

# EKONOMETRİK MODELLERDE EKSİK GÖLGE DEĞİŞKENLERİN GENETİK ALGORİTMALAR İLE TESPİT EDİLMESİ

Ar. Gör. Mehmet Hakan SATMAN

*İstanbul Üniversitesi İktisat Fakültesi, Ekonometri Bölümü,  
Yöneyim Anabilim Dalı*

## ÖZET

Genetik algoritmalar, biyolojideki doğal seçim ilkelerini taklit eden ve optimizasyon problemlerini çözmeye yarayan hesaplamalı yöntemlerdir. Hem kombinasyonel hem de reel sayı optimizasyon problemlerinde kullanılabilen genetik algoritmaların en büyük üstünlüğü problemin global optimumunu belirleyebilmesidir. Ekonometride, modelde yer alması gerektiği halde modele dahil edilmeyen değişken sorunu parametrelerin sapmalı tahmin edilmesine sebep olmaktadır. Modele dahil edilmeyen bir değişken varsa bu değişkenin tespit edilmesi bir optimizasyon problemi olarak düşünülebilir ve genetik algoritmalar ile çözülebilir. Bu makalede modelden dışlanmış tek bir gölge değişkenin belirlenmesi problemi modelden dışlanmış birden çok gölge değişken problemine genişletildi. Yapılan Monte Carlo simülasyonları sonucunda dışlanmış değişken sapmasının azaltılabileceği ve bu değişkenlerin yaklaşık olarak belirlenebileceği gösterildi.

**Anahtar Kelimeler:** Genetik Algoritma, Eksik Değişken Sapması

## ABSTRACT

Genetic algorithms are computational methods for solving optimization problems which mimic the natural selection principles in biology. Having ascendant abilities on searching global optima is the major advantage of genetic algorithms which can be used for both combinatorial and real valued search problems. In econometrics, omitting a relevant variable from the model affects omitted variable bias and parameters are estimated smaller or higher than what they must be. Searching for an omitted variable if exists, can be thought as an optimization problem and solved by genetic algorithms. In this article we extend the problem of searching an omitted dummy variable task to problem of searching more than one omitted dummy problem. Monte Carlo simulation results show that method can be used to reduce omitted variable bias and determine the omitted variables approximately.

**Keywords:** Genetic Algorithms, Omitted Variable Bias.

## GİRİŞ

Genetik algoritmalar, bir optimizasyon probleminin biyoloji bilimindeki doğal seçim ilkelerine göre çözülmesi için geliştirilmiş hesaplama yöntemleridir. Bu yöntemler, rasgele belirlenmiş aday çözümlerin kromozomlar (chromosome) ile ifade edilip, bu kromozomların çaprazlanması (crossing-over) ve mutasyonlanması (mutation) ile yeni döl (offspring) çözümlerin oluşturulması, ortamı belirleyen amaç fonksiyonuna (fitness function) iyi uyum sağlayan genlere (gene) sahip kromozomların yaşatılması ve uyum sağlayamayanların da topluluktan (population) atılması ilkelerine dayanmaktadır.

Genetik algoritmaların diğer optimizasyon yöntemlerinden en önemli farkı, bir topluluğa dayanması, diğer bir deyişle problemin çözüm kümesinin aranması işleminin birden fazla topluluk üyesi tarafından paralel bir şekilde yapılması, bu sayede de yerel optimum noktalara takılmadan problemin global optimum çözüm noktasının elde edilebilmesidir. Diğer bir çok optimizasyon yönteminde, optimize edilecek olan fonksiyonların türevlenebilir olması gerekliliğine karşın genetik algoritmalar ile türevlenemeyen amaç fonksiyonları da optimize edilebilir.

Ekonometride modelde yer alması gerektiği halde modele dahil edilmeyen değişkenler, tahmin edilen parametre tahmincilerinin sapmalı (biased) olmasına, gözlem sayısı arttırıldığında dahi bu sapmanın ortadan kalkmamasına (inconsistent) sebep olmaktadır. Modelde yer alması gerektiği halde modele dahil edilmeyen değişken, bir gölge değişken (dummy) olabileceği gibi reel sayı değerlerine sahip bir değişken de olabilir.

Tahmin edilen modelin kalıntıları içerisinde acaba istatistiksel olarak anlamlı bir gölge değişken çıkarılabilir mi? sorusu aslında türevlenemeyen bir amaç fonksiyonuna sahip bir optimizasyon problemi olarak düşünülebilir. Klasik algoritmalarla binlerce yıl sürebilecek bu arama işleminin genetik algoritmalar ile birkaç dakikaya indirilebileceği bilgisi literatüre

girmiştir. Bazı durumlarda ise birden fazla değişken modele dahil edilmesi gerektiği halde dahil edilmemiş olabilir. Bu çalışmanın da peşinde olduğu asıl soru, modele dahil edilmesi gerektiği halde dahil edilmeyen birden fazla gölge değişkenin yaklaşık olarak belirlenip belirlenemeyeceği olacaktır.

Bu çalışmanın I. bölümünde genetik algoritmalar tanıtılacak, II. bölümde modele eklenmeyen tek bir gölge değişkenin genetik algoritmalarla nasıl belirlenebileceği tartışılacak, III. bölümde modelde birden fazla gölge değişkenin eksik bırakıldığında genetik algoritmalar ile tespit edilebilirliği tartışılacak ve her iki yöntemin de başarısının test edilmesi için bir Monte Carlo simülasyon çalışması yapılacaktır.

## I. GENETİK ALGORİTMALAR

Doğada ortama iyi uyum sağlayan canlılar hayatta kalabilmektedir. Tüm bu canlılar hayatta kalabilmek için doğanın belirlemiş olduğu birden çok amaç fonksiyonunu optimize etmeye çalışırlar. Genetik algoritmalar, Darwin 'in Türlerin Kökeni (The Origin of Species) adlı eserinde en iyi uyum sağlayanın yaşaması ifadesiyle ünlenen genetik süreçlere, özellikle de doğal seçim ilkesine dayanmaktadır (Dawid, 1999: 41). Holland'ın 1975'te yazdığı *Adaptation in Natural and Artificial Systems* adlı eseriyle genetik algoritmaların temelleri atılmış, rasgele gibi görünen canlılardaki adaptasyon sürecinin olasılıklı süreçleri matematiksel olarak gösterilmiştir (Holland, 1975). Pozitif ve uygulamalı bilimlerde, özellikle de mühendislik dallarında sıkça kullanılmaya başlayan genetik algoritmaların ekonometri biliminde de bazı reel sayı ve kombinasyonel problemlerin çözümünde de kullanılmaya başlanması tesadüf değildir. Genetik algoritmalar, kümeleme analizinde (Hruschka ve Ebecken, 2003), ARIMA modellerinin tespitinde (Ong vd., 2005), model seçiminde (Balcombe, 2005, Hasmenia, Niaki, 2006), modelden dışlanmış değişkenin sebep

olduğu tahminci sapmasının araştırılmasında (Sessions, Stevans, 2006), doğrusal olmayan ekonometrik modellerin parametre tahminlerinde (Akyol, 2006), İMKB endeks verileri için uygun ARMA modelinin aranmasında (Satman, 2007) kullanılarak ekonometri biliminde de kullanılmaya başlanmıştır.

Genetik algoritmalar bir veya birden çok amaç fonksiyonunun optimize edilmesi için kullanılan yöntemler bütünüdür. Bu yöntemlere göre öncelikle rasgele belirlenmiş p adet aday çözüm genler ve kromozomlarla ifade edilip bir topluluk oluşturulur. Gelenek olarak her bir gen 1 ve 0 değerleri ile ifade edilebileceği gibi, reel sayılarla veya sembollerle de ifade edilebilir (Mitchell, 1999: 117).

Tablo 1.1 'de 3 aday çözüm ve bu aday çözümlerin kromozom gösterimi verilmiştir.

Tablo 1.1. 3 Aday Çözüm ve Kromozom Gösterimleri

Aday Çözüm	Kromozom Değeri
12	00001100
1	00000001
7	00000111

Tablo 1.1'deki aday çözümlerin tümünün de kromozom uzunluğu 8 olarak belirlenmiştir. Bu uzunluk aday çözümlerin değer aralığına göre küçültülüp büyütülebilir. 00001100 olan kromozom değeri ise 12 sayısının ikili sayı sisteminde gösterimidir.

İki kromozom arasında uygulanan çaprazlama işlemi genel olarak iki kromozomun da aynı noktadan kesilip, birinci kromozomun ilk parçası ile ikinci kromozomun ikinci parçası birleştirilerek yeni bir kromozom oluşturulması şeklindedir. Yeni oluşan kromozoma döl kromozom adı verilir. Benzer şekilde ikinci kromozomun ilk parçası ile birinci kromozomun ikinci parçası birleştirilerek başka bir döl oluşturulabilir. Tablo 1.1'deki 1. ve 2. aday

çözümlerin çaprazlanıp döl kromozomların elde edilme süreci Tablo 1.2 'de gösterilmiştir.

Tablo 1.2. Çaprazlama İşlemi ve Döl Kromozomların Oluşturulması

Aday Çözüm	Kromozom	Çaprazlama Noktası	Döl	Yeni Aday Çözüm
12	00001100	00001 100	00001001	9
1	00000001	00000 001	00000100	4

Çaprazlama işlemi, kromozomların tek bir noktadan kesilerek yapılabileceği gibi; kesme işleminin birden çok noktaya uygulanmasıyla da gerçekleştirilebilir.

Çaprazlama sonucu oluşan döl kromozomların herhangi bir geninin değerinin ters çevrilmesiyle, yani 1 olan değer 0, 0 olan değer 1 yapılmasıyla mutasyon işlemi gerçekleştirilmiş olur. Tablo 1.3'te Tablo 1.2'de oluşturulmuş olan döl kromozomların mutasyon işlemine tabi tutulma süreci gösterilmiştir.

Tablo 1.3. Mutasyon İşlemi

Aday Çözüm	Kromozom	Mutasyon Noktası	Yeni Kromozom	Yeni Aday Çözüm
9	00001001	00001001	01001001	73
4	00000100	00000100	00010100	20

Görüldüğü üzere çaprazlama işlemi iki aday çözüm civarında yeni çözümler oluşmasına sebep olmaktadır. Mutasyon işlemi ise değeri değiştirilen genin konumuna göre döl kromozomları, ebeveyn kromozomların yakınma veya uzağa düşürebilmektedir. Tablo 1.3'te 9 değerini alan aday çözümün soldan ikinci geninin değeri değiştirilmiştir. İkili sayı sisteminde düşünüldüğünde, yapılan bu ufak değişiklik, bu kromozomun tam sayı karşılığı olan 9 değerini 73'e çıkarmıştır. Tam tersine bu kromozomun

eğer sağdan birinci geninin değeri değiştirilseydi 9 olan tamsayı değeri 8 'e düşecekti. Bu bağlamda çaprazlama ve mutasyon işlemleri optimizasyon problemlerinde farklı görevleri üstlenmektedir. Çaprazlama işlemi iki aday çözüme ait kromozomların genlerini bir araya getirerek yerel optimum noktaları aramakta, mutasyon işlemi ise çözüm uzayında farklı noktalara sıçrayıp global optimum noktaları aramaktadır.

Genetik algoritmalarda çaprazlama ve mutasyon işlemleri rasgele seçilmiş kromozomlara uygulanmamaktadır. Amaç fonksiyonunu en iyi sağlayan, bir diğer deyişle, ortama en iyi uyum sağlayan kromozomlar çaprazlanmakta, iyi genlerin bir sonraki topluluğu oluşturma sürecinde rol oynaması olasılıkları arttırılmaktadır. Bu da genetik algoritmalarda seçim (selection) işlemidir.

Genetik algoritma operatörleri yukarıdaki gibi tanımlanmışken, genel bir genetik algoritma Tablo 1.4'te olduğu gibi tanımlanabilir.

**Tablo 1.4.** Genel bir Genetik Algoritma

Adım 1: p adet kromozom içeren bir topluluk oluştur.
Adım 2: Kromozomlara ait değerlerin amaç fonksiyonuna ne kadar iyi uyum sağladığını hesapla
Adım 3: Adım 2'de hesaplanan uyum değerlerine göre iki kromozom seç ve çaprazla
Adım 4: Adım 3'de çaprazlanan döş kromozomlara mutasyon operatörünü uygula
Adım 5: Döl kromozom sayısını p olana kadar Adım 3 ve Adım 4'ü tekrarla
Adım 6: Döl kromozomların oluşturduğu topluluğu Adım 1'de oluşturulan topluluğun üzerine kopyala
Adım 7: Yakınsama sağlandıysa dur. Sağlanmadıysa Adım 2'ye sıçra

Tablo 1.4.'te gösterilen genetik algoritma, genetik algoritmalar için genel bir form oluşturmakta olup, problemin doğasına göre çeşitli özelleştirme yollarına gidilebilmektedir. Örneğin, Adım 3 ve Adım 4'te belirtilen çaprazlama ve mutasyon işlemleri her kromozom için yapılmayabilir veya bu işlemlerin yapılma sıklığı

belirli oranlara tabi tutulabilir. Bu oranlara çaprazlama ve mutasyon olasılıkları adı verilir. Adım 3'te gösterilen uyum değerlerine göre çaprazlama işlemi ise seçim operatörünü göstermektedir ki, bu işlem de problemin doğasına göre çeşitlendirilebilir. Örneğin rulet makinesi seçiminde (Roulette Wheel selection) bütün uyum değerleri bir rulet makinesinin hanelerine yerleştirilir ve bu hanelere sanal olarak bir top düşürülür. Topun düştüğü haneye ait kromozom çaprazlama için seçilir. Rulet makinesinin bu hanelerinin genişliğinin aday çözüme ait uyum değerine göre değiştiğini belirtelim. Bir diğer sık kullanılan seçim yöntemi ise turnuva seçilimidir (Tournament selection). Bu seçim yöntemine göre rasgele iki kromozom seçilir, uyum değeri yüksek olan kaydedilir. Rasgele iki kromozom seçme işlemi bir kez daha yapılır ve yüksek uyum değeri olan kaydedilir. Kaydedilen bu iki kromozom turnuvada başarılı olmuştur ve çaprazlanmaya hak kazanır. Genellikle çaprazlanma olasılıkları %80 gibi büyük değerler olarak seçilirken mutasyon olasılıkları %1 gibi ufak tutulur.

## II. EKSİK DEĞİŞKEN PROBLEMİ

$k > 3$  iken gerçek model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.1)$$

ve tahmin edilen model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k \quad (2.2)$$

olarak yazılabilir. (2.1) 'deki  $\varepsilon$ , elemanları  $\varepsilon_i$ ,  $i = 1, 2, \dots$  olan, sıfır ortalamalı, bağımsız hata terimidir. (2.2)'de gösterilen tahmin denklemindeki parametre tahminçileri, regresyon denkleminin en küçük kareler yöntemi ile hesaplandığı varsayımı ile, bu haliyle (2.1)'de gösterilen modelin parametrelerinin sapmasız ve en küçük varyanslı birer tahminçisidir. Eğer (2.1) 'deki  $X_k$ , modelden ihraç edilirse yeni tahmin denklemini

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_{k-1} X_{k-1} \quad (2.3)$$

olacaktır.  $\beta_k = 0$  olmadıkça  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{k-1}$ 'um gerçek değerlerinden büyük mü küçük mü tahmin edildiği bilinemez (Güriş ve Çağlayan, 2005:616). Bir diğer deyişle (2.3) 'deki parametre tahminlerini sapmalı tahmin edilirler.

İhraç edilen değişken eğer bir gölge değişkense bu problem genetik algoritmalar ile çözülebilir. Sessions ve Stevans (2006) eğer modelde eksik bir gölge değişken varsa bunun belirlenebileceğini, reel sayı değerleri alan bir değişkenin ise belirlenmesi çabasında bu değişkenin ölçme hataları içerdiğini göstermiştir.

$X_k$  değişkeni bir gölge değişkense

$$X_{ki} \in \{0,1\} \quad , \quad i = 1,2,\dots,n \quad (2.4)$$

şeklinde gösterilebilir. Burada n, gözlem sayısını göstermektedir. Gözlem sayısının 20 olduğu düşünülürse modele eklenmeyen  $X_k$  değişkeni örneğin

$$11100001111000101011 \quad (2.5)$$

gibi olabilir. Gölge değişken özellikleri de göz önüne alınırsa

$$00011110000111010100 \quad (2.6)$$

değişkeni için hesaplanan parametre tahmincisinin işareti (2.5)'de gösterilen değişken için hesaplanan parametre tahmincisinin işaretinin tersi olacak fakat istatistiksel değerleri aynı kalacaktır. Öte

yandan eğer (2.5)'deki değişken, modelden ihraç edilen gerçek değişken ise bu değişken haricindeki herhangi başka bir gölge değişken daha iyi bir tahmin denklemini vermeyecektir. Tahmin denkleminin de kalitesi, eğer diğer bağımsız değişkenler ve bağımlı değişken sabit tutulursa, belirlilik katsayısı  $R^2$  ile ölçülebilir. Öyleyse tahmin denkleminin  $R^2$  değerini en büyüleyecek, istatistiksel olarak anlamlı bir gölge değişken, modelde olması gerektiği halde modele dahil olmayan gölge değişkendir.

Gözlem sayısı 20 iken yazılabilecek tüm mümkün gölge değişkenlerin sayısı  $2^{20} = 1,048,576$  olacaktır. Tüm mümkün gölge değişkenlerin modele eklenip, tahmin denkleminin hesaplanıp, içlerinden en uygun modelin belirlenmesi süreci bu haliyle çok uzun sürecektir. Her bir gölge değişken için hesaplama süresinin 3 saniye olduğu durumda, gereken toplam süre 36 gündür. Gözlem sayısının 30 olduğu durumda tüm hesaplamalar 102 yıl sürecektir, 50 olduğu durumda ise işlem süresi insanoğlunun ortaya çıkışından günümüze kadar olan sürenin üstüne çıkacaktır. 250 elemanlı, oluşturulabilecek tüm mümkün gölge değişken sayısı evrendeki atom sayısından fazla olmaktadır. Genetik algoritmalar ile bu işlemlerin süreleri makul düzeye indirilebilir.

Genetik algoritma oluşturulmadan önce kromozom uzunluğuna karar vermek gerekir. Buradaki problemde kromozom uzunluğu gözlem sayısına eşittir. Amaç fonksiyonu ise

$$R^2 \rightarrow En \text{ büyük} \quad (2.7)$$

olarak tanımlanabilir (Sessions ve Stevans, 2006). Rasgele seçilecek olan p adet n uzunluklu kromozom, aday gölge değişkenleri temsil edecek, bu gölge değişkenlerden  $R^2$  değerini en büyüleyenler bir sonraki topluluktaki aday gölge değişkenleri çaprazlama ve mutasyonla oluşturma hakkına sahip olacaklardır. Bu, doğal seçim

ilkeleri diliyle, ortamı belirleyen  $R^2$  değerine en iyi uyum sağlayanlarının hayatlarını ve soylarını idame ettirme olasılıklarının diğerlerine göre yüksek olmasıdır.

### III. BİRDEN FAZLA EKSİK DEĞİŞKEN PROBLEMİ

Gerçek model (2.1) eşitliğinde gösterildiği gibi iken eğer iki adet modelde yer alması gereken değişken modele dahil edilmemişse tahmin modeli

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_{k-2} \quad (3.1)$$

olarak yazılabilir. Bölüm 2'deki tahmin modelinde olduğu gibi parametre tahminçileri olan  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{k-2}$  tahminçileri sapmalı tahmin edilebilirler. Sapmanın ortadan kaldırılabilmesi için modelde yer alması gereken değişkenlerin modele dahil edilmeleri gerekir. Bölüm 2'de bahsedildiği üzere gözden kaçırılan veya iktisat teorisinin değinmediği bir gölge değişken sapmaya sebep olabileceği gibi birden fazla değişken de modele dahil edilmemiş olabilir. Problem bu haliyle büyümekte, daha karmaşık bir problem olan, birden fazla eksik değişkenin araştırılması problemine dönüşmektedir.

Modele dahil edilmeyen  $m$  adet gölge değişken olduğunda aranması gereken 1 ve 0'ların sayısı  $n \times m$  olmaktadır. Her bir bilinmeyen işlem süresini üstel olarak arttırması sebebiyle tüm mümkün durumların taranmasını gerektiren kaba kuvvet (Brute Force) yöntemler daha da imkansız hale gelmektedir.

Bölüm 2'de değinilen genetik algoritmanın genelleştirilmiş formu, bu tür bir probleme çözüm olarak sunulabilir. Amaç fonksiyonunun yine tahmin edilen regresyonun  $R^2$  değerinin en büyüklemesi olarak belirlenmesi ve bu amaç fonksiyonunu en iyi sağlayan gölge değişkenlerin oluşturduğu kromozomların yaşatılması

yöntemiyle gerçek gölge değişkenler belirlenmeye çalışılabilir. Birden fazla değişkenin arandığı durumda özel bazı karmaşık durumlar meydana geleceğinden, değişkenler, tek değişkenin aranması problemine göre daha çok hatayla belirlenebilirler.

Örneğin model

$$Y = 5 + 5D_1 + 5D_2 \quad (3.2)$$

şeklinde deterministik bir model,  $D_1$  ve  $D_2$  gölge değişkenler olsun.  $D_1 = 1$  ve  $D_2 = 0$  iken  $Y = 10$  olmaktadır. Benzer şekilde  $D_1 = 0$  ve  $D_2 = 1$  iken yine  $Y = 10$  olur. Dolayısıyla  $Y = 10$  olduğu bilgisi tek başına hangi gölge değişkenin 1 hangisinin ise 0 değeri alması gerektiği bilgisini vermez. Bu sorun parametre değerlerinin birbirine çok yakın olduğu durumda meydana gelebilmektedir. Bu da modele dahil edilmeyen birden fazla gölge değişkenin belirlenmesinde bir sorun yaratabilmektedir.

Diğer bir dezavantaj ise gölge değişkenlerin (2.5) ve (2.6) 'daki gibi ters belirlenme durumudur. Ters belirlenme durumunda ilgili değişkene dair parametre tahminçilerinin istatistiksel özellikleri değişmez, tahmin regresyonunun  $R^2$  değeri aynı kalır fakat parametre ters işaret ile tahmin edilir. Bu dezavantajdan kurtulmak için genetik algoritma ile belirlenen gölge değişkenlerin tersleri alınarak tahmin modelleri yeniden hesaplanıp uygun model seçilebilir.

Yöntemin güvenilirliğinin test edilmesi için bir simülasyon çalışması yapılmıştır. Öncelikle modelde yer alması gereken tek bir gölge değişken çıkartılıp genetik algoritmanın bu değişkeni belirlemesi test edilmiş sonra da bu çalışmanın asıl konusu olan birden fazla gölge değişkenin modelden çıkartılıp genetik algoritma tarafından belirlenebilmesi test edilmiştir.

İlk simülasyon modeli, gözlem sayısı  $n = 100$  iken ,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 D + \varepsilon \quad (3.3)$$

şeklinde oluşturulmuştur. Burada parametreler  $\beta_i = 5$  ,  $i = 0,1,2,3,4$  olarak verilmiştir. Her bir  $X_i$  ise ortalaması 7 ve varyansı 16 olan bir normal dağılımdan bağımsız olarak çekilmiştir. Veri üretme süreci içerisinde bağımsız değişkenlerin çekildiği dağılımın özellikleri, simülasyon sonucunu değiştirmeyecektir. (Wisnowski, 1999:69).  $D$  elemanları

$D_i \in \{0,1\}$  olan gölge değişken ve  $\varepsilon$ , 0 ortalama ve 1 varyansla normal dağılan hata terimidir. Tüm parametre ve değişkenler böyle tanımlanmışken  $Y$  değişkeni ise (3.3)'te olduğu gibi türetilmiştir. Model bu haliyle bir çok önemli ekonometrik model varsayımını sağlar .

Simülasyon için yazılan program ise genetik algoritma kullanarak sadece  $Y$  ve  $X$  değerleri veriyken gölge değişkeni, tahmin regresyonunun  $R^2$  değerini en büyükleyerek tahmin etmeye çalışmıştır. Simülasyon 100 kere tekrar edilmiş ve sonuçlar Tablo 3.1'de gösterilmiştir. Simülasyonun daha fazla tekrarlanması sonuçları belirgin şekilde değiştirmemektedir.

Tablo 3.1. Tek Eksik Değişken Simülasyonu için Tahmincilerin OHK Değerleri

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$ (Gölge)
Tüm ve Gerçek Değişkenlerle Tahmin Edilen Model	0,246984	0,00147	0,001329	0,001262	0,090073
Gölge Değişken İhraç Edilmişken Tahmin Edilen Model	7,57294	0,010285	0,010833	0,009682	YOK
Genetik Algoritma ile Belirlenen Gölge Değişkenle Tahmin Edilen Model	0,256063	0,00163	0,001356	0,001396	0,089844

Tablo 3.1'deki değerler simülasyon sonuçlarından hesaplanan ortalama hata kareleri (OHK – Mean Square Error) göstermektedir. OHK, parametre tahmincilerinin sapma ve varyanslarının bir fonksiyonudur ve

$$OHK(\hat{\beta}) = \text{Varyans} + \text{Sapma}^2 \quad (3.4)$$

olarak hesaplanır. Genetik algoritmanın modelden ihraç edilen gölge değişkeni ne kadar doğru belirlediği ölçüsü için OHK kullanılmıştır. Tabloda görüldüğü gibi ihraç edilen gölge değişken, OHK değerlerinin yükselmesine sebep olmuştur. Genetik algoritma ile belirlenen gölge

değişkenin tekrar modele eklenmesiyle OHK değerleri küçülmüş, gerçek gölge değişkenin modelde yer aldığı duruma yaklaşmıştır

İkinci simülasyon modeli, birinci simülasyon modeline benzer olarak iki gölge değişkenle yapılmıştır.

Simülasyon modeli

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 D_1 + \beta_5 D_2 + \varepsilon \quad (3.5)$$

olarak belirlenmiştir. Tüm değişkenler birinci simülasyon modelinde olduğu gibi belirlenmiş fakat modele birinci simülasyon modelinden farklı olarak bir gölge değişken daha eklenmiştir. Simülasyon sonuçları Tablo 3.2.'deki gibidir.

**Tablo 3.2.** İki Eksik Değişken Simülasyonu için Tahmincilerin OHK Değerleri

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$ (Gölge)	$\hat{\beta}_5$ (Gölge)
Tüm ve Gerçek Değişkenlerle Tahmin Edilen Model	0,273559	0,00131208	0,00125847	0,0015686	0,0737118	0,0838205
Gölge Değişkenler İhraç Edilmişken Tahmin Edilen Model	24,9454	0,0194383	0,0178876	0,0210105	YOK	YOK
Genetik Algoritma ile Belirlenen Gölge Değişkenle Tahmin Edilen Model	1,14918	0,003515554	0,003484841	0,00325021	0,607537	0,514738

Simülasyon sonunda OHK değerleri yanında tahmincilerin aldığı değerlerin de görülmesinde fayda olabilir. Tablo 3.3. 'te simülasyon sonunda elde edilen parametre tahmincilerinin 100 iterasyon için aldıkları ortalama değerler gösterilmektedir. Veri üretme sürecinde tüm parametreler eşit ve 5 olarak verildiğinden her bir hanedeki değer 5'e yakınlığı yöntemin başarısının değerlendirilmesi

için sübjektif bir kanıt olarak görülebilir. Gölge değişkenlerin modelden çıkartılmasıyla parametre tahmincileri, özellikle de  $\beta_0$  'm tahmincisi, 5'den uzaklaşmıştır. Buradaki etki, modele dahil edilmeyen gölge değişkenlerin diğer bağımsız değişkenlerle arasında bir ilişki bulunmamasından dolayı eksik değişken sapmasının sabit parametreye yansması sebebiyle olmuştur.

**Tablo 3.3.** İki Eksik Değişken Simülasyonu için Tahmincilerin Ortalama Değerleri

	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$ (Gölge)	$\hat{\beta}_5$ (Gölge)
Tüm ve Gerçek Değişkenlerle Tahmin Edilen Model	4,95653	5,00196	4,99986	5,00381	4,98859	5,00882
Gölge Değişkenler İhraç Edilmişken Tahmin Edilen Model	9,64347	5,01369	5,00552	5,02011	YOK	YOK
Genetik Algoritma ile Belirlenen Gölge Değişkenle Tahmin Edilen Model	5,02338	5,00063	5,00094	5,00931	4,88097	4,85717

Tablo 3.2 ve Tablo 3.3'te görüldüğü gibi, modelde yer alması gereken iki gölge değişken modelden çıkartıldığında parametre tahmincilerinin OHK değerleri yükselmiş, parametre tahmincileri gerçek parametre değerleri olan 5'ten uzaklaşmış, genetik algoritma ile bulunan gölge değişkenler modele eklendiğinde OHK değerleri küçülmüş ve parametre tahmincileri gerçek değerlerine yaklaşmıştır. Gölge değişkenlerin, bir eksik gölge değişken durumunda olduğundan daha doğru

belirlenememesi dezavantajı dışında, yöntem, daha az sapmalı parametre tahmincileri elde etme konusunda başarılıdır.

## SONUÇ

Ekonometrik bir modelin başarısı büyük oranda modelde yer alması gereken değişkenlerin modele dahil edilmesi ve modelde yer almaması gereken değişkenlerin de modelden ihraç edilmesine bağlıdır. Bu iki hata içinden daha kötü



sonuçlar doğuran modelde olması gerektiği halde modele dahil edilmeyen değişken sorunu, tahmin edilen parametre tahmincilerinin sapma olmasına neden olmaktadır. Ayrıca bu sapma, örneklem büyütülerek ortadan kalkmayabilir. Bu bağlamda tahminciler hem sapmalı hem de tutarsız olurlar. Modele dahil edilmeyen değişkenin bir gölge değişken olup olmadığı, eğer gölge değişkense hangi değerleri aldığı, genetik algoritma kullanılarak, yaklaşık olarak kestirilebilir. Bu çalışmada, literatüre ek olarak, birden fazla gölge değişkenin modelden ihraç edildiği durumda bu değişkenlerin kestirilip kestirilemeyeceği incelenmiştir. Sonuç olarak görülmüştür ki, her iki durumda da ihraç edilen gölge değişkenler yaklaşık olarak kestirilebilir fakat ihraç edilen gölge değişkenlerin sayısı arttıkça, belirsizlik artacağından, kestirim gücü azalacaktır. Genellikle kaç adet değişkenin modelden ihraç edildiği bilinmediğinden öncelikle tek bir dışlanan değişken vardır varsayımı yapıp birinci yöntem denenip sonra birden fazla değişken arama yöntemine gidilebilir. AIC, BIC veya düzeltilmiş  $R^2$  değerlerine bakarak genetik algoritmanın belirlediği kaç değişkenin modelde yer alacağına karar verilebilir. Bu çalışmada önerilen yöntemler, ekonometrik model kurma ve değişken seçme işlemlerinin tamamen mekanik hale getirilmesi için değildir. Modelde varoluş sebebi belirlenemeyen bir değişkenin salt  $R^2$ 'yi en büyüklüyor diye modele dahil edilmesi pek bilimsel sayılamaz. Fakat genetik algoritma ile belirlenen anlamlı bir değişken, modelde eksik değişken olduğu şüphesini doğurarak araştırmacıyı daha doğru modeller kurmaya yönlendirebilir.

#### KAYNAKÇA

AKYOL, A.P. (2006), **Doğrusal Olmayan Ekonometrik Modellerin Genetik Algoritma Yaklaşımı ile Parametre Tahmini**, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi

- BALCOMBE, Kelvin G. (2005), "Model Selection Using Information Criteria and Genetic Algorithms", **Computational Economics**, 25, ss.207-228.
- DAWID, Herbert (1999), **Adaptive Learning by Genetic Algorithms – Analytical Results and Applications to Economic Models**, Springer, New York.
- GÜRİŞ Selahattin, ÇAĞLAYAN Ebru (2005), **Ekonometri – Temel Kavramlar**, Der Yayınları
- HASHEMINIA, Hamed, NIAKI, Seyed T.A. (2006), "A genetic algorithm approach to find the best regression/econometric model among the candidates", **Applied Mathematics and Computation**, 183, ss.337-349.
- HRUSCHKA, E.R., EBECHEN, N.F.F. (2003), "A genetic algorithm for cluster analysis", **Intelligent Data Analysis**, 7, ss. 15-25
- HOLLAND, John H. (1975), **Adaptation in Natural and Artificial Systems**, University of Michigan Press. (Second Edition: MIT Press, 1992).
- MITCHELL, Melanie (1999), **An Introduction to Genetic Algorithms**, MIT Press, 1999, England.
- ONG, C.S., HUANG, J.J., TZENG, G.H. (2005), "Model identification of ARIMA family using genetic algorithms", **Applied Mathematics and Computation**, 164, ss.885-912.
- SATMAN, M. Hakan (2007), "Durağan Zaman Serilerinde Uygun ARMA Modelinin Genetik Algoritmalar ile Bulunması ve İMKB Uygulaması", **İktisat Fakültesi Mecmuası**, Basım Aşamasında
- SESSIONS, David N., STEVANS, Lonnie K. (2006) , "Investigating omitted variable bias in regression parameter estimation: A genetic algorithm approach", **Computational Statistics & Data Analysis**, 50, ss. 2835-2854
- WISNOWSKI, James Walter, **Multiple Outliers in Linear Regression: Advances in Detection Methods, Robust Estimation, and Variable Selection**, Arizona State University, Doktora Tezi, 1999