

**BİRİNCİ DERECE DEN TAM SAYI DEĞERLİ
OTOREGRESYON MODELLERİNİN İNCELENMESİ**

**ANALYSIS OF FIRST-ORDER INTEGER-VALUED
AUTOREGRESSIVE MODELS**

HATİCE NUR KARAKAVAK

PROF. DR. CEM KADILAR

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı için Öngördüğü

YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

ÖZET

BİRİNCİ DERECEDEDEN TAM SAYI DEĞERLİ OTOREGRESYON MODELLERİNİN İNCELENMESİ

Hatice Nur KARAKAVAK

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Prof. Dr. Cem KADILAR

Kasım 2023, 65 sayfa

Poisson dağılımı genellikle tam sayı değerli otoregresyon modellerinde hataların dağılımı için kullanılır. Ancak, dağılımın ortalaması varyansına eşittir ve eşit yayılan seriler için uygundur. Gerçek tam sayı değerli zaman serileri genellikle aşırı veya yetersiz yayılmaktadır. Bu nedenle, literatürde Binom inceltmesine dayalı, farklı dağılımlara sahip hatalarla birçok birinci dereceden tam sayı değerli otoregresyon (INAR(1)) modelleri geliştirilmiştir. Bu tez çalışması kapsamında, tam sayı değerli zaman serilerinin modellenmesi için literatürde yer alan INAR(1) modelleri ayrıntılı olarak incelenmiştir. İncelenen modeller biri aşırı, diğeri yetersiz yayılım gösteren iki farklı gerçek tam sayı değerli zaman serisi üzerinde belirli kriterler ile karşılaştırılmış ve seriler için en uygun INAR(1) modelleri belirlenmiştir.

Anahtar Kelimeler: Zaman serileri, Tam sayı değerli zaman serileri, Birinci dereceden tam sayı değerli otoregresyon modelleri, INAR(1)

ABSTRACT

ANALYSIS OF FIRST-ORDER INTEGER-VALUED AUTOREGRESSIVE MODELS

Hatice Nur KARAKAVAK

Master of Science, Department of Statistics

Supervisor: Prof. Dr. Cem KADILAR

November 2023, 65 pages

The Poisson distribution is often used for the distribution of errors in integer-valued autoregressive models. However, the mean of the distribution is equal to its variance and is suitable for equally distributed series. Real integer-valued time series are often over- or under-dispersion. Therefore, many first-order integer-valued autoregressive (INAR(1)) models based on binomial thinning with differently distributed errors have been developed in the literature. In this thesis, INAR(1) models in the literature for modeling integer-valued time series are examined in detail. The models are compared with certain criteria on two different real integer-valued time series, one of which is overdispersion and the other is underdispersion, and the most appropriate INAR(1) models are determined for the series.

Keywords: Time series, Integer valued time series, First order integer valued autoregressive models, INAR(1)

TEŐEKKÜR

Tez alıőmam sũresince destek ve katkılarını esirgemeyen, deneyim ve bilgileriyle her zaman yol gũsteren, tecrũbelerinden faydalandıđım deđerli danıőmanım ve hocam Sayın Prof. Dr. Cem Kadılar'a,

alıőmamın her aőamasında yanımda olan, yardım ve destekleriyle beni teővik eden deđerli hocam Sayın Do. Dr. Hatice Őncel ekim'e,

Tezimde bazı verilerinden yararlandıđım ve bursiyer olarak gũrev yaptıđım, Sayın Prof. Dr. Gamze Őzel Kadılar yũrũtũcũlũđũnde devam eden 121F208 nolu "Yapay Zeka ve Olasılıksal Model Tabanlı Deprem Tehlike Haritası" baőlıklı 1001 projesi ile destek sađlayan TũBİTAK'a,

Projede yer almama vesile olan ve akademik aıdan desteklerini hibir zaman esirgemeyen deđerli hocam Sayın Prof. Dr. Gamze Őzel Kadılar'a,

Lisans ve yũksek lisans eđitimim boyunca beni bu alanda yetiőtiren Hacettepe Őniversitesi İstatistik bũlũmũndeki deđerli hocalarıma teőekkũrlerimi sunarım.

Hayatımın her alanında olduđu gibi tez alıőmam sũresince de yanımda olan, sevgi ve desteklerini her zaman hissettiđim annem Hafize Karakavak'a, babam Hasan Karakavak'a ve kardeőtım Őmer Karakavak'a sonsuz teőekkũr ederim.

Hatice Nur Karakavak

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER.....	iv
ŞEKİLLER DİZİNİ.....	vi
SİMGELER VE KISALTMALAR.....	ix
1. GİRİŞ	1
2.1. Birinci Dereceden Tam Sayı Değerli Otoregresyon (INAR(1)) Süreçleri.....	5
2.2. INAR(1) Sürecinin Dağılım Özellikleri.....	7
2.3. INAR(1) Süreçlerinde Parametre Tahmini	11
2.3.1. Yule-Walker Tahmini	11
2.3.2. Koşullu En Küçük Kareler Tahmini.....	12
2.3.3. Koşullu Maksimum Olabilirlik Tahmini.....	13
2.4. INAR(1) Süreçlerinde Model Tanımlama.....	14
2.5. INAR(1) Süreçlerinde Tahmin.....	15
2.6. INAR(1) Süreçlerinde Model Yeterliliği	16
3. LİTERATÜRDE YER ALAN BAZI INAR(1) MODELLERİ	18
3.1. P-INAR(1) Modeli	18
3.2. G-INAR(1) Modeli.....	20
3.3. PL-INAR(1) Modeli.....	23
3.4. NB-INAR(1) Modeli	26
3.5. DP-INAR(1) Modeli	29
3.6. GP-INAR(1) Modeli	32

3.7. ZIP-INAR(1) Modeli	35
4. GERÇEK VERİ UYGULAMALARI	39
4.1. Artçı Sayısı Verisi.....	40
4.2. Deprem Sayısı Verisi	51
5. SONUÇLAR.....	59
6. KAYNAKLAR	61
EKLER	64
EK 6 - Tez Çalışması Orjinallik Raporu.....	64
ÖZGEÇMİŞ	65

ŞEKİLLER DİZİNİ

Şekil 1: (a) $N(2,2)$ 'den üretilmiş sürekli parametrelili zaman serisi, (b) Bernoulli(0.5) ve Poisson(1) dağılımlarından üretilmiş kesikli parametrelili zaman serisi	6
Şekil 2: Uygulama alanı.....	39
Şekil 3: AFAD-DDA ve KOERI deprem katalogları arasında eşleşen depremlerin deprem büyüklükleri açısından karşılaştırmaları.....	40
Şekil 4: AFAD-DDA ve Tan (2021) deprem katalogları arasında eşleşen depremlerin deprem büyüklükleri açısından karşılaştırmaları.....	41
Şekil 5: AFAD-DDA kataloğu depremleri derinlik bilgisi.....	42
Şekil 6: KOERI kataloğu depremleri derinlik bilgisi.....	42
Şekil 7: Tan (2021) kataloğu depremleri derinlik bilgisi.....	42
Şekil 8: AFAD-DDA kataloğu depremleri oluş zamanları.....	42
Şekil 9: KOERI kataloğu depremleri oluş zamanları.....	42
Şekil 10: Tan (2021) kataloğu depremleri oluş zamanları.....	42
Şekil 11: Harita üzerinde uygulama verileri.....	43
Şekil 12: Artçı sayısı serisinin zaman serisi, ACF ve PACF grafikleri.....	44
Şekil 13: (a) Orijinal seri (b) logaritmik dönüşüm uygulanan seri için kutu grafikleri..	45
Şekil 14: Logaritmik dönüşüm uygulanan artçı sayısı serisinin zaman serisi, çubuk, ACF ve PACF grafikleri.....	45
Şekil 15: PL-INAR(1) modeli uygulanarak hesaplanan hataların ACF, PACF ve Q-Q grafikleri.....	49
Şekil 16: Logaritmik dönüşüm uygulanan seri ve tahmin serisinin birlikte grafiği.....	50
Şekil 17: Harita üzerinde uygulama verileri.....	51
Şekil 18: Deprem sayısı serisinin zaman serisi, ACF ve PACF grafikleri.....	52
Şekil 19: (a) Orijinal seri (b) logaritmik dönüşüm uygulanan seri için kutu grafikleri..	52
Şekil 20: Logaritmik dönüşüm uygulanan deprem sayısı verisinin zaman serisi, çubuk, ACF ve PACF grafikleri.....	53

Şekil 21: DP-INAR(1) modeli uygulanarak hesaplanan hataların ACF, PACF ve Q-Q grafikleri.....	57
Şekil 22: Logaritmik dönüşüm uygulanan seri ve tahmin serisinin birlikte grafiği.....	58

TABLULAR DİZİNİ

Tablo 1: Artçı sayısı serisi tanımlayıcı istatistikleri.....	41
Tablo 2: Artçı sayısı serisi için CML tahminleri, AIC ve BIC bilgi kriterleri, ortalama, varyans ve HKOK hesaplamaları.....	43
Tablo 3: Deprem sayısı serisi tanımlayıcı istatistikleri.....	49
Tablo 4: Deprem sayısı serisi için CML tahminleri, AIC ve BIC bilgi kriterleri, ortalama, varyans ve HKOK hesaplamaları.....	51

SİMGELER VE KISALTMALAR

Simgeler

\circ	Binom inceltme operatörü
Y_i	Bağımsız ve aynı dağılımlı rasgele değişkenlerin Bernoulli serisi
X_t	Negatif olmayan tam sayı değerli zaman serisi
ε_t	Birinci dereceden tam sayı değerli otoregresyon sürecine ait hata serisi
I	Yayılm endeksi
μ_ε	Hata serisinin ortalaması
σ_ε^2	Hata serisinin varyansı
μ_X	Sürecin ortalaması
σ_X^2	Sürecin varyansı
φ_k	Sürecin olasılık yaratıcı fonksiyonu
γ_k	Sürecin otokovaryans fonksiyonu
ρ_k	Sürecin otokorelasyon fonksiyonu

Kısaltmalar

INAR(1)	Birinci dereceden tam sayı deęerli otoregresyon
YW	Yule-Walker
CLS	Koşullu en küçük kareler
CML	Koşullu maksimum olabilirlik
ACF	Otokorelasyon fonksiyonu
PACF	Kısmi otokorelasyon fonksiyonu
MA	Hareketli ortalama
AR	Otoregresyon
ADF	Geliştirilmiş Dickey Fuller
Mw	Depremin moment büyüklüğü
HKOK	Hata kareler ortalamasının karekökü
AIC	Akaike bilgi kriteri
BIC	Schwarz Bayesyen bilgi kriteri

1. GİRİŞ

Zaman serileri analizi, eşit zaman aralıklarıyla ve kronolojik sırada düzenlenmiş zaman serilerindeki değişimi, trendleri, mevsimsel dalgalanmaları, aykırı değerleri modelleme ve geleceğe dair öngörülerde bulunma işlemlerini kapsayan stokastik bir süreçtir. Zaman serileri, verilerin kaydedilme şekline bağlı olarak sürekli zaman serileri ve kesikli zaman serileri olarak ikiye ayrılır. Düzenli ve eşit aralıklarla kaydedilebilen verilere kesikli zaman serileri denir ve genellikle zaman serileri analizi kesikli zaman serileri üzerinden ele alınır.

Zaman serileri analizi, istatistik, mühendislik, ekonomi, finans, sağlık, meteoroloji, sismoloji gibi birçok bilim dalında yaygın olarak kullanılır. Alışılmış stokastik süreçler sürekli marjinal dağılımları varsayar ve bu nedenle sayım verilerini modellemek için uygun değildir. Ancak, gerçek hayat problemlerinde genellikle olay veya nesne sayısı olarak tam sayı değerli zaman serileri ile sıklıkla karşılaşmaktadır.

Bir mağazadaki günlük müşteri sayısı, borsada yapılan günlük işlem sayısı, üretim esnasında karşılaşılan hatalı ürün sayısı, aylık yağışlı gün sayısı, hastanede yatan hasta sayısı veya bir hastanın gün içerisinde geçirdiği nöbetlerin sayısı, belirli bir bölgede meydana gelen depremlerin sayısı ya da orman yangını sayısı gibi birçok alanda elde edilen veriler tam sayı değerli zaman serilerine örnek olarak verilebilir.

Literatürde zaman serilerini modellemek ve öngöründe bulunmak için sıklıkla kullanılan ayrıştırma yöntemleri, regresyon analizi, üstel düzleştirme yöntemleri ve Box-Jenkins modelleri genellikle hisse senedi fiyatları, döviz kurları, atmosferik basınç, rüzgar hızı, kan basıncı ya da kalp atış hızı, sosyo-ekonomik göstergeler, belirli bir bölgede meydana gelen depremlerin büyüklük veya derinlik bilgileri gibi sürekli değerli elde edilebilen serileri analiz etmek için kullanılır ve tam sayı değerli zaman serilerinin istatistiksel analizi için uygun modelleme yöntemleri değildir.

Yukarıda bahsedilen modelleme yöntemleri tam sayı değerli zaman serilerinin analizinde sağlıklı sonuçlar vermeyeceğinden, tam sayı değerli zaman serilerinin

istatistiksel analizi için yeni yöntemlere ihtiyaç duyulmuştur. İlk başlarda bu alana çok ilgi gösterilmemiş olsa da 1980'li yıllarda ilk olarak McKenzie (1985) ve Al-Osh ve Alzaid (1987) tarafından Steutel ve Van Harn (1979) Binom inceltme (thinning) operatörüne dayalı birinci dereceden tam sayı değerli otoregresyon (INAR(1)) süreci tanıtılmıştır.

Daha sonraki yıllarda uygun modelleme yöntemleri için yapılan birçok girişim olsa da INAR modelleri tam sayı değerli zaman serilerinin modellenmesinde ana model olarak kabul edilmektedir. Ancak, INAR(1) modeli tam sayı değerli zaman serilerini modellemek için her zaman uygun olmadığından, literatürde INAR(1) modelindeki inceltme operatörü değiştirilerek veya modeldeki hataların dağılımı değiştirilerek birçok farklı yeni alternatif model tartışılmıştır.

Binom inceltme operatörüne (\circ) dayalı birinci dereceden tam sayı değerli otoregresyon sürecinin formu aşağıdaki gibidir:

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \quad t = 0,1,2, \dots \quad (1)$$

Burada, ε_t 'nin ortalaması μ_ε ve sonlu varyansı σ_ε^2 olarak ifade edilir ve ilişkisiz, negatif olmayan tam sayı değerli değişkenlerin hata serisidir. McKenzie (1985) ve Al-Osh ve Alzaid (1987) tarafından tanıtılan ilk modelde ε_t 'nin Poisson dağılımına sahip olduğu varsayılmaktadır. Poisson dağılımının ortalaması, esnekliği sınırlayan varyansına eşittir. Ancak uygulamalarda veriler her zaman eşit şekilde yayılmamaktadır. Bu nedenle, Poisson dağılımlı bir ε_t 'ye sahip INAR(1) modeli tam sayı değerli zaman serilerini modellemek için her zaman uygun bir model olmayabilir.

Ortalaması varyansından büyük olan yetersiz (underdispersion) yayılıma sahip veya varyansı ortalamasından büyük olan aşırı (overdispersion) yayılıma sahip verileri modellemek için literatürde farklı çözüm önerileri tartışılmıştır. Bunlardan ilki, INAR(1) modelindeki binom inceltme işlemini değiştirmektir. İkinci ve daha yaygın olarak karşımıza çıkan yöntem ise, INAR(1) modelini ilgilenilen seriye uygun hale getirmek için modeldeki ε_t 'lerin dağılımını değiştirmektir. Bu çalışma kapsamında da,

ε_t 'lerin dağılımı değiştirilerek geliştirilen INAR(1) modelleri incelenecektir. Aşağıda literatürde yer alan ve ayrıntılı olarak inceleyeceğimiz bazı INAR(1) modellerinden kısaca bahsedilmiştir.

Jung, Ronning ve Tremayne (2005), Negatif Binom dağılımının, aşırı yayılım olgusunu yakalamak için uygulamalı çalışmalarda en sık kullanılan dağılım olduğunu vurgulayarak, Negatif Binom hatalarına sahip INAR(1) modelinin aşırı yayılıma sahip seriler için uygun bir model olduğunu belirtmişlerdir. Daha sonrasında, Wamwea, Mwelu ve Odin (2022), Kenya'daki doğrulanmış COVID-19 enfekte vakaların kümülatif sayısını modellemek ve tahmin etmek için Negatif Binom hatalarıyla birinci dereceden tam sayı değerli otoregresyon modelinin performansını incelemişlerdir.

Jazi, Jones ve Lai (2012a), aşırı yayılmış tam sayı değerli zaman serilerini modellemek için Geometrik dağılıma sahip hatalarla INAR(1) sürecinin yeni bir alternatifini tanıtmışlardır. Çalışma kapsamında geliştirilen yeni sürecin bazı matematiksel özelliklerini sunmuş ve modelin klasik INAR(1)'e göre üstünlüğü bazı gerçek zaman serileriyle göstermişlerdir.

Aşırı yayılmanın sık görülen bir belirtisi, sıfır sayım sıklığının bir Poisson dağılımından beklenenden daha büyük olmasıdır (Jazi, Jones ve Lai, 2012b). Bu doğrultuda, Jazi, Jones ve Lai (2012b) Sıfır Yığmal Poisson hataları ile yeni bir birinci dereceden tam sayı değerli otoregresyon sürecini tanıtmışlardır. Sürecin önemli yapısal özelliklerini türettikten sonra, hayvan sağlığı laboratuvarı değerleri ile ilgili bazı gerçek zaman serilerini analiz ederek önerilen modelin üstünlüğünü göstermişlerdir.

Livio ve ark. (2018) Binom inceltme operatörüne dayalı Poisson-Lindley hataları ile birinci dereceden negatif olmayan tam sayı değerli bir otoregresyon sürecini tanıtmışlardır. Tanıtılan modelin, özellikle aşırı yayılım sergileyen zaman serileri için uygun olduğunu ve bu nedenle geliştirilen diğer modellerle rekabet edebileceğini bir uygulama ile göstermişlerdir.

Uygulamalarda tam sayı değerli zaman serileri, genellikle yetersiz yayılım veya aşırı yayılım olgusunu gösterir. Bourguignon, Rodrigues ve Santos-Neto (2019), tam sayı

değerli zaman serilerini eşit yayılım, yetersiz yayılım ve aşırı yayılım ile modellemek için Çift Poisson ve Genelleştirilmiş Poisson hataları ile birinci dereceden tam sayı değerli otoregresyon sürecinin iki farklı uzantısını geliştirmişlerdir. Önerdikleri modellerin üstünlüğünü, yetersiz ve aşırı yayılıma sahip iki farklı gerçek zaman serisi üzerinde göstermişlerdir.

Bu çalışma kapsamında, literatürde tartışılan tüm bu modeller, matematiksel eşitlikleri ve dağılım özellikleriyle ayrıntılı olarak incelenmiştir. Sonrasında, elde edilen gerçek tam sayı değerli zaman serileri için her bir model uygulanarak seriler için en uygun modeli belirleyen iki farklı uygulama sunulmuştur.

Bu doğrultuda, çalışmanın ikinci bölümde tam sayı değerli zaman serilerinin analiz ve modellenmesine ilişkin olarak birinci dereceden tam sayı değerli otoregresyon süreçleri, hataların dağılımından bağımsız bir şekilde incelenmiş ve süreç adımlarına yer verilmiştir.

Çalışmanın üçüncü bölümünde, birinci dereceden tam sayı değerli otoregresyon süreçlerindeki hataların dağılımları değiştirilerek elde edilen ve literatürde tartışılan alternatif INAR(1) süreçleri, matematiksel eşitlikleri ve dağılım özellikleriyle birlikte ayrıntılı olarak incelenmiştir.

Çalışmanın dördüncü bölümünde, üçüncü bölümde incelenen modeller, elde edilen iki farklı gerçek tam sayı değerli deprem serileri üzerinde uygulanarak her iki seri için de en uygun modele karar verilen uygulama bölümü sunulmuştur.

2. GENEL BİLGİLER

Bu bölümde tam sayı değerli zaman serilerinin analiz ve modellenmesine ilişkin Birinci Dereceden Tam Sayı Değerli Otoregresyon (INAR(1)) Süreçleri ayrıntılı olarak incelenecektir.

2.1. Birinci Dereceden Tam Sayı Değerli Otoregresyon (INAR(1)) Süreçleri

Tam sayı değerli zaman serilerini analiz etmek için yaygın kullanılan modellerden biri olan ve birinci dereceden tam sayı değerli otoregresyon süreci (McKenzie, 1985; Al-Osh ve Alzaid, 1987) olarak adlandırılan model, Binom inceltme (\circ) operatörüne (Stutel ve Van Harn, 1979) dayalı olarak tanıtılmıştır.

Negatif olmayan tam sayı değerli bir rasgele değişken X ve $\alpha \in (0,1)$ ile tanımlanan bir sabit verildiğinde, Binom inceltme (\circ) operatörü;

$$\alpha \circ X = \sum_{i=1}^X Y_i \quad (2)$$

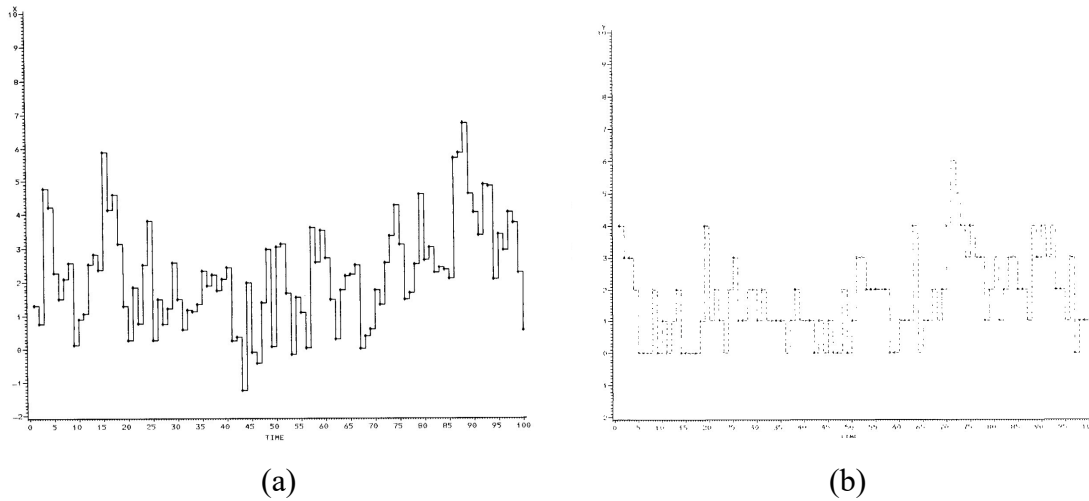
şeklinde tanımlanır. Burada, Y_i , $P(Y_i = 1) = 1 - P(Y_i = 0) = \alpha$ ile bağımsız ve aynı dağılımlı rasgele değişkenlerin bir Bernoulli serisidir ve X 'ten bağımsızdır. İnceltme (\circ) operatörünün tanımından, $\alpha \circ X$ sadece 0 ile X arasında tam sayı değerleri aldığı anlaşılmaktadır. Başka bir deyişle, α 'nın sınır değerleri 0 ve 1 için, $0 \circ X = 0$ ve $1 \circ X = X$ olmaktadır.

Binom inceltme işleminin daha iyi yorumlanabilmesi için, belirli bir t zamanında X büyüklüğünde bir kitle düşünelim. Aynı kitleyi daha sonra $t + 1$ zamanında gözlemlersek kitle büyüklüğü küçülmüş olabilir, çünkü değerlerin bir kısmı t ile $t + 1$ zamanları arasında eksilmiş olabilir. Eğer t 'den $t + 1$ 'e kitlede kalma olasılığı tüm değerler için α 'ya eşit ise kitlede kalanların sayısı $\alpha \circ X$ ile gösterilir (Weiß, 2018).

Bir gecikmeli bağımlılığa sahip tam sayı değerli durağan seriler için tasarlanan INAR(1) süreci, korelasyon yapısı ve dağılım özellikleri bakımından sürekli değerli AR(1) sürecine oldukça benzerdir. İstatistiksel kolaylık sağlamak için hata serisinin normal dağılıma sahip olduğu varsayılan sürekli değerli otoregresif hareketli ortalama (ARMA) sürecine benzer şekilde modellenmiştir (Al-Osh ve Alzaid, 1987).

İnceltme operatörüne dayalı INAR(1) süreci (1) numaralı eşitlikte verilmiştir (McKenzie, 1985; Al-Osh ve Alzaid, 1987). Burada, X_t , t zamanında kitlede incelenen olay sayısını, $\alpha \circ X_{t-1}$ ise $t - 1$ zamanında kitlede incelenen olay sayısını temsil eder. Süreçte inceltme işlemleri birbirlerinden ve ε_t 'den bağımsız olarak her t zamanında yeniden gerçekleşmektedir.

Al-Osh ve Alzaid (1987), INAR(1) süreci ile AR(1) süreci arasındaki farkı gözlemlemek için, Binom inceltme (\circ) operatörü tanımında yer alan Y_i 'nin $\alpha = 0.5$ ile Bernoulli rasgele değişkenleri olarak alındığı ve ε_t 'nin $\lambda = 1.0$ parametresiyle Poisson dağıldığı (1) numaralı modelden oluşturulan örnekleme dayalı simülasyon çalışması ile ε_t 'nin $\mu = \lambda$ ve $\sigma^2 = \lambda(1 + \alpha)$ ile normal dağıldığı kabul edilen standart AR(1) sürecine karşılık gelen simülasyon çalışmasını karşılaştırmışlardır. Karşılaştırma sonucunda, INAR(1) sürecinin gerçekleşmelerinin tümü negatif olmayan tam sayı değerleri iken AR(1) sürecinin gerçekleşmelerinin bir kısmı negatif değerler almıştır. Ayrıca, INAR(1) sürecinde gerçekleşmelerin birçoğu ortalamaya çok yakın değerlere sahipken, standart AR(1) sürecinde böyle bir durum gözlemlenmemiştir.



Şekil 1: (a) N(2,2)'den üretilmiş sürekli parametrelili zaman serisi, (b) Bernoulli(0.5) ve Poisson(1) dağılımlarından üretilmiş kesikli parametrelili zaman serisi (Al-Osh ve Alzaid, 1987)

2.2. INAR(1) Sürecinin Dağılım Özellikleri

Hatalar (ε_t) , bağımsız ve aynı dağılımlı zaman serisi olduğunda ve $E[\varepsilon_t] = \mu_\varepsilon$, $Var[\varepsilon_t] = \sigma_\varepsilon^2$ ile gösterildiğinde, $\alpha \in (0,1)$ için, yinelemeyi takip eden gözlem süreci (X_t) (1) numaralı modelde yer almaktadır. Bu modelde, tüm inceltme işlemleri birbirinden ve (ε_t) 'den bağımsız olarak yapılıyorsa ve t zamanındaki inceltme işlemlerinin yanı sıra ε_t 'de $(X_s)_{s < t}$ 'den bağımsızsa bu ifadenin bir INAR(1) süreci olduğu söylenebilir (Weiß, 2018).

Eşitlik (2)'de her Y_i , $Y_i \sim Binom(1, \alpha)$ dağılım koşulunu sağladığından ve Binom dağılımı toplamsal bir dağılım olduğundan, $\alpha \circ X$, X 'in değeri göz önüne alındığında koşullu binom dağılımına sahiptir. Yani, $\alpha \circ X | X \sim Binom(X, \alpha)$ şeklinde ifade edilebilir. Buradan, toplam beklenti yasası kullanılarak

$$E[\alpha \circ X] = E[E[\alpha \circ X | X]] = E[\alpha \cdot X] = \alpha\mu \quad (3)$$

yazılabilir. Dolayısıyla, Binom inceltme $\alpha \circ X$ ve çarpma $\alpha \cdot X$ aynı ortalamaya sahiptir. Bu durum değiştirilmiş bir AR(1) özyinelemesi için binom inceltmeyi kullanmayı avantajlı hale getirir. Ancak, diğer birçok özellik bakımından farklılık gösterebilirler. Örnek olarak, toplam varyans yasası aşağıdaki gibi ifade edilir:

$$\begin{aligned} V[\alpha \circ X] &= V[E[\alpha \circ X | X]] + E[V[\alpha \circ X | X]] \\ &= V[\alpha \cdot X] + E[\alpha(1 - \alpha) \cdot X] \\ &= \alpha^2 \sigma^2 + \alpha(1 - \alpha)\mu. \end{aligned} \quad (4)$$

Yani, binom inceltme $\alpha \circ X$ ve çarpma $\alpha \cdot X$ aynı varyansa sahip değildir.

INAR(1) süreci aynı zamanda homojen bir Markov zinciridir ve bu sürecin 1-adım geçiş olasılıkları

$$\begin{aligned} P_{ij} &= P[X_t = i | X_{t-1} = j] = P(\alpha \circ X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \\ &= \sum_{m=0}^{\min(i,j)} P(B_j^\alpha = m) P(\varepsilon_t = i - m) \end{aligned}$$

$$= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1-\alpha)^{j-m} P(\varepsilon_t = i-m) \quad (5)$$

biçiminde elde edilir (Jin-Guan ve Yuan, 1991). Burada, $\alpha \in (0,1)$ ve $j \in \mathbb{N}$ olmak üzere, $B_j^\alpha \sim \text{Binom}(j, \alpha)$ dağılımına sahiptir.

Ayrıca 1-adım geçiş olasılıklarından yola çıkarak sürecin birleşik olasılık fonksiyonu da

$$\begin{aligned} f(i_1, i_2, \dots, i_T) &= P(X_1 = i_1, X_2 = i_2, \dots, X_T = i_T) \\ &= P(X_1 = i_1) P(X_2 = i_2 | X_1 = i_1) \dots P(X_T = i_T | X_{T-1} = i_{T-1}) \\ &= P(X_1 = i_1) \prod_{k=1}^{T-1} \left[\sum_{m=0}^{\min(i_k, i_{k+1})} \binom{i_k}{m} \alpha^m (1-\alpha)^{i_k-m} P(\varepsilon_{k+1} = i_{k+1} - m) \right] \end{aligned} \quad (6)$$

şeklinde elde edilir.

Durağan bir INAR(1) sürecinin regresyon davranışı incelenirken, $V(X_t) < \infty$ olduğu varsayılır. X_t ' nin X_{t-1} üzerindeki koşullu ortalaması ve koşullu varyansı için, Eşitlik (1) üzerinden basit bir hesaplama ile

$$E[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon, \quad (7)$$

$$V[X_t | X_{t-1}] = \alpha(1-\alpha)X_{t-1} + \sigma_\varepsilon^2 \quad (8)$$

biçiminde elde edilmektedir (Alzaid ve Al-Osh, 1988). Elde edilen eşitliklerin her ikisi de X_t 'nin dağılımından bağımsız olarak, X_{t-1} 'in doğrusal bir fonksiyonudur. Bu, doğrusal bir regresyona ve sabit bir koşullu varyansa sahip olan standart AR(1) sürecinin tersidir.

$\varphi_X(s)$ ve $\varphi_\varepsilon(s)$ sırasıyla X_t ve ε_t ' nin olasılık yaratıcı fonksiyonunu gösterdiğinde, X_t 'nin durağan marjinal dağılımı, çeşitli marjinal dağılım türlerine izin veren

$$\varphi_X(s) = \varphi_X(1 - \alpha(1-s))\varphi_\varepsilon(s) \quad (9)$$

eşitlik ile elde edilir (Alzaid ve Al-Osh, 1988). Eşitlik (9), Eşitlik (3)'te olduğu gibi toplam beklenti yasası uygulanarak elde edilmiştir. Ayrıca, Eşitlik (1)'in marjinal dağılımı, hata serisi ε_t cinsinden şu şekilde ifade edilir:

$$X_t = \sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i} , \quad (10)$$

$\alpha \in (0,1)$ için, (10) numaralı eşitlikten X_t 'nin ε_t serisine bağımlılığının bu serinin zaman gecikmesiyle üstel olarak azaldığı görülebilir (Al-Osh ve Alzaid, 1987). Böylece, X_t 'nin olasılık yaratıcı fonksiyonu aşağıdaki gibidir:

$$\varphi_X(s) = \prod_{i=0}^{\infty} \varphi_{\varepsilon}(1 - \alpha^i + \alpha^i s). \quad (11)$$

Eşitlik (1)' de tanımlanan X_t sürecinin ortalaması ve varyansı;

$$\begin{aligned} E[X_t] &= \alpha E[X_{t-1}] + \mu_{\varepsilon} \\ &= \alpha^t E[X_0] \mu_{\varepsilon} \sum_{j=0}^{t-1} \alpha^j , \end{aligned} \quad (12)$$

$$\begin{aligned} Var[X_t] &= \alpha^2 Var[X_{t-1}] + \alpha(1 - \alpha)E[X_{t-1}] + \sigma_{\varepsilon}^2 \\ &= \alpha^{2t} Var[X_0] + (1 - \alpha) \sum_{j=1}^t \alpha^{2j-1} E[X_{t-j}] + \sigma_{\varepsilon}^2 \sum_{j=1}^t \alpha^{2(j-1)} \end{aligned} \quad (13)$$

eşitlikleriyle elde edilir. Eşitlik (12) ve (13)'ten görüleceği üzere, X_t sürecinin başlangıç değeri olan X_0 'ın beklenen değer ve varyansının;

$$E[X_0] = \frac{\mu_{\varepsilon}}{1 - \alpha} , \quad (14)$$

$$Var[X_0] = \frac{\alpha \mu_{\varepsilon} + \sigma_{\varepsilon}^2}{1 - \alpha^2} \quad (15)$$

şeklinde olması beklenir (McKenzie, 1985; Al-Osh ve Alzaid, 1987).

Freeland (1998), X_0 'ın ortalaması $\mu_\varepsilon/1 - \alpha$ ile dağıldığında, X_t 'nin marjinal dağılımının da $\mu_\varepsilon/1 - \alpha$ ile dağıldığı önermiş ve önerisini kanıtlamıştır.

Freeland'ın önermesinin sonucu olarak, μ_ε ve $\sigma_\varepsilon^2 < \infty$ ise INAR(1) süreçlerinin ortalaması ve varyansı şu şekildedir:

$$E[X_t] = \mu_X = \frac{\mu_\varepsilon}{1 - \alpha}, \quad (16)$$

$$Var[X_t] = \sigma_X^2 = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1 - \alpha^2} \quad (17)$$

Böylece, yayılım indeksi

$$I = \frac{I_\varepsilon + \alpha}{1 + \alpha} = \frac{\sigma_X^2}{\mu_X} \quad (18)$$

biçiminde elde edilir. Burada I_ε , hata serinin yayılım indeksini ifade ederken, I sürecin yayılım indeksini ifade eder. I 'nın değeri, X_t 'nin yayılım davranışını belirler ve ε_t 'ye bağlıdır. Yani, ε_t aşırı, eşit veya yetersiz yayılmışsa, X_t 'nin de aşırı, eşit veya yetersiz yayıldığı söylenebilir.

$$I = \begin{cases} I = 1, & \text{INAR(1) süreci eşit yayılımlıdır} \\ I < 1, & \text{INAR(1) süreci yetersiz yayılımlıdır} \\ I > 1, & \text{INAR(1) süreci aşırı yayılımlıdır} \end{cases}$$

Durağan bir INAR(1) sürecinin otokovaryans ve otokorelasyon fonksiyonları sırasıyla

$$Cov(X_t, X_{t-k}) = \gamma_k = \alpha^k \sigma^2, \quad k \geq 0; \quad (19)$$

$$Corr(X_t, X_{t-k}) = \rho_k = \alpha^k, \quad k \geq 0 \quad (20)$$

şeklinde ifade edilir. (19) ve (20) numaralı eşitliklerden otokorelasyon fonksiyonunun, k gecikmesiyle üstel olarak azaldığını ve AR(1)'deki Yule-Walker denklemiyle aynı forma sahip olduğunu görebiliriz. Ancak, standart AR(1) süreci için hesaplanan otokorelasyon fonksiyonunun aksine, INAR(1) süreci için otokorelasyon fonksiyonu her zaman pozitifdir (Al-Osh ve Alzaid, 1987).

2.3. INAR(1) Süreçlerinde Parametre Tahmini

INAR(1) modeli, inceltme parametresi α ve hataların marjinal dağılımını karakterize eden diğer parametrelerle belirlenir. Bu parametrelerin gerçek değerleri pratikte bilinmemektedir. Ancak, belirli bir zaman serisi verilerinden yararlanılarak parametre değerlerinin tahmin edilmesi gerekir. Literatürde yer alan çalışmalarda genellikle koşullu en küçük kareler (CLS), Yule-Walker (YW) ve koşullu maksimum olabilirlik (CML) olmak üzere üç farklı tahmin yöntemi kullanılmaktadır.

2.3.1. Yule-Walker Tahmini (YW)

X_1, X_2, \dots, X_T 'nin Eşitlik (1)'de tanımlanan X_t sürecinden geldiği kabul edildiğinde, örneklem ortalaması $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$ ve örneklem otokorelasyon fonksiyonu;

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2} \quad (21)$$

şeklinde tanımlanır. Burada, $\rho_k = \alpha^k$ olduğu bilinmektedir. Böylece α 'nın Yule-Walker tahmini şu şekilde elde edilir:

$$\hat{\alpha}_{YW} = \hat{\rho}(1) = \frac{\sum_{t=1}^{T-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2} . \quad (22)$$

X_t sürecinin beklenen değeri $E[X_t]$ ile gösterilir ve hataların dağılımına bağlı olarak elde edilir. $E[X_t]$ ve $\hat{\alpha}_{YW}$ kullanılarak hataların marjinal dağılımını karakterize eden diğer parametrelere ilişkin tahmin yapılabilir.

2.3.2. Koşullu En Küçük Kareler Tahmini (CLS)

Koşullu En Küçük Kareler tahmin işlemi, Klimko ve Nelson (1978) tarafından geliştirilmiştir ve bu işlem koşullu beklenen değere ilişkin sapmaların karelerinin toplamının minimizasyonuna dayanmaktadır.

INAR(1) sürecinde X_{t-1} bilindiğinde X_t 'nin koşullu ortalaması

$$E[X_t|X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon \equiv g(\theta, X_{t-1}) \quad (23)$$

olup burada θ tahmin edilecek parametreler kümesidir. Bu tahmin yönteminde amaç, X_t ile $g(\theta, X_{t-1})$ arasındaki karesel sapmaları toplamak ve bu toplamı en aza indirecek şekilde model parametrelerini tahmin etmektir.

$$Q_n(\theta) = \sum_{t=1}^n [(X_t - g(\theta, X_{t-1}))^2]. \quad (24)$$

Böylece, α ve μ_ε 'nin CLS tahmin edicileri şu şekilde ifade edilebilir (Klimko ve Nelson, 1978; Al-Osh ve Alzaid, 1987):

$$\hat{\alpha} = \frac{\sum_{t=1}^n X_t X_{t-1} - \frac{\sum_{t=1}^n X_t \sum_{t=1}^n X_{t-1}}{n}}{\sum_{t=1}^n X_{t-1}^2 - \frac{(\sum_{t=1}^n X_{t-1})^2}{n}} \quad (25)$$

ve

$$\hat{\mu}_\varepsilon = \frac{1}{n} \left(\sum_{t=1}^n X_t - \hat{\alpha} \sum_{t=1}^n X_{t-1} \right). \quad (26)$$

Koşullu en küçük kareler tahmin edicilerin dağılımı, asimptotik olarak Yule-Walker denklemlerine dayalı tahmin edicilerin dağılımına eşdeğerdir ve dolayısıyla herhangi biri diğerinden daha üstün değildir (Freeland ve McCabe, 2005).

2.3.3. Koşullu Maksimum Olabilirlik Tahmini (CML)

INAR(1) sürecine ait birleşik olasılık fonksiyonundan yararlanarak olabilirlik fonksiyonu

$$\begin{aligned} f(i_1, i_2, \dots, i_T) &= P(X_1 = i_1) \prod_{t=1}^{T-1} P(X_{t+1} = x_{t+1} | X_t = x_t) \\ &= f(x_1) \prod_{k=1}^{T-1} \left[\sum_{m=0}^{\min(i_t, i_{t+1})} \binom{x_t}{m} \alpha^m (1 - \alpha)^{i_t - m} P(\varepsilon_{t+1} = x_{t+1} - m) \right] \end{aligned} \quad (27)$$

biçiminde gösterilir. Marjinal dağılım genel olarak anlaşılması zor bir dağılım olduğundan, gözlemlenen X_1 'i koşullandırmak, aslında katkısını göz ardı etmek ve parametreleri CML ile tahmin etmek daha basit bir yaklaşımdır.

Bu nedenle, X_1 değişkenini koşullandırmak için koşullu log olabilirlik fonksiyonu

$$L(\theta) = \sum_{t=1}^{T-1} \log P(X_{t+1} = x_{t+1} | X_t = x_t) \quad (28)$$

biçiminde yazılır. Burada, θ tahmin edilecek parametreler kümesidir. θ 'nın koşullu maksimum olabilirlik tahminleri, koşullu log olabilirlik fonksiyonu $L(\theta)$ maksimize edilerek elde edilen $\hat{\theta}_{CML}$ değerleridir.

Literatürde yer alan uygulamalı çalışmalarda, koşullu maksimum olabilirlik tahminlerinin, koşullu en küçük kareler tahmini ve Yule-Walker tahmininden daha iyi performans gösterdiği sonucuna varılmış ve gerçek veri uygulamalarında INAR(1) sürecinin bilinmeyen parametre tahminleri koşullu maksimum olabilirlik tahmini ile gerçekleştirilmiştir (Huang ve Zhu, 2021; da Cunha, Bourguignon ve Vasconcellos, 2021; Bourguignon, Rodrigues ve Santos-Neto, 2019; Jazi, Jones ve Lai, 2012a; Lívio ve ark., 2018). Bu nedenle, bu çalışmanın uygulama bölümünde gerçek veri üzerinden yapılacak olan analizlerin parametre tahmininde koşullu maksimum olabilirlik tahmini kullanılacaktır.

2.4. INAR(1) Süreçlerinde Model Tanımlama

Bölüm 2.3'te gerçek bir tam sayı değerli zaman serisine bir INAR(1) modeli uydurmak için üç farklı standart yaklaşım incelenmiştir. Ancak, elde edilen tahminler yalnızca seriler gerçekten bir INAR(1) sürecinden kaynaklanıyorsa anlamlıdır. Bu nedenle, seriler için en uygun model belirlenmelidir.

Bu doğrultuda ilk olarak verilen serilerin bağımlılık yapısına bakılmalıdır. Verilerin bağımlılık yapısı otokorelasyon fonksiyonu (ACF) ve kısmi otokorelasyon fonksiyonu (PACF) grafikleri kullanılarak analiz edilir. Otokorelasyon bir değişkenin bir dönemdeki değeri ile geçmiş dönemlerdeki değerleri arasındaki ilişkiyi ölçer. Tüm geçmiş değerlere ait otokorelasyon değerleri Eşitlik (21)'de açıklanan otokorelasyon fonksiyonunu oluşturur. Kısmi otokorelasyon ise, diğer geçmiş dönemlerdeki değerler sabitken iki dönemdeki değerler arasındaki ilişkiyi ölçer ve tüm geçmiş değerlere ait kısmi otokorelasyon değerleri kısmi otokorelasyon fonksiyonunu oluşturur.

ACF ve PACF grafikleri genellikle sırasıyla hareketli ortalama (MA) ve otoregresyon (AR) derecesini belirlemek için kullanılır. Daha önce de belirtildiği gibi, INAR(1) süreci korelasyon yapısı ve dağılım özellikleri bakımından sürekli değerli AR(1) sürecine oldukça benzerdir. Dolayısıyla, uygulamalarda PACF grafiği dikkate alınır ve grafiğin sadece ilk gecikmesine ait ilişki miktarı istatistiksel olarak önemli ise ve ACF grafiğindeki ilişki miktarları gecikme sayısı arttıkça PACF grafiğindeki ilişki değerlerine göre daha yavaş azalıyorsa bir INAR(1) modeli bu seri için uygundur denilebilir.

Seriler daha sonrasında durağanlık açısından test edilmelidir. Uygulamalarda, ACF grafiği ile serinin durağan olup olmadığının kontrolü de yapılabilmektedir. ACF grafiğinde gecikmelere ait önemli ilişkiler mevcut ise serinin durağan bir seri olmadığı söylenebilir. Ancak, sadece ACF grafiğinden yola çıkılarak yapılan bu yorum istatistiksel anlamda yeterli olmayacaktır. Bu nedenle, durağanlık tespiti için birim kök testleri geliştirilmiştir. Geliştirilen testler arasından en yaygın olarak kullanılan Geliştirilmiş Dickey Fuller (ADF) testidir. ADF testini gerçekleştirirken dikkate alınan hipotezler genellikle

H_0 : Seri durağan değildir.

H_S : Seri durağandır.

biçimindedir. ADF testi sonucunda hesaplanan $P - deęeri < \alpha$ olduęunda H_0 hipotezi reddedilir. Burada, α anlamlılık düzeyidir.

Analiz edilen seri durağan ve seri için birkaç farklı modelin uygun olduęu düşünülüyorsa, Akaike Bilgi Kriteri (AIC) ve Schwarz Bayesyen Bilgi Kriteri (BIC) gibi bilgi kriterleri dikkate alınmalıdır. Bu doęrultuda, L incelenen INAR(1) modelinin olabilirlik fonksiyonunun maksimum deęeri, k modelin parametre sayısı ve T serideki gözlem sayısı olduęunda AIC ve BIC bilgi kriterleri řu řekilde hesaplanır:

$$AIC = -2 \ln(L) + 2k, \quad (29)$$

$$BIC = -2 \ln(L) + \log(T) * k \quad (30)$$

2.5. INAR(1) Süreçlerinde Tahmin

Gözlemlenen INAR(1) süreci için model kurulduęu varsayıldıęında, bu modelin ana uygulamalarından biri, sürecin gelecekteki sonuçlarını tahmin etmektir. Yani, x_1, x_2, \dots, x_T gözlemledikten sonra, $k \geq 1$ için X_{T+k} ' in tahmin edilmesi beklenir.

Gerçek deęerli süreçler için en yaygın nokta tahmini türü koşullu ortalamadır (Weiß, 2018). Toplam beklenti yasasını koşullu ortalama denklemi ile yinelemeli olarak uygulayarak, h -adım ileri koşullu ortalama

$$E[X_{t+k}|X_t] = \alpha^k X_t + \mu_\varepsilon \frac{1 - \alpha^k}{1 - \alpha} = \alpha^k X_t + \mu_X (1 - \alpha^k) \quad (31)$$

biçiminde elde edilir. Eşitlik (31)'de koşullu ortalamanın yalnızca X_t ' ye baęlı olduęu, Markov özellięi nedeniyle daha önceki gözlemlere baęlı olmadığı açıkça görülmektedir.

Koşullu ortalama tahmin yönteminin en önemli dezavantajı, X_t 'nin negatif olmayan bir tam sayı değeri olması gerekirken, tahmin sonucunda genellikle tam sayı olmayan bir değer elde edilir. Bu nedenle, tam sayı değerli seriler için, negatif olmayan tam sayı değerli tahminler üreten tahmin tekniklerine ihtiyaç vardır (Freeland ve McCabe, 2004).

Bu amaçla, INAR(1) modeli için X_t 'nin h -adım ileri koşullu dağılımının yani, h -adım ileri geçiş olasılıklarının hesaplanması düşünülebilir. Bu tahmin yöntemi de yine Markov özelliği nedeniyle sadece X_t 'ye bağlıdır. Ancak bu yöntem, X_t için bir dağılım mevcut olduğunda tutarlı bir nokta tahmini olarak kullanılabilir. Yani, $\varepsilon^{(k)}$ 'nin dağılımı mevcut olduğunda, h -adım ileri geçiş olasılıkları $P_{ij}^{(k)}$ INAR(1) sürecinin 1-adım geçiş olasılıkları denkleminde uyarlanarak hesaplanabilir (Weiß, 2018)

$$P_{ij}^{(k)} = \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^{km} (1 - \alpha^k)^{j-m} P(\varepsilon_t = i - m). \quad (32)$$

Kısaca, Eşitlik (32)'de, Eşitlik 5'te yer alan sürecin 1-adım geçiş olasılıkları denkleminde, α yerine α^k getirilmiştir.

Literatürdeki uygulamalı çalışmalarda genel olarak INAR(1) modeli tahmin sürecinde koşullu ortalama tahmin yöntemi kullanılmış ve elde edilen tahmin sonuçları sürekli değerler olsa da model karşılaştırmaları hata kareler ortalamasının karekökü hesaplanarak yapılmıştır. Yine bu çalışmada yer alacak gerçek veri uygulamalarında da INAR(1) model tahminleri koşullu ortalama tahmin yöntemiyle incelenecektir.

2.6. INAR(1) Süreçlerinde Model Yeterliliği

Analiz edilen seriler için uygun bulunan INAR(1) modelinin seriler için gerçekten yeterli olup olmadığının kontrol edilmesi oldukça önemlidir. Bu amaçla, uygun bulunan modelden hesaplanan hata serisi analiz edilmelidir. Hata serisinin bir akgürültü serisi olup olmadığı yani, serinin herhangi bir otokorelasyona sahip olup olmadığı yine ACF ve PACF grafikleriyle incelenebilir. Ancak, daha kesin sonuçlar için serilerin

otokorelasyona sahip olup olmadığını analiz eden Ljung-Box testi uygulanabilir. Ljung-Box testini gerçekleştirirken dikkate alınan hipotezler genellikle

H_0 : Hatalar arasında ilişki yoktur.

H_5 : Hatalar arasında ilişki vardır.

biçimindedir. Ljung-Box testi sonucunda hesaplanan P – değeri $> \alpha$ olduğunda H_0 hipotezi kabul edilir.

Ayrıca, merkezi limit teoremine göre veriler yeterince büyükse tüm dağılımlar normal dağılıma doğru eğilim gösterir. Bu durumda, büyük bir seri için uygun bulunan modelin hata serisi de normal dağılıma yakınsamalıdır. Hata serisi için çizilen Q-Q grafiği, serinin kabaca normal dağılıp dağılmadığı hakkında bir ön bilgi verebilir.

3. LİTERATÜRDE YER ALAN BAZI INAR(1) MODELLERİ

Bu bölümde, literatürde yer alan ve ε_t serilerin dağılımı değiştirilerek geliştirilen bazı alternatif INAR(1) modelleri dağılımları ve istatistiksel özellikleri bakımından ayrıntılı olarak incelenecektir.

3.1. P-INAR(1) Modeli

INAR(1) sürecinin en popüler ve ilk örneği, McKenzie (1985) ve Al-Osh ve Alzaid (1987) tarafından tanımlanan Poisson-INAR(1) modelidir. Bu modelde, bir INAR(1) sürecinin hatalarının Poisson dağılımına sahip olduğu varsayılır. Poisson dağılımı genellikle tam sayı değerli verilerin modellenmesi için bir ölçüt görevi görür. Dağılımın ana özelliklerinden biri, varyansının her zaman ortalamasına eşit olduğu anlamına gelen eşit yayılım özelliğidir (Weiß, 2018). Ancak pratikte, veriler Poisson dağılımına göre aşırı veya yetersiz yayılmış olabilir.

Negatif olmayan tam sayı değerli zaman serisi verilerinin aşırı veya yetersiz yayılım durumları için literatürde sürecin hatalarının dağılımı değiştirilerek elde edilen birçok yeni INAR(1) modeli tanıtılmıştır. Ancak, bu modelleri incelemeye başlamadan önce, Poisson dağılımının ve ana model olan P-INAR(1) sürecinin temel özellikleri üzerinde durulacaktır.

Tam sayı değerli verilerin model dağılımı için genellikle belirli bir zaman aralığında kullanılan ve kesikli bir olasılık dağılımı olan Poisson dağılımının olasılık fonksiyonu

$$P(Z = z) = \frac{e^{-\lambda} \lambda^z}{z!}, \quad z = 0, 1, 2, \dots; \lambda > 0 \quad (33)$$

biçimindedir. Dağılımın olasılık yaratıcı fonksiyonu ise

$$\varphi_Z(s) = E[s^Z] = \exp[\lambda(s - 1)] \quad (34)$$

şeklindedir.

Poisson dağılımının olasılık fonksiyonundan yararlanılarak elde edilen dağılımın ortalaması ve varyansı birbirine eşittir. Eşitlik ile ifade edilirse,

$$E[Z] = Var[Z] = \lambda. \quad (35)$$

Poisson dağılımı eşit yayılıma sahip olduğu için ortalama ve varyansa göre yayılım indeksi her zaman 1'e eşittir ve aşağıdaki gibi gösterilmektedir.

$$ID(Z) = \frac{Var[Z]}{E[Z]} = 1. \quad (36)$$

X_t , Eşitlik (1)'e göre tanımlanmış bir INAR(1) süreci ve ε_t hataları bağımsız ve aynı dağılımlı bir $Poisson(\lambda)$ serisi ise, bu durum bir P-INAR(1) modelini ifade eder:

$$\begin{cases} X_t = \alpha \circ X_{t-1} + \varepsilon_t, & t \geq 1, \\ \varepsilon_t \sim Poisson(\lambda), & \end{cases} \quad (37)$$

Burada, $0 < \alpha < 1$ ve $\varepsilon_t, t \geq 1$ için Eşitlik (2)'de açıklanan, bir Bernoulli serisi olan Y_i ve X_{t-1} 'den bağımsızdır.

P-INAR(1) süreci bir Markov zinciridir ve geçiş olasılıkları:

$$\begin{aligned} P_{ij} &= P(X_t = i | X_{t-1} = j) \\ &= P(\alpha \circ X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \\ &= \sum_{m=0}^{\min(i,j)} P(\alpha \circ X_{t-1} = m | X_{t-1} = j) P(\varepsilon_t = i - m) \\ &= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1 - \alpha)^{j-m} \frac{e^{-\lambda} \lambda^{i-m}}{(i-m)!}, \quad i, j = 0, 1, 2, \dots \end{aligned} \quad (38)$$

INAR(1) sürecinin genel dağılım özelliklerinden yararlanılarak elde edilen P-INAR(1) sürecinin koşullu ortalama, koşullu varyans, ortalama, varyans, kovaryans ve otokorelasyon fonksiyonu sırasıyla verilmiştir:

$$E[X_t|X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon = \alpha X_{t-1} + \lambda, \quad (39)$$

$$V[X_t|X_{t-1}] = \alpha(1 - \alpha)X_{t-1} + \sigma_\varepsilon^2 = \alpha(1 - \alpha)X_{t-1} + \lambda, \quad (40)$$

$$\mu_X = E[X_t] = \frac{\mu_\varepsilon}{1 - \alpha} = \frac{\lambda}{(1 - \alpha)}, \quad (41)$$

$$\sigma_X^2 = Var[X_t] = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1 - \alpha^2} = \frac{\lambda(1 + \alpha)}{(1 - \alpha^2)}, \quad (42)$$

$$\gamma_k = Cov(X_k, X_{k+1}) = \alpha^k \sigma^2, \quad (43)$$

$$\rho_k = Corr(X_k, X_{k+1}) = \alpha^k. \quad (44)$$

Buna göre, X_t 'nin yayılım indeksi ortalama ve varyansa dayalı olarak

$$I_X = \frac{I_\varepsilon + \alpha}{1 + \alpha} = \frac{\sigma_X^2}{\mu_X} = 1 \quad (45)$$

biçimindedir. Bu durumda, P-INAR(1) sürecinin, Poisson dağılımında olduğu gibi eşit yayılım özelliğine sahip olduğu ve eşit yayılım özelliği sergileyen seriler için uygun bir model olduğu söylenebilir.

3.2. G-INAR(1) Modeli

Poisson dağılımının eşit yayımlı zaman serileri için uygun olduğunu, ancak pratikte tam sayı değerli bazı zaman serileri aşırı veya yetersiz yayılıma sahip olabileceğini belirtmiştik. Bu nedenle, literatürde ilk olarak McKenzie (1986) tarafından, negatif binom ve geometrik marjinal dağılımlara sahip otoregresif hareketli ortalama süreçleri geliştirilmiştir. Alzaid ve Al-Osh (1988), daha sonrasında INAR(1) süreçlerini aşırı

yayılmış tam sayı değerli zaman serileri için geometrik marjinal dağılımla yeniden değerlendirilmişlerdir. Sonrasında ise Jazi, Jones ve Lai (2012a), yine aşırı yayılmış tam sayı değerli zaman serisi verilerini modellemek amacıyla binom inceltmesine dayalı INAR(1) modelinin geometrik hatalarla yeni bir modifikasyonunu tanıtmışlardır.

Aşırı yayılıma sahip zaman serisi verilerini modellemek için uygun olarak görülen geometrik dağılımın olasılık fonksiyonu:

$$P(Z = z) = p(1 - p)^z, \quad 0 < p < 1; z = 1, 2, 3, \dots \quad (46)$$

Olasılık fonksiyonundan yararlanılarak dağılımın olasılık yaratıcı fonksiyonu:

$$\varphi_Z(s) = E[s^Z] = \frac{p}{1 - (1 - p)s} \quad (47)$$

şeklindedir.

Yine dağılımın ortalaması ve varyansı sırasıyla

$$E[Z] = \frac{1 - p}{p} \quad (48)$$

ve

$$Var[Z] = \frac{1 - p}{p^2} \quad (49)$$

şeklindedir.

Ortalama ve varyansa göre yayılım indeksi

$$ID(Z) = \frac{Var[Z]}{E[Z]} = \frac{1}{p} > 1 \quad (50)$$

biçimindedir. Dolayısıyla, dağılım aşırı yayılıma sahip veriler için uygundur.

X_t Eşitlik (1)'e göre tanımlanmış bir INAR(1) süreci ve ε_t hataları bağımsız ve aynı dağılımlı bir *Geometrik*(p) zaman serisi ise, bu durum bir G-INAR(1) modelini ifade eder:

$$\begin{cases} X_t = \alpha \circ X_{t-1} + \varepsilon_t, & t \geq 1, \\ \varepsilon_t \sim \text{Geometrik}(p), \end{cases} \quad (51)$$

Burada, $0 < \alpha < 1$, $0 < p < 1$ ve ε_t , $t \geq 1$ için Bernoulli serisi Y_i ve X_{t-1} 'den bağımsızdır.

$\varepsilon_t \sim \text{Geometrik}(p)$ dağılımına sahip G-INAR(1) modeli, geçiş olasılıklarına sahip durağan bir Markov zinciri oluşturur:

$$\begin{aligned} P_{ij} &= P(X_t = i | X_{t-1} = j) \\ &= P(\alpha \circ X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \\ &= \sum_{m=0}^{\min(i,j)} P(\alpha \circ X_{t-1} = m | X_{t-1} = j) P(\varepsilon_t = i - m) \\ &= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1 - \alpha)^{j-m} p (1 - p)^{i-m}, \quad i, j = 0, 1, 2, \dots \end{aligned} \quad (52)$$

G-INAR(1) sürecinin koşullu ortalama, koşullu varyans, ortalama, varyans, kovaryans ve otokorelasyon fonksiyonları

$$E[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon = \alpha X_{t-1} + \frac{1-p}{p}, \quad (53)$$

$$V[X_t | X_{t-1}] = \alpha(1-\alpha)X_{t-1} + \sigma_\varepsilon^2 = \alpha(1-\alpha)X_{t-1} + \frac{1-p}{p^2}, \quad (54)$$

$$\mu_X = E[X_t] = \frac{\mu_\varepsilon}{1-\alpha} = \frac{1-p}{p(1-\alpha)}, \quad (55)$$

$$\sigma_X^2 = \text{Var}[X_t] = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1-\alpha^2} = \frac{1-p}{p^2(1-\alpha)}, \quad (56)$$

$$\gamma_k = Cov(X_k, X_{k+1}) = \alpha^k \sigma^2, \quad (57)$$

$$\rho_k = Corr(X_k, X_{k+1}) = \alpha^k \quad (58)$$

şeklindedir.

Buna göre, X_t 'nin yayılım indeksi

$$I_X = \frac{I_\varepsilon + \alpha}{1 + \alpha} = \frac{\sigma_X^2}{\mu_X} = \frac{1}{p} > 1 \quad (59)$$

olarak hesaplanır.

3.3. PL-INAR(1) Modeli

Negatif olmayan tam sayı değerli zaman serilerini modellemek için Lívio ve ark. (2018) PL-INAR(1) ile gösterilen binom inceltme operatörüne (Steutel ve van Harn, 1979) dayalı Poisson-Lindley (PL) hataları ile yeni bir INAR(1) modeli önermişlerdir. Modelde Poisson-Lindley dağılımına sahip hataların kullanılmasının birçok avantajı vardır. Poisson-Lindley dağılımı, birleşik Poisson ailesine aittir ve Negatif Binom dağılımı gibi tek tepe değerli, aşırı dağılmış ve sonsuz bölünebilirlik gibi ortak özellikleri de taşır (Ghitany ve Al-Mutairi, 2009).

Olasılık fonksiyonu

$$P(Z = z) = \frac{\theta^2(z + \theta + 2)}{(\theta + 1)^{z+3}}, \quad \theta > 0; z \in \mathbb{N} \quad (60)$$

olduğunda kesikli bir rasgele değişken Z , $\theta > 0$ parametresiyle Poisson-Lindley (PL) dağılımına (Sankaran, 1970) sahiptir. Kısaca $PL(\theta)$ dağılımı olarak adlandırılan bu dağılımın olasılık yaratıcı fonksiyonu

$$\varphi_Z(s) = E[s^Z] = \frac{\theta^2}{\theta + 1} \frac{2 + \theta - s}{(\theta + 1 - s)^2} \quad (61)$$

olarak verilmektedir.

$PL(\theta)$ dağılımının ortalaması ve varyansı ise sırasıyla

$$E[Z] = \frac{\theta + 2}{\theta(\theta + 1)} \quad (62)$$

ve

$$Var[Z] = \frac{\theta^3 + 4\theta^2 + 6\theta + 2}{\theta^2(\theta + 1)^2} = E[Z] \left[1 + \frac{\theta^2 + 4\theta + 2}{\theta(\theta + 1)(\theta + 2)} \right] \quad (63)$$

olarak hesaplanır.

Ortalama ve varyansa göre dağılımın yayılım indeksi 1'den büyük bir değer olarak elde edilir. Bu da Poisson-Lindley dağılımının aşırı yayılmış sayım verilerini modellemede uygun bir dağılım olduğunu gösterir.

$$ID(Z) = \frac{Var[Z]}{E[Z]} > 1. \quad (64)$$

X_t Eşitlik (1)'e göre tanımlanmış bir INAR(1) süreci ve ε_t hataları bağımsız ve aynı dağılımlı bir $PL(\theta)$ zaman serisi ise, bu durum Eşitlik 65'teki gibi bir PL-INAR(1) modelini ifade eder:

$$\begin{cases} X_t = \alpha \circ X_{t-1} + \varepsilon_t, & t \geq 1, \\ \varepsilon_t \sim \text{Poisson-Lindley}(\theta) \end{cases} \quad (65)$$

Burada, $0 < \alpha < 1$, $\theta > 0$ ve ε_t , $t \geq 1$ için Bernoulli serisi Y_i ve X_{t-1} 'den bağımsızdır.

Eşitlik (65)'te tanımlanan X_t süreci, geçiş olasılıkları olan durağan bir Markov zinciridir ve PL-INAR(1) sürecinin geçiş olasılıkları:

$$\begin{aligned}
P_{ij} &= P(X_t = i | X_{t-1} = j) \\
&= P(\alpha \circ X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \\
&= \sum_{m=0}^{\min(i,j)} P(\alpha \circ X_{t-1} = m | X_{t-1} = j) P(\varepsilon_t = i - m) \\
&= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1 - \alpha)^{j-m} \frac{\theta^2 (i - m + \theta + 2)}{(\theta + 1)^{i-m+3}}, \quad i, j = 0, 1, 2, \dots \quad (66)
\end{aligned}$$

olarak hesaplanır.

PL-INAR(1) sürecinin koşullu ortalama, koşullu varyans, ortalama, varyans, kovaryans ve otokorelasyon fonksiyonu ise sırasıyla

$$E[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon = \alpha X_{t-1} + \frac{\theta + 2}{\theta(\theta + 1)}, \quad (67)$$

$$V[X_t | X_{t-1}] = \alpha(1 - \alpha)X_{t-1} + \sigma_\varepsilon^2 = \alpha(1 - \alpha)X_{t-1} + \frac{\theta^3 + 4\theta^2 + 6\theta + 2}{\theta^2(\theta + 1)^2}, \quad (68)$$

$$\mu_X = E[X_t] = \frac{\mu_\varepsilon}{1 - \alpha} = \frac{\theta + 2}{(1 - \alpha)(\theta + 1)\theta}, \quad (69)$$

$$\sigma_X^2 = Var[X_t] = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1 - \alpha^2} = \frac{\theta + 2}{\theta(\theta + 1)(1 - \alpha^2)} \left[1 + \alpha + \frac{\theta^2 + 4\theta + 2}{\theta(\theta + 1)(\theta + 2)} \right], \quad (70)$$

$$\gamma_k = Cov(X_k, X_{k+1}) = \alpha^k \sigma^2, \quad (71)$$

$$\rho_k = Corr(X_k, X_{k+1}) = \alpha^k \quad (72)$$

şeklindedir.

Buna göre, X_t 'nin yayılım indeksi de Poisson-Lindley dağılımında olduğu gibi ortalama ve varyansa dayalı olarak 1' den büyük bir değer olarak Eşitlik 73'teki gibi elde edilir:

$$I_X = \frac{I_\varepsilon + \alpha}{1 + \alpha} = \frac{\sigma_X^2}{\mu_X} > 1. \quad (73)$$

Bu nedenle, PL-INAR(1) sürecinin aşırı yayılmış tam sayı değerli zaman serileri için uygun olduğu söylenebilir.

3.4. NB-INAR(1) Modeli

Poisson dağılımının Gamma dağılımıyla bir karışımı olduğu bilinen Negatif Binom dağılımı, aşırı yayılıma sahip sayım verilerinin modellenmesinde yaygın olarak kullanılmaktadır (Joe ve Zhu, 2005). Buradan yola çıkarak Jung, Ronning ve Tremayne (2005), Negatif Binom hatalarla yeni bir INAR(1) modelinin aşırı yayılıma sahip zaman serilerini modellemek için uygun olduğunu öne sürmüşlerdir. İki parametreye sahip olan Negatif Binom dağılımının bir parametresi limitte sonsuza giderken Poisson dağılımına yakınsar. Parametre değeri ne kadar büyürse, Negatif Binom dağılımı Poisson dağılıma o kadar yaklaşır. Bu durum, Negatif Binom dağılımının Poisson dağılımına göre daha güvenilir bir dağılım olduğunu göstermektedir.

İki farklı parametreye sahip olan Negatif Binom dağılımının olasılık fonksiyonu

$$\tilde{p}(z; r, p) \equiv P(Z = z) = \binom{z + r - 1}{z} (1 - p)^z p^r, \quad r > 0; 0 < p < 1; z = 0, 1, \dots \quad (74)$$

biçimindedir. Olasılık fonksiyonundan yararlanılarak dağılımın olasılık yaratıcı fonksiyonu ise

$$\varphi_Z(s) = E[s^Z] = \left(\frac{p}{1 - (1 - p)s} \right)^r, \quad |s| < \frac{1}{p} \quad (75)$$

şeklindedir.

Dağılımın ortalaması ve varyansı sırasıyla

$$E[Z] = \frac{r(1-p)}{p} \quad (76)$$

ve

$$Var[Z] = \frac{r(1-p)}{p^2} \quad (77)$$

olarak hesaplanır.

Negatif Binom dağılımının varyansının ortalamasına oranlanmasıyla elde edilen yayılım indeksi

$$ID(Z) = \frac{Var[Z]}{E[Z]} = \frac{1}{p} > 1 \quad (78)$$

şeklinde elde edilmektedir.

Yayılım indeksi her zaman 1' den büyüktür ve bu da Negatif Binom dağılımının aşırı yayılmış verileri modellemek için uygun bir dağılım olduğunu göstermektedir.

X_t Eşitlik (1)'e göre tanımlanmış bir INAR(1) süreci ve ε_t hataları bağımsız ve aynı dağılımlı bir Negatif Binom zaman serisi ise, bu durum bir NB-INAR(1) modelini ifade eder, yani;

$$\begin{cases} X_t = \alpha \circ X_{t-1} + \varepsilon_t, & t \geq 1, \\ \varepsilon_t \sim \text{Negatif Binom}(r, p) \end{cases} \quad (79)$$

biçimindedir.

Burada, $0 < \alpha < 1$, $r > 0$, $p \in [0,1]$ ve ε_t , $t \geq 1$ için Bernoulli serisi Y_i ve X_{t-1} 'den bağımsızdır.

NB-INAR(1) modeli durağan bir Markov zinciridir ve geçiş olasılıkları

$$\begin{aligned}
P_{ij} &= P(X_t = i | X_{t-1} = j) \\
&= P(\alpha \circ X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \\
&= \sum_{m=0}^{\min(i,j)} P(\alpha \circ X_{t-1} = m | X_{t-1} = j) P(\varepsilon_t = i - m) \\
&= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1 - \alpha)^{j-m} \binom{(i-m) + r - 1}{(i-m)} (1-p)^{(i-m)} p^r, \quad i, j = 0, 1, 2, \dots \quad (80)
\end{aligned}$$

biçimindedir.

NB-INAR(1) sürecinin koşullu ortalama, koşullu varyans, ortalama, varyans, kovaryans ve otokorelasyon fonksiyonu sırasıyla

$$E[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon = \alpha X_{t-1} + \frac{r(1-p)}{p}, \quad (81)$$

$$V[X_t | X_{t-1}] = \alpha(1-\alpha)X_{t-1} + \sigma_\varepsilon^2 = \alpha(1-\alpha)X_{t-1} + \frac{r(1-p)}{p^2}, \quad (82)$$

$$\mu_X = E[X_t] = \frac{\mu_\varepsilon}{1-\alpha} = \frac{r(1-p)}{p(1-\alpha)}, \quad (83)$$

$$\sigma_X^2 = Var[X_t] = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1-\alpha^2} = \frac{r(1-p)(\alpha p + 1)}{p^2(1-\alpha^2)}, \quad (84)$$

$$\gamma_k = Cov(X_k, X_{k+1}) = \alpha^k \sigma^2, \quad (85)$$

$$\rho_k = Corr(X_k, X_{k+1}) = \alpha^k \quad (86)$$

olarak elde edilmektedir.

Buna göre, X_t 'nin yayılım indeksi:

$$I_X = \frac{I_\varepsilon + \alpha}{1 + \alpha} = \frac{\sigma_X^2}{\mu_X} = \frac{\alpha p + 1}{\alpha p + p} > 1 \quad (87)$$

olarak hesaplanmaktadır.

3.5. DP-INAR(1) Modeli

Bourguignon, Rodrigues ve Santos-Neto (2019) eşit yayılım, yetersiz yayılım veya aşırı yayılıma sahip negatif olmayan tam sayı değerli zaman serilerini modellemek için DP-INAR(1) ile gösterilen çift Poisson hataları ile yeni bir binom inceltmeye dayalı INAR(1) süreci önermişlerdir.

Binom ve Poisson gibi tek parametrelili üstel ailelerde, varyans ortalamasının bir fonksiyonudur ancak çift üstel aileler, varyansı ortalamadan bağımsız olarak kontrol eden ikinci bir parametrenin eklenmesine izin verir (Efron, 1986). Böylelikle, Efron (1986) tarafından çift üstel aileye dayanılarak yeni bir dağılım önerilmiştir. Çift Poisson dağılımı $\mu > 0$ ve $\phi > 0$ olmak üzere iki parametre ile indekslenmektedir ve dağılımın olasılık fonksiyonu şu şekilde verilir:

$$p(z; \mu, \phi) \equiv P(Z = z) = Y(\mu, \phi) \tilde{p}(z; \mu, \phi) = Y(\mu, \phi) \phi^{\frac{1}{2}} e^{-\frac{\phi}{\mu}} \left[\frac{e^{-z} z^z}{z!} \right] \left(\frac{e\mu}{z} \right)^{\phi z}, \quad (88)$$

$$z = 0, 1, 2, \dots$$

Burada, $Y(\mu, \phi)^{-1} = \sum_{z=0}^{\infty} \tilde{p}(z; \mu, \phi) = \phi^{1/2} e^{-\phi/\mu} \sum_{z=0}^{\infty} \left[\frac{e^{-z} z^z}{z!} \right] \left(\frac{e\mu}{z} \right)^{\phi z} \approx 1 + \frac{1-\phi}{12\mu\phi}$, dır ve normalleştirme sabiti $\left(1 + \frac{1}{\mu\phi} \right)$ şeklindedir.

Efron (1986), $\sum_{z=0}^{\infty} \tilde{p}(z; \mu, \phi)$ sonsuz serisi için kapalı bir form yaklaşımı elde etmiş ve μ büyük olduğunda ($\mu = 10$) $\tilde{p}(z; \mu, \phi)$ 'nin oldukça iyi performansa sahip olduğunu göstermiştir. Ancak, μ küçük olduğunda aynı durum söz konusu değildir. Bu nedenle,

çift Poisson dağılımının kullanılması için farklı yaklaşımlar önerilmiştir (Zhu, 2012; Zou, Geedipally ve Lord, 2013).

Dağılımın olasılık yaratıcı fonksiyonu:

$$\varphi_Z(s) = E[s^Z] = \frac{Y(\mu s, \phi)}{Y(\mu, \phi)} \quad (89)$$

şeklindedir.

Çift Poisson dağılımının ortalaması ve varyansı yaklaşık değer olarak

$$E[Z] \cong \mu \quad (90)$$

ve

$$Var[Z] \cong \frac{\mu}{\phi} \quad (91)$$

olarak elde edilir.

Dağılımın ortalama ve varyansına göre yayılım indeksi

$$ID(Z) = \frac{Var[Z]}{E[Z]} = \frac{1}{\phi} \quad (92)$$

şeklinde hesaplanmaktadır.

Eşitlik (92)'den Çift Poisson dağılımının $\phi < 1$ olduğunda aşırı yayılıma, $\phi > 1$ olduğunda yetersiz yayılıma sahip olduğu görülür. $\phi = 1$ olması durumunda ise, çift Poisson dağılımı Poisson dağılımına benzer.

X_t , Eşitlik (88)'de verilen olasılık fonksiyonuyla çift Poisson dağılımına sahip bağımsız ve aynı dağılımlı bir zaman serisi ise, bu süreç DP-INAR(1) süreci olarak adlandırılır. Süreç

$$\begin{cases} X_t = \alpha \circ X_{t-1} + \varepsilon_t, & t \geq 1, \\ \varepsilon_t \sim \text{Çift Poisson}(\mu, \phi) \end{cases} \quad (93)$$

şeklinde tanımlanmaktadır.

Burada, $0 < \alpha < 1$, $\mu > 0$, $\phi > 0$ ve ε_t , $t \geq 1$ için Bernoulli serisi Y_i ve X_{t-1} 'den bağımsızdır.

Böylece, sürecin geçiş olasılıkları:

$$\begin{aligned} P_{ij} &= P(X_t = i | X_{t-1} = j) \\ &= P(\alpha \circ X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \\ &= \sum_{m=0}^{\min(i,j)} P(\alpha \circ X_{t-1} = m | X_{t-1} = j) P(\varepsilon_t = i - m) \\ &= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1 - \alpha)^{j-m} \left[\frac{e^{-(i-m)} (i-m)^{(i-m)}}{(i-m)!} \right] \left(\frac{e\mu}{(i-m)} \right)^{\phi(i-m)}, \\ & \quad i, j = 0, 1, 2, \dots \end{aligned} \quad (94)$$

olarak elde edilir.

DP-INAR(1) sürecinin koşullu ortalama, koşullu varyans, ortalama, varyans, kovaryans ve otokorelasyon fonksiyonu sırasıyla

$$E[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon = \alpha X_{t-1} + \mu, \quad (95)$$

$$V[X_t | X_{t-1}] = \alpha(1 - \alpha)X_{t-1} + \sigma_\varepsilon^2 = \alpha(1 - \alpha)X_{t-1} + \frac{\mu}{\phi}, \quad (96)$$

$$\mu_X = E[X_t] = \frac{\mu_\varepsilon}{1 - \alpha} = \frac{\mu}{(1 - \alpha)}, \quad (97)$$

$$\sigma_X^2 = Var[X_t] = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1 - \alpha^2} = \frac{\mu(1 + \alpha\phi)}{(1 - \alpha^2)\phi}, \quad (98)$$

$$\gamma_k = Cov(X_k, X_{k+1}) = \alpha^k \sigma^2, \quad (99)$$

$$\rho_k = Corr(X_k, X_{k+1}) = \alpha^k \quad (100)$$

şeklindedir.

Buna göre X_t 'nin yayılım indeksi:

$$I_X = \frac{I_\varepsilon + \alpha}{1 + \alpha} = \frac{\sigma_X^2}{\mu_X} = \frac{1 + \alpha\phi}{\phi + \alpha\phi} \quad (101)$$

olarak hesaplanır.

Eşitlik (101)'de görüldüğü gibi, DP-INAR(1) sürecinin de, $\phi < 1$ olduğunda aşırı yayılıma, $\phi > 1$ olduğunda yetersiz yayılıma sahip olduğu, $\phi = 1$ olduğunda ise eşit yayılım özelliğine sahip olduğu görülür. Ayrıca, μ parametresi sürecin yayılım indeksini değiştirmez.

3.6. GP-INAR(1) Modeli

DP-INAR(1) modelinin önerildiği çalışma da yine Bourguignon, Rodrigues ve Santos-Neto (2019), eşit yayılım, yetersiz yayılım veya aşırı yayılıma sahip negatif olmayan tam sayı değerli zaman serilerini modellemek için GP-INAR(1) ile gösterilen Genelleştirilmiş Poisson hataları ile Binom inceltmeye dayalı yeni bir INAR(1) süreci önermişlerdir.

Genelleştirilmiş Poisson dağılımı, $0 \leq \phi < 1$ ve $\mu > 0$ parametreleriyle negatif olmayan tam sayılar için tanıtılmış ve kapsamlı bir şekilde incelenmiştir (Consul ve Jain, 1973; Consul, 1989). Genelleştirilmiş Poisson dağılımının olasılık fonksiyonu şu şekilde verilir:

$$P(Z = z) = \frac{\mu(\mu + z\phi)^{z-1} e^{-(\mu+z\phi)}}{z!}, \quad z = 0, 1, 2, \dots \quad (102)$$

$0 < \phi < 1$ için karşılık gelen olasılık yaratıcı fonksiyonu şu şekilde verilir:

$$\varphi_Z(s) = E[s^Z] = e^{\mu[u(s,\phi)-1]} \rightarrow u = se^{\phi(u-1)}. \quad (103)$$

$\phi > 0$ için, Genelleştirilmiş Poisson dağılımı kesikli kendi kendine ayrıştırılabilir dağılımlar sınıfına aittir (Weiß, 2008) ancak $\phi < 0$ için Genelleştirilmiş Poisson dağılımı bu özelliğini kaybeder (Johnson, Kemp ve Kotz, 2005).

Dağılımın ortalaması ve varyansı sırasıyla

$$E[Z] = \frac{\mu}{1 - \phi} \quad (104)$$

ve

$$Var[Z] = \frac{\mu}{(1 - \phi)^3} \quad (105)$$

olarak elde edilir.

Dağılımın ortalama ve varyansına göre yayılım indeksi

$$ID(Z) = \frac{Var[Z]}{E[Z]} = \frac{1}{(1 - \phi)^2} \quad (106)$$

şeklinde hesaplanır.

Dolayısıyla, Genelleştirilmiş Poisson dağılımının $0 < \phi < 1$ olduğunda aşırı yayılıma, $-1 < \phi < 0$ olduğunda yetersiz yayılıma sahip olduğu görülür. $\phi = 0$ olması durumunda ise, Genelleştirilmiş Poisson dağılımı Poisson dağılımına benzer.

X_t , Eşitlik (102)'de verilen olasılık fonksiyonuyla Genelleştirilmiş Poisson dağılımına sahip bağımsız ve aynı dağılımlı bir zaman serisi ise, bu süreç GP-INAR(1) süreci olarak adlandırılır ve

$$\begin{cases} X_t = \alpha \circ X_{t-1} + \varepsilon_t, & t \geq 1, \\ \varepsilon_t \sim \text{Genelleştirilmiş Poisson}(\mu, \phi), \end{cases} \quad (107)$$

şeklinde tanımlanır.

Burada, $0 < \alpha < 1$, $0 \leq \phi < 1$, $\mu > 0$ ve ε_t , $t \geq 1$ için Bernoulli serisi Y_i ve X_{t-1} 'den bağımsızdır.

Böylece, sürecin geçiş olasılıkları

$$\begin{aligned} P_{ij} &= P(X_t = i | X_{t-1} = j) \\ &= P(\alpha \circ X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \\ &= \sum_{m=0}^{\min(i,j)} P(\alpha \circ X_{t-1} = m | X_{t-1} = j) P(\varepsilon_t = i - m) \\ &= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1 - \alpha)^{j-m} \frac{\mu(\mu + (i - m)\phi)^{i-m-1} e^{-(\mu + (i-m)\phi)}}{(i - m)!}, \quad i, j = 0, 1, 2, \dots \end{aligned} \quad (108)$$

biçiminde elde edilir.

GP-INAR(1) sürecinin koşullu ortalama, koşullu varyans, ortalama, varyans, kovaryans ve otokorelasyon fonksiyonu sırasıyla

$$E[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon = \alpha X_{t-1} + \frac{\mu}{1 - \phi}, \quad (109)$$

$$V[X_t | X_{t-1}] = \alpha(1 - \alpha)X_{t-1} + \sigma_\varepsilon^2 = \alpha(1 - \alpha)X_{t-1} + \frac{\mu}{(1 - \phi)^3}, \quad (110)$$

$$\mu_X = E[X_t] = \frac{\mu_\varepsilon}{1 - \alpha} = \frac{\mu}{(1 - \alpha)(1 - \phi)}, \quad (111)$$

$$\sigma_X^2 = Var[X_t] = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1 - \alpha^2} = \frac{\mu[1 + \alpha(1 - \phi)^2]}{(1 - \alpha^2)(1 - \phi)^3}, \quad (112)$$

$$\gamma_k = Cov(X_k, X_{k+1}) = \alpha^k \sigma^2, \quad (113)$$

$$\rho_k = Corr(X_k, X_{k+1}) = \alpha^k \quad (114)$$

olarak hesaplanır.

Buna göre, X_t 'nin yayılım indeksi ortalama ve varyansa dayalı olarak

$$I_X = \frac{I_\varepsilon + \alpha}{1 + \alpha} = \frac{\sigma_X^2}{\mu_X} = \frac{1 + \alpha(1 - \phi)^2}{(1 + \alpha)(1 - \phi)^2} \quad (115)$$

Şeklinde olur. Dolayısıyla, bu sürecin $\phi = 0$ olduğunda eşit yayılıma, $\phi < 0$ olduğunda yetersiz yayılıma ve $\phi > 0$ olduğunda aşırı yayılıma sahip olduğu sonucuna ulaşılır. Ayrıca, μ parametresinin sürecin yayılım indeksini değiştirmedeği görülmektedir.

3.7. ZIP-INAR(1) Modeli

Jazi, Jones ve Lai (2012b) tarafından aşırı yayılıma sahip kesikli zaman serilerini modellemek için Sıfır Yiğmal Poisson hataları ile yeni bir birinci dereceden tam sayı değerli otoregresyon süreci tanıtılmıştır. Genel olarak, aşırı yayılımın iki nedeni olduğu düşünülür. Bunlardan ilki kitlenin heterojenliği iken ikincisi ise sıfırların fazla olmasıdır (Wagh ve Kamalja, 2018). Tam sayı değerli zaman serilerinde de sıfır sayım oranının Poisson dağılımından beklenilenden daha fazla olması durumuna sıklıkla rastlanılmaktadır (Jazi, Jones ve Lai, 2012b). Sıfır Yiğmal Poisson dağılımı, Poisson dağılımının bir genellemesidir (Johnson, Kemp ve Kotz, 1992; Lambert, 1992). Poisson dağılımı ortalama parametresi ile tek parametrelidir. Ancak, Sıfır Yiğmal

Poisson dağılımı ortalama parametresi dışında, dağılım parametresi olarak bilinen bir parametre ile daha karakterize edilir (Wagh ve Kamalja, 2018).

Bir rastlantı değişkeni olan Z 'nin, olasılık fonksiyonu

$$P(Z = z) = \begin{cases} p + (1 - p)e^{-\lambda}, & z = 0 \\ (1 - p) \frac{e^{-\lambda} \lambda^z}{z!}, & z = 1, 2, 3, \dots \end{cases}; 0 < p < 1; \lambda > 0 \quad (116)$$

ise $ZIP(p, \lambda)$ olarak gösterilen, p ve λ parametreleriyle Sıfır Yığılmalı Poisson (ZIP) dağılımına sahip olduğu söylenilebilir. Bir başka gösterimle,

$$P(Z = z) = pI_{\{0\}}(z) + (1 - p) \frac{e^{-\lambda} \lambda^z}{z!}; z = 1, 2, 3, \dots; 0 < p < 1; \lambda > 0 \quad (117)$$

olarak elde edilir. Burada gösterge fonksiyonu $I_A(x)$, eğer $x \in A$ sifira eşitse, 1'e eşittir.

Sıfır Yığılmalı Poisson dağılımının olasılık yaratıcı fonksiyonu şu şekildedir:

$$\varphi_Z(s) = E[s^Z] = p + (1 - p)e^{\lambda(s-1)}, \quad 0 < s < 1. \quad (118)$$

Dağılımın olasılık fonksiyonundan, sırasıyla ortalama ve varyansı ise

$$E[Z] = \lambda(1 - p) \quad (119)$$

ve

$$Var[Z] = \lambda(1 - p)(1 + p\lambda) \quad (120)$$

şeklinde elde edilir.

Dağılımın ortalama ve varyansına göre yayılım indeksi

$$ID(Z) = \frac{Var[Z]}{E[Z]} = (1 + p\lambda) \quad (121)$$

olarak hesaplanır.

Eşitlik (121)'den dağılımın sadece $p > 0$ ise aşırı yayılıma sahip olacağı görülür. Ancak, $P(Z = 0) \geq 0$ olması koşulu altında p parametresi negatif değerler de alabilir. Bu da şu anlama gelir:

$$\max \left\{ -1, -\frac{e^{-\lambda}}{1 - e^{-\lambda}} \right\} \leq p \leq 0.$$

Bu durum yetersiz yayılıma sahip bir dağılım olan Sıfır Sönük (deflated) Poisson dağılıma karşılık gelir.

X_t , Eşitlik (1)'e göre tanımlanmış bir INAR(1) süreci ve ε_t hataları bağımsız ve aynı dağılmış bir $ZIP(p, \lambda)$ zaman serisi ise, bu durum bir ZIP-INAR(1) modelini ifade eder, yani;

$$\begin{cases} X_t = \alpha \circ X_{t-1} + \varepsilon_t, & t \geq 1, \\ \varepsilon_t \sim \text{Sıfır Yiğmalı Poisson}(p, \lambda) \end{cases} \quad (122)$$

şeklindedir.

Burada, $0 < \alpha < 1$, $0 < p < 1$, $\lambda > 0$ ve ε_t , $t \geq 1$ için Bernoulli serisi Y_i ve X_{t-1} 'den bağımsızdır.

ZIP-INAR(1) modeli,

$$\begin{aligned} P_{ij} &= P(X_t = i | X_{t-1} = j) \\ &= P(\alpha \circ X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \end{aligned}$$

$$\begin{aligned}
&= \sum_{m=0}^{\min(i,j)} P(\alpha \circ X_{t-1} = m | X_{t-1} = j) P(\varepsilon_t = i - m) \\
&= \sum_{m=0}^{\min(i,j)} \binom{j}{m} \alpha^m (1 - \alpha)^{j-m} \left[p I_{\{0\}}(i - m) + (1 - p) \frac{e^{-\lambda} \lambda^{(i-m)}}{(i - m)!} \right], \quad i, j = 0, 1, 2, \dots
\end{aligned} \tag{123}$$

geçiş olasılıklarına sahip durağan tam sayı değerli bir Markov zinciri oluşturur.

ZIP-INAR(1) sürecinin koşullu ortalama, koşullu varyans, ortalama, varyans, kovaryans ve otokorelasyon fonksiyonu ise sırasıyla

$$E[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon = \alpha X_{t-1} + \lambda(1 - p), \tag{124}$$

$$V[X_t | X_{t-1}] = \alpha(1 - \alpha)X_{t-1} + \sigma_\varepsilon^2 = \alpha(1 - \alpha)X_{t-1} + \lambda(1 - p)(1 + p\lambda), \tag{125}$$

$$\mu_X = E[X_t] = \frac{\mu_\varepsilon}{1 - \alpha} = \frac{\lambda(1 - p)}{1 - \alpha}, \tag{126}$$

$$\sigma_X^2 = Var[X_t] = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1 - \alpha^2} = \frac{\lambda(1 - p)(1 + \alpha + p\lambda)}{1 - \alpha^2}, \tag{127}$$

$$\gamma_k = Cov(X_k, X_{k+1}) = \alpha^k \sigma^2, \tag{128}$$

$$\rho_k = Corr(X_k, X_{k+1}) = \alpha^k \tag{129}$$

şeklindedir. Buna göre, X_t 'nin yayılım indeksi

$$I_X = \frac{I_\varepsilon + \alpha}{1 + \alpha} = \frac{\sigma_X^2}{\mu_X} = \frac{1 + \alpha + p\lambda}{1 + \alpha} \tag{130}$$

olur. Bu durumda, Sıfır Yığılmalı Poisson dağılımının yayılım indeksinde olduğu gibi $p > 0$ olduğunda X_t aşırı yayılır. p parametresinin negatif olduğu durumlarda ise X_t yetersiz yayılacaktır.

4. GERÇEK VERİ UYGULAMALARI

Bu bölümde, INAR(1) süreçlerinin uygulaması iki farklı gerçek veri seti üzerinde gerçekleştirilmiştir. Her iki veri seti için de Bölüm 3'te bahsedilen INAR(1) model sonuçları karşılaştırmalı olarak incelenmiştir. Ayrıca, karşılaştırmanın sonuçları yine bu bölümde tartışılmıştır.

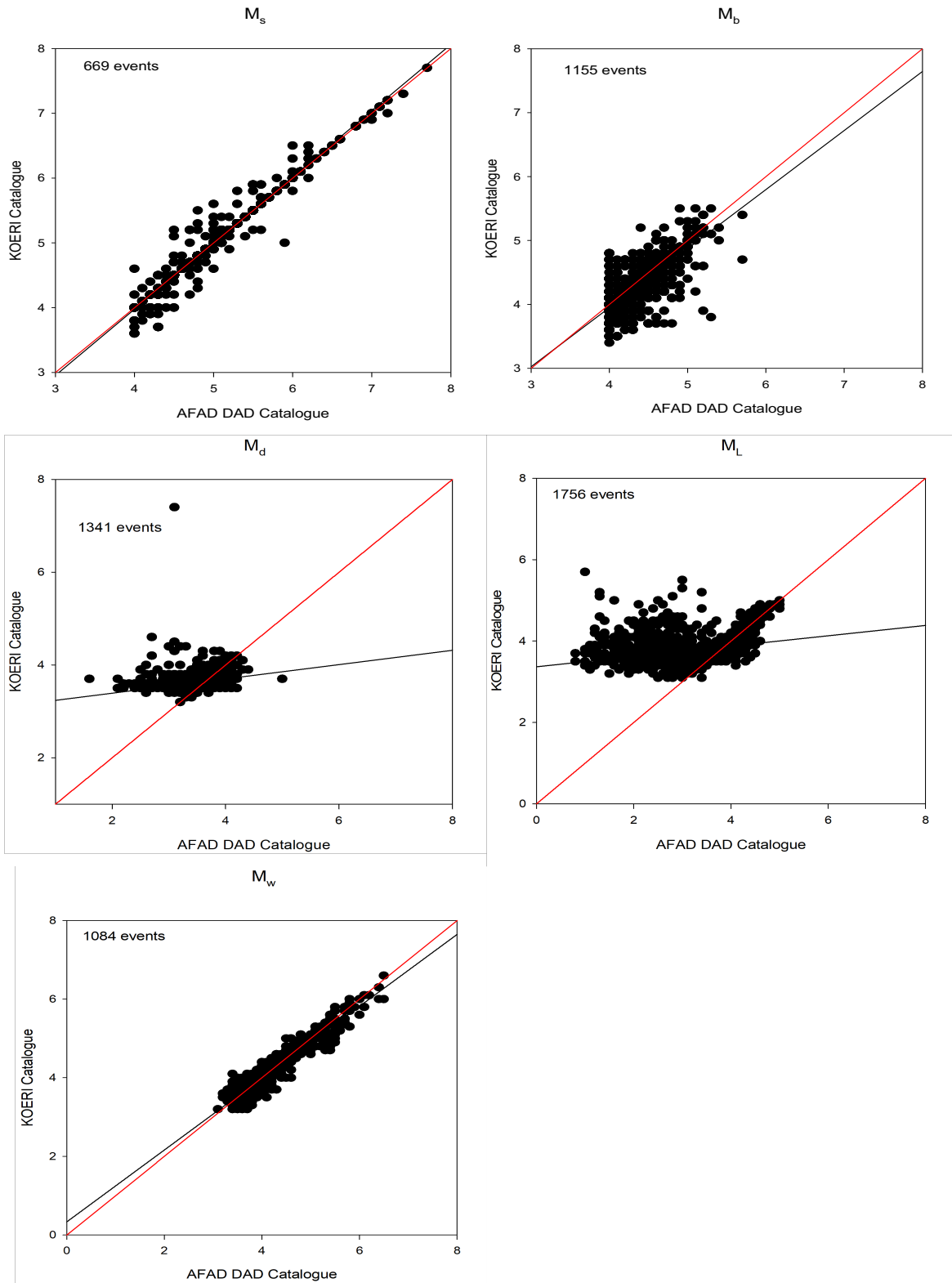
Uygulamalarda kullanılan deprem verileri, TÜBİTAK 1001 girişiminin bir parçası olan "Yapay Zekâ ve Olasılıksal Model Tabanlı Deprem Tehlike Haritası" projesinden elde edilmiştir ve Şekil 2'de gösterilen, Ege Bölgesi ve çevresindeki 200 km'lik tampon bölgeyi kapsamaktadır. Şekilde de görüldüğü üzere bölgede çok sayıda bağımsız fay hattı bulunmaktadır. Bu nedenle, bölgenin sismik açıdan çeşitli analizlerle incelenmesi hayati önem taşımaktadır.



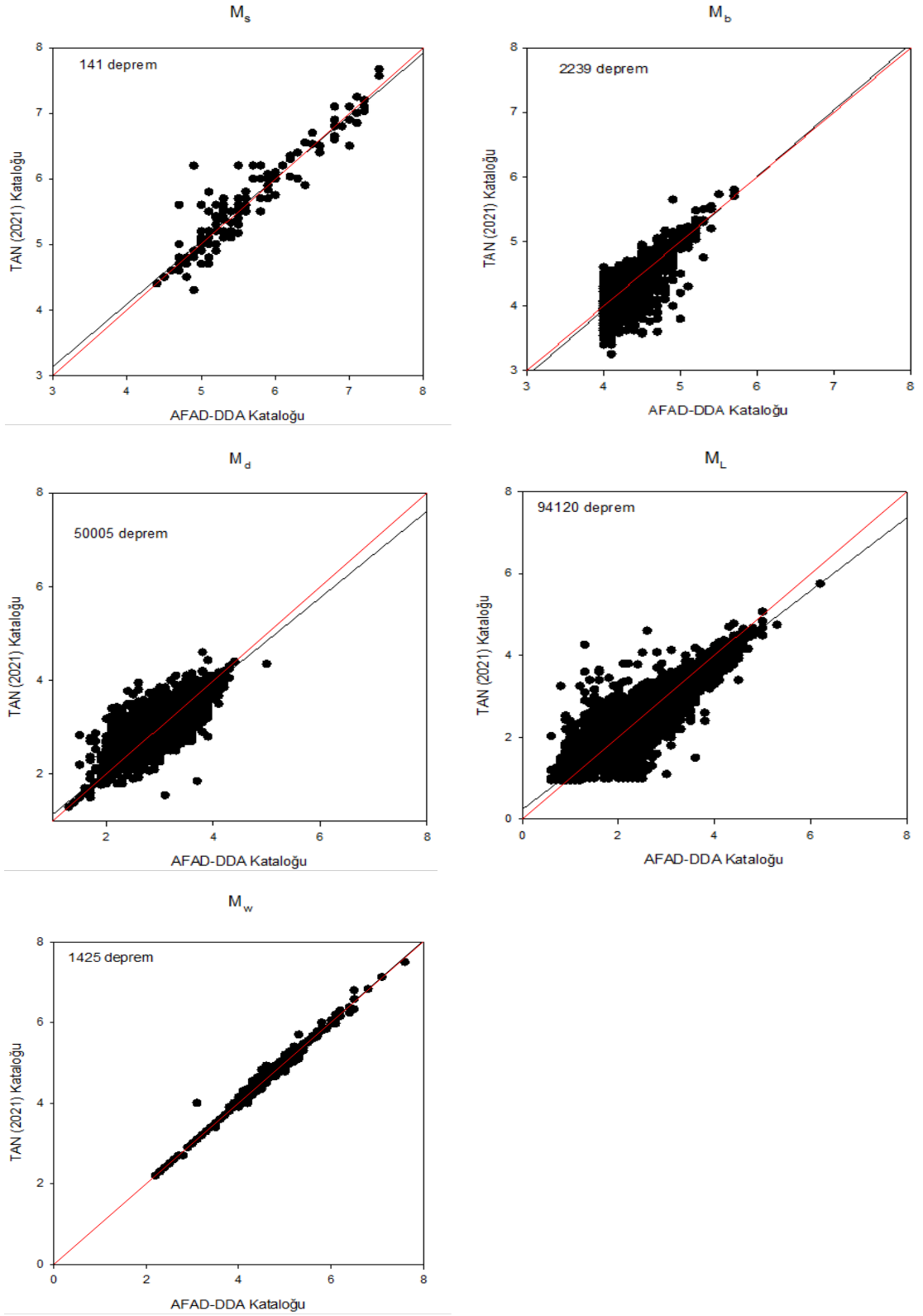
Şekil 2: Uygulama alanı

Çalışma kapsamında kullanılan veriler, Afet ve Acil Durum Yönetimi Başkanlığı (AFAD), Boğaziçi Üniversitesi Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü Bölgesel Deprem Tsunami İzleme Merkezi (KOERI) ve Tan (2021) kataloğu olmak üzere üç Türk deprem kataloğu incelenerek elde edilmiştir. Temel alınan katalog AFAD-DDA kataloğudur ve tutarlılığı sağlamak amacıyla Kadiroğlu ve Kartal (2016), KOERI ve Tan (2021)'in büyüklük dönüşüm ilişkileri uygulanmıştır. Bu dönüşümler

sayesinde oluşturulan deprem kataloğu homojenize edilmiş ve tüm büyüklükler tek bir ölçek (Moment Büyüklüğü (M_w)) ile ifade edilmiştir.

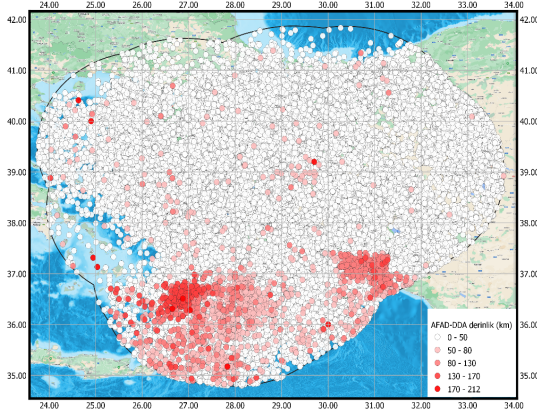


Şekil 3: AFAD-DDA ve KOERI deprem katalogları arasında eşleşen depremlerin deprem büyüklükleri açısından karşılaştırmaları

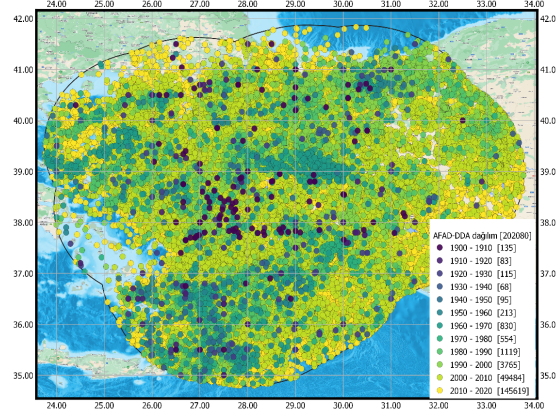


Őekil 4: AFAD-DDA ve Tan (2021) deprem katalogları arasında eŐleŐen depremlerin deprem bŕyŕklŕkleri aŐısından karŐılaŐtırmaları

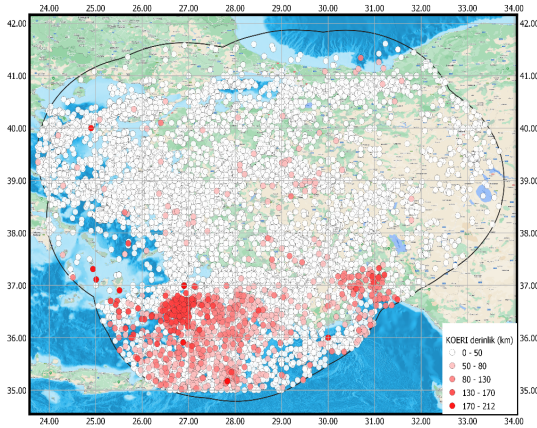
Şekil 3 ve Şekil 4’te AFAD-DDA ve Koeri deprem katalogları arasında eşleşen 7749 deprem ile AFAD-DDA ve Tan (2021) katalogları arasında eşleşen 153533 depremin büyüklükleri açısından karşılaştırmaları verilmiştir. Deprem katalogları eşleştirme kriterleri, depremler arası zaman farkı 3 dakikadan az olacak şekilde ve depremler arası konum farkı 0.5 dereceden az olacak şekilde belirlenmiştir.



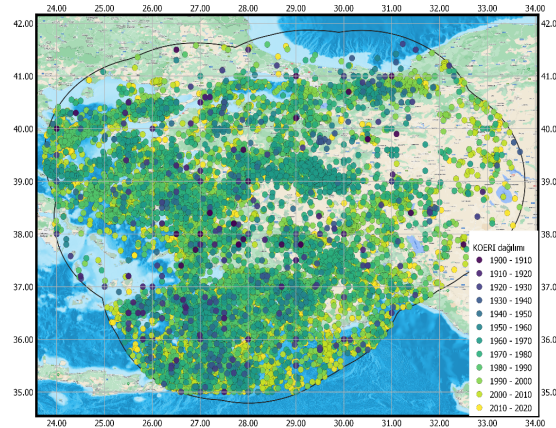
Şekil 5: AFAD-DDA katalogu depremleri derinlik bilgisi



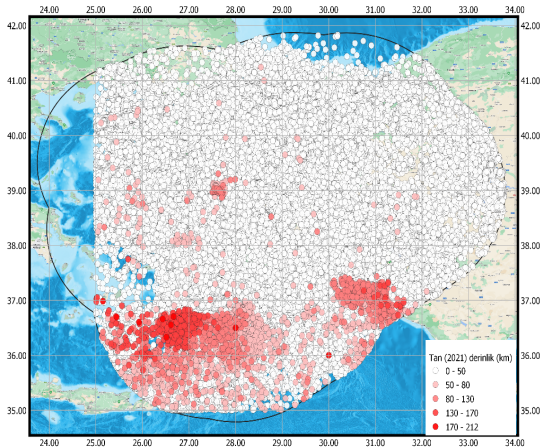
Şekil 8: AFAD-DDA katalogu depremleri oluş zamanları



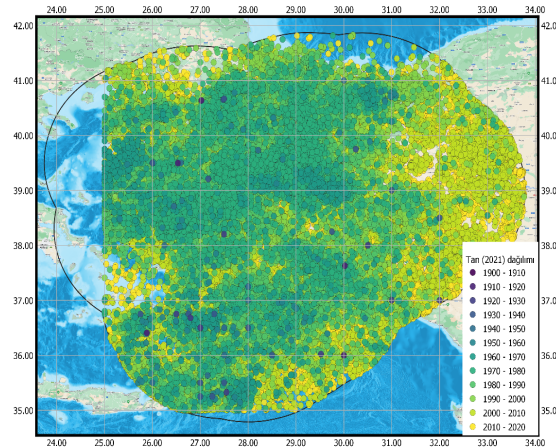
Şekil 6: KOERI katalogu depremleri derinlik bilgisi



Şekil 9: KOERI katalogu depremleri oluş zamanları



Şekil 7: Tan (2021) katalogu depremleri derinlik bilgisi

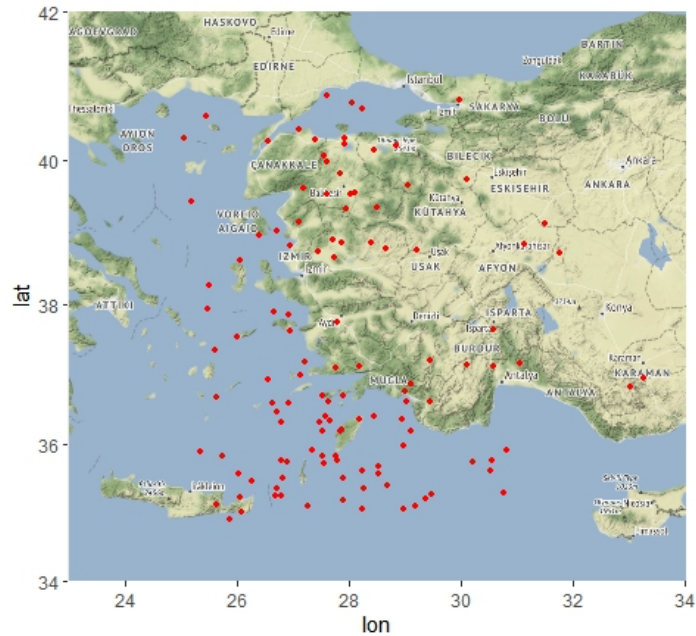


Şekil 10: Tan (2021) katalogu depremleri oluş zamanları

Şekil 5 ve Şekil 10 arasındaki şekillerde, incelenen 3 Türk deprem kataloğuna ilişkin depremlerin derinlik bilgisi ve oluş zamanları harita üzerinde gösterilmiştir. AFAD-DDA kataloğu ve KOERI kataloğu 1900 ile 2021 yılları arasında gerçekleşen depremleri kapsarken, Tan (2021) kataloğu 1905 ile 2018 yılları arasında gerçekleşen depremleri kapsamaktadır. AFAD-DDA kataloğunda 202.080'i deprem çalışma sahası içinde olmak üzere toplam 333.381 deprem, KOERI kataloğunda 11.772'si deprem çalışma sahası içinde olmak üzere toplam 18.163 deprem ve Tan (2021) kataloğunda 252.594'ü deprem çalışma sahası içinde olmak üzere toplam 337.429 deprem kaydedilmiştir.

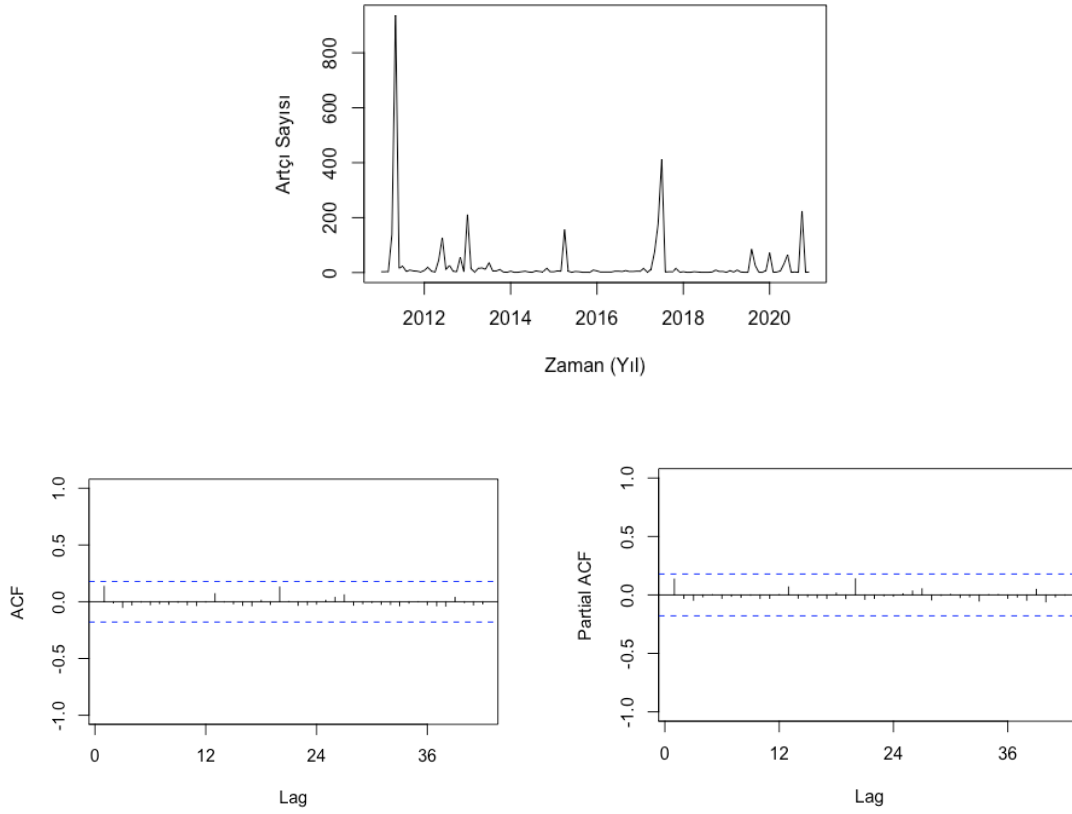
4.1. Artçı Sayısı Verisi

Uygulama kapsamında kullanılan veriler, 2011 Ocak ayı ile 2020 Aralık ayı içerisinde, belirlenen bölgede meydana gelen, aylık en büyük moment büyüklüğüne (M_w) sahip ana depremlerin artçı sayısı verileridir. Elde edilen veriler tam sayı değerli zaman serisi tanımına uymaktadır. Bu nedenle, çalışmanın amacı elde edilen seriyi modellemek ve seri için ileriye yönelik tahminlerde bulunabilmek için en uygun birinci dereceden tam sayı değerli otoregresyon modelini belirlemektir. Uygulamada kullanılan veriler, Şekil 11'de harita üzerinde gösterilmiştir.



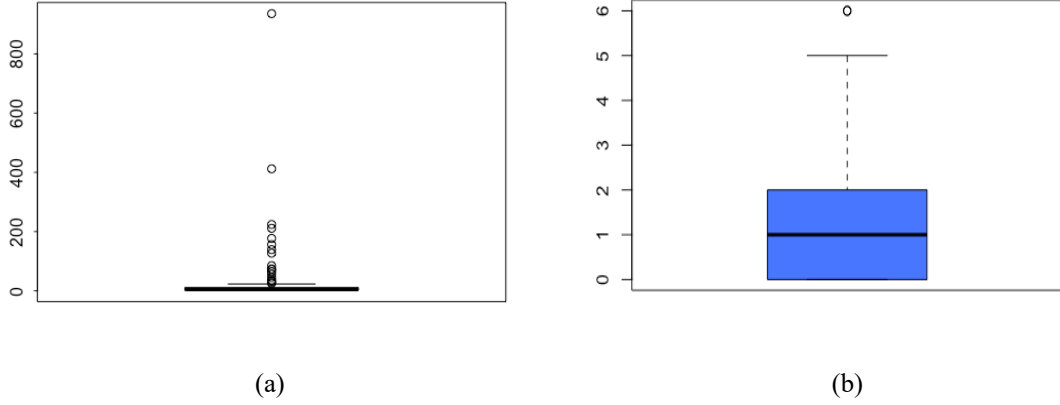
Şekil 11: Harita üzerinde uygulama verileri

Veriler, Ocak 2011'den başlayıp Aralık 2020'de sona eren 120 gözlemden ($T = 120$) oluşmaktadır. Verilere ait zaman serisi grafiği, ACF ve PACF grafikleri Şekil 12'de verilmiştir.



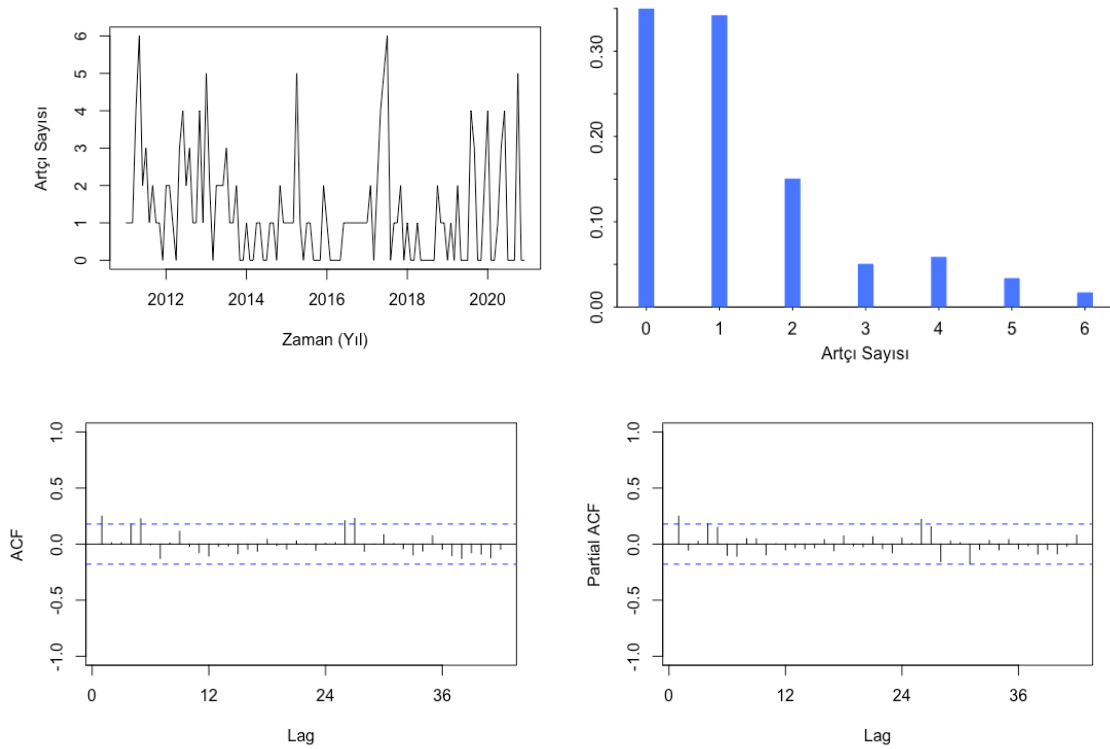
Şekil 12: Artçı sayısı verisinin zaman serisi, ACF ve PACF grafikleri

Zaman serisi grafiğinden de görüleceği üzere, seride çok fazla uç değer vardır ve uç değerler otoregresyon modellerinin parametre tahminlerinin sağlıklı bir şekilde yapılmasına engel olmaktadır. Bu sorunun önüne geçebilmek amacıyla, seriye logaritmik dönüşüm uygulanmıştır. Şekil 13'te uç değerlerin varlığını daha net görebilmek amacıyla orijinal serinin ve logaritmik dönüşüm uygulanan serinin kutu grafikleri gösterilmiştir.



Şekil 13: (a) Orijinal seri (b) logaritmik dönüşüm uygulanan seri için kutu grafikleri

Analizlerin daha sağlıklı yapılabilmesi için, uygulamaya bu aşamadan sonra logaritmik dönüşüm uygulanan seri ile devam edilecektir. Şekil 14'te serinin zaman serisi grafiği, çubuk grafiği, otokorelasyon fonksiyonu ve kısmi otokorelasyon fonksiyonu grafikleri verilmiştir. ACF grafiğinde sadece ilk gecikmeye ait ilişki istatistiksel olarak anlamlıdır. Buradan, serinin durağan olduğu yorumunu yapabiliriz. Bu aşamada modelleme işlemine geçilebilir ve artçı sayısı serisi için birinci dereceden otoregresyon AR(1) modellerin denenebileceği sonucuna varabiliriz.



Şekil 14: Logaritmik dönüşüm uygulanan artçı sayısı serisinin zaman serisi, çubuk, ACF ve PACF grafikleri

120 gözlemden oluşan artçı sayısı serisinin ilk 114 gözlemi, otoregresyon modellerinin parametre tahminlerini gerçekleştirmek ve model kriterlerini karşılaştırmak amacıyla kullanılmıştır. Son 6 gözlemi ise ileriye yönelik tahminler yapmak ve süreçlerin tahmin etme performansını karşılaştırmak için seriden kesilmiştir (Huang and Zhu, 2021).

Tablo 1’de serinin ilk 114 gözlemine ait tanımlayıcı istatistikler yer almaktadır.

Tablo 1: Logaritmik dönüşüm uygulanan artçı sayısı serisi tanımlayıcı istatistikleri

Minimum	Medyan	Ortalama	Varyans	Çarpıklık	Basıklık	Maksimum
0	1	1,316	2,023	1,358	4,454	6

Artçı sayısı serisinin ortalaması ve varyansı sırasıyla $\bar{X} = 1,316$ ve $S_{\bar{X}}^2 = 2,023$ olarak hesaplanmıştır. Dolayısıyla, yayılım indeksi $\hat{I}_{\bar{X}} = \frac{S_{\bar{X}}^2}{\bar{X}} = 1,537$ olarak elde edilir. Schweer ve Weiß’in (2014) aşırı yayılım testine göre, testin kritik değeri 1,1994’tür. Serinin yayılım indeksi bu kritik değeri aşmaktadır, bu nedenle serinin aşırı yayılıma sahip olduğunu ve eşit yayılım özelliği gösteren P-INAR(1) modelinin seri için uygun bir model olmadığını söyleyebiliriz.

Seri için INAR(1) modellerinin uygun olduğunu gösterdikten sonra, P-INAR(1) modeli, G-INAR(1) modeli, PL-INAR(1) modeli, NB-INAR(1) modeli, DP-INAR(1) modeli, GP-INAR(1) modeli ve ZIP-INAR(1) modeli karşılaştırmaları için parametre tahminleri koşullu maksimum olabilirlik tahmin yöntemi ile yapılmıştır.

Seriye uygulanan INAR(1) modellerinin ilk üçü iki parametrelidir ve son dördü ise üç parametrelidir. Parametre tahmininden sonra ilk olarak Akaike Bilgi Kriteri ve Bayesyen Bilgi Kriteri hesaplanmış, PL-INAR(1) modelinin AIC ve BIC değerinin diğer modellerden daha küçük olduğu görülmüştür.

Sonrasında, gerçek seriler ile modellerden tahmin edilen tahmin serileri için hata kareler ortalamasının karekökü (HKOK) hesaplanmıştır. $HKOK = \sqrt{\frac{1}{6} \sum_{k=1}^6 (X_{t+k} - \hat{X}_{t+k})^2}$

şeklinde tanımlanır. Hesaplamalar sonucu, ZIP-INAR(1) modelinin HKOK değerinin en küçük olduğu görülmüştür.

Diğer bir karşılaştırma yöntemi olarak, tahmin edilen parametreler ile Bölüm 3'te her bir model için açıklanan ortalama ve varyans formülleri ile ortalama ve varyans değerleri hesaplanmıştır. NB-INAR(1) model parametreleriyle hesaplanan ortalama değerinin ve DP-INAR(1) model parametreleriyle hesaplanan varyans değerinin gerçek ortalama ve varyansa en yakın değerler olduğu görülmüştür.

Son olarak, koşullu ortalama yöntemi $\left(E[X_{t+k}|X_t] = \alpha X_t + \mu_\varepsilon \frac{1-\alpha^k}{1-\alpha}\right)$ ile seriden kesilen son 6 veri tahmin edilmiştir. Yine modeller arasında bir karşılaştırma yapabilmek amacıyla, serilerin son 6 ayı için HKOK hesaplanmıştır. Hesaplama sonucunda, G-INAR(1) modelinin HKOK değerinin en küçük olduğu görülmüştür.

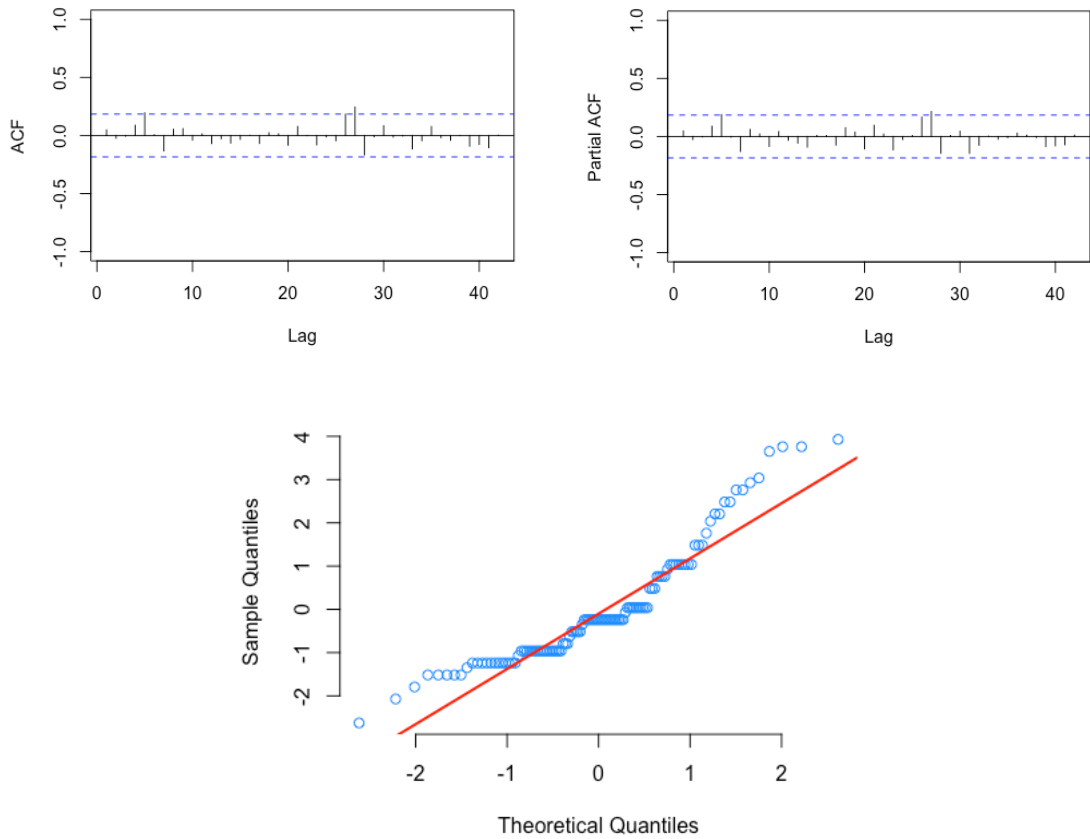
Tüm sonuçlar, HKOK hesaplamaları ile AIC ve BIC değerleri diğer modellere göre küçük elde edilen değerler ile ortalama ve varyans hesaplamalarında gerçek ortalama ve varyansa en yakın hesaplanan değerler koyu renkle gösterilmek üzere Tablo 2' de gösterilmiştir.

Tablo 2: Logaritmik dönüşüm uygulanan artçı sayısı serisi için CML tahminleri, AIC ve BIC bilgi kriterleri, ortalama, varyans ve HKOK hesaplamaları

Model	Parametreler	AIC	BIC	HKOK (tahmin serisi)	Ortalama	Varyans	HKOK (öngörü)
P-INAR(1)	$\alpha = 0.246$ $\lambda = 1.001$	353.705	359.177	1.354	1.328	1.328	2.040
G-INAR(1)	$\alpha = 0.288$ $p = 0.514$	347.968	353.440	1.351	1.328	2.584	1.851
PL-INAR(1)	$\alpha = 0.277$ $\theta = 1.461$	347.163	352.636	1.352	1.331	2.168	2.061
NB-INAR(1)	$\alpha = 0.253$ $r = 1.783$ $p = 0.643$	348.599	356.808	1.354	1.325	1.912	2.047
DP-INAR(1)	$\alpha = 0.272$ $\mu = 0.835$ $\phi = 0.519$	348.306	356.514	1.358	1.147	1.981	2.015
GP-INAR(1)	$\alpha = 0.249$ $\mu = 0.799$ $\phi = 0.198$	348.733	356.942	1.354	1.327	1.915	2.042
ZIP-INAR(1)	$\alpha = 0.307$ $\lambda = 1.380$ $p = 0.332$	349.148	357.356	1.350	1.330	1.797	2.080

Farklı karşılaştırma yöntemlerinde farklı modeller uygun model olarak görülse de model karşılaştırmalarında Akaike Bilgi Kriteri ve Bayesyen Bilgi Kriteri en doğru sonucu veren yöntemler olduğu söylenebilir. AIC ve BIC değerleri en küçük hesaplanan model ise PL-INAR(1) modelidir ve artçı sayısı verisine en uygun model olarak belirlenebilir.

Son olarak, belirlenen modelin gerçekten yeterli bir model olup olmadığının kontrol edilmesi için modelden hesaplanan hata serisi analiz edilmiştir. Bu doğrultuda, hata serisinin herhangi bir otokorelasyona sahip olup olmadığını görebilmek için serinin ACF ve PACF grafikleri incelenmiş ve yine serinin normal dağılıma yakınsayıp yakınsamadığını yorumlayabilmek için de Q-Q grafiği incelenmiştir. ACF, PACF ve Q-Q grafikleri Şekil 15’te sırasıyla gösterilmektedir.

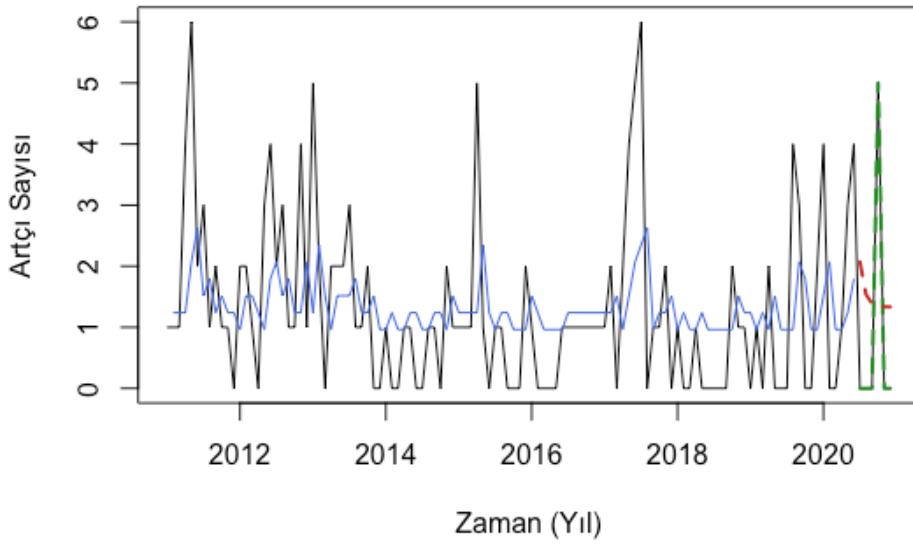


Şekil 15: PL-INAR(1) modeli uygulanarak hesaplanan hataların ACF, PACF ve Q-Q grafikleri

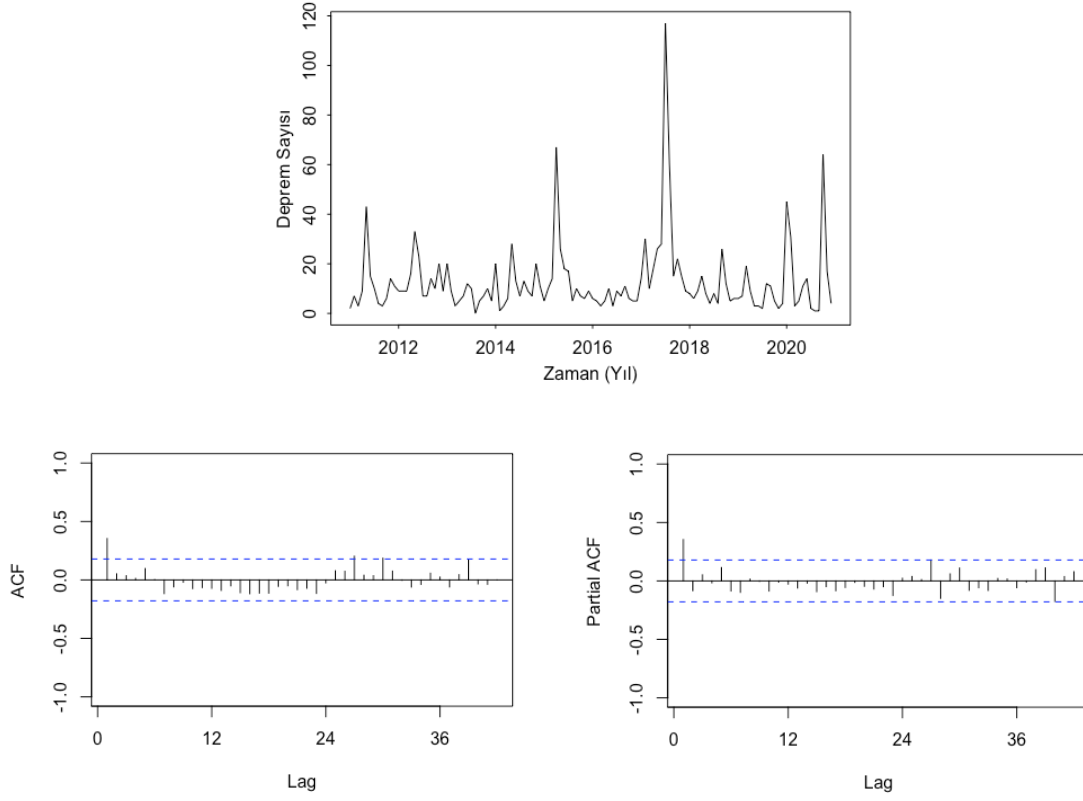
Hatalara ait ACF ve PACF grafiğinde gecikmelerdeki ilişkiler önemli değildir, yani sınırı aşmamaktadır. Bu durum, hatalarda otokorelasyon sorunu olmadığı anlamına

gelmektedir. Dolayısıyla, PL-INAR(1) modelinin artçı sayısı serisi için incelenen diğer modeller arasında en uygun model olduğunu söyleyebiliriz.

Şekil 16’da PL-INAR(1) modelinden elde edilen tahmin serisi ile artçı sayısı serisinin birlikte grafiği verilmiştir. Siyah renk artçı sayısı serisini, mavi renk ise tahmin serisini ifade etmektedir. Ayrıca yeşil renk ileriye yönelik tahminler yapmak ve modelin tahmin etme performansını test etmek için seriden kesilen son 6 veriyi gösterirken, kırmızı renk koşullu ortalama tahmin yöntemi ile tahmin edilen veriyi göstermektedir.

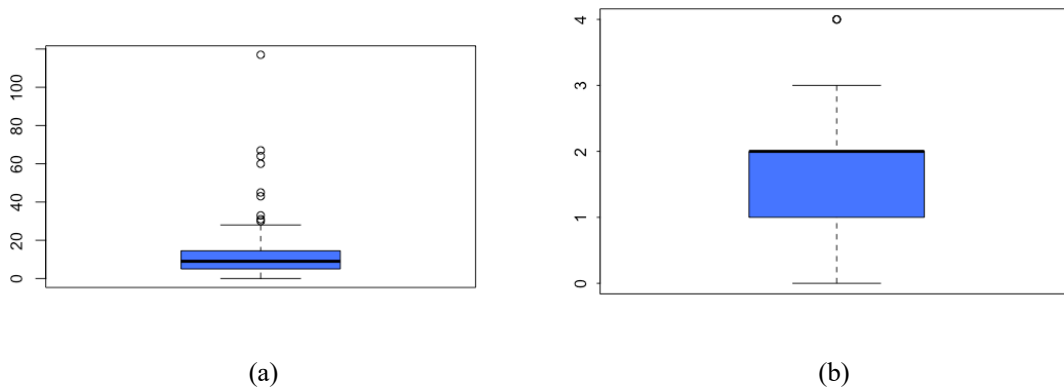


Şekil 16: Logaritmik dönüşüm uygulanan seri ve tahmin serisinin birlikte grafiği



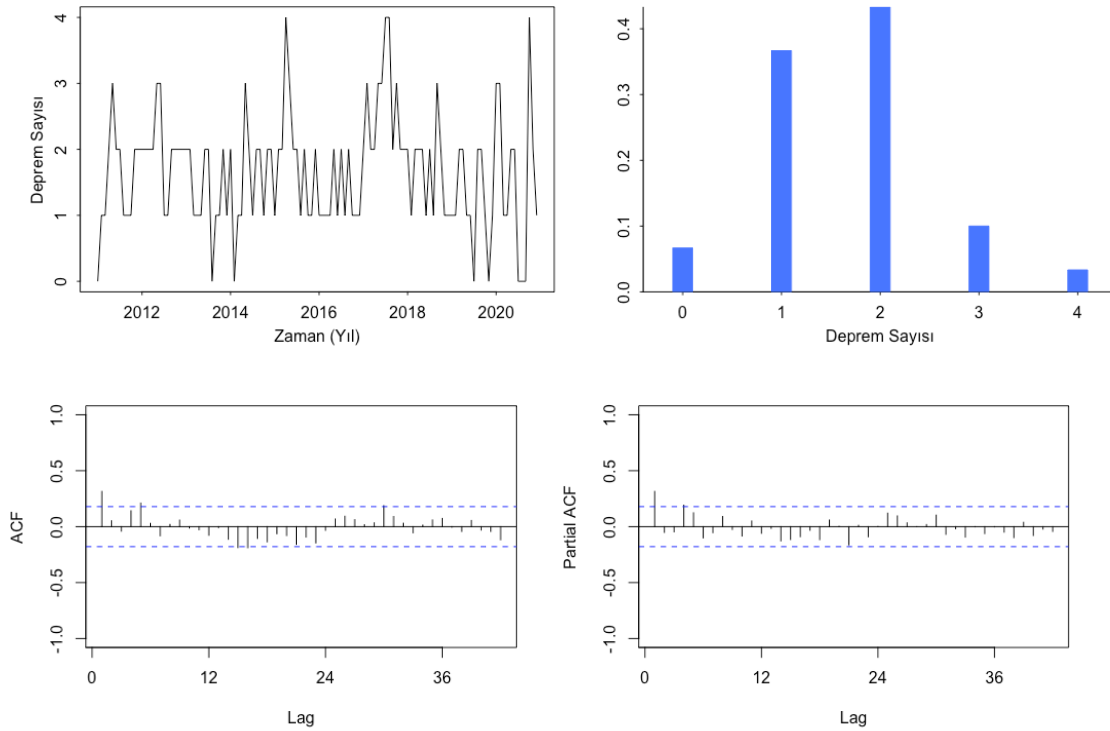
Şekil 18: Deprem sayısı serisinin zaman serisi, ACF ve PACF grafikleri

Zaman serisi grafiğinden de görüleceği üzere, seride çok fazla uç değer vardır ve uç değerler otoregresyon modellerinin parametre tahminlerinin sağlıklı bir şekilde yapılmasına engel olmaktadır. Bu sorunun önüne geçebilmek amacıyla, bir önceki uygulamada olduğu gibi seriye logaritmik dönüşüm uygulanmıştır. Şekil 19’da uç değerlerin varlığını daha net görebilmek amacıyla orijinal serinin ve logaritmik dönüşüm uygulanan serinin kutu grafikleri gösterilmiştir.



Şekil 19: (a) Orijinal seri (b) logaritmik dönüşüm uygulanan seri için kutu grafikleri

Analizlerin daha sağlıklı yapılabilmesi için, uygulamaya bu aşamadan sonra dönüşüm uygulanan seri ile devam edilecektir. Şekil 20’de serinin zaman serisi grafiği, çubuk grafiği, otokorelasyon fonksiyonu ve kısmi otokorelasyon fonksiyonu grafikleri verilmiştir. ACF grafiğinde sadece ilk gecikmedeki ilişki istatistiksel olarak anlamlıdır. Buradan, serinin durağan olduğu yorumunu yapabiliriz. Dolayısıyla, modelleme aşamasına geçilebilir ve artçı sayısı serisi için birinci dereceden otoregresyon AR(1) modelleri denenebilir.



Şekil 20: Logaritmik dönüşüm uygulanan deprem sayısı serisinin zaman serisi, çubuk, ACF ve PACF grafikleri

120 gözlemden oluşan deprem sayısı serisinin ilk 114 gözlemi, otoregresyon modellerinin parametre tahminlerini gerçekleştirmek ve model kriterlerini karşılaştırmak amacıyla kullanılmıştır. Son 6 gözlemi ise ileriye yönelik tahminler yapmak ve süreçlerin tahmin etme performansını karşılaştırmak için seriden kesilmiştir. Tablo 3’te serinin ilk 114 gözlemine ait tanımlayıcı istatistikler yer almaktadır.

Tablo 3: Logaritmik dönüşüm uygulanan deprem sayısı serisi tanımlayıcı istatistikleri

Minimum	Medyan	Ortalama	Varyans	Çarpıklık	Basıklık	Maksimum
0	2	1,693	0,675	0,425	3,361	4

Deprem sayısı serisinin ortalaması ve varyansı sırasıyla $\bar{X} = 1,693$ ve $S_{\bar{X}}^2 = 0,675$ olarak hesaplanmıştır. Dolayısıyla yayılım indeksi $\hat{I}_{\bar{X}} = \frac{S_{\bar{X}}^2}{\bar{X}} = 0,399$ olarak elde edilir. Serinin yayılım indeksi 1'den küçük bir değerdir, bu nedenle serinin yetersiz yayılıma sahip olduğunu ve eşit yayılım özelliği gösteren P-INAR(1) modelinin seri için uygun bir model olmadığını söyleyebiliriz.

Seri için INAR(1) modellerinin uygun olduğunu gösterdikten sonra, P-INAR(1) modeli, G-INAR(1) modeli, PL-INAR(1) modeli, NB-INAR(1) modeli, DP-INAR(1) modeli, GP-INAR(1) modeli ve ZIP-INAR(1) modeli karşılaştırmaları için parametre tahminleri koşullu maksimum olabilirlik tahmin yöntemi ile yapılmıştır.

Seriye uygulanan INAR(1) modellerinin ilk üçü iki parametrelidir ve son dördü ise üç parametrelidir. Parametre tahmininden sonra ilk olarak Akaike Bilgi Kriteri ve Bayesyen Bilgi Kriteri hesaplanmış, DP-INAR(1) modelinin AIC ve BIC değerinin diğer modellerden daha küçük olduğu görülmüştür.

Sonrasında, gerçek seriler ile modellerden tahmin edilen tahmin serileri için HKOK hesaplanmıştır. Hesaplamalar sonucu, ZIP-INAR(1) modelinin HKOK değerinin en küçük olduğu görülmüştür.

Diğer bir karşılaştırma yöntemi olarak, tahmin edilen parametreler ile Bölüm 3'te her bir model için açıklanan ortalama ve varyans formülleri ile ortalama ve varyans değerleri hesaplanmıştır. GP-INAR(1) model parametreleriyle hesaplanan ortalama değerinin ve ZIP-INAR(1) model parametreleriyle hesaplanan varyans değerinin gerçek ortalama ve varyansa en yakın değerler olduğu görülmüştür.

Son olarak, koşullu ortalama yöntemi $\left(E[X_{t+k}|X_t] = \alpha X_t + \mu_\varepsilon \frac{1-\alpha^k}{1-\alpha}\right)$ ile seriden kesilen son 6 veri tahmin edilmiştir. Yine modeller arasında bir karşılaştırma yapabilmek amacıyla, verilerin son 6 ayı için hata kareler ortalamasını karekökü değerleri hesaplanmıştır. Hesaplama sonucunda, DP-INAR(1) modelinin HKOK değerinin en küçük olduğu görülmüştür.

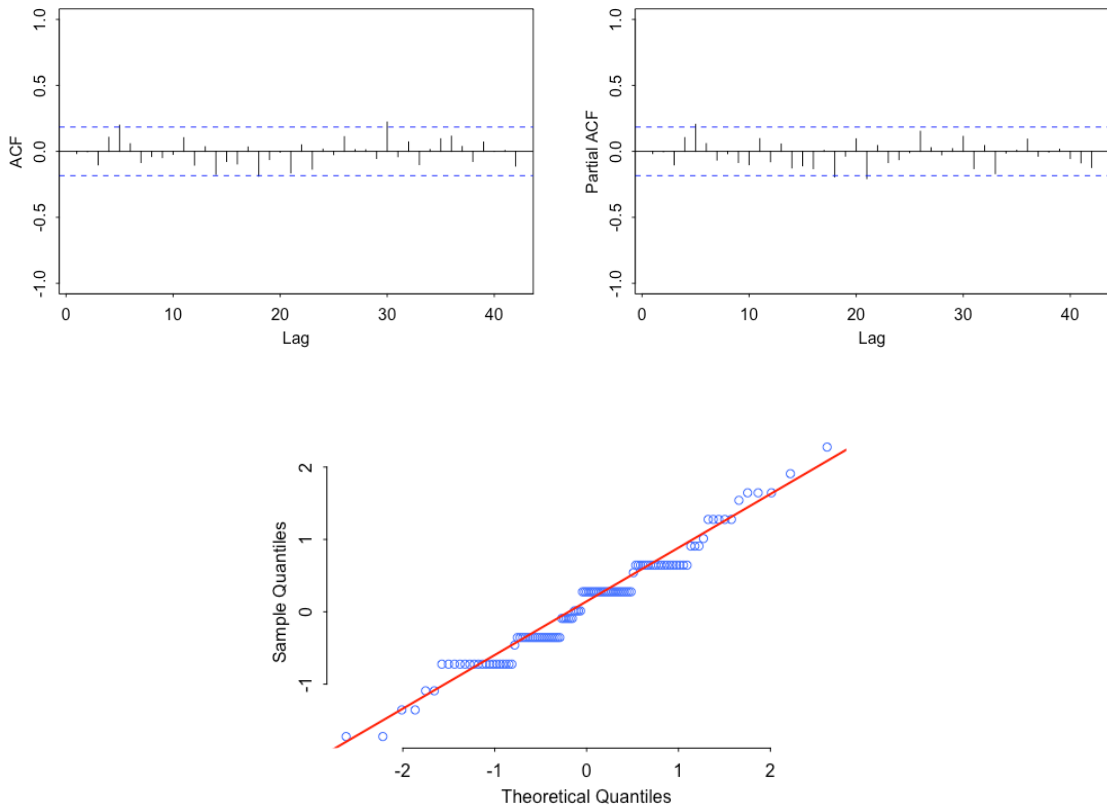
Tüm sonuçlar, HKOK hesaplamaları ile AIC ve BIC değerleri diğer modellere göre küçük elde edilen değerler ile ortalama ve varyans hesaplamalarında gerçek ortalama ve varyansa en yakın hesaplanan değerler koyu renkle gösterilmek üzere Tablo 4'te gösterilmiştir.

Tablo 4: Logaritmik dönüşüm uygulanan deprem sayısı serisi için CML tahminleri, AIC ve BIC bilgi kriterleri, ortalama, varyans ve HKOK hesaplamaları

Model	Parametreler	AIC	BIC	HKOK (tahmin serisi)	Ortalama	Varyans	HKOK (öngörü)
P-INAR(1)	$\alpha = 0.692$ $\lambda = 0.538$	283.646	289.118	0.802	1.747	1.747	1.633
G-INAR(1)	$\alpha = 0.731$ $p = 0.679$	294.669	300.141	0.814	1.757	2.588	1.640
PL-INAR(1)	$\alpha = 0.729$ $\theta = 2.653$	293.8964	299.369	0.814	1.772	1.025	1.641
NB-INAR(1)	$\alpha = 0.696$ $r = 111.400$ $p = 0.995$	285.864	294.073	0.804	1.841	0.943	1.650
DP-INAR(1)	$\alpha = 0.368$ $\mu = 0.988$ $\phi = 3.486$	259.898	268.107	0.759	1.579	0.748	1.556
GP-INAR(1)	$\alpha = 0.222$ $\mu = 2.222$ $\phi = -0.667$	267.094	275.303	0.760	1.713	0.151	1.576
ZIP-INAR(1)	$\alpha = 0.381$ $\lambda = 0.277$ $p = -2.832$	260.740	268.949	0.753	1.715	0.686	1.591

Farklı karşılaştırma yöntemlerinde farklı modeller uygun model olarak görülse de model karşılaştırmalarında Akaike Bilgi Kriteri ve Bayesyen Bilgi Kriteri en doğru sonucu veren yöntemler olduğu söylenebilir. AIC ve BIC değerleri en küçük hesaplanan model ise DP-INAR(1) modelidir ve artçı sayısı verisine en uygun model olarak belirlenebilir.

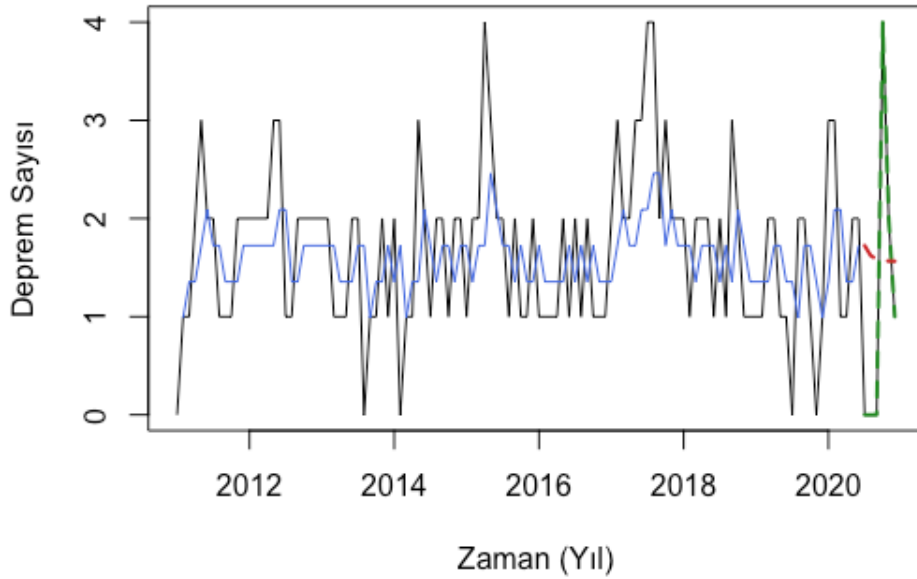
Son olarak, belirlenen modelin İstatistik Bilimi için geçerli bir model olup olmadığının kontrol edilmesi için modelden hesaplanan hata serisi analiz edilmiştir. Bu doğrultuda, hata serisinin herhangi bir otokorelasyona sahip olup olmadığını görebilmek için serinin ACF ve PACF grafikleri incelenmiş ve yine serinin normal dağılıma yakınsayıp yakınsamadığını yorumlayabilmek için de Q-Q grafiği incelenmiştir. ACF, PACF ve Q-Q grafikleri Şekil 21’de sırasıyla gösterilmektedir.



Şekil 21: DP-INAR(1) modeli uygulanarak hesaplanan hataların ACF, PACF ve Q-Q grafikleri

Hatalara ait ACF grafiğinde gecikmelerdeki ilişkiler önemli değildir yani güven sınırını ciddi bir şekilde aşmamaktadır. Bu nedenle, hatalar arasında herhangi bir ilişkiden bahsedilemez. Dolayısıyla, DP-INAR(1) modelinin deprem sayısı serisi için incelenen diğer modeller arasında en uygun model olduğunu söyleyebiliriz.

Şekil 22’de DP-INAR(1) modelinden elde edilen tahmin serisi ile deprem sayısı serisinin birlikte grafiği verilmiştir. Siyah renk deprem sayısı serisini, mavi renk ise tahmin serisini ifade etmektedir. Ayrıca yeşil renk ileriye yönelik tahminler yapmak ve modelin tahmin etme performansını test etmek için seriden kesilen son 6 veriyi gösterirken, kırmızı renk koşullu ortalama tahmin yöntemi ile tahmin edilen veriyi göstermektedir.



Şekil 22: Logaritmik dönüşüm uygulanan seri ve tahmin serisinin birlikte grafiği

5. SONUÇLAR

Zaman serileri tahmin ve öngörü işlemleri incelendiğinde, ayrıştırma yöntemi, zaman serilerinde regresyon analizi, üstel düzeltme yöntemi, Box-Jenkins modelleri gibi çeşitli yöntemler ön plana çıkmaktadır. Ancak, bu yöntemler genellikle sürekli değerli zaman serilerini modellemek ve öngöründe bulunmak için kullanılan yöntemlerdir. Tam sayı değerli zaman serilerini modellemek ve daha güvenilir öngörülerde bulunabilmek için farklı modellere ihtiyaç duyulmaktadır.

Bu çalışma kapsamında, tam sayı değerli zaman serilerinin analizi için literatürde sıklıkla karşımıza çıkan birinci dereceden tam sayı değerli otoregresyon modelleri incelenmiştir. Binom inceltme operatörüne dayalı olarak geliştirilen modeli daha iyi anlamak amacıyla öncelikle kısaca inceltme operatöründen bahsedilmiş ve sonrasında incelenen INAR(1) modelin standart AR(1) modeli ile farkları tartışılmıştır.

İlk olarak, McKenzie (1985) ve Al-Osh ve Alzaid (1987) tarafından Poisson dağılımına sahip hatalara dayalı olarak geliştirilen INAR(1) süreci, aynı zamanda 1 adım geçiş olasılıklarına sahip homojen bir Markov zinciridir. X_t ile ifade edilebilen sürecin, marjinal dağılım fonksiyonu, olasılık yaratıcı fonksiyonu, otokovaryans ve otokorelasyon fonksiyonları, X_t 'nin X_{t-1} üzerindeki koşulu ortalama ve varyansı, X_t 'nin ortalama, varyans ve yayılım indeksi gibi dağılım özellikleri ayrıntılı bir şekilde incelenmiştir.

Sürecin yayılım indeksinin değeri, X_t 'nin yayılım davranışını belirler ve hata serisine bağlıdır. Literatürde geliştirilen ilk model, Poisson dağılıma sahip hatalara bağlı olarak geliştirilmiştir ve ortalaması varyansına eşit olan eşit yayımlı seriler için uygun bir modeldir. Ancak, gerçek zaman serileri her zaman eşit yayılmayacağından modele alternatif olarak literatürde hataların dağılımı değiştirilerek yeni modeller geliştirilmiştir.

Geometrik dağılıma sahip hatalarla geliştirilen G-INAR(1) modeli ve Poisson-Lindley dağılımına sahip hatalarla geliştirilen PL-INAR(1) modeli iki parametrelidir ve varyansı

ortalamasından büyük olan aşırı yayılmış seriler için uygun olan modellerdir. Negatif binom dağılımına sahip hatalarla geliştirilen NB-INAR(1) modeli üç parametrelidir ve aşırı yayılmış seriler için uygun bir model iken, Çift Poisson, Genelleştirilmiş Poisson, Sıfır Yığmal Poisson hatalarla geliştirilen DP-INAR(1), GP-INAR(1) ve ZIP-INAR(1) modelleri üç parametrelidir ve parametre değerlerine bağlı olarak hem aşırı hem de ortalaması varyansından küçük olan yetersiz yayılmış seriler için uygun modellerdir.

INAR(1) modeli, inceltme operatöründen gelen α parametresi ve hataların dağılım özelliklerine bağlı olarak çeşitli parametrelerle belirlenir. Bu parametrelerin gerçek değerleri bilinmemektedir, ancak çeşitli yöntemlerle tahmin edilmesi gerekmektedir. Yule-Walker tahmini, koşullu en küçük kareler tahmini ve koşullu maksimum olabilirlik tahmini sık kullanılan parametre tahmin yöntemleridir. Bu çalışma kapsamında, yukarıda bahsedilen iki ve üç parametrelidir modellerin parametre tahminleri için koşullu maksimum olabilirlik tahmin yöntemi kullanılmıştır.

Çalışmada modeller, biri aşırı diğeri yetersiz yayılmış iki farklı gerçek zaman serisi üzerinde uygulanmıştır. Serilere gerekli dönüşümler uygulandıktan sonra koşullu maksimum olabilirlik tahmin yöntemiyle parametre değerleri tahmin edilmiştir. Sonrasında her bir model için tahmin serileri elde edilerek Akaike ve Bayesyen Bilgi kriterleri hesaplanmıştır. En küçük AIC ve BIC değerine sahip modelin bilimsel olarak geçerliliği hataların ACF, PACF ve Q-Q grafikleri incelenerek gösterilmiştir.

Sonuç olarak, hem aşırı hem yetersiz yayılan iki seri için de Poisson hatalarıyla geliştirilen ilk model uygun bulunmamıştır. Aşırı yayılan ilk seri için Poisson-Lindley hatalarıyla geliştirilen PL-INAR(1) modeli en uygun model olarak belirlenirken, yetersiz yayılan ikinci seri için Çift-Poisson hatalarıyla geliştirilen DP-INAR(1) modeli en uygun model olarak belirlenmiştir.

Bu tez çalışmasının devamı olarak, AR sürecine uymayan ama MA ya da ARMA sürecine uyan seriler için tezde bahsedilen yöntemler MA(1) ya da ARMA(1,1) modellerine uyarlanabilir.

6. KAYNAKLAR

- Al-Osh, M. A., Alzaid, A. A., First-order integer-valued autoregressive (INAR (1)) process, *Journal of Time Series Analysis*, 8(3), 261-275, 1987.
- Alzaid, A. A., Al-Osh, M. A., First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties, *Statistica Neerlandica*, 42(1), 53-61, 1988.
- Bourguignon, M., Rodrigues, J., & Santos-Neto, M., Extended Poisson INAR (1) processes with equidispersion, underdispersion and overdispersion, *Journal of Applied Statistics*, 46(1), 101-118, 2019.
- Consul, P. C., *Generalized Poisson Distributions: Properties and Applications*, Volume 99 of *Statistics: Textbooks and Monographs*. New York: Marcel Dekker Inc. 1989.
- Consul, P. C., Jain, G. C., A generalization of the Poisson distribution, *Technometrics*, 15(4), 791-799, 1973.
- da Cunha, E. T., Bourguignon, M., & Vasconcellos, K. L., On shifted integer-valued autoregressive model for count time series showing equidispersion, underdispersion or overdispersion, *Communications in Statistics-Theory and Methods*, 50(20), 4822-4843, 2021.
- Efron, B., Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association*, 81(395), 709-721, 1986.
- Freeland, R. K., *Statistical Analysis of Discrete Time Series with Application to the Analysis of Workers' Compensation Claims Data*, PhD Thesis, University of British Columbia, Vancouver, 1998.
- Freeland, R. K., & McCabe, B. P., Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20(3), 427-434, 2004.
- Freeland, R. K., McCabe, B., Asymptotic properties of CLS estimators in the Poisson AR (1) model, *Statistics & Probability Letters*, 73(2), 147-153, 2005.

- Ghitany, M. E., Al-Mutairi, D. K., Estimation methods for the discrete Poisson–Lindley distribution, *Journal of Statistical Computation and Simulation*, 79(1), 1-9, 2009.
- Huang, J., Zhu, F., A new first-order integer-valued autoregressive model with Bell innovations, *Entropy*, 23(6), 713, 2021.
- Jazi, M.A., Jones, G., & Lai, C. D., Integer valued AR(1) with geometric innovations, *Journal of the Iranian Statistical Society*, 11(2), 173-190, 2012a.
- Jazi, M. A., Jones, G., & Lai, C. D., First-order integer valued AR processes with zero inflated Poisson innovations, *Journal of Time Series Analysis*, 33(6), 954-963, 2012b.
- Jin-Guan, D., Yuan, L., The integer-valued autoregressive (INAR (p)) model, *Journal of Time Series Analysis*, 12(2), 129-142, 1991.
- Joe, H., Zhu, R., Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution, *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47(2), 219-229, 2005.
- Johnson, N. L., Kemp, A. W., & Kotz, S., *Univariate Discrete Distributions* (Vol. 444), John Wiley & Sons, 2005.
- Jung, R. C., Ronning, G., & Tremayne, A. R., Estimation in conditional first order autoregression with discrete support, *Statistical Papers*, 46, 195-224, 2005.
- Kadirioğlu. F. T., Kartal. R. F., The new empirical magnitude conversion relations using an improved earthquake catalogue for Turkey and its near vicinity (1900-2012), *Turkish Journal of Earth Sciences*, 25 (4), 300-310, 2016.
- Klimko, L. A., Nelson, P. I., On conditional least squares estimation for stochastic processes, *The Annals of Statistics*, 629-642, 1978.
- Lambert, D., Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14, 1992.
- Lívio, T., Khan, N. M., Bourguignon, M., & Bakouch, H. S., An INAR (1) model with Poisson–Lindley innovations, *Econ Bull*, 38(3), 1505-1513, 2018.

- McKenzie, E., Autoregressive moving-average processes with negative-binomial and geometric marginal distributions, *Advances in Applied Probability*, 18(3), 679-705, 1986.
- Mckenzie, E., Some simple models for discrete variate time series, *Journal of the American Water Resources Association*, 21(4), 645-650, 1985.
- Sankaran, M., The discrete poisson-lindley distribution, *Biometrics*, 145-149, 1970.
- Schweer, S., Weiß, C. H., Compound Poisson INAR (1) processes: stochastic properties and testing for overdispersion. *Computational Statistics & Data Analysis*, 77, 267-284, 2014.
- Steutel, F. W., Van Harn, K., Discrete analogues of self-decomposability and stability, *The Annals of Probability*, 893-899, 1979.
- Tan. O., Turkish Homogenized Earthquake Catalogue (TURHEC) [Data set], *Natural Hazards and Earth System Sciences (NHES)*, Zenodo. <https://doi.org/10.5281/zenodo.5056801>, 2021.
- Wagh, Y. S., Kamalja, K. K., Zero-inflated models and estimation in zero-inflated Poisson distribution, *Communications in Statistics: Simulation and Computation*, 47(8), 2248-2265, 2018.
- Wamwea, C., Mwelu, S., & Odin, M., Modelling COVID-19 cumulative number of cases in Kenya using a negative binomial INAR (1) Model. *Open Journal of Modelling and Simulation*, 11(1), 14-36, 2022.
- Weiß, C. H., *An Introduction to Discrete-Valued Time Series*, John Wiley & Sons, 2018.
- Weiß, C. H., Thinning operations for modeling time series of counts: a survey, *Advances in Statistical Analysis*, 92, 319-341, 2008.
- Zhu, F., Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models, *Journal of Mathematical Analysis and Applications*, 389(1), 58-71, 2012.
- Zou, Y., Geedipally, S. R., & Lord, D., Evaluating the double Poisson generalized linear model, *Accident Analysis & Prevention*, 59, 497-505, 2013.