

SİYASET, EKONOMİ ve YÖNETİM ARAŞTIRMALARI DERGİSİ



RESEARCH JOURNAL OF
POLITICS, ECONOMICS AND MANAGEMENT

April 2017, Vol:5, Issue:2

Nisan 2017, Cilt:5, Sayı:2

P-ISSN: 2147-6071

E-ISSN: 2147-7035

Journal homepage: www.siyasetekonomiyonetim.org



Kendinden Uyarımlı Eşik Otoregresif Modellerin Belirlenmesi İçin Genetik Algoritma Yaklaşımı¹

A Genetic Algorithm Approach for the Specification of Self-Exciting Threshold Autoregressive Models

Yrd. Doç. Dr. Serkan TAŞTAN

Cumhuriyet Üniversitesi, İİBF, Yönetim Bilişim sistemleri Bölümü, stastan@cumhuriyet.edu.tr

Prof. Dr. Nilgün ÇİL

İstanbul Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü, nilgun.cil@istanbul.edu.tr

MAKALE BİLGİSİ

ÖZET

Makale Geçmişi:

Geliş 05 Nisan 2016
Düzeltilme Geliş 12 Aralık 2016
Kabul 13 Nisan 2016

Anahtar Kelimeler:

Kendinden Uyarımlı Eşik Otoregresif,
SETAR, Genetik Algoritma, Model
Belirleme

Günümüzde doğrusal olmayan zaman serisi analizinde yaygın olarak kullanılan kendinden uyarımlı eşik otoregresif (SETAR) modeller; anlaşılması ve yorumlanması kolay, basit bir model biçimine sahip olsalar da söz konusu modeller belirlenirken tahmin edilmesi gereken birçok serbest parametre bulunmaktadır. Bu nedenle bu çalışmada SETAR modellerini belirleme süreci bir optimizasyon problemi olarak düşünülmüş ve ilgili probleme genetik algoritmalar ile çözüm aranmıştır. Bu bağlamda ele alınan probleme ilişkin genetik algoritma bileşenleri tanımlanmış ve algoritma için uygun parametreler belirlenmiştir. Önerilen yaklaşım simülasyon verileri ile değerlendirilerek, kullanılabilirliği gösterilmiştir.

© 2017 PESA Tüm hakları saklıdır

ARTICLE INFO

ABSTRACT

Article History:

Received 05 April 2016
Received in revised form 12 December
2016
Accepted 13 December 2016

Keywords:

Self-Exciting Threshold Autoregressive,
SETAR, Genetic Algorithms, Model
Specification

Even though, self-exciting threshold autoregressive (SETAR) models commonly used in nonlinear time series analysis nowadays have a simple, easy to understand and interpret model form, there are many free parameters to estimate, when building the models in question. Therefore in this paper the process of specification of SETAR models were considered as an optimization problem and the related problem was solved with genetic algorithms. In this context genetic algorithm components were defined for the discussed problem and appropriate parameters were determined for the algorithm. The usability of the proposed approach is shown by evaluating it with simulation data.

© 2017 PESA All rights reserved

¹ Bu makale, Prof. Dr. Nilgün ÇİL danışmanlığında, Serkan TAŞTAN tarafından hazırlanan "Genetik Algoritma Temelinde Rejim Değişikliği Modellerinin Belirlenmesi" başlıklı doktora tezinden türetilmiştir.

GİRİŞ

İlk defa Tong (1978) ve Tong ve Lim (1980) tarafından önerilen ve yine Tong (1990) tarafından detaylı olarak tartışılan eşik otoregresif (TAR) modellerde rejimler belirli bir eşik değeriyle ilgili olarak belirlenir (Franses ve Dijk, 2000: 71). Gerçekte eşik otoregresif modeller doğrusal dışı davranışlara izin verecek şekilde genişletilen otoregresif modellerdir. Diğer doğrusal olmayan modellere göre belirlenmesi, tahmin edilmesi ve yorumlanması daha kolay olan dolayısıyla günümüzde doğrusal olmayan zaman serisi analizinde yaygın olarak kullanılan (Hansen, 1997: 1) bu modeller; başlangıçta basit bir modelleme prosedürüne sahip olmamaları ile modelde yer alan eşik değişkeninin ve eşik değerlerinin tayinindeki yetersizlikler sebebiyle uygulamada yeteri kadar ilgi görmemiştir (Tsay, 1989: 231).

Her bir rejimde birinci derece otoregresif süreç içeren iki rejimli bir TAR modeli aşağıdaki gibi yazılabilir (Cryer ve Chan, 2008: 396):

$$y_t = \begin{cases} \phi_{1,0} + \phi_{1,1}y_{t-1} + \sigma_1\varepsilon_t & s_{t-d} \leq c \text{ ise} \\ \phi_{2,0} + \phi_{2,1}y_{t-1} + \sigma_2\varepsilon_t & s_{t-d} > c \text{ ise} \end{cases} \quad (1)$$

Burada ϕ 'ler otoregresif parametreleri, σ 'lar standart hataları göstermektedir. Gecikme ve eşik parametresi olarak adlandırılan d ve c sırasıyla pozitif bir tam sayı ve reel sayıdır. s_{t-d} , d dönem gecikmeli eşik değişkenidir ve gözlenen bir değişkenin deterministik fonksiyonu olduğu varsayılır (Jeremy, 2010: 192). ε_t ise bağımsız ve özdeş dağılım sıfır ortalama ve birim varyansa sahip rastsal hata terimidir.

(1)'deki model bağımlı değişkenin birden fazla gecikmeli değerlerini içerecek ve ikiden fazla rejime olanak verecek şekilde genişletilebilir. Model parametrelerinin eşik değişkeni s_{t-d} 'nin değerine göre değişmesine izin veren, $m-1$ eşik değeri $(c_1, c_2, \dots, c_{m-1})$ tanımlanarak elde edilen, m rejimli bir eşik otoregresif model aşağıdaki gibi gösterilebilir (Zivot ve Wang, 2006: 663):

$$y_t = X_t\phi^{(j)} + \sigma^{(j)}\varepsilon_t \quad c_{j-1} < s_{t-d} \leq c_j \text{ ise} \quad (2)$$

Burada $X_t = (1, y_{t-1}, y_{t-2}, \dots, y_{t-p})$, $j = 1, 2, \dots, m$ ve $-\infty = c_0 < c_1 < \dots < c_m = \infty$ 'dur.

TAR modellerinin doğrusal olmayan yapısında önemli bir yer tutan d dönem gecikmeli eşik değişkeni s_{t-d} , bağımlı değişken y_t 'nin dinamiklerini veya rejimini belirlediği düşünülen herhangi bir değişken olabilir. Ancak eğer rejimler zaman serisinin kendi gecikmeli değeri ile belirleniyorsa bu modeller kendinden uyarımlı eşik otoregresif modeller olarak adlandırılırlar.

SETAR modelleri ne kadar anlaşılması ve yorumlanması kolay, basit bir model biçimine sahip ise, eldeki mevcut veya deneysel verileri tam olarak açıklayan en iyi SETAR modelini belirlemek de o kadar zordur. Çünkü söz konusu model oluşturulurken tahmin edilmesi gereken birçok serbest parametre bulunmaktadır. Model yapısını belirleyen bu parametrelerin (yapısal parametrelerin) alabileceği değerlerin sayısına bağlı olarak çok fazla aday model bulunabilir ki bu durumda ortaya çıkan arama uzayının büyüklüğü tüm aday çözümlerin teker teker belirlenmesini ve makul sürelerde kontrol edilmesini olanaksız kılar (Baragona vd., 2004: 279). Örneğin, iki rejimli bir SETAR modeli için bilinmeyen yapısal parametrelerden; rejimlerin otoregresif dereceleri (p_1 ve p_2), gecikme parametresi

(d) için üst sınırın $p_1^{maks} = p_2^{maks} = 10$ ve $d^{maks} = 5$ şeklinde olduğu ve olası 50 eşik değeri (c) bulunduğu varsayılırsa, tam gecikme yapısında uygun modeli bulmak için toplamda 25000 model

denemek gerekir. Dahası eğer doğru gecikme yapısı otomatik olarak belirlenmek istenirse bu durumda 1047552250 tane SETAR modeli denemek zorundadır. Böyle bir durumda genel olarak değerlendirmemiz gereken $d^{maks}(2^{p_1^{maks}} - 1)(2^{p_2^{maks}} - 1)c$ tane SETAR modeli vardır. Tabii ki bu rakamlar rejim sayısının iki olarak bilindiği durum için geçerlidir. Eğer rejim sayısı bilinmiyorsa, sağlanan tek bilgi maksimum rejim sayısı ile ilgili ise en uygun modelin çok daha fazla sayıda alternatif arasında aranması gerekir. Bu yüzden pratikte uygun model seçilirken, belirlenmesi gereken parametrelerden biri ya da ikisi sabitken diğerleri belirlenmekte, yani ardışık bir seçim yöntemi izlenmektedir. Mevcut durum, beraberinde bazı öznel yargılar getirmektedir ki bu da elde edilen sonuçların güvenilirliğini azaltmaktadır (Ong vd., 2005: 887).

Rakip modeller arasında seçim yapmak amacıyla istatistikte yaygın olarak Akaike bilgi kriteri (AIC) ve Bayesian bilgi kriteri (BIC) kullanılmaktadır. Dolayısıyla prensipte SETAR modelleri; yapısal parametrelerin alabilecekleri değerlerinin tüm olası kombinasyonları dikkate alınarak bunlar içerisinde AIC veya BIC değerine göre en iyi olanın seçilmesi suretiyle belirlenebilirler. Bu bağlamda yapısal parametrelerin belirlenmesi, kesikli argümanlardan oluşan bir amaç fonksiyonunun optimum değerini bulmayı içeren bir optimizasyon problemi olarak düşünülebilir (Baragona ve Cucina, 2009: 3); ancak arama uzayındaki süresizlikler nedeniyle bu probleme geleneksel sayısal metotlar uygulanamaz (Maringer ve Meyer, 2008; Gilli ve Winker, 2009: 82). Bu noktada karşı karşıya olduğumuz durum en iyi değer için analitik araçlar olmadan taranması gereken geniş ve kesikli bir arama uzayıdır (Battaglia ve Protopapas, 2011a: 3). Özellikle çok fazla seçenek anlamına gelen böyle arama uzaylarında genetik algoritmalar arama ve problem çözme etkinliklerini göstermektedirler (Wu ve Chang, 2002: 322).

Literatürde, TAR veya SETAR modellerinin hem yapısal hem de otoregresif parametrelerini tahmin etme sürecinin genetik algoritmalar ile ele alındığı az sayıda çalışma bulunmaktadır. Wu ve Chang (2002), iki rejimli TAR modellerinin eşik ve gecikme parametrelerini belirlemek amacıyla genetik algoritmaları kullanmışlardır. TAR modellerini belirlemek amacıyla Tong (1983)'ün önerdiği AIC temelli algoritmaya karşı; döviz kuru verilerini kullanarak sınıadıkları, söz konusu parametreleri ve her bir rejimin otoregresif gecikmesini eş zamanlı olarak belirledikleri yöntemlerinin, arama süresi bakımından daha üstün olduğunu göstermişlerdir. Baragona vd. (2004), Parçalı Doğrusal Eşik Otoregresif (PLTAR) modellerini belirlemek ve tahmin etmek için basit bir genetik algoritma tabanlı yöntem önermişlerdir. Bu yöntem tüm model parametrelerini birlikte tahmin etmesine karşın ikiden fazla rejime imkân vermemektedir. Yöntemlerini simülasyon verileri ile değerlendirirken genetik algoritma ile belirlenmiş SETAR modelleri de kullanmışlardır. Baragona ve Cucina (2008), Çift Eşikli Genelleştirilmiş Otoregresif Koşullu Değişen Varyans (DTARCH) modelleri için melez bir algoritma önermişlerdir. Söz konusu algoritmada, DTARCH modellerin en önemli parametreleri olan rejim sayısını ve eşik değerlerini belirlemek için genetik algoritmaları kullanmışlardır. Battaglia ve Protopapas (2011a); çalışmalarında ele aldıkları zaman serileri için uygun iki rejimli SETAR, STAR veya PLTAR modellerinin parametrelerini, model türü de dahil olmak üzere genetik algoritma kullanarak belirlemişlerdir. Bunların dışında literatürde TAR modellerinde bilgi kriterleri kullanılarak; her bir rejimin otoregresif derecesi ve/veya gecikme parametresinin belirlendiği (Wong ve Li, 1998; De Gooijer, 2001), uygun rejim sayısının bulunduğu (Gonzalo ve Pitarakis, 2002; Strikholm ve Terasvirta, 2006; Peña ve Rodriguez, 2005; Hamaker, 2009) çeşitli çalışmalarda mevcuttur.

Bu çalışmada ise sunulacak genetik algoritma temelli yaklaşım ile SETAR modelleri -özellikle üç rejimli durum- ele alınacak yani çok fazla aday modelden oluşan büyük bir kümeden en uygun SETAR modeli veri odaklı bir belirleme stratejisi izlenerek seçilecektir. Bu stratejinin temelinde ekonometrik modellerin bir yaklaştırma olarak görülmesi (Hansen, 2005: 67) yatmaktadır. Bu çerçevede yapısal parametreler ve gerekli otoregresif parametreler birlikte belirlenecek bu sayede model kuranın şahsi yargılarına dayanan ayrıntılı ve zaman alıcı model belirleme süreci otomatik olarak gerçekleştirilecektir.

1. SETAR Modellerinin Tahmini

Uygulamada rejimlerin otoregresif (AR) dereceleri aynı değildir ve bilinmemektedir. Her bir rejimin otoregresif bileşenleri için uygun gecikme uzunluklarının belirlenmesinde sıklıkla uygulanan bir

yaklaşım; zaman serisi için kısmi otokorelasyon fonksiyonuna dayalı bir doğrusal AR(p) modelinin belirlenerek doğrusal olmayan modelin her rejiminin aynı sayıda gecikme içerdiği ve uygun derecenin p olduğunun varsayılmasıdır. Alternatif bir yaklaşım; otoregresif derecelerin her ne kadar olası tüm gecikme derecelerinin bileşiminin tahmin edilmesini gerektirse de, doğrudan bilgi kriterine dayalı olarak eş zamanlı seçilmesidir. Bu amaçla, AIC ve BIC kriterleri kullanılabilir. Bu bağlamda hangi kriter seçilirse seçilsin, gecikme uzunlukları için belirlenen üst sınırlar dahilinde SETAR modeli için uygun otoregresif dereceler, bilgi kriterini en küçükleyen derecelerdir.

Modelin diğer parametrelerinden eşik değeri, gecikme parametresi ve otoregresif parametrelerin birlikte tahmin edildiği bir yaklaşım Hansen (1997) tarafından sunulmuştur. Şöyle ki (1)'deki iki rejimli TAR modeli, $I \cdot J$ parantezin içerisindeki olay gerçekleşirse bir aksi takdirde sıfır değerini alan gösterge fonksiyon, $x_t = (1 y_{t-1} \dots y_{t-p})'$ ve $x_t(c) = (x_t' I(s_{t-1} \leq c) x_t' I(s_{t-1} > c))'$ olmak üzere alması olarak aşağıdaki gibi yeniden yazılabilir:

$$y_t = x_t(c)\varphi + \varepsilon_t \quad (3)$$

Burada $j=1,2$ için $\varphi_j = (\varphi_{j,0}, \varphi_{j,1}, \dots, \varphi_{j,p})$ ve $\varphi = (\varphi_1', \varphi_2')$ 'dir.

(3) bir regresyon eşitliği olduğundan dolayı ilgili parametreler φ , c ve σ^2 en küçük kareler (EKK) metodu ile tahmin edilebilir. Ayrıca hata terimlerinin normal dağıldığı varsayımı altında en küçük kareler tahmincileri, en çok olabilirlik tahmincilerine eşit olacaktır. Ancak söz konusu regresyon eşitliği doğrusal ve sürekli değildir bu yüzden EKK tahmincileri ardışık koşullu EKK metodu ile elde edilmelidir (Hansen, 1997: 3). Eşik parametresi c 'nin verili bir değeri için φ 'nin EKK tahmini,

$$\hat{\varphi}(c) = \left(\sum_{t=1}^n x_t(c) x_t(c)' \right)^{-1} \left(\sum_{t=1}^n x_t(c) y_t \right) \quad (4)$$

ilgili kalıntılar $\hat{\varepsilon}_t(c) = y_t - x_t(c)' \hat{\varphi}(c)$ ve kalıntı varyansı,

$$\hat{\sigma}^2(c) = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t(c)^2 \quad (5)$$

şeklinde elde edilir (Franses ve Dijk, 2000: 84).

Tüm mümkün eşik değerleri ($C = [c, \bar{c}]$) için (5)'deki kalıntı varyansının en küçüklenmesi ile c 'nin EKK tahmini bulunabilir:

$$\hat{c} = \arg \min_{c \in C} \hat{\sigma}^2(c) \quad (6)$$

(6)'daki en küçükleme problemi doğrudan arama metodu ile çözülebilir. Sonuçta otoregresif parametreler için ulaşılan nihai EKK tahminleri $\hat{\varphi} = \hat{\varphi}(\hat{c})$ ve benzer şekilde kalıntı varyansı da $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{c})$ şeklinde olacaktır. Ancak izlenen algoritma sonucunda otoregresif parametrelerin güvenilir tahminlerinin elde edilmesi için mümkün eşik değerleri kümesi C , her bir rejim yeterli gözleme sahip olacak şekilde belirlenmelidir. Bu doğrultuda yaygın tercih C 'nin, her bir rejimin gözlemlerin önceden belirlenmiş en az 0,10 veya 0,15'lik bölümünü içermesini sağlayacak şekilde belirlenmesidir (Andrews, 1993: 826; Franses ve Dijk, 2000: 84; Strikholm ve Terasvirta, 2006: 481).

Gecikme parametresi ise modeldeki diğer parametreler ile beraber tahmin edilebilir. Durum SETAR modelleri çerçevesinde ele alındığında $d \in [1, d]$ olmak üzere eşik değişkeni, y_{t-d} şeklinde belirlenir. Devamında (6)'ın gecikme parametresinin çeşitli değerleri için genişletilmesi gerekir. Bu durumda en küçükleme problemi $D = \{1, \dots, \bar{d}\}$ olmak üzere,

$$(\hat{c}, \hat{d}) = \arg \min_{c \in C, d \in D} \hat{\sigma}^2(c, d) \quad (7)$$

haline gelecektir (Franses ve Dijk, 2000: 87).

2. Genetik Algoritmalar

Genetik algoritma, çözüm alanını rastgele tarayarak optimum çözümü bulmaya çalışan (Şen, 2004: 65); bir çeşit alan bağımsız, hızlı, güvenilir ve güçlü arama ve optimizasyon tekniğidir (Michalewicz, 1996: 15). Yerel optimuma kolayca yakalanmalarını sağlayan popülasyon tabanlı paralel bir arama mekanizmasına sahip olan genetik algoritmalar; problem parametreleri için kullandıkları farklı şekillerdeki kodlamalar ve amaç fonksiyonunun türevlenebilir olmasının gerekmemesi gibi noktalarda geleneksel optimizasyon metodlarından ayrılmaktadırlar (Sivanandam ve Deepa, 2007: 34).

Genetik algoritmalar problemin çözümüne rastgele bireylerden bir popülasyon oluşturarak başlarlar. Popülasyondaki her bir birey probleme ilişkin olası bir çözümü gösterir ve kromozom olarak adlandırılır. Kromozomlar bir dizi halinde sıralanmış keyfi uzunluktaki bit dizilerinden yani genlerden oluşur. Genlerin kromozomlardaki konumu lokus olarak adlandırılır ve genlerin alabileceği her farklı değere alel denir. Olası çözümleri kromozom formunda kodlamak genetik algoritmaların ilk aşamasıdır. Kromozomlar hesaplama kolaylığı nedeniyle genellikle ikili sayı sistemi temel alınarak kodlanır ancak ondalık sayı sistemi kullanılarak da kodlama yapılabilir (Şen, 2004: 66).

Popülasyonda yer alan bireyler, bazı uygunluk ölçüleri kullanılarak ardışık tekrarlar sayesinde gelişirler ki her bir tekrar bir nesil olarak adlandırılır (Sakawa, 2002: 12). Her nesilde öncelikle uygunluk fonksiyonunun ilgili değeri hesaplanarak her bireyin problemin çözümüne ne kadar yakın olduğunu gösteren uygunluk değeri bulunur. Devamında sonraki neslin popülasyonu genetik operatörler kullanılarak oluşturulur. Genetik algoritmalarda kullanılan genetik operatörler seçim, çaprazlama ve mutasyondur. Popülasyonun ortalama uygunluğunu artırmak amacıyla seçim işleminde yüksek uygunluk değerlerine sahip bireylere daha çok yaşama şansı verilmek suretiyle sonraki neslin popülasyonuna eklenecek bireyler belirlenir (Blickle ve Thiele, 1995: 9). Çaprazlama seçim sürecinde seçilen iki eski bireyin rastgele birleşimlerinden yeni iki birey oluşturulmasıdır. Genetik algoritmalarındaki en güçlü genetik operatör olan çaprazlama (Bingül vd., 2000: 96) bireylerin iyi özelliklerinin bir araya gelmesini kolaylaştırır. Kaybolan genetik materyalin geri getirilmesinde ve genetik bilginin rastgele karıştırılmasında rol oynayan (Sivanandam ve Deepa, 2007: 56) mutasyon ise seçilen bireylerin bir veya daha fazla geninin değiştirilmesidir. Genetik algoritma, sonlandırma ölçütü sağlanıncaya kadar devam eder. Algoritmayı sonlandırmak için farklı ölçütler kullanılabilir, ancak genellikle algoritma sabit bir iterasyon sayısına ulaştıktan sonra sonlandırılır.

$P(t)$ t. nesildeki popülasyonu göstermek üzere genetik algoritmaların genel yapısı Şekil 1'deki gibi tanımlanır (Michalewicz, 1996: 2; Sakawa, 2002: 12; Şen, 2004: 64).

Şekil 1: Genetik Algoritmanın Genel Yapısı**Procedure: Genetik Algoritma****begin**

t:=0;

P(t)'yi başlat;

P(t)'yi değerlendir;

while not (sonlandırma şartları) **do****begin**

t:=t+1;

P(t-1)'den P(t)'yi seç;

P(t) içinde çiftleri yenile;

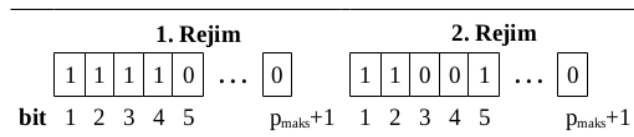
P(t)'yi değerlendir;

end**end****3. SETAR Modelleri İçin Genetik Algoritma Yaklaşımı**

Rejim sayısının bilindiği varsayımı altında SETAR modellerinin belirlenmesi bir optimizasyon problemi olarak değerlendirildiğinde ilgilenilen karar değişkenleri; her bir rejimde yer alan gecikmeli değişkenler, gecikme parametresi ve eşik değeridir. Bu probleme genetik algoritmalar yardımıyla çözüm aranırken ilgili modelin kromozom formunda ifade edilmesi gerekmektedir. Probleme yer alan tüm karar değişkenleri kesiklidir ve bağımlı değişkenin gecikmeli değerlerinin hangisinin modelde yer alıp almayacağını iki seçeneikli bir durumu ifade etmektedir. Dolayısıyla kodlama türü olarak ikili kodlama tercih edilmiştir. İkili kodlamada bit dizilerinin uzunlukları; otoregresif derecelerin ve gecikme parametresinin üst sınırlarına göre tayin edilecektir. Takip eden iki rejimli,

$$y_t = \begin{cases} \varphi_{1,0} + \varphi_{1,1}y_{t-1} + \varphi_{1,2}y_{t-2} + \varphi_{1,3}y_{t-3} + \varepsilon_t & y_{t-d} \leq c \text{ ise} \\ \varphi_{2,0} + \varphi_{2,1}y_{t-1} + \varphi_{2,4}y_{t-4} + \varepsilon_t & y_{t-d} > c \text{ ise} \end{cases} \quad (8)$$

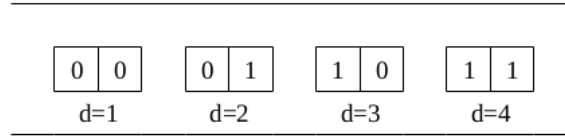
modelinde otoregresif parametreler önsel olarak sağlanan bir maksimum gecikme uzunluğu $p_{maks} \geq 4$ dahilinde her biri $p_{maks} + 1$ bit uzunluğunda iki gen yardımıyla Şekil 2'deki gibi gösterilebilir.

Şekil 2: Otoregresif Parametrelerin Kodlanması

Şekilde, ilk gendeki 1, 2, 3 ve 4. bit ile ikinci gendeki 1, 2 ve 5. bit bir değerini almıştır. Bu söz konusu bitlerin temsil ettiği otoregresif parametrelerin modelde yer aldığı anlamına gelmektedir. Her bir gendeki ilk bit sabit terimi ifade ederken, ikinci bittenden itibaren (i+1). bit i. otoregresif parametreyi temsil etmektedir. Modelde yer almayan diğer parametreler ise ilgili bitlerde sıfır değerini alacak şekilde temsil edilmişlerdir.

Gecikme parametresi d, tam sayı olması itibarıyla $1 \leq d \leq d_{maks}$ koşuluyla kolaylıkla ondalık sayı sisteminden ikilik sayı sistemine çevrilerek kodlanabilir. Örneğin, $d_{maks} = 4$ için iki bit uzunluğunda bir gen kullanılacak ve her bir gecikme parametresi Şekil 3'deki gibi kodlanacaktır.

Şekil 3: Gecikme Parametresinin Kodlanması



Kodlama sürecinin en hassas kısmı eşik değerinin kodlanmasıdır. İki rejimli bir SETAR modelinde gözlemler arasından seçilen bir tane eşik değeri vardır. İlgili gende olası eşik değerlerinin sayısına bağlı olarak belirlenen bit uzunluğunda kodlanan söz konusu eşik değeri, öncelikle sıfır ve bir arasında bir reel sayıya (r) dönüştürülür. Elde edilen bu değer ilk rejimde bulunan gözlemlerinin toplam gözlem sayısına oranını göstermektedir. Dolayısıyla gözlemler küçükten büyüğe doğru sıralandığında ilk rejimin içerdiği gözlem sayısı; N toplam gözlem sayısı ve a bir rejimde bulunması gereken en az gözlem sayısı olmak üzere,

$$n_1 = a + (N - 2a)r \quad (9)$$

olacak (Battaglia ve Protopapas, 2011b) ve n_1 gözlem eşik değeri olarak seçilecektir. Burada r 'nin uç değerleri olan sıfır ve bir için, sırasıyla ilk rejim ve ikinci rejim en az gözlem sayısı kadar gözlem içerecektir.

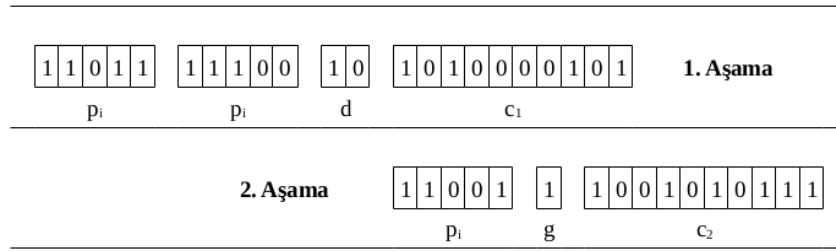
Belirlenen karar değişkenleri doğrultusunda ilgili SETAR modeli en küçük kareler yöntemiyle tahmin edilerek modelin uyumu bir başka ifadeyle karar değişkenlerinin sağlandığı kromozomun uygunluğu bilgi kriterleri hesaplanarak değerlendirilebilir. Bilgi kriterini en küçükleyen modelin uygun model olması ve bazı durumlarda bilgi kriterinin hem pozitif hem de negatif değerler alabilmesi nedeniyle uygunluk değeri olarak kullanılmadan önce bir s kromozomu ile belirlenen x modelinden hesaplanan bilgi kriteri takip eden şekilde dönüştürülmelidir:

$$f(s) = \begin{cases} C_{maks} - BK(x), & \text{eger } BK(x) < C_{maks} \\ 0, & \text{d.d.} \end{cases} \quad (10)$$

Burada C_{maks} bilgi kriterinin mevcut popülasyona kadar ki en büyük değeridir.

TAR literatüründe ikiden fazla rejim içeren modellerin her seferinde bir eşik değeri bulunması suretiyle ardışık olarak belirlenebileceği gösterilmiştir (Hansen, 1999; Gonzalo ve Pitarakis, 2002). Şöyle ki, önce iki rejim içeren bir TAR modelinin gecikme parametresi ve eşik değeri tahmin edilir. Devamında bu tahminlere dayanarak daha fazla rejim içeren nihai model belirlenir ise bu sayede nihai modele ait gerçek gecikme parametresi ve eşik değerleri için tutarlı tahminler sağlanabilir (Hansen, 1999: 560). Bu noktadan hareketle yukarıda iki rejimli durum için açıklanan kodlama stratejisi genişletilerek, üç rejimli durum için iki aşamalı bir genetik algoritma yaklaşımı sunulmuştur. İlk aşamada; iki rejimli durumdaki kodlamaya sahip genetik algoritma ile iki rejimim gecikme yapısı, ilk eşik değeri ve gecikme parametresi bulunmuştur. Sonrasında ikinci aşamada; ilk aşamada elde edilen parametreler kullanılarak ikinci eşik değeri ve kalan rejimin gecikme yapısı belirlenerek nihai modele ulaşılmıştır. İkinci aşamada kullanılan genetik algoritmanın kromozom yapısı Şekil 4'deki gibidir.

Şekil 4: Üç Rejimli SETAR Modeli İçin İki Aşamalı Kromozom Kodlaması



Modelin gecikme yapısı, belirlenen maksimum gecikme uzunluklarına bağlı olarak iki rejimli durumdaki gibi kodlanmıştır. Mevcut durum için kodlamaya fazladan bir bit uzunluğunda bir gen (g) eklenmiştir. İlgili genin ve eşik değerinin değerine bağlı olarak ilk ve ikinci aşamada belirlenen gecikme yapılarının hangi rejimlere ait olduğuna karar verilmiştir. Buna göre gen bir değerini aldığı anda ikinci aşamada belirlenen gecikme yapısı orta rejim için kullanılırken, sıfır değerini aldığı anda yeni eşik değerinin ilk eşik değerinden büyük veya küçük olmasına bağlı olarak sırasıyla üst ve alt rejimler için kullanılmıştır. İkinci eşik değeri ise kromozomda yine iki rejimli durumdaki gibi gösterilmiş ve takip eden eşiklik kullanılarak dönüştürülmüştür:

$$n_2 = \begin{cases} a + (n_1 - 2a)r, & \text{eger } n_1 < 2a \\ n_1 + a + (N - 2a - n_1)r, & \text{eger } n_1 > N - 2a \\ c_2 + I[c_2 > n_1 - a](2a - 1), & d.d. \end{cases} \quad (11)$$

Eşitlikte $c_2 = a + (N - 4a + 1)r$ ve $I[\cdot]$ parantezin içindeki ifade gerçekleştiğinde bir diğer durumda sıfır değerini alan gösterge fonksiyondur.

4. Simülasyon Çalışması

Bu noktaya kadar açıklanan kodlama stratejisi kullanılarak oluşturulan genetik algoritmanın, SETAR modellerini belirleme noktasındaki performansını değerlendirmek amacıyla bir Monte Carlo simülasyonu gerçekleştirilmiştir. Bir simülasyon çalışmasında genetik algoritmanın performansı, belirli sayıda gerçekleştirilen denemelerde seçilen bir bilgi kriterine göre gerçek modelin elde edilme frekansı hesaplanarak ölçülebilir. Bu amaçla öncelikle bilinen bir modelden veri üretilir ve devamında bilgi kriteri bağlamında model seçim işlemi için genetik algoritma kullanılır. Ancak burada üzerinde durulması gereken önemli bir konu, gerçek modelin her zaman en küçük bilgi kriteri değerini sağlamadığıdır (Balcombe, 2005; Hamaker, 2009; Battaglia ve Protopapas, 2011a). Böyle bir durumda genetik algoritmanın ulaştığı model gerçek model olmasa da bilgi kriteri açısından en küçük değeri sağlayan model olabilir. Dolayısıyla genetik algoritmanın performansının doğru olarak belirlenebilmesi için simülasyonlarda gerçek modelin seçilme frekansı yerine en iyi yakınsayan modelin seçilme frekansı dikkate alınmalıdır ki bu da tüm olası modellerin tahmin edilmesi ve her birine ilişkin bilgi kriterlerinin hesaplanmasını gerektirmektedir. Bu hesaplama açısından pahalı olmasına rağmen, kavramsal olarak araştırılması kolay bir değerlendirmedir (Balcombe, 2005: 212). Makul süreler dahilinde gerçekleştirilebildiği sürece genetik algoritmanın performansı bu yol izlenerek belirlenebilir. Fakat arama uzayının büyüklüğünün buna imkan vermediği durumlarda genetik algoritma, seçilen model gerçek modele ilişkin hesaplanan bilgi kriterinden daha küçük ya da eşit bir bilgi kriteri sağlıyorsa başarılı sayılabilir. Ancak bu şekilde ulaşılan sonuçların ilk yol izlenerek ulaşılan sonuçlar kadar kesin olmayacağı açıktır. Bu yüzden simülasyon verileri üretilirken tüm olası modellerin denenebileceği kadar küçük bir modelin kullanılması daha uygun olacaktır. Yine de özellikle ikiden

fazla rejim içeren durumlarda küçük bir model kullanılsa da arama uzayı çok geniş olabilir ve ilk yol izlenemeyebilir.

Simülasyon çalışmasında veri üretme sürecinde $\varepsilon_t \sim N(0,1)$ olmak üzere aşağıdaki model kullanılmıştır:²

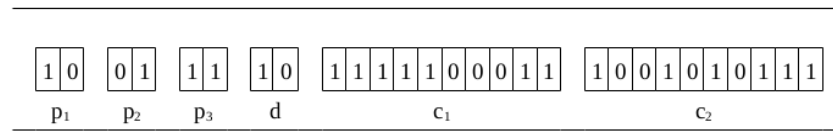
$$y_t = \begin{cases} 2,7 + 0,8y_{t-1} + \varepsilon_t, & y_{t-2} < 5 \\ 6 + 0,9y_{t-2} + 1,5y_{t-4} + \varepsilon_t, & 5 < y_{t-2} \leq 12 \\ 1 + 0,7y_{t-1} - 0,2y_{t-2} + 0,3y_{t-4} + \varepsilon_t, & y_{t-2} > 12 \end{cases} \quad (12)$$

Bu model kullanılarak 100, 200 ve 400 efektif gözlem içeren, üç veri seti oluşturulmuştur.³ Deneyle hesaplama maliyeti nedeniyle tüm olası modeller tahmin edilememiş bu nedenle tam gecikme yapısı kullanılarak arama uzayı daraltılmıştır. Bu şekilde belirlenen genel modelden, %5 anlamlılık seviyesinde anlamsız parametreler atılarak gerçek gecikme yapısını içeren nihai modele ulaşılmıştır. Ayrıca her veri seti için (12); bilinen gecikme yapısı, gecikme parametresinin asıl değeri ve eşik değeri olarak seriden asıl eşik değerine en yakın gözlem kullanılmak suretiyle tahmin edilmiştir. Her iki yolla elde edilen modeller için k tüm rejimlerdeki toplam parametre sayısı olmak üzere,

$$AIC = N \ln \hat{\sigma}^2 + 2k \quad (13)$$

formülünden hareketle AIC hesaplanmış daha küçük AIC değerine sahip model algoritmanın başarısını değerlendirmek adına ölçüt olmuştur. Bu şekilde elde edilen ölçüt modeller Tablo 1'de görülmektedir.

Şekil 5: Üç Rejimli Genel SETAR Modeli İçin Kromozom Kodlaması



Üç rejimli durum için öncelikle ölçüt modellerin yuvalandığı modellerin belirlenmesinde genetik algoritmanın kullanılabilirliği araştırılmıştır. Bu amaçla çözümlerin Şekil 5'deki gibi kodlandığı ve Tablo 2'deki tasarıma sahip genetik algoritma kullanılarak eş zamanlı olarak her bir rejimdeki maksimum gecikme uzunluğu, gecikme parametresi ve eşik değerleri bulunmuştur. Eşik değerleri ve gecikme parametresi kromozomlarda iki rejimli durumdaki gibi ifade edilirken, gecikme parametreleri doğrudan ondalık sayı sisteminden ikili sayı sistemine çevrilerek kodlanmıştır. Farklı popülasyon büyüklüklerinin kullanıldığı ve her veri seti için genetik algoritmanın 100 kez çalıştırıldığı deneyin sonuçları Tablo 3'de verilmiştir.

² Strikholm ve Terasvirta (2006)'nın araştırmalarında kullandıkları model, mevcut duruma göre gecikme yapısı boşluklar içerecek şekilde yeniden uyarlanmıştır.

³ Tüm simülasyonlarda her veri seti için aslında N+112 gözlem oluşturulmuş ve her veri setinde ilk 100 veri başlangıç değerlerinin etkisini azaltmak amacıyla dışlanmıştır. Sonraki 12 veri maksimum gecikme uzunluğu 12 olmak üzere uygun gecikme yapısının oluşturulmasında kullanılmıştır.

Tablo 1: Üç Rejimli SETAR Modeline İlişkin Tahmin Edilen Simülasyon Modelleri

	N					
	100		200		400	
$\phi_{1,0}$	2,6999	(0,1422)	2,7546	(0,1287)	2,8481	(0,1003)
$\phi_{1,1}$	0,8077	(0,0187)	0,7796	(0,0177)	0,7816	(0,0137)
$\phi_{1,2}$	-	-	-	-	-	-
$\phi_{1,3}$	-	-	-	-	-0,0226	(0,0094)
$\phi_{1,4}$	-	-	-	-	-	-
$\phi_{2,0}$	5,9289	(0,5582)	5,6475	(0,4722)	5,8966	(0,3150)
$\phi_{2,1}$	-	-	-	-	-	-
$\phi_{2,2}$	-0,8875	(0,0678)	-0,8351	(0,0588)	-0,9017	(0,0389)
$\phi_{2,3}$	0,0814	(0,0397)	-	-	-	-
$\phi_{2,4}$	1,4308	(0,0391)	1,4833	(0,0270)	1,5113	(0,0160)
$\phi_{3,0}$	8,4141	(1,6053)	12,0767	(1,6643)	6,5280	(1,2856)
$\phi_{3,1}$	-0,0898	(0,0396)	-0,1747	(0,0442)	0,0513	(0,0197)
$\phi_{3,2}$	-1,0279	(0,0926)	-1,1656	(0,0861)	-0,9824	(0,0569)
$\phi_{3,3}$	-	-	-0,1039	(0,0449)	0,0784	(0,0391)
$\phi_{3,4}$	1,5433	(0,0767)	1,4014	(0,0495)	1,4772	(0,0720)
d	2		2		2	
c_1	4,7325		4,9733		4,9850	
c_2	11,7943		11,4781		12,1155	
$\hat{\sigma}^2$	0,5846		0,9221		0,9366	
AIC	-33,6773		3,7921		-4,1727	

Tablo 3'de farklı popülasyon büyüklükleri için genetik algoritmanın hedeflenen AIC değerine ulaşmak için ihtiyaç duyduğu minimum, maksimum ve ortalama iterasyon sayıları görülmektedir. Burada algoritmanın performansını, veri setinin büyüklüğünden çok incelenen serinin karmaşıklığının etkilediği söylenebilir. Şöyle ki 200 gözlem içeren veri setinde her popülasyon büyüklüğünde algoritma 100 denemenin tamamında hedeflenen değere ulaşırken; 100 ve 400 gözlem içeren veri setlerinde 50 kromozomdan oluşan popülasyon, algoritmanın tüm denemelerde başarılı olması için yeterli olmamıştır. Dahası veri seti büyüklüğündeki artışın algoritmanın performansına tahmin edilen model sayısı noktasında olumsuz bir etkisi olmamıştır. Deney sonuçlarına göre genetik algoritmanın uygun modeli belirleyebilmesi için veri setinin büyüklüğünden bağımsız olarak, 150 kromozomdan oluşan bir popülasyonda en çok 400 iterasyon işletilmesi yeterlidir. Bu da toplamda 60500 modelin tahmin edilmesi anlamına gelmektedir ki bu rakam 100, 200 ve 400 gözlem içeren veri setlerinde toplam model sayısının, sırasıyla %7,11 , %1,81 ve %0,45'ine denk düşmektedir. Dolayısıyla hesaplama maliyeti açısından önerilen yaklaşımın ızgara aramasından daha üstün olduğu açıktır.

Tablo 2: Üç Rejimli Genel SETAR Modeline İlişkin Genetik Algoritma Tasarımı

p_{maks}	4
d_{maks}	4
Kromozom Uzunluğu	28 bit
Uygunluk Fonksiyonu	AIC (13)
Başlangıç Popülasyonu	500
Elitizm	2
Seçim Türü	Turnuva (Turnuva Büyüklüğü=2)

Çaprazlama Türü	Tek noktalı çaprazlama
Çaprazlama Olasılığı	0,80
Mutasyon Olasılığı	0,06

Tablo 3: Üç Rejimli Genel SETAR Modeline İlişkin Simülasyon Sonuçları

N _{pop}	N=100			N=200			N=400		
	Min	Maks	Ort	Min	Maks	Ort	Min	Maks	Ort
50	2	992	152,77	13	276	69,23	5	987	176,62
100	5	720	76,08	6	121	33,90	17	512	87,65
150	6	366	40,65	2	132	34,50	11	375	65,39
200	4	395	38,74	4	93	32,72	12	122	53,46

Not: İstatistikler başarılı denemelere göre hesaplanmıştır. 50 popülasyon büyüklüğünde, 100 ve 400 gözlem içeren veri setlerinde genetik algoritma sırasıyla 83 ve 91 kez başarılı olmuştur.

Tablo 4: Üç Rejimli SETAR Modeline İlişkin Genetik Algoritma Tasarımı

P_{maks}	4
d_{maks}	4
Kromozom Uzunluğu	22 bit (1. Aşama) 16 bit (2. Aşama)
Uygunluk Fonksiyonu	AIC (13)
Başlangıç Popülasyonu	500
Elitizm	2
Seçim Türü	Turnuva (Turnuva Büyüklüğü=2)
Çaprazlama Türü	İki noktalı çaprazlama
Çaprazlama Olasılığı	0,80
Mutasyon Olasılığı	0,05

Sonraki deneyde doğrudan ölçüt modele ulaşma süreci açıklanan iki aşamalı yaklaşım ve kodlama stratejisiyle araştırılmıştır. Bu bağlamda algoritmanın ilk aşamasında; iki rejimli SETAR modelini belirlemeye yönelik bir ön çalışmasının sonuçlarından hareketle, 100 kromozomdan oluşan bir popülasyon kullanılmış ve algoritma 80 iterasyon işletilmiştir. İkinci aşama için uygun popülasyon büyüklüğü ve iterasyon sayısını belirlemek adına genetik algoritma, her veri seti için söz konusu parametrelerin farklı değerleri kullanılarak 100 kez çalıştırılmıştır. Bu noktada hedef, genetik algoritma ile ölçüt modellere ilişkin AIC değerinden daha düşük bir AIC değerine ya da en azından ölçüt modelin AIC değerine ulaşmaktır. Gerçekleştirilen deneyin her iki aşamasında da Tablo 4'de verilen tasarıma sahip genetik algoritma kullanılmış, elde edilen sonuçları Tablo 5'de verilmiştir.

Tablo 5: Üç Rejimli SETAR Modeline İlişkin Simülasyon Sonuçları

N _{pop}	İter.	N					
		100		200		400	
30	20	97	(93)	99	(99)	89	(0)
	30	98	(94)	100	(100)	98	(0)
	40	98	(95)	100	(100)	100	(0)
	50	98	(95)	100	(100)	100	(0)
50	20	96	(94)	100	(100)	95	(0)
	30	97	(97)	100	(100)	100	(0)
	40	100	(100)	100	(100)	100	(0)
	50	100	(100)	100	(100)	100	(0)

Not: Parantez içerisinde verilen değerler ölçüt modelden daha iyi uyuma sahip model sayısını göstermektedir.

Genetik algoritmanın en azından ölçüt modeli bulması için ikinci aşamada popülasyon büyüklüğü 50 iken 40 iterasyon yeterli olmaktadır. Bu toplamda 11000 modelin tahmin edilmesi anlamına gelmektedir. Hem popülasyon büyüklüğü hem de iterasyon sayısı 50 olarak belirlendiğinde, 400 gözlem içeren veri seri için algoritma tüm denemelerde ölçüt modeli belirleyebilmiş ancak ölçüt modelden daha iyi bir modele ulaşamamıştır. Tüm olası modelleri denememiz mümkün olmadığı için söz konusu modelin veriye en iyi uyan model olup olmadığını kesin olarak bilemeyiz, yine de ölçüt modele tahmin edilmesi gereken model sayısının sadece %0,087'sinin denenerek ulaşılması genetik algoritma temelli yaklaşımın başarısını göstermektedir. Diğer taraftan 100 ve 200 gözlem içeren veri setlerinde AIC kriterine göre veriye ölçüt modellerden daha iyi modeller yaklaştırılmıştır. Algoritmanın ulaştığı en iyi modeller Tablo 6'da verilmiştir. Yine bu modellerin tüm olası modeller arasındaki en iyi yaklaşımlar olduğundan emin olamayız fakat daha az hesaplama maliyeti ile elde edilen modellerin ölçüt modellere göre daha iyi uyuma sahip olması genetik algoritma bağlamında önerilen yaklaşımın etkinliğini açıkça ortaya koymaktadır.

Tablo 6: Üç Rejimli SETAR Modeline İlişkin Tahmin Edilen En İyi Modeller

N	$\varphi_{1,j}$	$\varphi_{2,j}$	$\varphi_{3,j}$	d	c_1	c_2	$\hat{\sigma}^2$	AIC
100	0-1	1-3-4	0-2-3-4	2	3,7812	4,7325	0,5750	-37,3344
200	0-1	0-2-4	0-1-2-4	2	4,9733	7,3468	0,9265	2,7415

SONUÇ

Bu çalışmada, özellikle üç rejimli durum bağlamında SETAR modellerinin belirlenmesi amacıyla; genetik algoritma tabanlı, tüm yapısal parametrelerin birlikte belirlendiği, daha nesnel ve tutarlı bir model belirleme yaklaşımı önerilmiştir. Önerilen genetik algoritma temelli yaklaşım simülasyon verileri kullanılarak sınanmıştır. Ulaşılan sonuçlar yaklaşımın hesaplama maliyeti açısından ızgara aramasından daha üstün olduğunu göstermiştir. Dahası simülasyon verilerine en iyi uyan modelin bilinmediği mevcut durumda, algoritma çoğu kez ölçüt olarak kabul edilen modelden daha iyi modellere yakınsayabilmiştir. Bu noktada ilgilenilen veri setinin içerdiği gözlem miktarının, algoritmanın performansına tahmin edilen model sayısı noktasında olumsuz bir etkisinin olmadığı aksine performansın daha çok incelenen serinin karmaşıklığına bağlı olduğu tespit edilmiştir. Üç rejimli modele bağlı olarak ortaya çıkan geniş arama uzayının, model belirleme sürecine yönelik olarak yarattığı olumsuz etkilerin iki aşamada iki farklı genetik algoritma kullanılmak suretiyle önüne geçilebileceği gösterilmiştir. Üç rejimli durum için önerilen algoritma tasarımı daha fazla rejim içeren durumlara, mevcut aşama sayısının artırılması yoluyla uyarlanabilir. Ancak çalışmada kullanılan genetik algoritma parametrelerinin ve seçime esas teşkil eden AIC bilgi kriterinin her veri seti uygun olmayabileceği gerçeği, önerilen yaklaşımın kullanılması noktasında dikkat edilmesi gereken bir husustur.

KAYNAKÇA

- Andrews, Donald W. K. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point", *Econometrica*, Volume: 61, Issue: 4, p. 821–856.
- Balcombe, Kelvin G. (2005), "Model Selection Using Information Criteria and Genetic Algorithms", *Computational Economics*, Volume: 25, Issue: 3, p. 207–228.
- Baragona, Roberto, Francesco Battaglia ve Domenico Cucina (2004), "Fitting piecewise linear threshold autoregressive models by means of genetic algorithms", *Computational Statistics & Data Analysis*, Volume: 47, Issue: 2, p. 277–295.

- Baragona, Roberto ve Domenico Cucina (2008), "Double threshold autoregressive conditionally heteroscedastic model building by genetic algorithms", *Journal of Statistical Computation and Simulation*, Volume: 78, Issue: 6, p. 541-558.
- Baragona, Roberto ve Domenico Cucina (2009), "Genetic search for threshold parameters in time series threshold models: algorithms and computer programs", TECHNICAL REPORT 10/2009, Department of Statistics, Probability and Applied Statistics, Sapienza University of Rome, Italy.
- Battaglia, Francesco ve Mattheos K. Protopapas (2011a), "Time-varying multi-regime models fitting by genetic algorithms", *Journal of Time Series Analysis*, Volume: 32, Issue: 3, p. 237-252.
- Battaglia, Francesco ve Mattheos K. Protopapas (2011b), "Multi-regime models for nonlinear nonstationary time series", *Computational Statistics*, Volume: 27, Issue: 2, p. 319-341.
- Bingul, Z. A. S. Sekmen S. Palaniappan ve S. Zein-Sabatto (2000), "Genetic algorithms applied to real time multiobjective optimization problems", Southeastcon 2000. Proceedings of the IEEE, p. 95-103.
- Blickle, Tobias ve Lothar Thiele (1995), "A Mathematical Analysis of Tournament Selection", Proceedings Of The Sixth International Conference On Genetic Algorithms, p. 9-16.
- Cryer, Jonathan D. ve Kung-Sik Chan (2008), *Time Series Analysis: With Applications in R (2nd ed.)*, Springer.
- De Gooijer, Jan G. (2001), "Cross-validation Criteria for Setar Model Selection", *Journal of Time Series Analysis*, Volume: 22, Issue: 3, p. 267-281.
- Franses, Philip H. ve Dick van Dijk (2000), *Nonlinear Time Series Models In Empirical Finance*, Cambridge University Press.
- Gilli Manfred ve Peter Winker (2009), "Heuristic Optimization Methods In Econometrics", BELSLEY, David A. ve Erricos J. KONTOGHIOGHES (Ed.), *Handbook of Computational Econometrics*, p. 81-119, John Wiley and Sons.
- Gonzalo, Jesus ve Jean-Yves Pitarakis (2002), "Estimation And Model Selection Based Inference In Single And Multiple Threshold Models", *Journal of Econometrics*, Volume: 110, Issue: 2, p. 319-352.
- Hamaker, E. L. (2009), "Using Information Criteria To Determine The Number Of Regimes In Threshold Autoregressive Models", *Journal of Mathematical Psychology*, Volume: 53, Issue: 6, p. 518-529.
- Hansen, Bruce (1997), "Inference in TAR Models", *Studies in Nonlinear Dynamics ve Econometrics*, Volume: 2, Issue: 1, p. 1-14.
- Hansen, Bruce (1999), "Testing for Linearity", *Journal of Economic Surveys*, Volume: 13, Issue: 5, p. 551-576.
- Hansen, Bruce (2005), "Challenges for Econometric Model Selection", *Econometric Theory*, Volume: 21, Issue: 1, p. 60-68.
- Jeremy, Piger (2010), "Econometrics: Models of Regime Changes", MEYERS, Robert A. (Ed.), *Complex Systems in Finance and Econometrics (1st ed.)*, p. 190-202, Springer.
- Maringer, Dietmar G. ve Mark Meyer (2008), "Smooth Transition Autoregressive Models - New Approaches to the Model Selection Problem", *Studies in Nonlinear Dynamics & Econometrics*, Volume: 12, Issue: 1, p. 1-21.
- Michalewicz, Zbigniew (1996), *Genetic Algorithms + Data Structures = Evolution Programs (3rd ed.)*, Springer.

- Ong, Chorng-Shyong, Jih-Jeng Huang ve Gwo-Hshiung Tzeng (2005), "Model Identification Of ARIMA Family Using Genetic Algorithms", *Applied Mathematics and Computation*, Volume: 164, Issue: 3, p. 885–912.
- Peña, Daniel ve Julio Rodriguez (2005), "Detecting Nonlinearity In Time Series By Model Selection Criteria", *International Journal of Forecasting*, Volume: 21, Issue: 4, p. 731–748.
- Sakawa, Masatoshi; (2002), *Genetic Algorithms And Fuzzy Multiobjective Optimization*, Springer.
- Sivanandam, S. N. ve S. N. Deepa (2007), *Introduction To Genetic Algorithms*, Springer.
- Strikholm, Birgit ve Timo Teräsvirta (2006), "A Sequential Procedure For Determining The Number Of Regimes In A Threshold Autoregressive Model", *The Econometrics Journal*, Volume: 9, Issue: 3, p. 472–491.
- Şen, Zekai (2004), *Genetik Algoritmalar Ve En İyi Yöntemleri*, İstanbul, Su Vakfı.
- Tong, Howell (1978), "On A Threshold Model", CHEN, C, (Ed.), *Pattern Recognition And Signal Processing*, p. 575-586, NATO ASI Series E: Applied Sc. (29), Sijthoff ve Noordhoff, Netherlands
- Tong, Howell (1983), *Threshold Models in Non-linear Time Series Analysis*, New York, Springer.
- Tong, Howell (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Pres.
- Tong, Howell ve K. S. Lim (1980), "Threshold Autoregression, Limit Cycles and Cyclical Data", *Journal of the Royal Statistical Society. Series B (Methodological)*, Volume: 42, Issue: 3, p. 245-292.
- Tsay, Ruey S. (1989), "Testing and Modeling Threshold Autoregressive Processes", *Journal of the American Statistical Association*, Volume: 84, Issue: 405, p. 231-240.
- Wong, C. S. ve W. K. Li (1998), "A Note On The Corrected Akaike Information Criterion For Threshold Autoregressive Models", *Journal of Time Series Analysis*, Volume: 19, Issue: 1, p. 113-124.
- Wu, Berlin ve Chih-Li Chang (2002), "Using Genetic Algorithms To Parameters (d,r) Estimation For Threshold Autoregressive Models", *Computational Statistics & Data Analysis*, Volume: 38, Issue: 3, p. 315-330.
- Zivot, Eric ve Jiahui Wang (2006), *Modeling Financial Time Series With S-PLUS*, Birkhäuser.