

GENETİK ALGORİTMALAR İLE ARIMA MODELLERİNİN BELİRLENMESİ

Serkan TAŞTAN*

Özet

Box-Jenkins yönteminde ARIMA modelleri; belirleme, tahmin ve uygunluk testi olmak üzere üç aşamalı bir sürece göre oluşturulmaktadır. Bunlar arasında, modelin otoregresif ve hareketli ortalama derecelerinin tespit edildiği belirleme aşaması en kritik olandır. Bu çalışmada, büyük ölçüde model kuranın öznel yargılarına dayanan belirleme aşaması genetik algoritmalar çerçevesinde ele alınmıştır. Önerilen yaklaşım üretici fiyat endeksi ve kişi başına elektrik tüketimi serilerine uygulanmıştır. Uygun modelin veriye dayalı olarak seçildiği bu yaklaşım ile daha iyi uyuma sahip olduğu gibi daha az parametre içeren daha cimri modeller belirlenebilmiştir. Ayrıca tahmin edilen model sayısı bağlamında düşünüldüğünde, önerilen yaklaşım hesaplama maliyeti bakımından etkindir.

Anahtar Kelimeler: Genetik algoritmalar, ARIMA, model seçimi

Specification of ARIMA Models with Genetic Algorithms

Abstract

In the Box-Jenkins methodology, ARIMA models are built according to a process that consists of three stages namely specification, estimation and diagnostic checking. Among them the identification stage in which autoregressive and moving average orders of model are determined is the most crucial. In this paper the identification stage mostly based on the subjective decisions has been handled with genetic algorithms. The proposed approach is applied on the producer price index and per capita electric power consumption time series. With this data-driven approach better fitted and also more parsimonious models can be determined. Furthermore, when considering in the context of estimated model number the proposed approach is efficient.

Keywords: Genetic algorithms, ARIMA, model selection

* Yrd. Doç. Dr., Cumhuriyet Üniversitesi, İİBF, Yönetim Bilişim Sistemleri Bölümü, Sivas,
stastan@cumhuriyet.edu.tr.

GİRİŞ

Otoregresif hareketli ortalama modeli (ARMA) ve otoregresif bütünleşik hareketli ortalama modeli (ARIMA), sırasıyla durağan ve durağan olmayıp fark alma işlemi ile durağanlaştırılan tek değişkenli zaman serilerinin analizinde uzun süredir başarıyla kullanılmaktadır. Bu modellerin yaygın kullanımının altında, güçlü teorik temelleri ve çeşitli alanlardan farklı zaman serilerine başarıyla uygulanabilmeleri yatmaktadır. Regresyon modelleri ile benzerlikleri nedeniyle regresyon analizindeki kavramlar ışığında kolayca yorumlanabilmeleri bir diğer tercih nedeni olmaktadır (Baragona ve Battaglia, 2009:353). Box-Jenkins yöntemine göre, ARIMA modelleri oluşturulurken belirleme, tahmin ve uygunluk testi olmak üzere üç aşamalı bir süreç takip edilmektedir. Belirleme aşaması, zaman serisinin durağanlığının araştırılması ile otoregresif ve hareketli ortalama derecelerinin tespit edilmesini içerir. Bu aşamada belirlenen model ikinci aşamada tahmin edilir; üçüncü aşamada ise modelin uygunluğu, tahmin edilen modelin kalıntıları üzerinden test edilir. Bu aşamalardan model belirleme aşaması kritik öneme sahiptir; çünkü doğru belirlenmeyen modeller için sonraki aşamada gerçekleştirilen parametre tahminleri de doğru olmayacak (Abo-Hammour vd., 2013:2) ve devamında üçüncü aşama sonucunda modelin yeniden belirlenmesi gerekecektir (Ong vd., 2005:886).

Model belirlemede genellikle otokorelasyon ve kısmi otokorelasyon fonksiyonlarından yararlanır. Ancak, bu karışık bir süreçtir ve büyük ölçüde araştırmacının yetenek ve tecrübelerine dayanır. Öyle ki, çoğu zaman serinin içerdiği karışık otoregresif ve hareketli ortalama etkileri nedeniyle ilgili derecelerinin açıkça belirlenmesi mümkün olmaz (Ong vd., 2005:886). Dahası, farklı modeller aynı korelasyon ve kısmi korelasyon örüntülerine sahip olabilirler (Minerva ve Poli, 2001:335). Bu durum, belirleme sürecinin bazı kişisel yargılar içermesine ve dolayısıyla belirlenen modelin güvenilirliğinin azalmasına sebep olmaktadır. Ayrıca, Minerva ve Poli (2001)'nin belirttiği üzere, bu şekilde model belirleme sürecinin genellikle çok zaman alması gibi bir boyutu da vardır ki göz ardı edilmemesi gerekir.

Bir diğer model belirleme yöntemi ise bilgi kriterlerine dayanmaktadır. Bu yöntemde, seçilen bilgi kriterinin değerini en küçükleyen otoregresif ve hareketli ortalama derecelerine yönelik bir ardışık arama yürütülmektedir. Popüler ekonometri ve istatistik yazılımlarının güncel sürümlerinde bu yöntem artık varsayılan bir özellik olarak yer almaktadır. Pratik olmasına karşın, bu yöntemle arama tam modeller bir başka deyişle yerel optimumlar üzerinden işletilmekte dolayısıyla tüm model uzayının taranması ve optimum derecelerinin bulunması mümkün olmamaktadır. Ayrıca cimrilik ilkesi gereği, aynı dinamikleri yansıtabilen iki ARIMA modelinden parametre sayısı az olan tercih edilmelidir. Sezgisel olarak modelin az parametreye sahip olması özellikle

öngöründe faydalıdır (Yavuz, 2014:263). Bu bağlamda, tam modelde anlamsız değişkenleri dışlayacak şekilde birtakım ardışık kısıtlamalar getirilerek nihai model elde edilebilir. Ancak bu genelden özele ardışık azaltma yaklaşımı genellikle ele alınan problem için karışık ve problemlili olabilir (Balcombe, 2005:207). Bu noktada, uygun model için yürütülen arama doğrudan nihai modele yönelik olarak genişletilebilir. Ancak bu yaklaşım seçilen sınırlara bağlı olarak olası model sayısının çok hızlı bir şekilde artması nedeniyle pratik olmaktan uzaktır. Örneğin; böyle bir aramada otoregresif ve hareketli ortalama dereceleri için üst sınır 10 olarak belirlenirse, toplamda $2^{10+10} - 1$ (yaklaşık bir milyon) modelin tahmin edilmesi gerekmektedir. Her bir modelin tahmin edilmesi ve bilgi kriterleri kullanılarak karşılaştırması gerektiği düşünülünce, bu yaklaşımın çok fazla hesaplama yükü ve zamanı gerektireceği açıktır ki temel problemde budur. O halde, burada ihtiyaç duyulan bu tür aramaları daha az hesaplama yükü ve daha kısa zamanda yapabilecek bir sezgisel algoritma tabanlı yaklaşımdır. Genetik algoritma bu amaçla kullanılabilir; bu tür sonlu, kesikli ve büyük arama uzaylarında arama ve problem çözme etkinliğine sahip bir optimizasyon tekniğidir. Ayrıca ele alınan model kolaylıkla genetik algoritmalar ile ifade edilebilecek niteliktedir.

Literatürde, bu çalışmada olduğu gibi ARMA/ARIMA modellerini belirleme sürecinin genetik algoritmalar ile ele alındığı çalışmalar mevcuttur. Ong vd. (2005) ARIMA ve mevsimsel ARIMA modellerini belirlemek amacıyla genetik algoritmaları kullanmışlardır. Önerdikleri yöntemi gerçek zaman serileri kullanarak, diğer model belirlemede kullanılan örüntü tanıma yöntemleriyle karşılaştırmışlar ve genetik algoritmalara dayalı yöntemin daha başarılı olduğu sonucuna varmışlardır. Baragona ve Battaglia (2009) istatistiksel veri analizinde genetik algoritmanın kullanımını inceledikleri çalışmalarında, ARMA modellerinin belirlenmesi konusuna da yer vermişlerdir. Ağırlıklı olarak probleme ilişkin mevcut genetik algoritma yaklaşımlarına değinmişler ve konuyu gerçek bir zaman serisi kullanarak örneklendirmişlerdir. Abo-Hammour vd. (2012) ARMA modellerinin dereceleri ve parametrelerinin eş zamanlı olarak belirlemek için genetik algoritmaları kullanmışlardır. Yaklaşımlarının etkinliğini gerçekleştirdikleri simülasyonlar yardımıyla göstermişlerdir. Bu çalışma daha sonra Abo-Hammour vd. (2013) tarafından dinamik doğrusal sistemlerin tahminine yönelik olarak genişletilmiştir. Simülasyon ve gerçek verilerle desteklenen, genetik algoritmalar ile nihai ARMA modelinin belirlendiği bir diğer çalışmada Gaetan (2000) tarafından yapılmıştır. Satman (2007) ARMA modellerini genetik algoritmalar ile belirlediği çalışmasında İMKB endeksleri üzerine bir uygulama yapmıştır. Kullandığı endekslerden sadece İMKB Ulusal-50 Endeksi için güvenilir sonuçlara ulaşabilmiştir. Minerva ve Paterlini (2002) model seçim problemini ele aldıkları çalışmalarında; doğrusal regresyon modelleri ve kümeleme analizi ile beraber ARMA

modellerini incelemişler, önerdikleri yaklaşımın hesaplama zamanının geleneksel yöntemlerden az da olsa yüksek olduğunu bildirmişlerdir. Minerva ve Poli (2001) ARMA modellerini belirlemek için önerdikleri genetik algoritmada, diğer çalışmalardan farklı olarak değişen uzunlukta kromozomlara dayalı bir kodlama kullanmışlardır. Simülasyon çalışmaları yaklaşımlarının gerçek veri üretme sürecini yüksek oranda belirleyebildiği göstermiştir.

Bu çalışmada, çok fazla aday modelden oluşan büyük bir kümeden en uygun ARIMA modelinin belirlenmesine yönelik bir genetik algoritma temelli yaklaşım sunulmuştur. Ele alınan probleme ilişkin genetik algoritma tasarımı belirlenmiş ve bu tasarım kullanılarak, bir yandan karşılaşılan hesaplama yükü ve zamanı azaltılırken diğer yandan yerel optimuma takılmadan belirli kriterler dahilinde optimal modelinin belirlenmesine çalışılmıştır.

I. ARIMA MODELLERİ İÇİN GENETİK ALGORİTMA

ARIMA modelleri genel olarak ARIMA(p, d, q) biçimde ifade edilir ve aşağıdaki gibi yazılabilir:

$$y_t = c + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + e_t + \beta_1 e_{t-1} + \dots + \beta_q e_{t-q} \quad (1)$$

Burada, p ve q sırasıyla otoregresif ve hareketli ortalama derecelerini, d ise fark alma derecesi göstermektedir. y_t d. dereceden farkı alınmış zaman serisini, e_t hata terimlerini gösterirken c ise bir sabittir. Ayrıca $\alpha_1, \dots, \alpha_p$ otoregresif katsayılar ve β_1, \dots, β_q ise hareketli ortalama katsayılarıdır. Fark alma derecesi d, bir ya da daha büyük olabileceği gibi (1)'de zaman serisi farkı alınmadan düzey değerleriyle de yer alabilir, bu durumda (1) bir ARMA(p, q) modelini gösterecektir. Uygun fark alma derecesi ekonometrik yöntemlerle model belirleme aşamasından önce tespit edilir. Dolayısıyla ARIMA modelleri belirlenirken, esasen otoregresif ve hareketli ortalama dereceleri ile beraber zaman serisinin ve hata teriminin sıfırdan farklı gecikmeleri seçilmelidir. Bu amaçla, her bir aday model belirli bir uygunluk ölçüsüne göre değerlendirilmelidir. Bir uygunluk ölçüsü olarak Akaike bilgi kriteri (AIC) ve Bayesian bilgi kriteri (BIC) kullanılabilir. Nitekim bu kriterler rakip modeller arasında seçim yapmak amacıyla istatistik ve ekonometride yaygın olarak kullanılmaktadırlar. Dolayısıyla prensipte ARIMA modelleri; seçilen maksimum otoregresif ve hareketli ortalama derecelerine bağlı olarak tüm olası gecikme kombinasyonları dikkate alınarak, bunlar içerisinde AIC veya BIC gibi bir bilgi kriterine göre en iyi olanın seçilmesi suretiyle belirlenebilirler. Bu bağlamda ARIMA modellerinin belirlenmesi, kesikli argümanlardan oluşan bir amaç fonksiyonunu içeren bir optimizasyon problemi

olarak düşünülebilir ancak geleneksel optimizasyon metotları arama uzayındaki süreksizlikler nedeniyle bu problem için uygun değildir. Diğer taraftan, genetik algoritmalar problem parametreleri için kullandıkları farklı şekillerdeki kodlamalar ve amaç fonksiyonunun türevlenebilir olmasının gerekmemesi gibi noktalarda geleneksel optimizasyon metotlarından ayrıldıkları gibi (Sivanandam ve Deepa, 2007:34) etkinlik ve hesaplama maliyeti bakımından da ele alınan problem için uygun bir tekniktir.

Genetik algoritmalar; bir çeşit genel amaçlı, hızlı, güvenilir ve güçlü arama ve optimizasyon tekniği olarak ifade edilebilirler (Michalewicz, 1996:15). Genetik algoritmalar, en iyinin korunumu ve doğal seçilim gibi doğal olguların benzetim yoluyla bilgisayarlara uygulanması ile elde edilir (Nabiyev, 2005:633). Genetik algoritmalar problemin çözümüne rastgele bireylerden oluşturdukları bir popülasyon ile başlarlar. Popülasyondaki her bir birey, probleme ilişkin olası bir çözümü gösterir ve kromozom olarak adlandırılır. Kromozomlar bir dizi halinde sıralanmış keyfi uzunluktaki bit dizilerinden oluşur. Genetik algoritmaların ilk aşaması olası çözümleri yani aday ARIMA modellerini kromozom formunda kodlamaktır. ARIMA modelinin uygun gecikmelerinin belirlenmesi, bir başka ifadeyle zaman serisinin ve hata teriminin gecikmeli değerlerinden hangilerinin modelde yer alıp almayacağına karar verilmesi, iki seçeneqli bir durumu ihtiva etmesi nedeniyle ikili kodlama problem için en doğal kodlamadır. İkili kodlamada tüm gecikmeler kromozomlarda her birinin sıfır ve bir değerlerinden birini alabildiği bitlerden oluşan; sabit uzunlukta ve sıralı, ayrık, dizileri aracılığıyla kodlanabilirler. Bu tarz bir kodlamada, bit dizilerinin uzunlukları konusunda temel belirleyiciler otoregresif ve hareketli ortalama derecelerinin üst sınırları olacaktır. Bu sınır gözlem sayısına veya önsel bir kabule bağlı olarak belirlenebilir (Baragona ve Battaglia, 2009:355). Her bir gendeki ilk bit sabit terimi ifade ederken, ikinci bitten itibaren p bit otoregresif terimleri, $(p+2)$. bitten itibaren takip eden q bit hareketli ortalama terimlerini temsil etmektedir. Her hangi bir bitin bir değerini alması ilgili bitin temsil ettiği otoregresif veya hareketli ortalama teriminin modelde yer aldığı anlamına gelmektedir. Modelde yer almayan diğer gecikmeler ise ilgili bitlerde sıfır değerini alacak şekilde temsil edilmişlerdir.

Bu çalışmada, diğer çalışmalardan farklı olarak genetik algoritmanın başlangıç popülasyonu; her bir gecikmenin teker teker yer aldığı modelleri içerecek ve daha cimri modellerin oluşturulmasını sağlayacak, yani sıfır değerli bitlerin daha yüksek olasılıkla kromozomlarda yer alacağı şekilde kontrollü olarak oluşturulmuştur. Sıfır değerli bir bit 0,6 olasılıkla kromozomda yer alabilirken aynı olasılık bir değerli bir bit için 0,4'tür. Genetik algoritmalarda popülasyonu oluşturan bireyler bazı uygunluk ölçülerinin kullanılması yoluyla nesil adı verilen ardışık tekrarlar sayesinde gelişirler (Sakawa, 2002:12). Her nesilde öncelikle uygunluk fonksiyonunun ilgili değeri hesaplanarak her kromozomun uygunluk değeri belirlenir. Bu değer, kromozomun

problemin çözümüne ne kadar yakın olduğunun bir göstergesidir. ARIMA modeli tahmin edildikten sonra modelin uyumu dolayısıyla kromozomun uygunluğu bilgi kriterleri hesaplanarak değerlendirilir. Bilgi kriterini en küçükleyen model en uygun modeldir, üstelik bazı zaman serileri için bilgi kriteri hem pozitif hem de negatif değerler alabilmektedir. Bu nedenle, çalışmada bir modelden hesaplanan bilgi kriteri uygunluk değeri olarak kullanılmadan önce (1000 – Bilgi Kriterinin Değeri) şeklinde dönüştürülmüştür. Tahmin aşamasında bir yakınsama problemi ile karşılaşılacak modeller ya da tamamı sıfır değerli bitlerden oluşan kromozom için uygunluk değeri sıfır olarak atanmıştır.

Bir popülasyonda yer alan her kromozomun uygunluk değeri hesaplandıktan sonra, bir sonraki neslin popülasyonu genetik operatörler kullanılarak oluşturulur. Genetik algoritmalarda kullanılan genetik operatörler seçim, çaprazlama ve mutasyondur. Seçim işleminde yüksek uygunluk değerlerine sahip kromozomlara daha çok şans verilerek sonraki neslin popülasyonuna eklenecek kromozomlar belirlenir ki burada amaç popülasyonun ortalama uygunluğunu artırmaktır (Blickle ve Thiele, 1995:9). Seçim sürecinde popülasyondan genetik işlemlerin uygulanacağı ebeveyn kromozomlar seçilir ve eşleme havuzu adı verilen deneme popülasyonuna kopyalanır. Ele alınan problem için tasarlanan genetik algoritmada elitizm ve turnuva seçimi kullanılmıştır. Elitizmde en iyi iki kromozom hiç bir değişikliğe uğramadan sonraki nesle kopyalanır. Turnuva seçiminde ise popülasyondan iki kromozom rastgele seçilir ve bunların uygunluk değerleri karşılaştırılır. En iyi kromozom sonraki nesli oluşturmak için eşleme havuzuna eklenir.

Çaprazlama, seçim sürecinde seçilen iki eski kromozomdan gelen bilgi parçalarının birleştirilmesiyle sonraki nesil için iki yeni kromozom oluşturulmasıdır. Temel amacı çözüm uzayını aramak olan çaprazlama operatörü, kromozomların iyi özelliklerinin bir araya gelmesini kolaylaştırmaktadır. Yaygın olarak kullanılan çaprazlama yöntemlerinden biri tek noktalı çaprazlamadır. Bu yöntemde, eşleme havuzundan iki kromozom rastgele eşleştirildikten sonra rastgele bir çaprazlama noktası belirlenir. Devamında eşleştirilen kromozomlarda çaprazlama noktasından sonra gelen bitler karşılıklı olarak değiştirilerek iki yeni kromozom elde edilir. Çaprazlama belirlenen bir olasılık doğrultusunda gerçekleştirilir.

Mutasyon ise seçilen kromozomların bir veya daha fazla bitinin değiştirilmesidir. Mutasyon hem kaybolan genetik materyalin geri getirilmesinde hem de genetik bilginin rastgele karıştırılmasında rol oynar (Sivanandam ve Deepa, 2007:56). İkili kodlamada mutasyon, her bir bitin değerinin düşük bir olasılıkla ters çevrilmesi yani sıfır değerinin bir veya bir değerinin sıfır yapılmasıdır.

Seçim, çaprazlama ve mutasyon adımları sabit bir iterasyon sayısına

ulaşılincaya kadar tekrar eder. Bu şekilde kromozomlar ile ifade edilen olası çözümlerin, zaman içinde iyileşmesi sağlanır.

II. UYGULAMA

Bu kısımda üretici fiyat endeksi (ÜFE) ve kişi başına elektrik tüketimi (KBET) serileri analiz edilerek uygun ARIMA modeli genetik algoritma yaklaşımı ile belirlenecektir. ÜFE ve KBET verileri sırasıyla Türkiye İstatistik Kurumu (TÜİK) ve Dünya Bankası'nın web sitelerinden alınmıştır. Aylık olarak ölçülen ÜFE verisi 2006:1-2005:6 dönemini kapsamaktadır. KBET serisi ise 1960-2012 dönemini kapsayan yıllık verilerden oluşmaktadır.

Öncelikle serilerinin durağanlığı genişletilmiş Dickey-Fuller (ADF) birim kök testi ile araştırılmıştır. Birim kök testlerinde sabitli-trendli model kalıbı kullanılmış ve uygun gecikme uzunlukları Schwarz Bilgi Kriteri'ne göre belirlenmiştir. Tablo 1'de verilen birim kök testi sonuçları incelendiğinde, her iki serinin seviyede durağan olmadıkları görülmektedir. ÜFE serisinin birinci farkı, KBET serisinin ise ikinci farkı durağandır. Farkları alınan serilerinin grafikleri Şekil 1'de verilmiştir. Ayrıca, Şekil 1'de iki serinin otokorelasyon fonksiyonu (ACF) ve kısmi otokorelasyon fonksiyonu (PACF) grafikleri de yer almaktadır.

Tablo 1. ADF birim kök testinin sonuçları

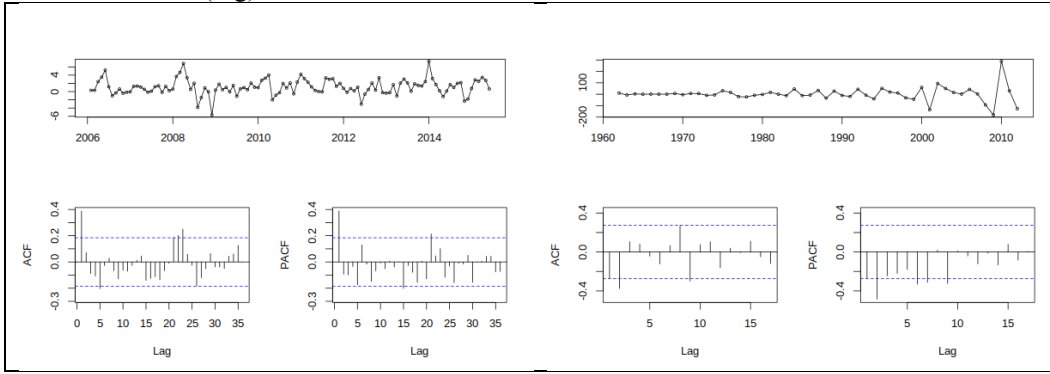
Değişken	Seviye	1. Fark	2. Fark
ÜFE	-3,07 (1)	-6,95 ** (0)	-
KBET	1,56 (8)	-1,86 (8)	-3,75 (8) *

Not: Gecikmeler parantez içerisinde verilmiştir. * ve ** sırasıyla %5 ve %1 seviyesinde anlamlılığı göstermektedir.

Farkı alınan serilerin ACF ve PACF grafikleri, genetik algoritmanın aramayı gerçekleştireceği maksimum otokorelasyon derecesi (p^*) ve hareketli ortalama derecesini (q^*) belirlemek için kullanılmıştır. Eğer seçilen maksimum değerler dahilinde uygun bir ARIMA modeli belirlenemez ise bu durumda aramanın alternatif, daha büyük p^* ve q^* değerleri için genişletilmesi yoluyla uygun bir model belirlenmeye çalışılmıştır. ÜFE serisinin ACF grafiği incelendiğinde 1 ve 5. gecikmeler ile 22 ve 23. gecikmelerin güven aralıklarının dışında olduğu görülmektedir. PACF grafiğinde ise 1, 15 ve 21. gecikmeler anlamlıdır. Diğer taraftan, KBET serisinin ilgili grafikleri incelendiğinde; ACF grafiğinde 2 ve 9. gecikmelerin,

PACF grafiğinde ise 2, 6, 7 ve 9. gecikmelerin anlamlı olduğu görülmektedir. Buna göre, ÜFE serisi için $p^*=q^*=5$ şeklinde belirlenmiş ve uygun bir ARIMA modeli tahmin edilebilmiştir. KBET serisi için ise p^* ve q^* için öncelikle 6 ve 7 değerleri denenmiş ancak uygun ARIMA modeli sadece $p^*=q^*=9$ için bulunabilmiştir. Her iki seri için uygunluk kriteri olarak AIC seçilmiş, mutasyon olasılığı 0,1 ve çaprazlama olasılığı 0,9 olarak belirlenmiştir. ÜFE serisinde algoritma 30 kromozomdan oluşan bir popülasyon için 10 iterasyon işletilmiştir; KBET serisi için popülasyon büyüklüğü 50 ve iterasyon sayısı 20 olarak belirlenmiştir. Genetik algoritma her bir seri için iki kez çalıştırılmış ve ulaşılan en iyi modeller Tablo 2’de, bu modellerin kalıntı grafikleri ise Şekil 2’de verilmiştir.

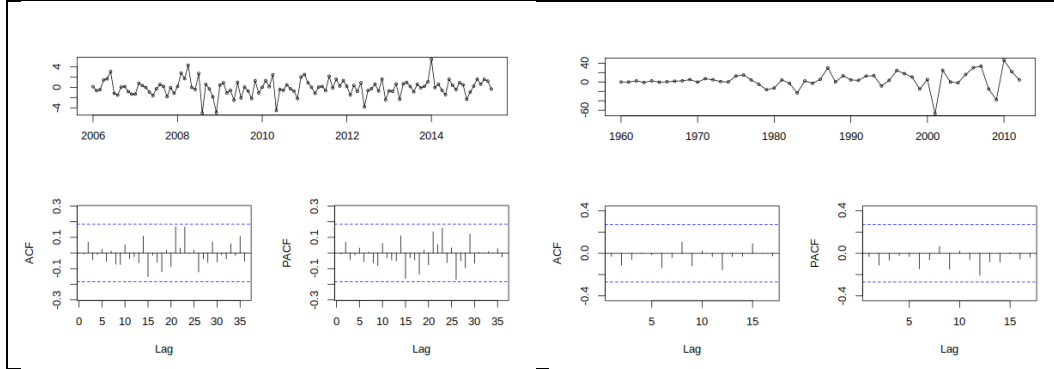
Şekil 1. Tüketici Fiyat Endeksi Serisinin 1.Farkı (sol) ve Kişi Başına Elektrik Tüketimi Serisinin 2.Farkı (sağ)



ÜFE serisi için belirlenen ARIMA(5, 1, 5) modeli, sabit terimin yanı sıra 1, 2 ve 5. otoregresif terimler ile 1 ve 5. hareketli ortalama terimlerini içermektedir. KBET serisi için belirlenen model ARIMA(8, 2, 9) modelidir. Bu modelde; 2, 3, 7 ve 8. otoregresif terimler ile 1, 4, 8 ve 9. hareketli ortalama terimleri bulunurken, ikinci derecen fark alınması sebebiyle sabit terim bulunmamaktadır. Her iki modelin kalıntılarına ilişkin otokorelasyon ve kısmi otokorelasyon grafikleri güven sınırları dışında terim içermemektedir. Ayrıca, tahmin edilen ARIMA modelinin başarısı için önem arz etmesi nedeniyle (Yavuz, 2014:265), kalıntı serisinin temiz dizi olup olmadığı Portmanteau testleriyle sınanmıştır. Bu amaçla KBET serisi için 10 gecikmeye, ÜFE serisi için ise 24 gecikmeye göre Ljung-Box istatistiği hesaplanmıştır. Hem kalıntı grafiklerine göre hem de Ljung-Box istatistiğine göre her iki zaman serisi için belirlenen ve tahmin edilen ARIMA modelleri uygundur ve içerdikleri parametre sayılarına göre cimri modeller oldukları da söylenebilir. Bu nedenle, genetik algoritma ile belirlenen ilgili modeller öngörü amacıyla kullanılabilirler.

Tablo 2. Belirlenen ARIMA Modelleri

	ÜFE ($p^*=q^*=5$)	KBET ($p^*=q^*=9$)
Sabit	Var	Yok
Otoregresif terimler	1 - 2 - 5	2 - 3 - 7 - 8
Fark derecesi	1	2
Hareketli ortalama terimleri	1 - 5	1 - 4 - 8 - 9
AIC	440,57	528,64
AICc	441,64	533,03
BIC	459,66	546,02
$\hat{\sigma}^2$	2,59	317,90
Ljung-Box	X^2	20,09
	p	0,33
	s.d.	18
		2

Şekil 2. Tüketici Fiyat Endeksi Serisi İçin Tahmin Edilen Modelin Kalıntıları (sol) ve Kişi Başına Elektrik Tüketimi Serisi İçin Tahmin Edilen Modelin Kalıntıları (sağ) Grafikleri

SONUÇ

Bu çalışmada, ARIMA modelinde yer alacak otoregresif ve hareketli ortalama terimlerinin tespit edildiği belirleme aşaması genetik algoritmalar çerçevesinde ele alınmıştır. Bu amaçla, üretici fiyat endeksi ve kişi başına elektrik tüketimi serileri için uygun ARIMA modelleri genetik algoritmalar ile belirlenmiştir. Belirleme aşaması, bir taraftan büyük ölçüde model kuranın öznel yargılarına dayanırken, diğer taraftan

ARIMA modellerini oluşturma sürecinin en kritik aşamasıdır. Dolayısıyla, ayrıntılı ve zaman alıcı bu süreç, daha öznel olarak ele alınmış ve önerilen genetik algoritma tasarımı ile otomatik olarak gerçekleştirilmiştir. Önerilen yaklaşımda korelogram ve bilgi kriteri yöntemleri birlikte kullanılmıştır. Şöyle ki; uygun ARIMA modeli araştırılırken maksimum otoregresif ve hareketli ortalama derecelerine korelograma bakılarak karar verilmiş, devamında bu sınırlar dahilinde aday modellerin uygunluğu AIC kriteri aracılığı ile değerlendirilmiştir. Maksimum otoregresif ve hareketli ortalama derecelerinin seçimi dışında, hangi otoregresif veya hata terimlerinin modelde yer alacağına yönelik hiçbir öznel yargı sürece dahil edilmemiştir. Dolayısıyla, uygun modelin seçimi veriye dayalı olarak yapılmıştır. Bu sayede, daha iyi uyuma sahip olduğu gibi daha az parametre içeren daha cimri modeller belirlenebilmiştir. Genetik algoritma kullanılarak belirlenen modelin kalıntı serisinin temiz dizi olacağının garantisi yoktur. Bu noktada, istenen özelliği sağlayacak bir model bulma amacıyla arama uzayı genişletilebilir. Nitekim ÜFE serisi için seçilen ilk maksimum değer kullanılarak uygun bir model bulunsa da KBET serisi için ancak seçilen üçüncü maksimum değer kullanılarak uygun bir modele ulaşılmıştır. Son olarak tahmin edilen model sayısı bağlamında düşünüldüğünde, önerilen yaklaşım hesaplama maliyeti bakımından etkindir. Ancak diğer sezgiseller gibi genetik algoritmalarında her zaman optimum çözümü bulmaları kesin değildir, arama uzayının çok geniş olduğu durumlarda yerel optimum değerlere yakınsayabilirler. Bu nedenle, ARIMA modelleri belirlenirken maksimum otoregresif ve hareketli ortalama dereceleri için nispeten büyük değerler seçildiği durumlarda, birkaç tekrarlı denemeler yapılması daha iyi modeller elde edilmesinden önemlidir.

KAYNAKLAR

- ABO-HAMMOUR, Za'er S.; Othman M. K. ALSMADI; Adnan M. AL-SMADI; Maha I. ZAQOUT and Mohammad S. SARAIH (2012), "ARMA Model Order and Parameter Estimation Using Genetic Algorithms.", *Mathematical and Computer Modelling of Dynamical Systems*, Vol 18, No 2; 201–21.
- ABO-HAMMOUR, Za'er S.; Othman M. K. ALSMADI; Shaher MOMANI and Omar ABU ARQUB (2013), "A Genetic Algorithm Approach for Prediction of Linear Dynamical Systems.", *Mathematical Problems in Engineering* 2013.
- BALCOMBE, Kelvin G. (2005), "Model Selection Using Information Criteria and Genetic Algorithms.", *Computational Economics*, Vol 25, No 3; 207–28.
- BARAGONA, Roberto and Francesco BATTAGLIA (2009), "Evolutionary Computing in Statistical Data Analysis." In *Foundations of Computational*

Intelligence Volume 3, Studies in Computational Intelligence, eds. Ajith Abraham, Aboul-Ella Hassanien, Patrick Siarry, and Andries Engelbrecht. Springer Berlin Heidelberg; 347–86.

BLICKLE, Tobias and Lothar THIELE (1995), “A Mathematical Analysis of Tournament Selection.” Proceedings Of The Sixth International Conference On Genetic Algorithms; 9–16.

GAETAN, Carlo (2000), “Subset ARMA Model Identification Using Genetic Algorithms.” Journal of Time Series Analysis, Vol 21, No 5; 559–70.

MICHALEWICZ, Zbigniew (1996), Genetic Algorithms + Data Structures, Springer.

MINERVA, Tommaso and Sandra PATERLINI (2002), “Evolutionary Approaches for Statistical Modelling.” In Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress, Vol 2; 2023–2028.

MINERVA, Tommaso and IRENE Poli (2001), “Building ARMA Models with Genetic Algorithms.” In Applications of Evolutionary Computing, Lecture Notes in Computer Science, ed. Egbert J. W. Boers. Springer Berlin Heidelberg; 335–42.

NABIYEV, Vasif Vagifoğlu (2005), Yapay Zeka, Seçkin Yayınevi.

ONG, Chornng-Shyong; Jih-Jeng HUANG and Gwo-Hshiung TZENG (2005), “Model Identification of ARIMA Family Using Genetic Algorithms.”, Applied Mathematics and Computation, Vol 164, No 3; 885–912.

SAKAWA, Masatoshi (2002), Genetic Algorithms and Fuzzy Multiobjective Optimization, Springer.

SATMAN, Mehmet Hakan (2007), “Durağan Zaman Serilerinde Uygun Arma Modelinin Genetik Algoritmalar İle Bulunması ve İMKB Verileri Üzerine Bir Uygulama”, İktisat Fakültesi Mecmuası, Cilt 57, Sayı 1; 21-38.

SIVANANDAM, S. N. and S. N. DEEPA (2007), Introduction to Genetic Algorithms (1st ed.), Springer.

YAVUZ, Nilgün Çil (2014), Finansal Ekonometri, Der Yayınları.