

# MODELLEMEDE YENİ BİR YAKLAŞIM: OTOMATİK MODEL SEÇİMİ

Yrd. Doç. Dr. Melda AKIN\*

## Özet

Hem bilgisayarların bellek ve işlem kapasitesinin artması hem de etkin yöntemlerin geliştirilmesi sayesinde otomatik model seçimi yakın zamanda bir seçenek haline gelmiştir. Literatürde önerilen model seçimiyle ilgili çoğu tartışmalı, çok sayıda yaklaşım arasında PcGets, RETINA ve Genetik Yordamlar Kullanarak Model Seçme (GYKMS) öne çıkmaktadır. Bu yöntemler model uzayında çeşitli başlangıç modellerinden hareketle bir arama yaparak en doğru modeli seçmeye çalışmaktadır. Bu arama sırasında çeşitli istatistiksel testlerin yanında değişik aşamalarda modeller arasında tercih yaparken başta Akaike Enformasyon Ölçütü (AIC) olmak üzere değişik enformasyon ölçütleri de kullanılmaktadırlar. Yordamlar bir veri kümesini açıklamaya çalışan, farklı ama bağdaşmaz olmayan yaklaşımlar olarak değerlendirilebilir. Öncül bilgiye, modelin kullanma amacına ve araştırmacının deneyimine göre uygulama bazında biri diğerlerinden daha yararlı olabilir.

**Anahtar Sözcükler:** Model Seçimi, Genelden Özele Modelleme, Enformasyon Ölçütleri, Çapraz-geçerleme, Genetik Yordamlar, Kestirim, Esnek Modelleme, Yapay Sinir Ağları

## Abstract

Recently automatic model selection has become an option both as a result of the increase in the processing power and the memory capacity of the computers and the development of effective methods. Among the many, often debated, model selection methods proposed in the literature PcGets, RETINA and model selection using genetic algorithms are the more prominent ones. These methods, starting with an initial model try to select the best model doing a search in the model space. During various stages of search, in order to choose among models, in addition to the various statistical tests they use information criteria including Akaike Information Criterion (AIC). The procedures can be considered as different but not irreconcilable approaches that try to explain a data set.

---

\* İstanbul Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü

*Depending on the prior information, the aim of the model and the experience of the researcher, one of them may be better than the others on the basis of application.*

*Keywords: Model Selection, General-to-Specific Modeling, Information Criteria, Cross-validation, Genetic Algorithms, Forecasting, Flexible Modeling, Neural Networks*

## 1. GİRİŞ

Model seçimi ve seçilen modelin değerlendirilmesi ekonometrik modellemenin temelini oluşturur. Seçilen modellerin değerlendirilmesi için aralarında Durbin-Watson testinin de olduğu bir çok istatistiksel yöntemler geliştirilmiştir. Model seçiminin nasıl yapılacağı ise daha tartışmalı bir konudur. Otomatik model seçimi, hem bilgisayarların bellek ve işlem kapasitesinin artması hem de etkin yöntemlerin geliştirilmesi sayesinde yakın zamanda bir seçenek haline gelmiştir. Literatürde önerilen model seçimiyle ilgili çok sayıda yaklaşım vardır. Bunların arasında üç tanesi PcGets (Krolzig ve Hendry, 2001), RETINA (Perez-Amaral, Gallo ve White, 2003) ve Genetik Yordamlar Kullanarak Model Seçme (GYKMS) (Balcombe, 2005) öne çıkmaktadır.

Bu yöntemler model uzayında çeşitli başlangıç modellerinden hareketle bir arama yaparak en doğru modeli seçmeye çalışmaktadır. Bu arama sırasında çeşitli istatistiksel testlerin yanında, değişik aşamalarda modeller arasında tercih yaparken, başta Akaike Enformasyon Ölçütü (AIC) (Akaike, 1973) olmak üzere değişik enformasyon ölçütleri de kullanılmaktadır. PcGets'in arkasında genelden özele giderek model seçme ile ilgili bir kuram vardır. Başlangıç model kullanıcı tarafında seçilir ve çeşitli testlerle değerlendirme yapılan bir arama süreci sonucunda son model elde edilir. RETINA ise özelden genele gitmeyi amaçlayan ve yapay sinir ağlarında kullanılan tekniklerden (Akin, 2001) (Akin, 2002) esinlenmiş bir yaklaşımdır. Doğrusal olmayan dönüşümleri otomatik olarak üretir ve ayırık altörneklemeler kullanarak örneklem dışı testlerle seçim yapar. GYKMS ise model uzayında model seçmede arama tekniği olarak genetik yordamları kullanmaktadır.

Genelde otomatik model seçimine, özelde ise PcGets yaklaşımına karşı patika bağımlılığı, tekrarlanan testler, seçim etkileri, test öncesi sorunlar, veri

madenciliği, teori olmadan ölçme ve belirleme eksikliği gibi itirazlar yapılmıştır. Hendry (2000) bütün bu itirazları çürütürerek otomatik model seçiminin teorik açıdan haklı çıkarılabileceğini göstermiştir.

Bu derleme makalede ekonometride ve diğer sosyal bilimlerde modellemede kullanılacak bu yeni yaklaşımların özellikleri verilip karşılaştırmaları yapılarak tanıtılması amaçlanmaktadır.

## 2. MODEL SEÇİMİNDE KULLANILAN ENFORMASYON ÖLÇÜTLERİ

Literatürde model seçiminde en çok kullanılan enformasyon ölçütleri arasında Akaike (AIC) (Akaike, 1973), Schwarz-Bayes (BIC) (Schwarz, 1978), Hannan-Quinn (HQ) (Hannan ve Quinn, 1979) yer almaktadır. Her üç ölçüt de iki terimden oluşmakta olup ölçütler arasında sadece ikinci terimler arasında farklılık bulunmaktadır.

$$AIC(k) = \ln(\sigma^2) + \frac{2k}{T} \quad (1)$$

$$BIC(k) = \ln(\sigma^2) + \frac{2\ln(\ln(T))k}{T} \quad (2)$$

$$HQ(k) = \ln(\sigma^2) + \frac{\ln(T)k}{T} \quad (3)$$

Burada  $k$  parametre sayısını,  $T$  ise örneklem büyüklüğünü göstermektedir. İlk terimdeki  $\sigma^2$  hata varyansının en büyük olasılık kestirimini göstermektedir.

Genetik yordam yaklaşımında ise bunlara ek olarak aşağıda verilen posterior enformasyon kestirim ölçütü (PICF) (Philips, 1995) de kullanılmaktadır.

$$\ln(pf^k) = \sum_{t=K+1}^T \frac{\ln f_{t,k}}{2} + \sum_{t=K+1}^T \frac{v_{t,k}^2}{2f_{t,k}} \quad (4)$$

Burada  $K < T$  olmak üzere  $k$ ' dan büyük olacak şekilde seçilir.  $v_{t,k}$  bir adım sonraki kestirim hatası ve  $f_{t,k}$  ise kestirim varyansdır.

### 3. PCGETS YORDAMI

Bu yordam Hendry tarafından önerilen genelden özele modelleme yönteminden hareketle üretilmiştir. Bilgisayar kullanılarak otomatik model seçecek hale getirilmesi Hoover ve Perez (1999) tarafından önerilmiştir. PcGets programı ile ilgili ilk bilgi Hendry ve Krolzig (2004)'de verilmiştir. Bu programın bu makalede incelenen ikinci yöntem olan RETINA ile örnek bazı veriler üstünde karşılaştırılması ise Perez-Amaral, Gallo ve White (2005) ve Castle (2005) de verilmiştir.

PcGets, bir genel kısıtsız modelin (GKM) basit ve içerilmemiş bir gösteriminin otomatik olarak seçilmesini hedefleyen bir yordamdır. GKM potansiyel olarak ilgili bütün faktörleri içeren çok genel bir modeldir. Bir bağımsız değişken  $y_t$ , bir örneklem,  $t=1, \dots, T$  ve  $n$  adet aday regresör değişken  $z'_t = (z_{1,t}, \dots, z_{n,t})$  için GKM aşağıdaki şekilde verilebilir:

$$y_t = \sum_{i=1}^n \gamma_i z_{i,t} + v_t, \quad v_t \sim IN[0, \sigma_v^2] \quad (5)$$

GKM'nin kestirimlerinin sapmasız, ancak bazı  $\gamma$ 'ler sıfır olduğu zaman verimsiz olduğu gösterilebilir (Campos, Hendry ve Krolzig, 2003).

Veri Üreten Süreç (VÜS) denkleminde sadece  $m \leq n$  değişken vardır:

$$y_t = \sum_{j=1}^m \beta_j z_{j,t} + \varepsilon_t, \quad \varepsilon_t \sim IN[0, \sigma_\varepsilon^2] \quad (6)$$

Dolayısıyla Denklem (1)'deki GKM, VÜS'ü de içerir.

Genelden özele giden yordam, bu başlangıç GKM'den yola çıkarak özel bir model seçer. Doğrusal bir çerçevede bu yaklaşımın özellikleri Krolzig ve Hendry (2001) tarafından incelenmiştir.

PcGets Yordamının dört aşaması vardır. Bunların arasında GKM'nin kestirilmesi ve test edilmesi, arama öncesi süreci, çok patikalı arama yordamı ve arama sonrası değerlendirme yer almaktadır. Yordam Şekil 1'de verilmiştir. Başlangıçta GKM teoriye ve elde bulunan bilgilere göre yerel veri üreten süreci (YVÜS) kapsayabilecek kadar genel bir şekilde seçilir. Bu varsayım her zaman geçerli olmayabilir. GKM'de olması gereken gerekli bazı değişken ihmal edilmiş olabilir. Ancak YVÜS'ün temel bir bileşeni modelde yer almazsa bu kendisini yapılacak testlerde hatalı belirleme olarak gösterecektir. GKM üstünde eşleşikliği göstermek için bir dizi hatalı-belirleme testi yapılır. Burada eşleşiklik modelin bilinmeyen YVÜS'e her açıdan uyması anlamına gelmektedir. Eğer eşleşiklik hatası varsa arama süreci başlamaz. Bu durumda araştırmacının GKM'yi yeniden belirlemesi gerekir. Bu nedenle potansiyel olarak ilgili bütün değişkenleri içerecek kadar genel bir modelle başlamak esastır.

Arama öncesi testlerle önemsiz değişkenler elenir. Bu aşamadan sonra çok patikalı arama stratejisi makul olan her elemeyi yaparak her patikada arama yapar. Bu aramaya  $t$  ve  $F$  testleri yapılarak daha fazla indirgeme yapılamayınca kadar devam edilir. Her indirmeden sonra eşleşiklik denetimi yapılır. Sonuçta elde edilen modellere *uç modeller* adı verilir. Tek eşleşik ve baskın bir model seçilene kadar uç modellerin birleşimi alınarak tekrar tekrar GKM oluşturulur. Eğer bu şekilde tek bir model ortaya çıkmazsa model seçimi enformasyon ölçütü kullanılarak yapılır. Sonuçta seçilen modelin altörneklem güvenilirliği örtüşen iki örnekte test edilecek değerlendirilir.

PcGets'de liberal strateji, muhafazakar strateji, uzman kullanıcı stratejisi ve hızlı modelleyici olmak üzere çeşitli arama stratejileri kullanılabilir. Liberal arama stratejisi seçilmeme olasılığını daha düşük alırken muhafazakar strateji de silinmeme olasılığı daha düşüktür. Önemli ve önemsiz değişkenler arasındaki dengeye göre değişik stratejiler seçilebilir. Eğer regresör sayısı azsa ve önemli değişkenlerin  $t$ -değerleri düşükse liberal strateji önerilmektedir. Yöntem hakkında fazla bilgisi olmayan kullanıcılar hızlı modellemeyi kullanarak GKM'yi oluşturabilirler. PcGets gözlem sayısından fazla değişken olduğu

durumlarda da kullanılabilir. Bu, özellikle doğrusal olmayan modellerde işe yaramaktadır.

- 
- Aşama I- GKM'nin Seçilmesi
- 1) Teoriye dayanarak göreceli olarak ortogonal bir GKM seç.
  - 2) Bütün örneklem üzerinde GKM'yi kestir.
- Aşama II- GKM'nin Seçilmesi
- 3) Hatalı belirleme (kalıt ve parametre) testlerini seç. GKM'yi tekrarlanan örneklem içi testlerle indirge. Temel bir model elde et.
- Aşama III- Modelin Seçilmesi
- 4) Farklı indirgeme aramaları yaparak uç modeller elde et.
  - 5) GKM'yi uç modellerin birleşimi olarak alarak önceki adımları tekrar et.
- Aşama IV- Modelin Değerlendirilmesi
- 6) Seçimin güvenilirliğini denetlemek için son modeldeki her değişkenin önemi örtüşen iki altkümede değerlendir.
- 

**Şekil 1:** PcGets Yordamı

#### 4. RETINA YORDAMI

RETINA yordamı, veri tarafından yönlendirilen, kestirime yönelik bir modelleme yaklaşımıdır. Potansiyel olarak kestirimde yararlı olacak bir enformasyon kümesinin seçiminde yardımcı olması dışında herhangi bir kurama gerek duyulmaz. Yordam geliştirilirken, esneklik, seçici arama, doğrusal bağlantının (collinearity) denetimi ve örneklem dışı kestirim gücü ana prensipler olmuştur ve yapay sınır ağlarında kullanılan yöntemlerden esinlenilmiştir.

Bağımlı değişkenin koşullu ortalamasının bilinmediği, modelleme süreci içinde doğrusal olmayan ilişkiler ve etkileşimler için test yapılması gerektiği durumlarda özellikle kullanışlıdır. Yordam basitliği sağlamak için altörneklem üstünde çapraz geçirme yapılmasına dayanmaktadır. RETINA yordamında dört aşama vardır. Yordam Şekil 2'de verilmiştir. Bu yordamda hem doğrusal olmayan ilişkileri hem de etkileşimleri göz önüne almak için aralarında kareler, tersler v.b. bulunan çeşitli dönüşümler kullanılarak dönüştürülmüş değişkenler kümesi  $\zeta(X) = \{W_1, \dots, W_m\}$  elde edilmektedir. Son model seçilirken AIC kullanılmaktadır.

## Aşama 0-Ön işlemler

## 1) Veri Oluşturma ve Sıralama

- Dönüştürülmüş değişkenler kümesini ( $\zeta(X) = \{W_1, \dots, W_m\}$ ) oluştur.
- Örnekleme rasgele olarak üç altörnekleme böl.

## Aşama I- Bir "Aday" Modelin Ayırıştırılması

## 2) Birinci Altörneklemdaki Verileri Kullanarak

- Sadece birinci altörnekleme  $\zeta(X)$  içindeki değişkenleri bağımlı değişken ile olan örneklem korelasyonlarına göre sırala.  $W_{(1)}$  Y ile en yüksek korelasyona sahip değişken,  $W_{(2)}$  ikinci en yüksek korelasyona sahip değişken, v.b. olsun.
- Bir sabit ve  $W_{(j)}$ 'i içeren değişik regresör kümelerini göz önüne al: Her regresör kümesi  $\zeta_\lambda(X)$  bir "doğrusal bağlantı eşiği"  $0 \leq \lambda \leq 1$  tarafından endekslenir ve  $W_{(j)}$  ( $j=2, \dots, m$ )'nin halen kümede bulunan değişkenlerle regresyonunun  $R^2$ 'si  $\lambda$ 'dan küçük veya eşitse  $W_{(j)}$ 'i  $\zeta_\lambda(X)$  'nin içine eklenerek oluşturulur.
- Regresör kümelerinin sayısı  $\lambda$ 'nın 0 ile 1 arasında seçilen değerleri sayısı,  $\nu$  tarafından belirlenir.

## 3) Birinci ve İkinci Altörneklemdaki Verileri Kullanarak

- Sadece birinci altörneklemdaki verileri kullanarak Y'nin her regresör kümesi  $\zeta_\lambda(X)$  ile regresyonunu yaparak  $\nu$  adet model kestir ve sadece ikinci altörneklemdaki verileri kullanarak bir örneklem-dışı kestirim ölçütünü (çaprazgeçerlenmiş ortalama kareli kestirim hatası) hesapla.
- En iyi örneklem dışı başarıma  $\zeta_{\lambda^*}(X)$  karşılık gelen bir "aday" model seç.

## Aşama II- Arama Stratejisi

## 4) İkinci ve Üçüncü Altörneklemlerin Her İkisindeki Verileri Kullanarak

- Daha basit bir model ara: bir sabit ve  $\zeta_{\lambda^*}(X)$  içindeki bütün regresörleri içeren bütün modelleri regresörlerin eklenme sıralarına ve bu sefer 2.a'daki yöntemle ikinci örneklem kullanılarak üretilen sıralarına göre birer birer kestir.
- Bir başarıım ölçüsü (çaprazgeçerlenmiş ortalama kareli kestirim hatası, ki buna modeldeki parametre sayısının bir fonksiyonu olan bir ceza terimi de eklenebilir) hesaplayarak modellerin (üçüncü altörneklemdaki verileri kullanarak) örneklem dışı başarıım değerlendirmesini yap.

## Aşama III- Model Seçimi

- I. Ve II. Aşamaları Alt kümelerin Sırasını Değiştirerek Tekrar Et; Her Altörnekleme Sıralaması için Bir Aday Üret
- Bütün Örneklem Üzerinde En Yüksek Başarıma Sahip Modeli Seç

Şekil 2: Retina Yordamı

## 5. GENETİK YORDAMLAR KULLANARAK MODEL SEÇME YORDAMI

PcGets ve RETINA yaklaşımlarından daha farklı olarak arama yönteminde özellikle genetik yordamları (Akin, 1997) kullanan yeni bir yaklaşım da Balcombe tarafından (2005) önerilmiştir. Genetik yordamlar biyolojik evrim kuramından esinlenen daha geniş bir yordamlar ailesi olan Evrimsel Yordamlar ailesinin bir üyesidir. Evrimsel yordamlar paralel, yönlendirilmiş rastgele arama yapan bir yöntem olarak nitelendirilebilir. Aralarında farklılıklar göstermekle beraber her yaklaşımda kuşaklar boyunca evrim gösteren çözümler kümesi vardır. Her kuşakta eldeki problemin çözümünde kullanılacak olası birden çok çözüm birlikte bulunur. En uygunun ayakta kalması ilkesi kullanılarak bir sonraki kuşak oluşturulur. Bireysel çözümler “genlerden” oluşan kromozomlar şeklinde ifade edilir. Burada genler kullanılan yordama göre çözümü oluşturan parametrelerin ikili veya gerçel gösterimi olabilir. Her çözümün uygunluğu yani eldeki problemi çözmedeki başarımı bir uygunluk fonksiyonu kullanılarak belirlenir. Bu uygunluk fonksiyonunun en yüksek değeri aldığı kromozom en iyi çözüm olarak kabul edilir. Bir kuşaktaki çözümlerden bir sonraki kuşaktaki çözümler aralarında seçme, çaprazlama ve mutasyonun olduğu bazı işlemciler kullanılarak elde edilir. Seçme işlemcisi bir sonraki kuşakta genlerinden bir kısmı kalacak “ebeveynleri” seçmede kullanılır. Uygunluğa göre sıralama v.b. gibi çok değişik şekillerde yapılabilir. Çaprazlama işlemcisi seçilen iki ebeveynin genlerinden parçalar kullanarak yeni “çocuk” çözümler üretir. Çaprazlama işlemine tabi tutulacak ebeveynler seçilenler arasından rasgele belirlenir. Genelde iki ebeveyninden iki çocuk çözüm üretilebilir. Mutasyon işlemcisi belirli bir olasılıkla seçilen bir veya birkaç gende değişiklik yapar. Eğer genetik yordamlarda olduğu gibi ikili bir gösterim kullanılıyorsa genin değeri sıfırda bir, bir ise sıfır yapılıdır. Bir sonraki kuşağı oluşturma aşamaları ya belirli sayıda kuşak geçinceye kadar ya da bir çözümün uygunluk değeri belirli bir eşik değeri aşana kadar devam eder. Kullanılan yordam Şekil 3’te verilmiştir.



## Aşama I-İlk Kuşağın Oluşturulması

- 1) İkili gösterim kullanarak rasgele bir küme ( $G_0$ ) model (sıfır ve birlerden oluşan bir vektör) oluştur. Her öğeye 0.5 olasılıkla sıfır ya da bir değeri ata.  $G_0$  aramanın boyutuna bağlı olarak seçilebilir ve yüksek olasılıklı modelleri de içerebilir.

## Aşama II- Ana Çevrim

- 2) Bütün modelleri kullanarak kestirim yap. Başarımlarına göre hepsine bir uygunluk değeri ver, modelleri bu değere göre sırala. Bu değer enformasyon ölçütü kullanılarak ya da öngörü başarımı gibi başka bir ölçüye göre verilebilir.
- 3) En düşük değere sahip  $D_0$  modeli kümeden çıkar. Bunun yanında genetik yordamlarda kullanılan çok değişik seçim işlemcileri de kullanılarak eleme yapılabilir. Geriye kalan  $H_0=G_0-D_0$  model bir sonraki kuşağı üretmede kullanılacaktır.
- 4)  $H_0$ 'dan bir sonraki kuşağın bireylerini üretme çaprazlama ya da mutasyon işlemcilerini ya da her ikisini birlikte belirli olasılıklarla kullanarak yapılabilir. Mutasyonda genellikle çok düşük seçilen belirli olasılıkla bazı modeller seçilir. Bunlardaki gene olasılıkla seçilen bazı sıfırlar bir, birler sıfır yapılıır. Çaprazlamada ise iki "ebeveyn" model seçilir. Bu şekilde  $H_0$ 'dan bir sonraki kuşak  $G_1$  üretilir.
- 5) Belirlenen sayıda model kalana kadar ya da belirli sayıda iterasyon yapıłana kadar 2. adıma git. Her iterasyonda  $H_i=G_i-D_i$  şeklinde elde et. Çaprazlama ve mutasyon kullanarak  $H_i \rightarrow G_{i+1}$ 'yi oluştur. Her kuşakta aynı  $D_i$  sayısında çözüm elenmeyebilir. Başlangıç kuşağında modellerin %90'ı elenirken sonraki kuşaklarda %25'i elenebilir.

Şekil 3: Genetik yordamlar kullanarak model seçme yordamı

Bir kromozomun nasıl oluşturulduğunu göstermek için Balcombe (2005)'deki örneği kullanalım:

$$y_i = \alpha y_{i-1} + \gamma_0 x_i + \gamma_1 x_{i-1} + \varepsilon_i \quad (7)$$

şeklinde bir genel modelimiz olsun. Burada  $y_i$  bağımlı,  $x_i$  bağımsız değişkeni  $\alpha$ ,  $\gamma_0$ ,  $\gamma_1$  ise parametreleri göstermektedir. Bu parametrelerin sıfır değeri aldığı durumlarda modelde bazı terimler yer almayacaktır. (1,1,1) her üç parametrenin de modelde tutulacağını gösterirken (1,0,0) otoregresif bir modele karşı gelebilir. Bu şekilde 8 permütasyon elde edilebilir. Bunun dışında  $\gamma_0 + \gamma_1 = 0$  gibi kısıtlar da eklenirse 16 olası model elde edilir. Bu durumda (1,1,1,0) gibi dört genli kromozomlar olacaktır. Kullanılan modele göre bu şekilde kromozomların uzunluğu belirlenir. Gen sayısı  $n$  ise  $2^n$  adet farklı model elde edilir. Genetik yordamların en önemli yararı bütün bu  $2^n$  modelin hepsini kestirmeye gerek

kalmadan çok daha küçük bir küme üzerinden arma yaparak en uygun modeli belirlemesidir.

Her ne kadar genetik yordamlar global arama yordamları olsalar da iterasyon sayısının yetersizliği yada kullanılan işlemcilerin uygun olmaması nedeniyle her zaman en uygun model bulunamayabilir. Bu nedenle işlem birden fazla kez değişik başlangıç kuşaklarıyla tekrarlanmalı ve her zaman aynı modelin elde edilip edilmediği denetlenmelidir. Elde edilemiyorsa kuşak sayısı artırılması,  $G_0$ 'ın büyütülmesi,  $D_i$ 'nin boyunun değiştirilmesi gibi yöntemler denenebilir.

## 6. YORDAMLARIN KARŞILAŞTIRILMASI

Bu bölümde üç yordam değişik açılardan karşılaştırılarak benzer ve farklı yönlerine işaret edilecektir.

### 6.1. Amaçlar

PcGets'de basit, gereğinden daha genel bir model olan GKM teori ile uyumlu olarak seçilir. Bu yaklaşımda seçilen GKM'nin VÜS'ü de içerdiği varsayılarak aramaya başlanır. İlgisiz değişkenler elenerek VÜS'ün ortaya çıkarılması hedeflenir. RETINA'da modelin baştan bilinmediği varsayılır ve örneklem dışı kestirimin başarısını en yüksek hale getirecek şekilde değişkenlere dönüşümler uygulanır. GYKMS'da bir GKM'den başlanarak onun en uygun alt modelinin belirlenmesine çalışılır.

### 6.2. Strateji

Hem PCGets hem de GYKMS genel modelden başlayarak modeli özelleştirmeye çalışır. RETINA sadece bir sabit içeren modelden başlayarak yeni değişkenleri dönüşümler yaparak ekler. Bunu yaparken bunların örneklem dışı kestirime katkılarını göz önüne alır ve daha önce eklenen değişkenlerle doğrusal bağlantısı olmamasına dikkat eder.

### 6.3. Temel Model

PcGets ve GYKMS'de başlangıç modeli araştırmacı tarafından belirlenir, belirleme için otomatik bir yöntem yoktur. Özgün değişkenlerin dönüşümleri de kullanılabilir. RETINA kendi modellerini otomatik olarak oluşturur.

### 6.4. Esneklik

Esneklik, ekonomide çok sık rastlanan bir durum olan bir bağımlı değişken  $y$ 'nin verilen  $x$ 'ler cinsinden fonksiyonel biçimi hakkındaki bilgi eksikliği ile baş edebilmek için gereklidir. Bu esnekliğe ulaşabilmek için girdi değişkenlerine doğrusal olmayan ve etkileşimleri içeren bir dizi dönüşüm uygulanabilir. PCGets'de ve GYKMS'de başlangıç modeli en büyük esnekliği belirler. RETINA'da girdilerin izin verilen dönüşümleri en büyük esnekliği belirler. Aday modelin esnekliği yöntem tarafından seçilir.

### 6.5. Arama

Aday regresör sayısı,  $m$ , arttıkça tüm  $2^m$  olası modelin değerlendirilerek model seçilmesi pratik olarak olanaksız hale gelir. Bu nedenle RETINA'da White (1998)'deki fikirler kullanılarak olası modellerin  $m$  ile orantılı bir alt kümesi seçilir. RETINA seçici bir arama yaparak  $2^m$  olası modelin hepsini değerlendirme ve model seçmeden kaçınır. Dönüşümlerin doğal göz önüne alınma sırasını belirlemek için bağımlı değişkenle korelasyonu gibi dönüşümlerin dikkat çeken özellikleri kullanılır. PcGets GKM'den başlayarak çoklu azaltma patikalarını izleyerek sistematik bir arama yapar. Son seçime ulaşana kadar her azaltmanın geçerliliğini istatistiksel testler yaparak denetler. Bütün patikalar araştırılınca son bir model elde edilene kadar modelleri birleşimleriyle karşılaştırır. GYKMS ise genetik yordamları paralel global arama yaparak en uygun modeli belirlemeye çalışır.

### 6.6. Doğrusal Bağlantı

PcGets göreceli olarak ortogonal bir ifade aramak için GKM'yi belirlemeye çalışır. RETINA'da hem ilk küme de hem de daha sonra yapılan doğrusal olmayan dönüşümler de doğrusal bağlantı denetimi öncelik taşır.

## KAYNAKÇA

- Akaike, H. (1973). 'Information theory and an extension of the maximum likelihood principle', B. N. Petrov, F. Csaki (Derl.) içinde, Second International Symposium on Information Theory, 267-281. Budapeşte: Academiai Kiado.
- Akın, H. L. (1997). 'Evrimsel Hesaplama:Yapay Sorulara Doğal Bir Yanıt,' **Endüstri&Otomasyon**, 9, 12-16.
- Akın, M. (2001). 'Yapay Sinir Ağları', **İ.Ü. S.B.F. Dergisi**, 25, 7-19.
- Akın, M. (2002). 'Yapay Sinir Ağlarının Ekonomide Uygulamaları', **İ.Ü. S.B.F. Dergisi**, 26, 7-16.
- Balcombe, K. G. (2005), 'Model Selection Using Information Criteria and Genetic Algorithms', **Computational Economics**, 25, 207-228.
- Campos, J. Hendry, D. F. ve Krolzig, H. M. (2003). 'Consistent Model Selection by an Automatic Gets Approach', **Oxford Bulletin of Economics and Statistics**, 65, 803-819.
- Castle, J. L. (2005), 'Evaluating PcGets and RETINA as Automatic Model Selection Algorithms', **Oxford Bulletin of Economics and Statistics**, 67, Supplement, 837-880.
- Hannan, E. J. ve Quinn, G. (1979). 'The determination of the order of an autoregression', **Journal of the Royal Statistical Society**, B, 41, 190-195.
- Hendry, D. F. (2000). 'Epilogue: The Success of General-To-Specific Model Selection', **Econometrics: Alchemy or Science?**, 467-, Oxford University Press.
- Hendry, D.F. ve Krolzig, H. M. (2004). 'Sub-sample model selection procedures in general-to-specific Modelling', R. Becker ve S. Hurn (derl.), **Contemporary Issues in Economics and Econometrics: Theory and Applications** içinde , 53-75, Edward Elgar.
- Hoover, K. D. ve Perez, S. J. (1999). 'Data mining reconsidered: encompassing and the general-to-specific approach to specification search', **Econometrics Journal**, 2, 167-191.
- Krolzig, H.-M. ve Hendry, D. F. (2001). 'Computer automation of general-to-specific model selection procedures', **Journal of Economic Dynamics and Control**, 25, 831-866.
- Perez-Amaral, T., Gallo, G. M. ve White, H. (2003). 'A flexible tool for model building: the relevant transformation of the inputs network approach (RETINA)', **Oxford Bulletin of Economics and Statistics**, 65, 821-838.
- Perez-Amaral, T., Gallo, G. M. ve White, H. (2005). 'A comparison of complementary automatic modelling methods: RETINA and PcGets', **Econometric Theory**, 21, 262-277.
- Phillips, P.C.B. (1995). 'Bayesian model selection and prediction with empirical applications', **Journal of Econometrics**, 69, 289-331.
- Schwarz, G. (1978). 'Estimating the dimension of a model', **Annals of Statistics**, 6, 461-464.
- White, H (1998), Artificial Neural Network and Alternative Methods for Assessing Naval Readiness, Teknik Rapor, NeuralNet R&D Associates, San Diego.