11. EKONOMETRİK MODEL SEÇİMİ

Klasik doğrusal regresyon modelinin klasik varsayımları arasında yer alan biri de regresyon modelinin doğru kurulmasıdır. Kurulan bu modelde kuruluş sapması veya hatası bulunmamalıdır.

Bir ekonometrik araştırmada hangi konu ile ilgileniliyorsa onun ile ilgili ekonometri modelinin kurulması gerekir. Modelleme aşamasında ortaya çıkan önemli sorunlar şu şekildedir:

1) Modele hangi değişkenleri katmak gerekir?

2) Modelin fonksiyonel kalıbı ne olmalıdır?

3) Modele giren bağımsız değişken, bağımlı değişken ve artık terim'e ilişkin olasılık varsayımları neler olmalıdır?. Bu sorular oldukça önemlidir çünkü önem arzeden değişken/değişkenleri modele dahil etmemek, yanlış fonksiyonel kalıbı seçmek ve modelin varsayımlarına ilişkin yanlış değerlendirmeler yapmak olarak değerlendirilir (Gujarati, 1999).

Ekonometrik çözümlemelerde doğru model belirlenirken deneme yanılma yöntemleri kullanılır. Seçilen model bir iktisadi kurama dayanmalı ve modelde yapılacak olan her değişim iktisadi temeller çerçevesine oturtulmalı ve elde edilen model tahmini iktisadi kuram çerçevesine dayandırılmalıdır.

Eğer modelin doğru kurulduğuna inanılıyorsa o zaman modelin ana kütle regresyon katsayılarının tahmin edilmesine sıra gelmiştir. Modelin teşhis ve tanı koyma istatistiklerinden belirtme katsayısı (R^2), t testi, F testi ve DW testi anlamlı sonuçlar veriyorsa, modelimiz artık güvenilir ve doğru kurulmuştur. Eğer bu tanı koyma istatistiklerinden biri veya birkaçı anlamlı sonuçlar vermiyorsa, otokorelasyon sorunu, çoklu bağlantı sorunu vb. sorunlar irdelenmelidir. Eğer bu sorunlar da çözüme kavuşmuyorsa o zaman modelin yanlış kurulduğu veya model kurma sapması şüphesi ortaya çıkar. Modele gereksiz değişkenler mi dahil edilmiş veya gerekli olan bir değişken model dışında mı bırakılmış, modelin fonksiyonel kalıbı yanlış mı belirlenmiş gibi soruların cevaplanması gereklidir (Gujarati, 1999).

Model kurma sapması veya hatası modele bir değişkenin eklenmemesi ile ilgili ise o ilgili değişkenlerin modele ilavesi yapılır ve model yeniden tahmin edilir. Bu yaklaşım ekonometrik modelleme çalışmalarında "aşağıdan yukarıya doğru yaklaşım" olarak ifade edilir. Uygulama çalışmalarında çoğu araştırmalar bu şekilde yapılır. Bu yöntem çok eleştiri alsa dahi birçok ekonometrist tarafından tercih edilen bir yöntem olarak kullanılmaya devam edilmektedir.

Bir model ne kadar basit olarak kurulursa ve ilgili model gerçeğe ne kadar yakın olarak oluşturulursa, bütün önemsiz ve rassal dalgalanmalar artık (hata) terimine yüklenirse, o model daha basit ve işlerlik kazanmış demektir.

Ekonometrik modellemede amaç, k sayıda bağımsız değişkenin bağımlı değişkendeki değişimi açıklama miktarı olarak ifade edildiğine göre, düzeltilmiş $\bar{R}^2$'nin yüksek olması modelin iyi olması yani uyum iyiliğini ifade eder. R^2'nin yüksek olması fazla abartılmamalı ve tahmin edilen katsayıların işaretleri kontrol edilmeli ancak bu şekilde modelin fit olup olmadığı kararı verilebilir. Örneğin arz ile ilgili bir fonksiyonda eksi çıkması gereken bir katsayının artı çıkması sonucunda o modele kuşku ile bakmak doğru bir yaklaşımdır. R^2 modelin kestirim gücünü ifade eder ancak bu güç belli bir örneklem içindeki kestirim gücüdür. Modellemede esas amaç ise örneklem dışındaki kestirim gücünü bulmaktır (Gujarati, 1999).

11.1. Model Kurma Aşamasında Ortaya Çıkan Hatalar

a. İlgili bir değişkeni dışlama hatası

İyi bir model olarak kabul ettiğimiz toplam maliyet fonksiyonuna ilişkin modelimiz aşağıdaki verilmiş olsun.

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_1 \ + \beta_3 X_1^3 + u_i$ (11.1)

Bu modelde, Y: Üretim maliyeti ve X_1: Üretim değeridir.

Ekonometrist elindeki donelere göre 11.1 modeli yerine aşağıdaki modeli kullanmaya karar verdiğini varsayalım:

$Y_i = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_1^2 + u_i$ (11.2)

Eğer araştırmacı 11.2 modelini doğru olarak kabul ederse o zaman ilgili bir değişken olan X_1^3 dışarıda bırakılmış olur. Sonuçta ise, model kurma hatası ya da bir değişkeni dışlama hatası ortaya çıkmış olur (Gujarati, 1999).

b. İlgisiz bir değişkenin modele konma hatası

Bir araştırmacının kullandığı model aşağıdaki gibi verilmiş olsun.

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \beta_4 X_1^4 + u_{1i}$ (11.3)

Gerçek olarak kabul edilen model (11.1)'de verilen model ise, (11.3)'de verilen modelde bir model kurma hatası ortaya çıkmıştır. Bu model kurma hatasının nedeni ise, ilgisiz bir değişkeni modele dahil etmektir. Çünkü (11.1)'de verilen doğru modelde β_4 katsayısının sıfır olması öngörülmüştür. Bu durumda hata terimi şu şekildedir:

$u_{1i} = u_1 - \beta_4 X_1^4 = u_1, \quad doğru\ model\ için\ \beta_4 = 0\ olduğu\ için$

(Gujarati, 1999)

c. Yanlış fonksiyon kalıbının seçilmesi

Araştırmacının aşağıdaki modeli (11.4)'ü kullanmak istediğini varsayalım.

$lnY_i = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + u_{1i}$ (11.4)

Başlangıçta verilen doğru modelimizi (11.1)'i hatırlayalım.

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + u_i$ (11.1)

Bu durumda, doğru modelimizi (11.1)'e göre, (11.4)'de bir model kurma sapması oluşmuştur. Bu sapmanın yanlış fonksiyon kalıbının kullanılmasından kaynaklandığını ifade edebiliriz.

Doğru modelimiz (11.1)'de bağımlı değişken doğrusal iken, (11.4)'de bağımlı değişkenin logaritmik doğrusal olduğunu görürüz (Gujarati, 1999).

d. Ölçme hatasının ortaya çıkması

Araştırmacının aşağıdaki modeli (11.5)'i kullandığını varsayalım.

$Y_i^* = \beta_0^* + \beta_1^* X_1^* + \beta_2^* X_1^{*2} + \beta_3^* X_1^{*3} + u_{1i}^*$ (11.5)

Burada dikkat edilmesi gereken nokta asıl kullanılması gereken bağımlı değişken Y_i ve bağımsız değişken X_i yerine, Y_i^* ve X_1^* değerleri kullanılmıştır. Dolayısıyla, $Y_i^* = Y_i + \varepsilon_i$'dir yani ε_i ölçme hatası olarak karşımıza çıkar. Bu duruma ölçme sapması hatası denir. Genel olarak çalışmalarda ikincil verileri kullanırız ve bu verilerin nasıl işlendiklerini bilmemiz genel itibariyle imkansızdır. Veriler; eksik kapsam, bazı gözlemleri içermeme ve yaklaştırma ya da benzetme hataları ile karşı karşıya kalmaktadır.

Sonuç olarak, bir model doğru olarak tespit edildikten sonra aşağıdaki işlemleri kontrol amacıyla yaparsak modelin sağlamasını yapmış oluruz.

Bunlar;

1. İlgili bir değişkeni dışlama hatasının kontrol edilmesi,
2. İlgisiz bir değişkenin modele dahil edilme hatasının kontrol edilmesi,
3. Yanlış fonksiyon kalıbının seçilip seçilmediğinin kontrol edilmesi,
4. Ölçme hatasının ortaya çıkıp çıkmadığının kontrol edilmesi

şeklinde sıralayabiliriz.

Model kurma arayışına Leamer şu şekilde yaklaşmıştır:

1. Model sınaması ile doğru bir modelin seçilmesi
2. Verileri yorumlamak amacı çerçevesinde, aralarında ilişki olan değişkenlerin yorumunun yapılması
3. Basit bir model kurarak modeli sadeleştirmek
4. Değişkenler arasında seçim yaparak yaklaşık değişken aramanın yapılması
5. Uygun veriler ile tahmin ve kestirimin yapılması
6. Mevcut bir modelin geliştirilmesi şeklindedir.

Model seçimine Hendry veya London School of Economics (LSE) yaklaşımı, yukarıdan aşağıya doğru veya genelden özele doğrudur. Modele çok sayıda açıklayıcı değişken dahil edilir ancak önemli değişkenler modelde kalana kadar indirgemeye devam edilir. LSE yöntemlerinin temelinde, iktisadi zaman serisi verileri vardır ve gözlemler zaman alt indisi şeklinde t ile endekslenir.

Model eğer aşağıdaki şekilde oluşturuldu ise,

$Y_t = \alpha X_t$ (11.6)

LSE yönteminde (11.6)'da verilen modele ulaşabilmek için, bağımlı ve bağımsız değişkenlerin t, t-1,...,t-m dönemindeki gecikmeli değerlerine göre regresyonunu almak gerekir. Bu yöntem, aşağıdaki modeli önermektedir:

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_m X_{t-m} + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \cdots + \delta_m Y_{t-m} + u_t$$

Bu ardışık bağlanımlı modelde bağımlı değişkenin gecikmeli değerleri açıklayıcı değişken olarak görülür ve bağımsız değişkenin, bağımlı değişken üzerindeki etkisinin zamana dağılmış olduğunu göstermektedir. Bu yüzden, bu model gecikmesi dağıtılmış bir modele örnek olarak verilebilir.

Bu tür modellerde gecikme sayısının ne kadar olacağına ilişkin karar vermek önemlidir. Hendry'nin yaklaşımına göre basit bir model aşağıdaki ölçütlere göre belirlenmelidir:

1. Modelden yapılan kestirimlerin anlamlı olması
2. İktisadi teoriler çerçevesine oturmalı
3. Açıklayıcı yani bağımsız değişkenlerin hata terimi ile ilgisi olmamalı
4. Katsayı değerleri durağan olmalı
5. Model kurma hatası olmaması için, modelden tahmini yapılan kestirimlerin rassal olmalıdır yani veri uyumu sağlanmalı
6. Model, diğer tüm modelleri içine almalı yani kapsamı içine almalıdır.

Bu ölçütler çerçevesinde, Hendry yöntemi sına, sına, sına yani kısaca SSS yöntemi olarak bilinmektedir. İşte, çeşitli modeller arasından doğru model yapısına ulaşmak için sürekli denemenin önemini vurgulamaktadır (Gujarati, 1999).

11.2. Rakip Modeller Arasında Seçim Süreci

Ekonometriciler alternatif ya da rakip modeller arasında çeşitli sınamalar gerçekleştirmektedir. Bu sınamaları ikiye ayırabiliriz:

1. Yuvalanmış (Nested) model sınamaları
2. Yuvalanmamış (Nonnested) model sınamaları (Gujarati, 1999).

11.2.1. Yuvalanmış (Nested) model sınamaları

Yuvalanmış model sınamasını açıklamak için aşağıdaki modelleri yazalım:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u_i \qquad \text{(Model 1)}$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i \qquad \text{(Model 2)}$$

Model 2'nin değişkenler itibariyle bakıldığında model 1 içinde yuvalandığını söyleyebiliriz. Eğer model 1 tahmin edilir ve $H_0: \beta_3 = 0$ yokluk hipotezi reddedilmez ise model 1, model 2'ye indirgenir. Bu hipotezin sınamasını t ya da F sınaması ile yapmak mümkündür.

Yuvalanmış modeller, belirli katsayılarına sıfır kısıtlaması getirilmiş kısıtlı regresyon modelleri olarak adlandırılır. Bu tür modellerin seçimi, bağımsız değişkenler arasından seçim yapma anlamına gelmektedir.

Yuvalanmış ve yuvalanmamış modellerin karşılaştırılmasında kullanılan standart seçim kriterleri aşağıda verilmiştir (Gujarati, 1999).

11.2.2. Yuvalanmamış (Nonnested) model sınamaları

Yuvalanmamış model sınamasını açıklamak için aşağıdaki modelleri yazalım:

$Y_i = \alpha_1 + \alpha_2 X_1 + u_i$ (Model 1)

$Y_i = \beta_1 + \beta_2 K_1 + v_i$ (Model 2)

Model 1 ve model 2 incelendiğinde bağımsız değişkenler X ve K'nın farklı değişkenler olduğu görülür. Bu tür modeller birbirinin özel durumu olarak türetilememektedir bu yüzden bu modellere yuvalanmamış modeller denir (Gujarati, 1999).

11.2.3. Model Seçim Kriterleri

Araştırmacılar tarafından başvurulan yaygın standart seçim kriterleri arasında determinasyon katsayısı (R-kare ölçütü), düzeltilmiş determinasyon katsayısı (Düzeltilmiş R-kare), Akaike Bilgi Ölçütü (AIC) ve Schwarz Bilgi Kriteri (SIC) olarak sayılabilir. Bütün bu ölçütler, Artık Kareler Toplamını (AKT)'yi minimize etmeye dayanır. Zaman serisi modellerinde gecikme uzunluğunu saptamada AIC ve SIC kullanılmaktadır.

11.2.3.1. Determinasyon katsayısı (R-kare ölçütü)

Regresyon denkleminin verilere uyumunun bir göstergesidir. Belirtme katsayısı, açıklanabilen değişimin toplam değişime oranlanmasıyla elde edilir. Aynı zamanda, bağımlı değişkendeki değişimin yüzde ne kadarının bağımsız değişken veya değişkenlerdeki değişmelerle açıklanabileceğini ifade etmektedir. Bu katsayı, k sayıda bağımsız değişkenin bağımlı değişkendeki değişimi açıklama miktarını vermektedir.

Toplam değişim aşağıdaki gibi,

$KT_y = BKT + AKT$

Toplam Değişim=Açıklanabilen Değişim + Açıklanamayan Değişim formülü ile ifade edilir.

Bunu daha açık bir biçimde şöyle de yazmak mümkündür:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Bu eşitlikten hareketle çoklu belirtme katsayısı,

$$R^2 = \frac{BKT}{KT_y}$$

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = r^2$$

Veya

$$R^2 = 1 - \frac{AKT}{KT_y}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

olarak da bulunur.

a) R-kare, basit doğrusal regresyon modellerinde olduğu gibi çoklu regresyon modelleri için de $0 \leq R^2 \leq 1$ olacaktır.

b) Modele ilave edilen her yeni bağımsız değişken R-kare'nin artmasına yol açar.

c) R-kare yüksek çıktı diye modelin örneklem harici veya dışı gözlemleri iyi açıklayacağına güvenilmesi hatalı olur.

d) R-kare değerlerinin karşılaştırılması için bağımlı değişkenlerin aynı olma zorunluluğu mevcuttur.

Örnek 11.1. 2001-2013 yılları arasındaki Şekerpancarı verimine etki eden sıcaklık ve gübre miktarları verilmiştir. Parametre tahmini yapılan çoklu regresyon modelinin determinasyon katsayısını hesaplayınız ve yorumlayınız?

Çizelge 42. Türkiye'de Şekerpancarı Verimi, Sıcaklık ve Gübre Miktarları

Yıl	**Verim (10 kg/da)**	**Gübre (10 Kg/da)**	**Sıcaklık(°C)**
2001	354	1,0	14,22
2002	444	1,1	13,21
2003	401	1,2	13,29
2004	429	1,0	13,40
2005	452	1,3	13,41
2006	446	1,4	13,59
2007	415	1,1	13,75
2008	482	1,2	13,54
2009	533	1,3	13,70

2010	545	1,2	15,20
2011	548	1,5	14,08
2012	532	1,4	14,20
2013	566	1,5	14,10

Kaynak: Türkiye İstatistik Kurumu, 2014, Meteroloji Genel Müdürlüğü, 2014.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{AKT}{KT_y} = \frac{\sum u_i^2}{\sum (y_i - \bar{y})^2} = \frac{17682,46}{53275,69} = 0,67$$

KT_y = BKT + AKT
BKT = KT_y - AKT
BKT = 53275,69-17682,46=35593,23

$$R^2 = \frac{BKT}{KT_y} = \frac{35593,23}{53275,69} = 0,67$$

Sonuç olarak, her iki formüle göre hesaplanan R^2 değerlerinden aynı sonuca ulaşılmıştır. Şekerpancarı verimindeki değişmelerin %67'i gübre miktarı ve sıcaklıktaki değişmeler tarafından açıklanmaktadır.

11.2.3.2. Düzeltilmiş determinasyon katsayısı (Düzeltilmiş R-kare)

Düzeltilmiş R-kare aşağıdaki gibi,

$$\bar{R}^2 = 1 - \frac{\sum u_i^2/(n-k)}{\sum y_i^2\,/(n-1)}$$

gösterilir.

Burada;
k: Sabit terim dahil modeldeki katsayıların sayısı
n: Gözlem sayısı

Düzeltilmiş ifadesi, formülasyondaki kareler toplamının serbestlik derecelerine göre düzeltilmiş olmasını ifade eder. Düzeltilmiş R-kare ile R-kare arasındaki ilişkiden aşağıdaki formülasyona ulaşırız.

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

a) k>1 olduğunda Düzeltilmiş R-kare < R-kare'dir. Açıklayıcı değişken (X) sayısı arttıkça Düzeltilmiş R-kare'nin R-kare'ye göre daha az arttığı görülür.

b) R-kare hiç bir zaman eksi değer almaz ama düzeltilmiş R-kare eksi değerli olabilir.

c) Modeller arası karşılaştırmalarda düzeltilmiş R-kare daha iyi bir ölçüdür ancak dikkat edilmesi gereken husus bağımlı değişkenlerin aynı olması zorunluluğudur.

Örnek 11.2. 2001-2013 yılları arasındaki Şekerpancarı verimine etki eden sıcaklık ve gübre miktarları verilmiştir. Parametre tahmini yapılan çoklu regresyon modelinin determinasyon katsayısını ve düzeltilmiş determinasyon katsayısını hesaplayınız ve yorumlayınız?

Çizelge 43. Türkiye'de Şekerpancarı Verimi, Sıcaklık ve Gübre Miktarları, 2001-2013

Yıl	Verim (10 kg/da)	Gübre (10 Kg/da)	Sıcaklık (°C)
2001	354	1,0	14,22
2002	444	1,1	13,21
2003	401	1,2	13,29
2004	429	1,0	13,40
2005	452	1,3	13,41
2006	446	1,4	13,59
2007	415	1,1	13,75
2008	482	1,2	13,54
2009	533	1,3	13,70
2010	545	1,2	15,20
2011	548	1,5	14,08
2012	532	1,4	14,20
2013	566	1,5	14,10

Kaynak: Türkiye İstatistik Kurumu, 2014, Meteroloji Genel Müdürlüğü, 2014.

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

n=13

k=4

$$R^2 = \frac{BKT}{KT_y} = \frac{35593,23}{53275,69} = 0,67$$

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} = 1 - (1 - 0,67)\frac{13-1}{13-4} = 0.56$$

R^2 yerine $\bar{R}^2$ kullanmanın daha iyi olduğu bellidir. Bunun nedeni ise, regresyonun uyumu konusunda özellikle açıklayıcı değişken sayısı gözlem sayısından az değilse aşırı iyimser bir tablo çizebilir (Theil, 1978).

11.2.3.3. Akaike Bilgi Ölçütü (AIC)

Akaike bilgi ölçütü, Hirotugu Akaike tarafından 1974 yılında geliştirilmiştir. Bu kriter, hem yuvalanmış hem yuvalanmamış modellerin ayrıca gecikmeli modellerde kullanılan yararlı bir kriterdir. Alternatif modeller için bu kriter hesaplanır ve değerlerini minimum yapan model en iyi model olarak tercih edilmektedir. Modeller karşılaştırılırken AIC değeri düşük olan seçilir. AIC değeri ne kadar düşükse o kadar iyidir.

Farklı modelleri i indisi ile ifade ettiğimizde bu kriter aşağıda gibi,

$$AIC_i = \ln(\hat{\sigma}_i^2) + \frac{2k_i}{n}$$

yazılır. Burada,

$$\hat{\sigma}_i^2 = \frac{AKT_i}{n}$$

formülü ile hesaplanır. Bu formülde S_e^2'den farkı ortaya koymak amacıyla $\hat{\sigma}$ ile ifade edilmiştir. AIC kriterinin küçük örneklerde iyi sonuç vermemesi nedeni ile bu kriter aşağıdaki gibi düzeltilerek yeniden yazılmıştır (Dikmen, 2012).

$$AIC_{di} = \ln(\hat{\sigma}_i^2) + \frac{n + k_i}{n - k_i - 2}$$

Örnek 11.3. 2001-2013 yılları arasındaki Şekerpancarı verimine etki eden sıcaklık ve gübre miktarları verilmiştir. Tahmin edilen modeller için AIC değerini hesaplayınız.

Çizelge 44. Türkiye'de Şekerpancarı Verimi, Sıcaklık ve Gübre Miktarları

Yıl	**Verim (10 Kg/da)**	**Gübre (10 Kg/da)**	**Sıcaklık(°C)**
2001	354	1,0	14,22
2002	444	1,1	13,21
2003	401	1,2	13,29
2004	429	1,0	13,40
2005	452	1,3	13,41
2006	446	1,4	13,59
2007	415	1,1	13,75
2008	482	1,2	13,54
2009	533	1,3	13,70
2010	545	1,2	15,20
2011	548	1,5	14,08
2012	532	1,4	14,20
2013	566	1,5	14,10

Y: Verim (10 Kg/da), X_1: Gübre (10 Kg/da), X_2: Sıcaklık(°C)

Kaynak: Türkiye İstatistik Kurumu, 2014, Meteroloji Genel Müdürlüğü, 2014.

Yukarıdaki veriler kullanılarak aşağıdaki model tahmin edilmiştir.

$Log\hat{Y}_i = 3{,}02 + 0{,}72 LogX_1 + 1{,}13 LogX_2$

$S_{\hat{\beta}_0}$=1,88

$S_{\hat{\beta}_1}$=0,20

$S_{\hat{\beta}_2}$=0,72

$\sum e_i^2$=0,09

$R^2 = 0{,}65$

$\bar{R}^2 = 0{,}58$

olarak kısıtsız model tahmin edilir. Bu modelde yer alan bağımsız değişkenleri kullanarak iki alt model oluşturulur ve yeni modeller kurulabilir. Burada bir bağımsız değişkenli iki model için AIC değeri hesaplanır ve uygun olan model seçilir. Bu modeller aşağıda verilmiştir:

Modellerden birisi,

$Log\hat{Y}_i = 5{,}99 + 0{,}78 LogX_1$

$S_{\hat{\beta}_0}$=0,05

$S_{\hat{\beta}_1}$=0,21

$\sum e_i^2$=0,11

$R^2 = 0{,}56$

$\bar{R}^2 = 0{,}52$

Modellerden diğeri,

$Log\hat{Y}_i = 1{,}88 + 1{,}63 LogX_2$

$S_{\hat{\beta}_0}$=2,68

$S_{\hat{\beta}_1}$=1,02

$\sum e_i^2$=0,20

$R^2 = 0{,}19$

$\bar{R}^2 = 0{,}11$

olarak tahmin edilmiştir. Tahmin edilen bu modeller için AIC kriteri aşağıdaki gibi hesaplanır.

İlk model için AIC kriteri,

$$AIC_{di} = \ln(\hat{\sigma}_i^2) + \frac{n + k_i}{n - k_i - 2}$$

$$\hat{\sigma}_i^2 = \frac{0.11}{13} = 0{,}008$$

$$AIC_{di} = \ln(0{,}008) + \frac{13+2}{13-2-2} = -3{,}12$$

İkinci model için AIC kriteri,

$$\hat{\sigma}_i^2 = \frac{0{,}20}{13} = 0{,}02$$

$$AIC_{di} = \ln(0{,}02) + \frac{13+2}{13-2-2} = -2{,}47$$

Birinci modelin AIC değeri daha düşük çıktığı için bu model tercih edilir.

11.2.3.4. Schwarz Bilgi Kriteri (SIC)

Schwarz kriteri,1978 yılında Gideon Schwarz tarafından önerilmiştir. Formülü aşağıdaki gibidir;

$$SIC = n^{k/n}\frac{\sum \hat{u}^2}{n} = n^{k/n}\frac{AKT}{n}$$

Logaritmik formda ise aşağıdaki formül yardımıyla elde edilir ,

$$lnSIC = \left(\frac{k}{n}\right) lnn + \ln(\frac{AKT}{n})$$

$SIC_i = \ln\left(\hat{\sigma}_i^2\right) + \frac{k_i \ln(n)}{n}$ olarak belirlenir (Dikmen, 2012) (Ucal, 2006).

Burada,

$\left(\frac{k}{n}\right) lnn$ penaltı (sınırlama) faktörü olarak isimlendirilir.

Schwarz Bilgi Kriteri (SIC), her zaman AIC'den daha düşük çıkmaktadır.

Örnek 11.4. 2001-2013 yılları arasındaki Şekerpancarı verimine etki eden sıcaklık ve gübre miktarları verilmiştir. Tahmin edilen modeller için SIC değerini hesaplayınız.

Çizelge 45. Türkiye'de Şekerpancarı Verimi, Sıcaklık ve Gübre Miktarları

Yıl	Verim (10 kg/da)	Gübre (10 Kg/da)	Sıcaklık(°C)
2001	354	1,0	14,22
2002	444	1,1	13,21
2003	401	1,2	13,29
2004	429	1,0	13,40
2005	452	1,3	13,41
2006	446	1,4	13,59
2007	415	1,1	13,75
2008	482	1,2	13,54
2009	533	1,3	13,70
2010	545	1,2	15,20
2011	548	1,5	14,08
2012	532	1,4	14,20
2013	566	1,5	14,10

Y: Verim (10 kg/da)
X_1: Gübre (10 Kg/da)
X_2: Sıcaklık (°C)
Kaynak: Türkiye İstatistik Kurumu, 2014, Meteroloji Genel Müdürlüğü, 2014.

Yukarıdaki veriler kullanılarak aşağıdaki model tahmin edilmiştir. Modellerdeki L harfi, logaritmik değerleri ifade eder.

$$Log\hat{Y}_i = 3{,}02 + 0{,}72 LogX_1 + 1{,}13 LogX_2$$

$S_{\hat{\beta}_0}$ =1,88

$S_{\hat{\beta}_1}$ =0,20

$S_{\hat{\beta}_2}$ =0,72

$\sum e_i^2$=0,09

$R^2 = 0{,}65$

$\bar{R}^2 = 0{,}58$

olarak kısıtsız model tahmin edilir. Bu modelde yer alan bağımsız değişkenleri kullanarak iki alt model oluşturulur ve yeni modeller kurulabilir. Burada bir bağımsız değişkenli iki model için SIC değeri hesaplanır ve uygun olan model seçilir. Bu modeller aşağıda verilmiştir:

Modellerden birisi,

$$Log\hat{Y}_i = 5{,}99 + 0{,}78 LogX_1$$

$S_{\hat{\beta}_0}$ =0,05

$S_{\hat{\beta}_1}$=0,21

$\sum e_i^2$=0,11

$R^2 = 0{,}56$

$\bar{R}^2 = 0{,}52$

Modellerden diğeri,

$Log\,\hat{Y}_i = 1{,}88 + 1{,}63\,LogX_2$

$S_{\hat{\beta}_0}$=2,68

$S_{\hat{\beta}_1}$=1,02

$\sum e_i^2$=0,20

$R^2 = 0{,}19$

$\bar{R}^2 = 0{,}11$

olarak tahmin edilmiştir. Tahmin edilen bu modeller için SIC kriteri aşağıdaki gibi hesaplanır.

İlk model için SIC kriteri,

$$SIC_i = \ln(\hat{\sigma}_i^2) + \frac{k_i \ln(n)}{n}$$

$$SIC_i = \ln(0{,}008) + \frac{2\ln(13)}{13} = -4{,}43$$

İkinci model için SIC kriteri,

$$SIC_i = \ln(0{,}016) + \frac{2\ln(13)}{13} = -3{,}74$$

olur. Birinci modelde SIC kriteri daha düşük çıktığından bu modeli tercih edebiliriz.

11.2.3.5. Davidson-MacKinnon J Testi

Davidson-MacKinnon J sınaması veya testini açıklamak için iki modeli karşılaştırmak istediğimizi varsayalım. Bunun için, C modelini, D modeli ile karşılaştırmak istediğimizi varsayalım. Bu testin aşamaları şu şekildedir:

1. $Y_i = \alpha_1 + \alpha_2 X_{2i} + u_i$ (Model C)

$Y_i = \beta_1 + \beta_2 Z_{2i} + v_i$ (Model D)

Burada, X ve Z'ler farklı değişkenler olup C ve D yuvalanmamış modellerdir. Yuvalanmamış model denmesinin sebebi, biri diğerinin özel bir durumu olarak türetilmesi sözkonusu değildir.

Model D'yi tahmin ettikten sonra buradan tahmin edilmiş Y değerlerini ($\hat{Y}_i^D$) buluruz.

2. Tahmin edilmiş Y değerini, Model C'ye ilave eder daha sonra

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 \hat{Y}_i^D + u_i$$

modeli tahmin edilir.

3. T testi kullanılarak $\alpha_3 = 0$ hipotezi sınaması yapılır.

4. Eğer $\alpha_3 = 0$ hipotezi reddedilmezse model C doğru model olarak kabul edilir. $Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 \hat{Y}_i^D + u_i$ modeline dahil olan ve model C'ye alınmamış değişkenlerin etkisini gösteren $\hat{Y}_i^D$ model C'nin açıklayıcı gücüne herhangi bir katkı yapmamıştır.

Eğer $\alpha_3 = 0$ hipotezi reddedilirse, model C doğru model değildir.

5. Birinci aşamada uygulanan yöntemin tersini şimdi uyguluyoruz. Model C'yi tahmin eder buradan tahmin edilmiş Y değerlerini ($\hat{Y}_i^C$) buluruz. Aşama 2'den 4'e kadar yeniden uygulama yaparız.

$$Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 \hat{Y}_i^C + u_i$$

Burada $\hat{Y}_i^C$ model C'den tahmin edilmiş Y değerlerini gösterir. T testi kullanılarak $\beta_3 = 0$ hipotezi sınaması yapılır. Eğer $\beta_3 = 0$ hipotezi reddedilmezse model D doğru model olarak kabul edilir. Eğer $\beta_3 = 0$ hipotezi reddedilirse model C doğru modeldir, model C model D'ye üstün tutulur (Gujarati, 1999).

Örnek 11.5. 2001-2013 yılları arasındaki Şekerpancarı verimine etki eden sıcaklık ve gübre miktarları verilmiştir.

Çizelge 46. Türkiye'de Şekerpancarı Verimi, Sıcaklık ve Gübre Miktarları

Yıl	Verim (10 kg/da)	Gübre (10 Kg/da)	Sıcaklık (°C)
2001	354	1,0	14,22
2002	444	1,1	13,21
2003	401	1,2	13,29
2004	429	1,0	13,40
2005	452	1,3	13,41
2006	446	1,4	13,59
2007	415	1,1	13,75
2008	482	1,2	13,54
2009	533	1,3	13,70
2010	545	1,2	15,20
2011	548	1,5	14,08
2012	532	1,4	14,20
2013	566	1,5	14,10

Y: Verim (10 kg/da)
X_1: Gübre (10 Kg/da)
X_2: Sıcaklık (°C)
Kaynak: Türkiye İstatistik Kurumu, 2014, Meteroloji Genel Müdürlüğü, 2014.

Modellerimiz aşağıdaki gibi,
Verim ve sıcaklık ile ilgili modelimiz,

$\hat{Y}_{t1} = -329{,}05 + 58{,}01 X_{2t}$

$S_{\hat{\beta}_0}$ =453,87

$S_{\hat{\beta}_1}$ =32,81

$\sum e_i^2$=41486,32

$R^2 = 0{,}221$

$\bar{R}^2 = 0{,}15$

Verim ve gübre ile ilgili modelimiz,

$\hat{Y}_{t2} = 111.73 + 289{,}78 X_{1t}$

$S_{\hat{\beta}_0}$ =98,28

$S_{\hat{\beta}_1}$ =78,19

$\sum e_i^2$=23691,21

$R^2 = 0{,}56$

$\bar{R}^2 = 0{,}52$ tahmin edilmiştir. İkinci modelden $\hat{Y}_{t2}$ değerleri hesaplanır ve aşağıdaki

Çizelge oluşturulduğunda,

Çizelge 47. $\hat{Y}_{t2}$ Değerleri

Yıl	$\hat{Y}_{t2}$
2001	401,5
2002	430,5
2003	459,5
2004	401,5
2005	488,4
2006	517,4
2007	430,5
2008	459,5
2009	488,4

2010	459,5
2011	546,4
2012	517,4
2013	546,4

$\hat{Y}_{t2}$ birinci modele bağımsız değişken olarak ilave edilecek ve yeni tahmin edilecek olan model aşağıdaki gibi,

$\hat{Y}_t = -542,00 + 42,18 X_2 + 0,91 \hat{Y}_{t2}$

$sh \quad (316,15) \; (22,88) \quad (0,25)$

$R^2 = 0,67$

F=10,07 olarak tahmin edilmiştir.

$$H_0 : \alpha_2 = 0$$
$$H_0 : \alpha_2 \neq 0$$

Test istatistiği,

$$t = \frac{\alpha_2}{S_{\alpha_2}} = \frac{0,91}{0,25} = 3,67$$

olarak bulunur. 0.05 hata payı için Çizelge değeri $t_{0.05;10}$=2.23'dir. t > $t_{0.05;10}$ olduğundan H_0 hipotezi reddedilir.