinde, parametrik olmayan ağırlık işlevi, yüksek frekanslı değişken için 40 saatlik bir gecikme uzunluğu hesaplanmıştır. MIDAS-SLS modelinin kestirdiği, ağırlık işlevinde gecikme uzunluğu ve ağırlık katsayıları arasındaki ilişki şekil 6.4'de sunulmuştur.

Aşağıda şekil 6.5'de MIDAS-SLS modelinden elde edilen (\hat{y}) ile açıklanan veri setinde yer alan bağımlı değişken (y)'nin değerleri arasındaki ilişki görülmektedir. Şekilde vaka sayıları (y) noktalar halinde, kestirim değerleri (\hat{y}) ise çizgiyle gösterilmiştir.



Şekil 6.5: MIDAS-SLS modeli vaka sayıları(y) ve kestirimi(\hat{y})

MIDAS-SLS modelinin uyum ve tahmin etme etkinliği, MIDAS-NLS'e göre daha yüksektir. Benzetim deneyinde tartışılan, büyük gecikme uzunlukları ve artan-azalan veri yapılarında, parametrik olmayan MIDAS-SLS yönteminin, parametrik MIDAS-NLS yöntemine göre daha etkin olduğu tespiti, uygulama bölümünde de teyit edilmiştir.

6.5 Sonuçlar üzerine değerlendirme

Pandeminin ilk dönemlerinde, COVID-19 vaka verileri yüksek belirsizlik (Ozdinc, Senel, Ozturkcan ve Akgul, 2020) ve gürültü faktörü içerir (Senel, Ozdinc ve Ozturkcan, 2021). Bu dönemde, hem vakaların tespit edilmesi için kullanılan tanı testlerinin yaygın bulunamayışı, hem de veri tabanlarının henüz geliştiriliyor olması etkili olabilmektedir (Senel, Ozdinc, Ozturkcan ve Akgul, 2020). Bilhassa solunum yoluyla yayılan virüslerin çok hızlı salgına dönüşme potansiyeli barındırması, erken evrelerde bu salgınların izlenmesini için yeterli veri oluşturulamaması sorunlara sebep olabilir. Öte yandan, alternatif bir yaklaşımla, gündeme dair sosyal medya paylaşımlarını baz alarak, vaka sayılarına dair kestirimler yapabilmek, eksik ve gürültülü içeren veri dönemine ilişkin açıklayıcı bilgiler sunabilir. MIDAS regresyon yöntemiyle elde edilen tahminler, halk sağlığı yönetiminde karar alıcılara tedbir geliştirip, hayati kararlara alabilecekleri önemli bir süre öncesinden uyarı verebilir. MIDAS-NLS ile 27

saat, MIDAS-SLS ile ise 40 saat öncesinden vaka sayılarının günden güne değişimine dair erken bilgi üretilebilmesi söz konusudur.

Uygulama üzerinden elde edilen sonuçlar, benzetim deneyinde gözlenen sonuçlara paralel bir seyir göstermektedir. Parametrik olmayan MIDAS-SLS modelinin, parametrik MIDAS-NLS modeline göre kestirim performansının daha yüksek olduğu gözlenmektedir. Uygulama verilerinde, yüksek frekanslı değişkenin uzun gecikme uzunluklarına sahip olması, artan azalan yapısı ve negatif-pozitif değerler içermesi, parametrik olmayan MIDAS-SLS modelinin daha başarılı sonuç üretmesi mümkün kılmaktadır. MIDAS-SLS modelinin uygulama veri seti üzerindeki performansı, COVID-19 vaka sayılarındaki değişimlerin uzun zaman öncesinden şimdinin tahmininin (nowcasting) yapabilmesi imkanını işaret etmektedir.



Bölüm 7

Sonuç ve Tartışma

MIDAS regresyon yöntemleri, yüksek frekanslı bağımsız değişkenin gecikmeli değerlerini ve değişkenliğini, kurulan modele yansıtılabilen bir ağırlıklandırma işleviyle dönüştürerek, bağımlı değişkenin boyutuyla eşleştirir. Yüksek frekanslı serinin, ağırlıklandırma işlevleriyle dönüştürülmesi, değişkenler arasındaki neden-sonuç ilişkisinin daha doğru kurulmasına olanak sağlamaktadır. Neden-sonuç ilişkilerini açıklayabilme yeteneğinin artması, yüksek frekanslı serinin değişkenliğinin ve gecikmeli değerlerinin, modele yansıtılabilmesiyle mümkün olmaktadır.

Yüksek frekanslı değişkenler yoluyla düşük frekanslı değişkenlerin modellenmesi, hem geleceğin tahmin edilmesi (forecasting), hem de şimdinin tahmininin (nowcasting) yapılabilmesine olanak sağlar. Şimdinin tahmini, henüz değeri bilinmeyen düşük frekanslı değişkenin, ilişki içinde olduğu düşünülen yüksek frekanslı değişkenlerin gecikmeli değerleriyle tahmin edilmesini sağlar. Özellikle ulusal gelir, işsizlik, fiyat endeksleri gibi değişkenler, derlenme zamanlarının uzunluğu nedeniyle, içinde yaşanılan zaman hakkında değil, geçmiş dönem hakkında bilgi verebilirler. Oysa, iktisat politikası oluşturulabilmesi için, bu değişkenlerin şimdiki durumunun tahmini gereklidir. MIDAS regresyon yöntemleri, şimdinin tahmininde bu nedenlerle önem kazanmış ve yaygın kullanıma ulaşmıştır.

Model parametrelerinin kestirimi, parametrik ya da parametrik olmayan MIDAS regresyon yöntemleriyle yapılabilir. Ağırlıklandırma işlevinin şeklinin bilindiği durumda parametrik MIDAS regresyon yöntemleri kullanılır. Eğer yüksek frekanslı serinin dağılımına ilişkin ön bilgi bulunmuyor ya da parametrik ağırlıklandırma işlevleri yeterli sonuçlar üretilmiyorsa, parametrik olmayan yöntemlere başvurulur.

Parametrik MIDAS regresyon yöntemleri, örneklem hacminin küçük ve yüksek frekanslı değişkenin gecikme uzunluklarının az olduğu durumlarda etkin çalışırlar. Parametrik yöntem, yüksek frekanslı seride önceden yapısı bilinen bir ağırlıklandırma işlevinin varlığına dayanır. Uygulamalarda, ağırlıklandırma için, Üssel Almon ve Beta işlevlerinin kullanımları yaygındır. Parametrik ağırlıklandırma işlevleri, yüksek frekanslı serinin gecikme katsayılarını, az sayıda parametreye indirgeyen yalınlaştırma görevi görürler. Parametrik ağırlıklandırma işlevleri, genelde gecikme

uzunluğu arttıkça azalan bir yapıdadırlar. Bu nedenle, kısa vadeli tahmin üretmekte başarılı sonuçlar verebilirler.

Parametrik MIDAS regresyon modelleri, genellikle doğrusal olmayan EKK (MIDAS-NLS) formundadır. MIDAS-NLS modellerinde parametre kestirimi, sayısal en uygun değer arayışı ile yapılmaktadır. Bu nedenle yeterli iterasyon uzunluğu yanında, seçilen parametrik ağırlıklandırma işlevinin formu da önemlidir. Eğer işlevin birinci ya da ikinci türevi alınabilir bir formda ise sayısal çözümlemede işlem süresi de azaltılabilir.

Parametrik olmayan MIDAS regresyon yöntemi, ağırlıklandırma işlevinin şekli hakkında önsel bilgi yoksa kullanılır. Ayrıca, yüksek frekanslı değişkenin gecikmeli değerleri uzunsa ve örneklem hacmi yeterliyse, parametrik MIDAS regresyon yöntemine göre daha başarılı kestirimler yapılabilir. Parametrik olmayan MIDAS regresyon yöntemi belli durumlarda daha iyi performans gösterebilir: (i) yüksek frekanslı değişkenin uzun vadeli gecikme değerlerinin kısa vadeli etkilerden daha büyük olduğu; (ii) artan-azalan; ve (iii) negatif-pozitif değerler alabildiği durumlar.

Bu tez kapsamında, MIDAS-NLS ve MIDAS-SLS modelleri üzerine benzetim deneyi tasarlanarak, uygulanmıştır. Benzetim deneyi sonuçlarına göre, MIDAS-SLS modeli uzun gecikme değerlerinde, artan-azalan ve negatif-pozitif değerler alabilen veri örüntülerinde, parametrik MIDAS-NLS yöntemine göre, model seçiminde daha yüksek başarı göstermiştir. Ancak, MIDAS-SLS modeli bu performansı sergilerken daha geniş bir örneklem hacmine ihtiyaç göstermektedir. Öte yandan, MIDAS-NLS ise daha küçük örneklem hacminde ve daha kısa vadeli gecikme uzunluklarında daha başarılı model seçimleri yapabilmektedir. Benzetim deneyi göreceli performans bulgularına göre, yüksek frekanslı değişkende gecikme uzunluğu arttıkça, MIDAS-NLS'in başarı oranı azalmaktadır.

Parametrik olmayan MIDAS regresyon yöntemlerine ilişkin tartışmalar ve geliştirmeler literatürde devam etmektedir. Parametrik MIDAS regresyon yöntemini kullanmak, yazılım paketleri sayesinde kolaylaşarak yaygınlaşmıştır. Parametrik yöntemde, sonuçları değerlendirmek ve model seçimi görece daha hızlı ve kolay yapılır. Öte yandan, parametrik olmayan MIDAS regresyon yöntemlerini kullanmak çeşitli zorluklar içerir. Parametrik olmayan model kurulumu kodlama gerektirir. Ek olarak, model kestirimlerinin yapılması iteratif işlemler gerektirdiğinden, sonuçlara ulaşmak hem maliyetlidir hem de daha uzun işlem zamanı ve yüksek donanım kapasitesi gerektirir.

Büyük veri (big data) ve akışkan veri (stream data) içeren durumlarda, parametrik olmayan MIDAS regresyon yöntemlerini kullanmak, daha iyi yapılandırılmış modelleri mümkün kılar. Akışkan verilerde, ağırlıklandırma işlevinin bilinmemesi ya da değişkenlik göstermesi nedeniyle parametrik olmayan yöntemler daha uygundur.

Bu tez kapsamında, MIDAS-NLS ve MIDAS-SLS modelleri, COVID-19 pandemisi çerçevesinde bir örnek durum seçilerek, uygulama verisiyle sınanmıştır. Örnek durum uygulamasında, Türkiye'de kaydedilmiş COVID-19 vaka (hasta) sayılarıyla, özellikle salgının ilk dönemlerinde, sosyal medyada toplumun COVID-19 konusunu tartışma sıklığı üzerinden, parametrik olmayan yöntemlerle etkin

kestirimlerin yapılabilirdiği gözlenmiştir. Geliştirilen modelin şimdiki durum tahmin etme yeteneđi, salgın hastalık gibi kriz durumlarının, ilk aşamalarında, halk sađlığı sisteminin yaklaşık 40 saat kadar erkenden uyarılmasını mümkün kılabilmektedir. Böylece kaynakların ve sürecin daha iyi yönetilebilmesi olanađı doğur.

Parametrik olmayan MIDAS-SLS modeli, splayn regresyonu temel alır. Çekirdek regresyon yöntemleri Nadaraya-Watson, Priestly-Choa ve Gasser-Müller yöntemleri, akışkan verinin analizinde önemli işlevler yerine getirebilirler. Çekirdek regresyon yöntemleriyle, MIDAS regresyon modeli oluşturmak mümkün görünmekle birlikte, bir dizi kuramsal ve uygulanabilirlik açısından çözülmesi gereken problem vardır. İleri dönem çalışmalarında, bu alanda kullanılabilir yöntemler ortaya konulabilir. Özellikle akışkan verinin zaman aralıklarındaki düzensizlik, çekirdek yoğunluk işlevleriyle dönüştürülebilir. Ancak, uygulanacak çekirdek MIDAS regresyonu için model seçim kriterlerinin uyarlanması da ihtiyaç vardır. Öte yandan, çekirdek regresyon yöntemleri ağır hesaplama yükü içerirler. Yüksek frekanslı deđişkenin gecikmeli deđerlerinin modele dahil olması, boyut sayısını artmasına yol açar. Bu da, her bakımdan yönetilmesi ve analiz edilmesi güç bir durum yaratabilmektedir. Parametrik olmayan MIDAS regresyon yönteminin, çekirdek regresyon yöntemlerine uyarlanması, gelecek çalışmalar açısından önem arz etmektedir.

Ek A

Kodlamalar

A.1 Parametrik Ağırlıklandırma İşlevleri

A.1.1 Almon Polinomu İşlevi

```
import numpy as np

def wf_almon_poly(coeffs, lags):
    nLag=len(lags)
    nCoeffs=len(coeffs)

    arr = [[i**j for j in range(1,nCoeffs)] for i in range(1,nLag+1)]
    plc = np.dot(np.matrix(arr), coeffs[1:]) + coeffs[0]

    return np.ravel(plc)
```

A.1.2 Üssel Almon İşlevi

```
def wf_almon_poly(coeffs, lags):
    nLag=len(lags)
    nCoeffs=len(coeffs)

    arr = [[i**j for j in range(1,nCoeffs)] for i in range(1,nLag+1)]
    plc = np.dot(np.matrix(arr), coeffs[1:]) + coeffs[0]

    return np.ravel(plc)
```

A.1.3 Beta İşlevi

```
def wf_beta(coeffs, lags):
    d = len(lags)
```

```

b = np.concatenate((coeffs,0), axis=None)
eps= sys.float_info.epsilon
xi = lags/(len(lags) - 1)
xi[0] = xi[0] + eps
xi[d-1] = xi[d-1]- eps
nb = np.power(xi, (b[1] - 1)) * np.power((1 - xi), (b[2] - 1))

if(np.sum(nb) < eps) :
if(np.abs(p[3])<eps) :
return np.repeat(0,b, axis=0)
else :
return p[0] * np.repeat(1/d,length(nb-1), axis=0)
else :
w = (nb / sum(nb)) + b[3]
return b[0] * w / np.sum(w)

```

A.1.4 Adımlama İşlevi

```

def step(p, d, a):
b = np.concatenate((0,a,d), axis=None)
return (np.repeat(p,np.diff(b), axis=0))

```

A.2 Rassal veri üretim süreci

A.2.1 Rassal Veri Seti Üretimi

```

import numpy as np
import pandas as pd
from scipy.optimize import minimize
from statsmodels.tsa.arima_process import ArmaProcess
import glob
import midas_funcs as mf
import comvar as cm

files = glob.glob('./data/*')
for f in files:
os.remove(f)

ar1 = np.array([1, -0.9])
ma1 = np.array([1, 0])
AR_object1 = ArmaProcess(ar1, ma1)

T = cm.T
m = cm.m
lags = cm.lags
NumOfSims = cm.NumOfSims

for numSim in trange(1,NumOfSims+1):
#X_hf = mf.RDG_ar1V3(T * m, 1)
X_hf = AR_object1.generate_sample(nsample=T*m)
#X_hf = np.random.randn(T*m)

```

```

X_lg = mf.dw_H2Lg(X_hf,m,lags)
u_t = (np.random.normal(0, 0.125, X_lg.shape[0])).reshape((X_lg.shape[0],1))

for wfunc in range(len(cm.wFuncs)):
    wname = list(cm.wFuncs.values())[wfunc]
    theta = list(cm.Coeffs.values())[wfunc]
    w_j = mf.wBeta(wname, lags, theta)
    beta_j = cm.alfal*w_j
    beta_0 = 0.5
    y = beta_0 + X_lg.dot(beta_j) + u_t
    X_lf = mf.dw_Lg2L(X_lg,cm.m)
    yX = np.concatenate((y,X_lf),axis=1)
    np.savetxt('./data/'+str(list(cm.wFuncs)[wfunc])+'_sim'+str(numSim)+'.csv',yX, delimiter=',')

```

A.2.2 Veri Üretimi İçin Ağırlıklandırma İşlevi Modelleri

A.2.2.1 Üssel Azalan Ağırlıklandırma İşlevi

```

def wf_exp_decline(coeffs, lags):
    nLag=len(lags)
    nCoeffs=len(coeffs)

    arr = [[i**j for j in range(1,nCoeffs+1)] for i in range(1,nLag+1)]
    plc = np.dot(np.matrix(arr), coeffs[0:])

    return np.ravel(np.exp(plc)/np.sum(np.exp(plc)))

def wf_hump_shaped(coeffs, lags):
    nLag=len(lags)
    nCoeffs=len(coeffs)

    arr = [[i**j for j in range(1,nCoeffs+1)] for i in range(1,nLag+1)]
    arr1=np.dot(np.matrix(arr),np.matrix('1 0; 0 -1'))
    plc = np.dot(arr1, coeffs[0:])

    return np.ravel(np.exp(plc)/np.sum(np.exp(plc)))

```

A.2.2.2 Tümsek Ağırlıklandırma İşlevi

```

def wf_hump_shaped(coeffs, lags):
    nLag=len(lags)
    nCoeffs=len(coeffs)

    arr = [[i**j for j in range(1,nCoeffs+1)] for i in range(1,nLag+1)]
    arr1=np.dot(np.matrix(arr),np.matrix('1 0; 0 -1'))
    plc = np.dot(arr1, coeffs[0:])

    return np.ravel(np.exp(plc)/np.sum(np.exp(plc)))

```

A.2.2.3 Doğrusal Azalan Ağırlıklandırma İşlevi

```
def wf_linear_weight(coeffs, lags):
    P = np.max(lags+1)
    nlag = len(lags)

    a0 = coeffs[0]
    a1 = coeffs[1]

    arr=[]
    for p in range(0,P):
        arr.append((a0+a1*p) /a0*P+a1*(P-1)*P/2)
    norm = [-1 + 2 * (x - min(arr)) / (max(arr) - min(arr)) for x in arr]
    return norm
```

A.2.2.4 Çevrimsel Ağırlıklandırma İşlevi

```
def wf_cyclical(coeffs, lags):
    p =np.max(lags)
    c1=coeffs[0]
    c2=coeffs[1]
    arr=[]
    for j in range(0,p+1):
        arr.append((c1/(p+1))*(np.sin(c2+(j*2*(np.pi)/p))))
    plc = arr/np.sum(arr)
    norm = [-1 + 2 * (x - min(arr)) / (max(arr) - min(arr)) for x in arr]
    return norm
```

Kaynakça

- Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*, 33(1), 178. doi: 10.2307/1911894
- Andreou, E., Ghysels, E. ve Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2), 246–261. doi: 10.1016/j.jeconom.2010.01.004
- Andreou, E., Ghysels, E. ve Kourtellos, A. (2013). Should Macroeconomic Forecasters Use Daily Financial Data and How? *Journal of Business & Economic Statistics*, 31(2), 240–251. doi: 10.1080/07350015.2013.767199
- Başarı, G. (2001). *Uygulamalı Parametrik Olmayan İstatistiksel Yöntemler*. Mimar Sinan Üniversitesi Fen-Edebiyat Fakültesi, İstanbul.
- Bai, J., Ghysels, E. ve Wright, J. H. (2013). State Space Models and MIDAS Regressions. *Econometric Reviews*, 32(7), 779–813. doi: 10.1080/07474938.2012.690675
- Baltagi, B. H. (2011). *Econometrics* (Fifth ed.). Berlin Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-20059-5
- Breitung, J., ve Roling, C. (2015). Forecasting Inflation Rates Using Daily Data: A Nonparametric MIDAS Approach. *Journal of Forecasting*, 34(7), 588–603. doi: 10.1002/for.2361
- Chen, E., Lerman, K. ve Ferrara, E. (2020). Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, 6(2), e19273.
- Chow, G. C., ve Lin, A.-I. (1971). Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics*, 53(4), 372–375. doi: 10.2307/1928739
- Clements, M., ve Galvão, A. (2008). Macroeconomic Forecasting With Mixed-Frequency Data. *Journal of Business & Economic Statistics*, 26, 546–554. doi: 10.1198/073500108000000015
- Clements, M. P., ve Galvão, A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, 24(7), 1187–1206. doi: 10.1002/jae.1075
- Davidson, R., ve MacKinnon, J. (1999). *Econometric theory and methods*. Oxford University Press.
- Dempster, A. P., Laird, N. M. ve Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x

- Dong, E., Du, H. ve Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases*, 20(5), 533–534. doi: 10.1016/S1473-3099(20)30120-1
- Foroni, C., Marcellino, M. ve Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 178(1), 57–82. doi: 10.1111/rssa.12043
- Foroni, C., ve Marcellino, M. G. (2013). *A Survey of Econometric Methods for Mixed-Frequency Data* (Tech. Rep. No. ID 2268912). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.2268912
- Forsberg, L., ve Ghysels, E. (2007). Why Do Absolute Returns Predict Volatility So Well? *Journal of Financial Econometrics*, 5(1), 31–67. doi: 10.1093/jfinec/nbl010
- Gasser, T., ve Müller, H.-G. (1979). Kernel estimation of regression functions. In T. Gasser ve M. Rosenblatt (Eds.), *Smoothing techniques for curve estimation* (pp. 23–68). Springer. Retrieved from 10.1007/BFb0098489 doi: 10.1007/BFb0098489
- Gasser, T., ve Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian journal of statistics*, 171–185.
- Ghysels, E., Kvedaras, V. ve Zemlys, V. (2016). Mixed Frequency Data Sampling Regression Models: The R Package midasr. *Journal of Statistical Software*, 72(4), 1–35. doi: 10.18637/jss.v072.i04
- Ghysels, E., ve Marcellino, M. (2018). *Applied economic forecasting using time series methods*. Oxford University Press.
- Ghysels, E., ve Qian, H. (2019). Estimating MIDAS regressions via OLS with polynomial parameter profiling. , 9, 1–16. Retrieved 2020-02-19, from <http://www.sciencedirect.com/science/article/pii/S2452306218300066> doi: 10.1016/j.ecosta.2018.02.001
- Ghysels, E., Santa-Clara, P. ve Valkanov, R. (2004). *The MIDAS Touch: Mixed Data Sampling Regression Models*. CIRANO.
- Ghysels, E., Santa-Clara, P. ve Valkanov, R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1), 59–95. doi: 10.1016/j.jeconom.2005.01.004
- Ghysels, E., Sinko, A. ve Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1), 53–90. doi: 10.1080/07474930600972467
- Ghysels, E., Valkanov, R. I. ve Rubia, A. (2009). Multi-Period Forecasts of Volatility: Direct, Iterated, and Mixed-Data Approaches. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1344742
- Giannone, D., Reichlin, L. ve Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665–676. doi: 10.1016/j.jmoneco.2008.05.010
- Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.
- Hardle, W., Müller, M., Sperlich, S. ve Werwatz, A. (2012). *Nonparametric and semiparametric models*. Springer.
- Hendry, D. F., ve Mizon, G. E. (1978). Serial Correlation as a Convenient Simplification, not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England. *Economic Journal*,

- 88(351), 549–563.
- Hutchinson, M. F., ve de Hoog, F. R. (1985). Smoothing noisy data with spline functions. *Numerische Mathematik*, 47(1), 99–106. doi: 10.1007/BF01389878
- Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. ve Lee, T.-C. (1988). *Introduction to the theory and practice of econometrics*. J. Wiley.
- Kalman, R. (1960). A new approach to linear filtering and prediction problems" transaction of the asme-journal of basic.. doi: 10.1115/1.3662552
- Kapetanios, G., Marcellino, M. G. ve Papailias, F. (2014). *Variable Selection for Large Unbalanced Datasets Using Non-Standard Optimisation of Information Criteria and Variable Reduction Methods* (Tech. Rep. No. ID 2444418). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.2444418
- Kuzin, V., Marcellino, M. ve Schumacher, C. (2013). Pooling Versus Model Selection for Nowcasting Gdp with Many Predictors: Empirical Evidence for Six Industrialized Countries. *Journal of Applied Econometrics*, 28(3), 392–411. doi: 10.1002/jae.2279
- Ozdinc, M., Senel, K., Ozturkcan, S. ve Akgul, A. (2020). Predicting the progress of covid-19: the case for Turkey. *Turkiye klinikleri - journal of medical sciences*, 40(2), 117.
- Parker, J. A. (2021). *Distributed-Lag Models*. Retrieved 2021-02-10, from https://www.reed.edu/economics/parker/312/tschapters/S13_Ch_3.pdf
- Priestley, M. B., ve Chao, M. (1972). Non-parametric function fitting. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(3), 385–392.
- Roling, C. (2014). *Three Essays in Econometrics* (Unpublished doctoral dissertation). Universitäts- und Landesbibliothek Bonn.
- Ruppert, D., Wand, M. P. ve Carroll, R. J. (2003a). *Semiparametric Regression* (No. 12). Cambridge University Press.
- Ruppert, D., Wand, M. P. ve Carroll, R. J. (2003b). *Semiparametric Regression*. Cambridge University Press. doi: 10.1017/CBO9780511755453
- Schumacher, C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting*, 32(2), 257–270. doi: 10.1016/j.ijforecast.2015.07.004
- Seabold, S., ve Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th python in science conference*.
- Senel, K., Ozdinc, M. ve Ozturkcan, S. (2021). Single parameter estimation approach for robust estimation of sir model with limited and noisy data: The case for COVID-19. *Disaster medicine and public health preparedness*, 15(3), e8–e22.
- Senel, K., Ozdinc, M., Ozturkcan, S. ve Akgul, A. (2020). Instantaneous r for covid-19 in Turkey: Estimation by bayesian statistical inference. *Turkiye klinikleri - journal of medical sciences*, 40(2), 127.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.
- Sinko, A., Sockin, M. ve Ghysels, E. (2010). *Matlab Toolbox for Mixed Sampling Frequency Data*

Analysis Using Midas Regression Models.

- Stock, J. H., ve Watson, M. W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*, 20(2), 147–162. doi: 10.1198/073500102317351921
- Stock, J. H., ve Watson, M. W. (2003). How did leading indicator forecasts perform during the 2001 recession? *SSRN Electronic Journal*. Retrieved from <https://papers.ssrn.com/abstract=2184936>
- Toker, S., Özbay, N. ve Månsson, K. (2022). Mixed data sampling regression: Parameter selection of smoothed least squares estimator. *Journal of Forecasting*, 41(4), 718–751. doi: 10.1002/for.2828
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272. doi: 10.1038/s41592-019-0686-2
- Zhang, R. (2007). *Model selection techniques for kernel-based regression analysis using information complexity measure and genetic algorithms.*