

PAPER DETAILS

TITLE: HATALARI ARDISIK BAGIMLI (OTOKORELASYONLU) OLAN REGRESYON
MODELLERININ TAHMIN EDILMESI

AUTHORS: Selahattin YAVUZ

PAGES: 123-140

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/29737>

HATALARI ARDIŞIK BAĞIMLI (OTOKORELASYONLU) OLAN REGRESYON MODELLERİNİN TAHMİN EDİLMESİ

Selahattin YAVUZ^(*)

Özet: Doğrusal regresyon modellerinde en küçük kareler yönteminin başarılı bir biçimde uygulanıp istenilen sonuçları verebilmesi için bazı varsayımların sağlanması gerekir. Bu varsayımlardan biri ardışık hata terimi değerlerinin birbirinden bağımsız olmasıdır. Yani hata terimi değerleri arasında otokorelasyon bulunmamasıdır. Bu varsayımın sağlanmaması durumunda en küçük kareler yöntemiyle bulunan tahminler sapmalı, t ve F gibi testler olduğundan büyük çıkıp güvenilirliğini yitirirler.

Bu çalışmada doğrusal regresyon modellerinde otokorelasyon sorunu, otokorelasyonun saptanması yöntemlerinden; grafik yöntemi, sıra testi, Durbin-Watson testi, Von-Neumann oran testi, Ki-kare otokorelasyon testi üzerinde durulduktan sonra bir doğrusal modelde otokorelasyon olması durumunda modelin nasıl tahmin edileceği yöntemi anlatılmıştır. Bu yöntem çok değişkenli bir regresyon modeline uygulanmıştır.

Anahtar Kelimeler: Regresyon, Otokorelasyon, Parametre Tahmini

Abstract: In linear regression models it is need to provide some hypothesis to perform the least square method successfully to have wanted results. One of the hypothesis is that values of the sequential residual terms have to be independent from each other. It means that there is not any autocorrelation among the values of the sequential residual terms. In case of not to provide this hypothesis; the predictions which have been obtained by the method of the least square must decline. The tests such as t and F lose their reliability by increasing more than natural.

In this study, autocorrelation problem in linear regression models, from autocorrelation determine methods; graphic method, rank test, Durbin-Watson test, Von-Neumann rate test, chi-square autocorrelation tests have been looked around and then the method that how to estimate the model if there is autocorrelation in linear model has been told. This method has been applied on a multivariate regression model.

Key Words: Regression, Autocorrelation, Estimate of Parameter

I.Giriş

Bilimin en temel amaçlarından biri değişkenler arasındaki ilişkileri araştırmaktır. Hem değişkenler arasındaki ilişkilerin derecelerinin, hem de bu ilişkilerin fonksiyonel şekillerinin belirlenmesinde istatistik bize yardımcı olur. İstatistikte değişkenler arasındaki ilişkinin derecesini gösteren katsayıya korelasyon katsayısı, değişkenler arasındaki ilişkinin fonksiyonel şeklini belirleyen denkleme ise regresyon denklemi adı verilir. Bir bağımlı (Y) ve bir bağımsız (X) değişkenden oluşan ilişki basit doğrusal regresyon ile incelenir. Bir bağımlı (Y) ve birden fazla bağımsız ($X_1, X_2, X_3, \dots, X_n$) değişkenlerden oluşan ilişki ise çoklu doğrusal regresyon ile incelenir (Gürsaka, 2002: 305).

^(*) Yrd.Doç. Dr. Kahramanmaraş Sütçü İmam Üniversitesi Elbistan MYO

Regresyon katsayılarının tahmini için en iyi yöntem en küçük kareler yöntemidir. Regresyon doğrusunun gözlem değerlerini iyi bir şekilde temsil etmesi için, bu gözlem noktalarını tam olarak ortalaması gerekir. Bu şekilde e_i kalıntıların kareler toplamı minimize edilmiş olacaktır. Bunun için en küçük kareler yönteminde gerçek ilişkiye bir terim olarak eklenen u değişkeni ile ilgili varsayımlardan biri u tesadüfi değişkeni arasında korelasyon sıfırdır. Yani $Kov(u_i, u_j) = 0, (i \neq j)$ 'dir. (Karagöz, 1999: 198).

Regresyon analizinde hata teriminin bağımsız, ortalaması sıfır, varyansı sabit olan normal dağılışı gösterdiği varsayılır. Uydurulan model için bu varsayımlar tutmadığı takdirde o model ile ilgili her türlü yorum şüphe ile karşılanır ve tartışmaya açıktır. Bu nedenle model kullanılmadan önce gerekli testlerden geçerek uydurma tekniğinin gerektirdiği varsayımlar kontrol edilmelidir (Şahinler, 2000: 61).

Bu çalışmada, doğrusal regresyon analizi ilgili temel bilgiler verildikten sonra regresyon modellerinde otokorelasyon sorunu ve otokorelasyonun varlığını tespit etmede kullanılan bazı yöntemler üzerinde durulmuştur. Regresyon modelinde otokorelasyon olması durumunda gerekli olan çözüm yöntemleri verildikten sonra bu yöntemler, minitab istatistiksel yazılımı da kullanılarak çok değişkenli bir doğrusal modele uygulanmıştır.

II. Regresyon Analizi

Regresyon; sözcük anlamıyla, bir şeyi başka bir şeye bağlama işi ve biçimidir. Bilimsel olarak regresyon terimi bir değişkenle başka bir (ya da birden çok) değişken arasında ilişki kurma işini ve ilişkinin biçimini anlatır.

İstatistiksel anlamda, iki değişken arasındaki ilişki, bunların değerlerinin karşılıklı değişimleri arasında bir bağıllık şeklinde anlaşılır. (X) değişkeninin değerleri değişirken, buna bağlı olarak (Y) değişkeninin değerleri de değişiyorsa, bu iki değişken arasında bir ilişki olduğu söylenebilir. Regresyon; bir bağımlı değişken(Y), diğeri de bağımsız değişken veya değişkenler(X) olmak üzere değişkenler arasındaki en uygun ilişkinin matematiksel bir fonksiyon şeklinde ifade edilmesidir (<http://web.sakarya.edu.tr/~adurmus>).

Regresyonda bir bağımsız değişken olması durumunda basit doğrusal regresyon, birden fazla bağımsız değişken olması durumunda ise çoklu doğrusal regresyon adını alır.

A. Basit Doğrusal Regresyon

Regresyon analizi bir bağımlı değişken ile bir bağımsız değişkenden oluşuyorsa basit doğrusal regresyon söz konusudur. Basit doğrusal regresyon modeli,

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad (1)$$

biçimindedir. Burada β_0 ve β_1 modelin bilinmeyen parametreleridir. ε hata terimi olup Y (bağımlı) ve X (bağımsız) değişkenlerdir (Ünver-Gamgam, 1996: 275).

Modelin parametrelerinin tahmini için farklı yöntemler kullanılmaktadır. Burada parametrelerin tahmini için en yaygın kullanılan yöntem olan en küçük kareler yöntemi tanıtılacaktır.

B. Parametre Tahmini

Regresyon modeli için parametrelerin tahmini en küçük kareler metoduna göre yapılır. Bu yöntemle göre,

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (2)$$

denklemindeki α ve β parametreleri öyle tahmin edilmelidir ki hata kareler toplamı en küçük olsun. Bunun için önce denklemden ε_i çekildiğinde,

$$\varepsilon_i = Y_i - \alpha - \beta X_i \quad (3)$$

eşitliği elde edilir. Bu eşitliğin her iki tarafının karesi alınıp toplanırsa,

$$\sum \varepsilon_i^2 = \sum (Y_i - \alpha - \beta X_i)^2 \quad (4)$$

elde edilir. Bu eşitliğin sağ tarafındaki α ve β parametreleri için ayrı ayrı kısmi türevi alınarak sıfıra eşitlenip çözülürse, β parametresinin tahmin edicisi olan b ,

$$b = \frac{\sum X_i Y_i - (\sum X_i Y_i) / n}{\sum X_i^2 - [(\sum X_i)^2] / n} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad (5)$$

formülüyle elde edilir. Benzer şekilde α parametresinin tahmin edicisi olan a ise,

$$a = \bar{Y} - b \bar{X} \quad (6)$$

Eşitliğinden elde edilir. Bu değerler $Y = a + bX$ denkleminde yerine yazılarak doğrusal regresyon modeli kurulmuş olur (Yıldız-Akbulut-Bircan, 1999: 219).

C. Çoklu Doğrusal Regresyon

Özellikle ekonomi ve işletmecilik alanlarında herhangi bir ekonomik değişkeni tek bir bağımsız değişkenle açıklamak mümkün değildir. Ekonomik değişkenler karmaşık değişkenlerdir. Birçok ekonomik değişken bir araya gelerek bir değişkeni etkileyebildikleri gibi, kendi aralarında da birbirlerini etkilemektedirler. Bu nedenle, tek bağımsız değişkenli regresyon analizi yapmak mümkün değildir. Bunun için birden fazla bağımsız değişkenli regresyon analizi yapılır. Birden fazla bağımsız değişkenli analize “çoklu regresyon analizi” adı verilir (Orhunbilge, 1996: 81). k sayıda bağımsız değişkenin olduğu doğrusal regresyon modeli,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (7)$$

biçimindedir. Burada,

Y : Bağımlı değişken,
 X : Bağımsız değişken,
 β_i : Tahmin edilecek parametreler,
 ε : Hata terimi

olmak üzere ε modelin stokastik olduğunu ifade eder ve modele dahil edilmeyen değişkenleri içerir. Çoklu doğrusal regresyon modelinin varsayımları aşağıdaki gibidir (Kalaycı, 2006: 259):

- ✓ Normal dağılım.
- ✓ Doğrusallık.
- ✓ Hata terimlerinin ortalaması sıfırdır.
- ✓ Sabit varyans.
- ✓ Otokorelasyon olmaması.
- ✓ Bağımsız değişkenler arasında çoklu bağlantı olmaması.

Çoklu regresyon modelinde yer alan $\beta_0 \beta_1 \beta_2 \dots \beta_i$ parametrelerin tahmin edicileri olan $b_0 b_1 b_2 \dots b_i$ olup tahmin edilecek model,

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k \quad (8)$$

biçimindedir. Çoklu doğrusal regresyon modelinin parametreleri tahmini basit regresyon modelinde olduğu gibi,

$$\sum \varepsilon_i^2 = \sum (Y_i - \hat{Y})^2 ; \hat{Y} : \text{Tahmin edilen değer} \quad (9)$$

ifadesinin minimum olması ilkesine dayanır. Minimum olması için sırasıyla $b_0 b_1 b_2 \dots b_i$ tahmin edicilerine göre kısmi türevler alınıp sıfıra eşitlenerek elde edilecek denklem sistemi çözümlenerek parametreler bulunur (Ünver-Gamgam, 1996: 319).

III. Regresyon Modellerinde Otokorelasyon Sorunu

Regresyonda temel varsayımlardan biri hata terimleri arasında ilişki olmamasıdır. Hata terimleri arasında ilişki olması otokorelasyonun varlığını ortaya çıkarır (Ünver-Gamgam, 1996: 345).

Otokorelasyon durumunda parametrelerin en küçük kareler tahmincileri sapmasız ve tutarlı olup, etkin değildir. Hata teriminin varyansının tahmincisi sapmalıdır ve bu yüzden parametrelerin varyansları da sapmalı olur. Pozitif otokorelasyon varsa sapma negatif olur. Yani varyanslar olduğundan küçük bulunur. Bunun sonucunda t test istatistiği değeri büyük çıkar. Böylece anlamsız bir katsayının anlamlı olma olasılığı artar. R^2 de yükselir. Dolayısıyla F değeri olduğundan büyük bulunur. Sonuç olarak t ve F testleri güvenilirliğini yitirip yanıltıcı sonuç verirler (<http://bilgiteknoloji.net/ekonometri/ekotest>).

EKK metodunda kabul ettiğimiz varsayımlardan biri rassal değişken u'nun ardışık değerlerinin birbirinden bağımsız olduğudur. u'nun herhangi bir dönemde aldığı değer, daha önceki herhangi bir dönemde aldığı değerden farklıdır. u_i ve u_j rassal değişkenler olsun. Bu durumda,

$$\text{Cov}(u_i, u_j) = E\{[u_i - E(u_i)][u_j - E(u_j)]\} = E(u_i, u_j) = 0 \quad (10)$$

şeklinde ifade edilir. Buna karşılık bahsedilen u rassal değişkeni için otokorelasyon varsa iki değer arasında ilişki vardır. Ve bu durum,

$$E(u_i, u_j) \neq 0, \quad (i \neq j) \quad (11)$$

biçiminde yazılır (Kutlar, 1998: 139).

Değişkenler arasındaki ilişkinin modeli yanlış belirlenmişse bu durum hata terimleri arasındaki korelasyona sebep olabilmektedir. Örneğin Y ve X değişkenleri arasındaki ilişki kuadratik iken doğrusal bir ilişkinin belirlendiği varsayılın. Gerçek ilişkideki hata terimi içsel bağıntısız olduğu halde, doğrusal ilişkiye tekabül eden X^2 bir terim içerecektir. Eğer X değerlerinde herhangi bir serisel korelasyon varsa, bileşik hata teriminde de serisel korelasyon olacaktır (Johnston, 1981: 180).

Hata terimleri arasındaki otokorelasyonun varlığı genel olarak aşağıdaki ana nedenlere dayanmaktadır (Koutsoyiannis, 1989: 207):

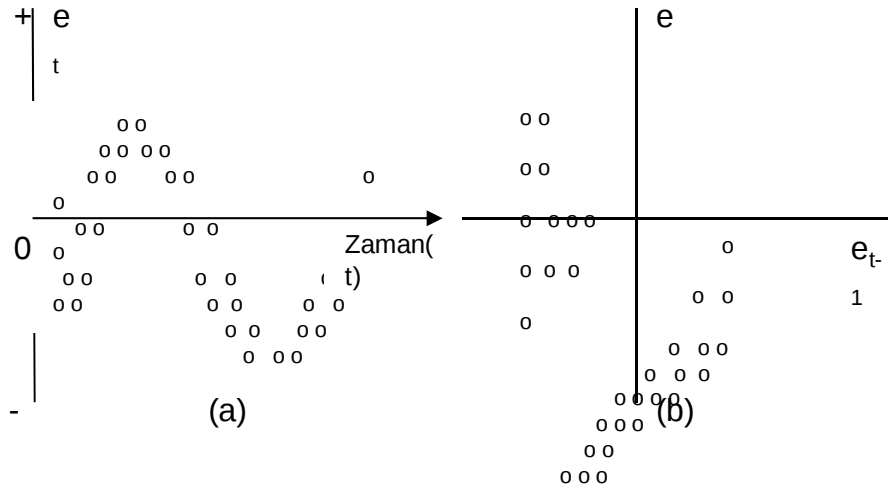
- ✓ Modele bazı açıklayıcı değişkenlerin alınmaması
- ✓ Modelin matematiksel kalıbının yanlış seçilmesi
- ✓ Bağımlı değişkenin ölçme hatalı olması
- ✓ Verilerin sistematik incelenmesi
- ✓ u 'nun yanlış spesifikasyonudur.

IV. Otokorelasyonun Saptanması

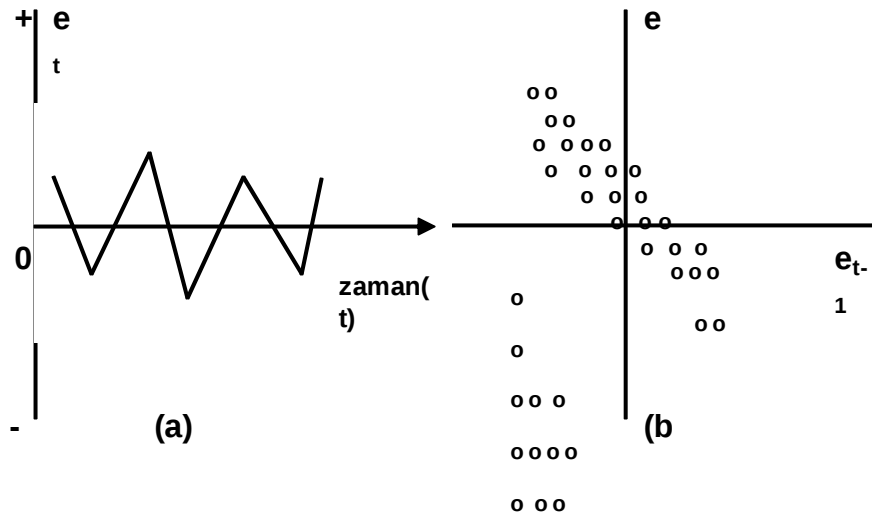
Bir modelde hata teriminin otokorelasyonlu olup olmadığını tespit etmek için çeşitli yöntemler bulunmuştur. Burada bu yöntemlerden grafik yöntemi, sıra testi, Durbin-Watson testi, Von-Neumann oran testi incelenecektir.

A. Grafik Yöntemi

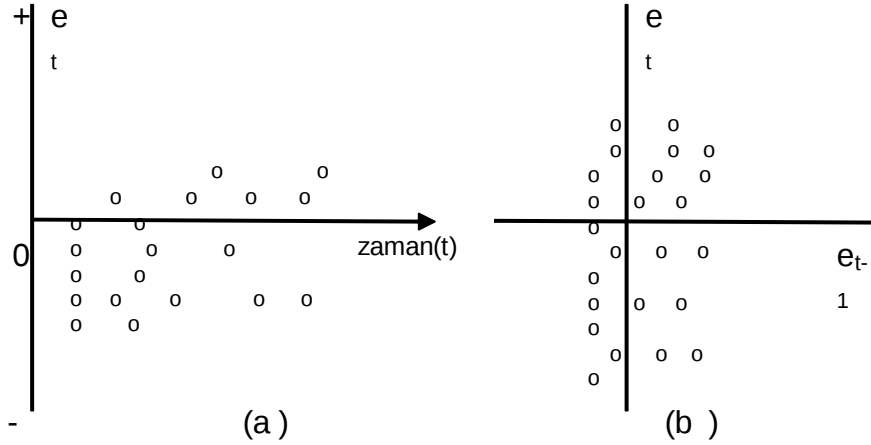
Otokorelasyonun söz konusu olup olmadığı örnek hata terimi e_t değerlerinden faydalanarak grafik yoluyla tespit edilebilir. Bunun için ya zaman ile e_t değerleri, ya da e_t ile e_{t-1} değerleri alınarak elde edilen grafiklerin durumu tetkik edilir. Aşağıda pozitif ve negatif otokorelasyon ve otokorelasyonun olmaması durumlarını gösteren bu grafikler yer almaktadır (Akkaya-Pazarlıoğlu, 2000: 448):



Şekil 1: Pozitif Otokorelasyon Durumu



Şekil 2: Negatif Otokorelasyon Durumu



Şekil 3: Otokorelasyon Olmaması Durumu

Çizilen şekillere bakıldığında Şekil-1 ve Şekil-2’de noktalar sistematik (düzenli) bir biçim göstermektedir. Şekil-3’te ise noktalar sistematik olmayan düzensiz bir görünüm arz etmektedir. Şekil-1(a)’da noktalar önce sürekli artış gösteriyor, sonra belli bir noktadan itibaren azalıp tekrar artıyor. Şekil-2(a)’da ise noktalar birden azalıp, yükseliyor. Bu sebepten Şekil-1(a)’da pozitif, Şekil-2(a)’da ise negatif otokorelasyon söz konusudur. Şekil-3(a)’da noktalar zaman eksenini etrafında paralel dağılmakta ve artan veya azalan bir seyirleri olmadığı için otokorelasyon söz konusu değildir (Akaya-Pazarlıoğlu, 2000: 449).

B.Sıra (Dizilim) Testi

Bazı modellerin hata terimleri incelendiğinde şu özellik dikkatimizi çekmektedir. Başlangıçta bazı hatalar eksidir, bir süre sonra artıya dönüşürler, son olarak tekrar eksiye dönüşürler. Eğer hatalar tümüyle rassal olsaydı böyle bir durum mümkün olabilir miydi? Sezgisel olarak bu pek mümkün görünmüyor. Bu sezginin doğru olup olmadığı sıra testi, bazen de Geary testi denen dağılımdan bağımsız (nonparametrik) bir test ile anlaşılabilir (Gujarati, 2001: 419).

Sıra testi ile hata teriminin otokorelasyonlu olup olmadığını incelemek için şu dört aşama takip edilir (Akkaya-Pazarlıoğlu, 2000: 450):

1. H_0 : Hata terimleri birbirinden bağımsızdır, otokorelasyon yoktur.
 H_1 : Hata terimleri birbiriyle bağımlıdır, otokorelasyon vardır.
2. $n > 20$ olması halinde %1 veya %5 anlamlılıkla normal dağılım tablosu Z değerleri iki taraflı teste göre alınır $Z_{tab} = \pm 2.58$ veya ± 1.96

3. $n > 20$ olması halinde birbirini takip eden işaretlerin sayısı D , aşağıdaki ortalama ve varyanslı normal dağılmaktadır:

$$E(D) = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad \text{ve} \quad \sigma_D^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \quad (12)$$

Burada n_1 , pozitif işaretli hata terimi sayısı; n_2 , negatif işaretli hata terimi sayısı; $n = n_1 + n_2$ toplam gözlem sayısını göstermektedir. Bu aşamada $E(D) \pm Z_{\text{tab}}\sigma_D$ aralığı sınır değerleri hesaplanır.

4. Eğer D birbirini takip eden işaret sayısı, $E(D) \pm Z_{\text{tab}}\sigma_D$ aralığı içinde kalıyorsa, örneğin %5 önem düzeyinde;

$$E(D) - 1.96\sigma_D \leq D \leq E(D) + 1.96\sigma_D \text{ aralığı}$$

ise H_0 hipotezi kabul edilir.

C. Durbin-Watson Testi

Hata terimindeki otokorelasyonu belirlemenin diğer bir yöntemi ise Durbin Watson istatistiği olarak bilinen,

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (13)$$

istatistiğinin kullanımıdır (Draper ve Smith, 1981). Burada e_t , $t=1,2,\dots,n$ için en küçük kareler regresyonundan elde edilen t 'nci kalıntı değerlerini göstermektedir. Eşitlikte hesaplanan d değeri d_L ve d_U şeklinde iki kritik cetvel değeri ile karşılaştırılır. Bu karşılaştırma sonucunda;

$0 < d < d_L$ ise pozitif otokorelasyon vardır

$d_L \leq d \leq d_U$ ise karar verilmemektedir

$d_U < d < 4 - d_U$ ise otokorelasyon yoktur

$4 - d_U \leq d \leq 4 - d_L$ ise karar verilmemektedir

$4 - d_L < d < 4$ ise negatif otokorelasyon vardır.

D. Von-Neumann Oran Testi

Regresyon analizinde anakütleyi temsil edebilecek örneğin mümkün olduğunca büyük olması tercih edilir. Fakat bazen az birimle de analiz yapma zorunluluğu ortaya çıkabilir. Bu durumda büyük birimler için kullanılan ($n > 15$) Durbin Watson testi uygulanamaz. Durbin-Watson testi yerine küçük birimler ($n < 15$) için geliştirilen Von-Neumann oran testi uygulanır. Von-Neumann test istatistiği,

$$V = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2 / (n - k - 1)}{\sum_{t=1}^n (e_t - \bar{e})^2 / (n - k)} \quad (14)$$

biçimindedir. Burada k analizdeki değişken sayısı $\bar{e}$ ise tanım gereği değeri sıfırdır. Hesaplanan V oranı tablo değerleri ile karşılaştırılır. Hesaplanan V değeri V_α ve V_α^* tablo değerleri arasında kalıyorsa otokorelasyonun olmadığına, V_α değerinden küçük ise pozitif otokorelasyon, V_α^* değerinden büyük ise negatif otokorelasyon olduğuna karar verilir (Orhunbilge, 1996: 179).

E.Ki-Kare Otokorelasyon Testi

Yalnızca DW-d testinin kararsız olduğu durumda kullanılır. Doğrudan 1. derece otokorelasyon araştırılmasında kullanılamaz. Parametrik olmayan bir testtir. (Yani DW-d testinde olduğu gibi n ve k gibi parametre değerlerine ihtiyaç duymaz)

Tablo 1: Otokorelasyon Testi İçin Hataları Karşılaştırma Tablosu

		$\hat{u}_t$		
		+	-	
$\hat{u}_{t-1}$	+	A	b	a+b
	-	C	d	c+d
		a+c	b+d	n-1

Hata terimleri, bir önceki hata terimleri ile karşılaştırılarak Tablo-1 gibi bir işaret tablosu oluşturulur. Örneğin hata terimi + iken bir önceki hata terimi – ise tablodaki c hücrelerine bir puan eklenir. Tüm liste geçildikten sonra işaret bileşimlerinin toplamı a, b, c, d hücrelerine doldurulmuş olur. Ki-kare değeri de bu hücrelerden faydalanarak,

$$K^2 = \frac{(ad - bc)^2 \cdot (n - 1)}{(a + b) \cdot (c + d) \cdot (a + c) \cdot (b + d)} \quad (15)$$

formülü yardımıyla hesaplanır ve 1 serbestlik dereceli tablo değeri ile karşılaştırılır. Eğer hesaplanan ki-kare değeri tablo değerden ($K_{hes}^2 < K_{tab}^2$) küçük ise H_0 hipotezi kabul edilir. Yani bu durumda otokorelasyon yoktur (http://bilgiteknoloji.net/ekonometri/ekotest/e03_otokorelasyon.asp).

V. Otokorelasyon İçin Çözümler

A. Hataları Ardışık Bağımlı Olan Regresyonun Tahmini

Regresyon hatalarının ardışık bağımlı olabileceği görülünce en küçük karelere tahminleriyle onlara dayanan çıkarsamalar güvenilmez hale gelir. Böyle durumlarda başka bir tahmin edici kullanılması tercih edilir. Kullanılacak tahmin sürecini başlatmak için,

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t \quad (16)$$

regresyon modeli yazılsın. Böylece $(t-1)$ döneminde,

$$Y_{t-1} = \alpha + \beta_1 X_{1,t-1} + \beta_2 X_{2,t-1} + \dots + \beta_k X_{k,t-1} + \varepsilon_{t-1} \quad (17)$$

modeli bulunur. Model, ardışık hatalar arasındaki korelasyon ρ ile çarpılırsa,

$$\rho Y_{t-1} = \alpha\rho + \beta_1 \rho X_{1,t-1} + \beta_2 \rho X_{2,t-1} + \dots + \beta_k \rho X_{k,t-1} + \rho \varepsilon_{t-1} \quad (18)$$

eşitliği yazılır. Eşitlik (1)'den eşitlik (3) çıkarıldığında,

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta_1 (X_{1t} - \rho X_{1,t-1}) + \beta_2 (X_{2t} - \rho X_{2,t-1}) + \dots + \beta_k (X_{kt} - \rho X_{k,t-1}) + u_t \quad (19)$$

eşitliği bulunur. Burada,

$$u_t = \varepsilon_t - \rho \varepsilon_{t-1} \quad (20)$$

ve u_t rassal değişkeni sabit varyanslı, ardışık bağımsızdır. Dolayısıyla eşitlik (19)'a $(Y_t - \rho Y_{t-1})$ bağımlı değişkenini $(X_{1t} - \rho X_{1,t-1})$, $(X_{2t} - \rho X_{2,t-1})$, ..., $(X_{kt} - \rho X_{k,t-1})$, gibi k tane bağımsız değişkenle ilişkilendiren bir regresyon modeli olarak bakılabilir. Bu modelin anakütle katsayıları, sabit terim α yerine $\alpha(1 - \rho)$ konulması dışında (16)'daki model ile aynıdır. Ancak (16) modeli ile (19) modeli arasında çok önemli bir fark vardır. İlkinde ε_t hata terimleri ardışık bağımlı, böylece modelin anakütle katsayılarının en küçük kareler tahminleri elverişsizdir. Ancak diğer modelin u_t hataları ardışık bağımlı olmadığından, alışıldık en küçük kareler çıkarsama süreçleri bu modele uygulandığında pekâlâ geçerlidir (Newbold, 2002: 653).

Bulunan yeni modelde hatalar hesaplanarak otokorelasyon testi yeniden uygulanır. Otokorelasyon kalmadığı anlaşılırsa işlem tamamlanmış olur. Aksi halde modeldeki hata terimleri otokorelasyondan arındırılınca kadar aynı aşamalar tekrar edilir (Orhunbilge, 1996: 180).

VI. Uygulama

Çalışmanın bu bölümünde, Amerikalı bilim adamı olan Robert J.Fisher'in "çok değişkenli doğrusal olmayan modellerde parametre sorunu" için ABD Tarım Bakanlığı'ndan derlemiş olduğu veriler, "hataları ardışık bağımlı (otokorelasyonlu) olan

regresyon modellerinin tahmin edilmesi” için kullanılmıştır. Gerekli hesaplamalar minitab istatistiksel yazılımı ile elde edilmiştir. Tablo-2’de verilen değişkenler için,

Y: Kişi başına tavuk eti tüketimi, libre(1 libre = 454 gram)

X1: Kişi başına reel harcanabilir gelir, \$

X2: Tavuk etinin reel perakende libre fiyatı, sent

X3: Domuz etinin reel perakende libre fiyatı, sent

X4: Tavuk eti yerine geçebilecek besinlerin bileşik reel libre fiyatı, sent olarak tanımlanmıştır.

Tablo 2: ABD’de 1960-1982 Yılları Arasında Tavuk Eti Talebi

Yıllar	Y	X1	X2	X3	X4
1960	27,8	397,5	42,2	50,7	65,8
1961	29,9	413,3	38,2	52,0	66,9
1962	29,8	439,2	40,3	54,0	67,8
1963	30,8	459,7	39,5	55,3	69,6
1964	31,2	492,9	37,3	54,7	68,7
1965	33,3	528,6	38,1	63,7	73,6
1966	35,6	560,3	39,3	69,8	76,3
1967	36,4	624,6	37,8	65,9	77,2
1968	36,7	666,4	38,4	64,5	78,1
1969	38,4	717,8	40,1	70,0	84,7
1970	40,4	768,2	38,6	73,2	93,3
1971	40,3	843,3	39,8	67,8	89,7
1972	41,8	911,6	39,7	79,1	100,7
1973	40,4	931,1	52,1	95,4	113,5
1974	40,7	1021,5	48,9	94,2	115,3
1975	40,1	1165,9	58,3	123,5	136,7
1976	42,7	1349,6	57,9	129,9	139,2
1977	44,1	1449,4	56,5	117,6	132,0
1978	46,7	1575,5	63,7	130,9	132,1
1979	50,6	1759,1	61,6	129,8	154,4
1980	50,1	1994,2	58,9	128,0	174,9
1981	51,7	2258,1	66,4	141,0	180,8
1982	52,9	2478,7	70,4	168,2	189,4

Kaynak: Gujarati N. Damodar, Temel Ekonometri, Çevirenler: Ümit Şenesen ve Gülay Göktürk Şenesen, Literatür Yayıncılık, İstanbul, s.228, 2001.

Tablo-2’de verilen değişkenler kullanılarak oluşturulan çoklu doğrusal regresyon modeli ve her parametrenin standart hatası,

$$Y = 37,9 + 0,0098X_1 - 0,539X_2 + 0,161X_3 + 0,028X_4$$

$$(4,311) (0,0038) (0,1614) (0,0724) (0,0742)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 2,041$ ve belirlilik katsayısı ise $R^2 = 93,7$ olarak bulunmuştur.

Regresyon denkleminin bir bütün olarak anlamlılığını test etmek için varyans analizinden yararlanılır. Varyans analizi için gerekli hesaplamalar Tablo-3’de verilmiştir.

Tablo 3: *Varyans Analizi Tablosu*

Değişim Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Ortalama Kare	F	p
Regresyon	4	1120,94	280,24	67,27	0,000
Hata	18	74,99	4,17		
Toplam	22	1195,93			

F testi için oluşturulacak hipotezler,

H_0 : Değişkenler arasında çoklu doğrusal ilişki yoktur.

H_1 : Değişkenler arasında çoklu doğrusal ilişki vardır.

biçimindedir.

%5 anlamlılık seviyesinde F tablo değeri $F_{4;18} = 2,93$ olarak bulunur. Tablo-2’de görüldüğü gibi hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında çoklu doğrusal ilişki vardır. Ayrıca %5 anlamlılık seviyesinde $p = 0,000$ değeri de anlamlı bulunmuştur.

Şimdi de bulunan model için otokorelasyon sorunu olup olmadığını araştıralım. Bunun için kurulacak hipotezler,

H_0 : $\rho = 0$ (otokorelasyon yoktur)

H_1 : $\rho \neq 0$ (otokorelasyon vardır)

biçimindedir. Tablo-4’de verilen değerler kullanıldığında Durbin-Watson test istatistiği,

Tablo 4: *Ardışık Hata Terimleri Değerleri*

e_t	e_{t-1}	e_t	e_{t-1}
-1,242875	-	0,837268	2,185759
-1,693341	-1,242875	2,947653	0,837268
-1,261684	-1,693341	0,782445	2,947653
-1,152869	-1,261684	-1,478243	0,782445
-2,141233	-1,152869	-1,988992	-1,478243
-1,545022	-2,141233	-0,136738	-1,988992
0,034369	-1,545022	2,967965	-0,136738
0,000358	0,034369	3,494993	2,967965
0,415572	0,000358	-1,041434	3,494993
1,459394	0,415572	-0,235437	-1,041434
1,402486	1,459394	-3,654699	-0,235437
2,185759	1,402486		

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{68,70756}{75,035806} = 0,92$$

olarak hesaplanır. $n = 23$ ve $k = 4$ olduğundan %5 önem seviyesinde Durbin-Watson tablosundan $d_L = 0,99$ ve $d_U = 1,79$ olarak bulunur. Burada $d = 0,92 < d_L = 0,99$ olduğundan otokorelasyon vardır. Yani hata terimleri birbirleriyle ardışık bağımlıdır.

Şimdi hatalardaki ardışık bağımlılığı hesaba katarak modelin anakütle katsayılarını bir kez daha tahmin edelim. Bunun için önce ρ değeri hesaplanır. Bu değer e_t hata terimleri ile e_{t-1} hata terimleri arasındaki korelasyon değeridir. Tablo-3'te verilen hata terimleri arasındaki korelasyon değeri $\rho = 0,494$ olarak hesaplanır. Bulunan bu değer hesaba katılarak yeni değişkenler oluşturulmalıdır. Bağımlı değişken $(Y_t - \rho Y_{t-1})$ olduğundan her bir değeri,

$$Y_2 - \rho Y_1 = 29,9 - (0,494)(27,8) = 16,1668$$

$$Y_3 - \rho Y_2 = 29,8 - (0,494)(29,9) = 15,0294$$

$$\dots\dots\dots Y_{23} - \rho Y_{22} = 52,9 - (0,494)(51,7) = 27,3602$$

olarak hesaplanır. Bağımsız değişkenler de aynı yöntemle bulunur. Örneğin $(X_{1t} - \rho X_{1,t-1})$ bağımsız değişkenin yeni değerleri,

$$X_{12} - \rho X_{1,1} = 413,3 - (0,494)(397,5) = 216,935$$

$$X_{13} - \rho X_{1,2} = 439,2 - (0,494)(413,3) = 235,0298$$

$$\dots\dots\dots X_{123} - \rho X_{1,22} = 2478,7 - (0,494)(2258,1) = 1363,199$$

olarak bulunur. Diğer bağımsız değişkenler de aynı şekilde bulunur. Bulunan bağımlı ve bağımsız yeni değişkenler Tablo-5'te verilmiştir.

Tablo 5: Dönüştürülmüş Değişken Değerleri

Yıllar	$Y_t - \rho Y_{t-1}$	$X_{1t} - \rho X_{1,t-1}$	$X_{2t} - \rho X_{2,t-1}$	$X_{3t} - \rho X_{3,t-1}$	$X_{4t} - \rho X_{4,t-1}$
1960	-	-	-	-	-
1961	16,1668	216,935	17,3532	26,9542	34,3948
1962	15,0294	235,0298	21,4292	28,312	34,7514
1963	16,0788	242,7352	19,5918	28,624	36,1068
1964	15,9848	265,8082	17,787	27,3818	34,3176
1965	17,8872	285,1074	19,6738	36,6782	39,6622
1966	19,1498	299,1716	20,4786	38,3322	39,9416
1967	18,8136	347,8118	18,3858	31,4188	39,5078
1968	18,7184	357,8476	19,7268	31,9454	39,9632
1969	20,2702	388,5984	21,1304	38,137	46,1186
1970	21,4304	413,6068	18,7906	38,62	51,4582
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
1982	27,3602	1363,199	37,5984	98,546	100,0848

Yapılan çoklu doğrusal regresyon denklemi ve her parametrenin standart hatası,

$$Y = 16,5 + 0,0101X_1 - 0,149X_2 + 0,0343X_3 + 0,01014X_4$$

$$(1,911) \quad (0,0032) \quad (0,1422) \quad (0,0645) \quad (0,0572)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 1,466$ ve belirlilik katsayısı ise $R^2 = 87,6$ olarak bulunmuştur.

Regresyon denkleminin bir bütün olarak anlamlılığını test etmek için varyans analizinden yararlanılır. Varyans analizi için gerekli hesaplamalar Tablo-6’te verilmiştir.

Tablo 6: Varyans Analizi Tablosu

Değişim Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Ortalama Kare	F	p
Regresyon	4	257,619	64,405	29,97	0,000
Hata	17	36,53	2,149		
Toplam	21	294,149			

F testi için oluşturulacak hipotezler,

H_0 : Değişkenler arasında çoklu doğrusal ilişki yoktur.

H_1 : Değişkenler arasında çoklu doğrusal ilişki vardır.

biçimindedir.

%5 anlamlılık seviyesinde F tablo değeri $F_{4;17} = 2,96$ olarak bulunur. Tablo-5’te görüldüğü gibi hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında çoklu doğrusal ilişki vardır. Ayrıca %5 anlamlılık seviyesinde $p = 0,000$ değeri de anlamlı bulunmuştur.

Şimdi de bulunan yeni model için otokorelasyon sorunu olup olmadığını araştırılır. Bunun için kurulacak hipotezler,

H_0 : $\rho = 0$ (otokorelasyon yoktur)

H_1 : $\rho \neq 0$ (otokorelasyon vardır)

biçimindedir. Durbin-Watson test istatistiği ise,

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{37,46257}{36,57879} = 1,02$$

olarak hesaplanır. $n = 22$ ve $k = 4$ olduğundan %5 önem seviyesinde Durbin-Watson tablosundan $d_L = 0,96$ ve $d_U = 1,80$ olarak bulunur. Burada $d_L < d < d_U$ olduğundan test sonuçsuz kalmıştır. Yani otokorelasyon olup olmadığı hakkında karar verilemez. Kararsızlık durumu için hata terimlerine ki-kare testi uygulanır.

Tablo 7: Modelin Ardışık Hata Terimleri Değerleri

e_t	e_{t-1}	işaret	e_t	e_{t-1}	işaret
-1,21053	-		1,244148	0,721252	++
-1,97354	-1,21053	--	0,659918	1,244148	++
-1,30013	-1,97354	--	-0,19071	0,659918	-+
-1,8354	-1,30013	--	-1,54265	-0,19071	--
-0,21964	-1,8354	--	-0,17418	-1,54265	--
0,961274	-0,21964	+-	0,286363	-0,17418	+-
0,063491	0,961274	++	1,893406	0,286363	++
0,044077	0,063491	++	2,479303	1,893406	++
1,219889	0,044077	++	-1,70617	2,479303	-+
1,708377	1,219889	++	-0,46902	-1,70617	--
0,721252	1,708377	++	-1,69693	-0,46902	--

Tablo 8: Hata Terimleri İçin İşaret Tablosu

		$\hat{u}_t$		
		+	-	
$\hat{u}_{t-1}$	+	a = 9	b = 2	a+b = 11
	-	c = 2	d = 8	c+d = 10
		a+c = 11	b+d = 10	n-1 = 21

Hata terimleri, bir önceki hata terimleri ile karşılaştırılarak Tablo-8'deki gibi bir işaret tablosu oluşturulur. Ki-kare değeri bu hücrelerden faydalanarak,

$$K^2 = \frac{(ad - bc)^2 \cdot (n-1)}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)} = \frac{(72 - 4)^2 (22 - 1)}{(11) \cdot (10) \cdot (11) \cdot (10)} = 8,0251$$

olarak hesaplanır. $K_{hes}^2 = 8,0251 > K_{tab}^2 = 3,841$ olduğundan H_0 hipotezi reddedilir. Yani hala otokorelasyon vardır.

Bu durumda Tablo-5'deki değişkenler esas alınarak bu değişkenlere yeniden dönüşüm yaptırılarak modelin anakütle katsayılarını bir kez daha tahmin edilmelidir. Bunun için önce ρ değeri hesaplanır. Bu değer e_t hata terimleri ile e_{t-1} hata terimleri arasındaki korelasyon değeridir ve $\rho = 0,456$ olarak bulunur. Bulunan bu değer hesaba katılarak yeni değişkenler oluşturulduğunda tahmin edilecek yeni model,

$$Y = 9,40 + 0,0096X_1 - 0,151X_2 + 0,0471X_3 - 0,0075X_4$$

$$(1,017) \quad (0,0030) \quad (0,1027) \quad (0,0528) \quad (0,0504)$$

biçimindedir. Regresyon denkleminin standart hatası $S = 1,297$ ve belirlilik katsayısı ise $R^2 = 74,4$ olarak bulunmuştur. Modelin anlamlılığı için hesaplanan F değeri 11,60 olarak ρ değeri ise 0,000 olarak bulunmuştur. %5 anlamlılık seviyesinde F tablo değeri $F_{4;16} = 3,01$ olarak bulunur. Hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler

arasında çoklu doğrusal ilişki vardır. Ayrıca %5 anlamlılık seviyesinde $p = 0,000$ değeri de anlamlı bulunmuştur.

Modelin hataları arasında otokorelasyon olup olmadığını araştırmak için hesaplanan Durbin-Watson test istatistiği değeri $d = 1,88$ olarak bulunmuştur. $n = 21$ ve $k = 4$ olduğundan %5 önem seviyesinde Durbin-Watson tablosundan $d_L = 0,93$ ve $d_U = 1,81$ olarak bulunur. Burada $d = 1,88 > d_U = 1,81$ olduğundan H_0 kabul edilir. Modelin hata terimleri arasında otokorelasyon olmadığına karar verilir. O halde hata terimleri arasında otokorelasyondan arındırılmış olan,

$$Y = 9,40 + 0,0096X_1 - 0,151X_2 + 0,0471X_3 - 0,0075X_4$$

$$(1,017) \quad (0,0030) \quad (0,1027) \quad (0,0528) \quad (0,0504)$$

modeli tahmin amacıyla kullanılabilir.

VII. Sonuç

Doğrusal regresyon analizinde hata terimi ile ilgili en önemli varsayımlardan biri hata terimleri arasında korelasyon olmamasıdır. Olması durumunda en küçük kareler tekniğinin bağımlı(Y) ve bağımsız(X) gözlemlere doğrudan doğruya uygulanması halinde bazı olumsuzluklar ortaya çıkabilmektedir. Bunlardan birincisi, modeldeki parametrelerin sapmasız tahminleri elde edilirse de bu tahminlerin örnek varyansları farklı bir tahmin metodu ile elde edileceklerinden daha büyük olabilir. İkincisi, en küçük kareler yönteminin regresyon katsayılarının örnek varyansları için uygulanması halinde bu varyansların önemli ölçüde düşük tahminleri elde edilir. Bu durum doğrusal model için elde edilen t ve F testlerinin güvenilirliklerini yitirebileceği anlamına gelmektedir.

Bu çalışmada bir bağımlı (Y) ve dört bağımsız (X_1, X_2, X_3, X_4) değişkenler alınarak oluşturulan çoklu doğrusal model ve her parametrenin standart sapması,

$$Y = 37,9 + 0,0098X_1 - 0,539X_2 + 0,161X_3 + 0,028X_4$$

$$(4,311) \quad (0,0038) \quad (0,1614) \quad (0,0724) \quad (0,0742)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 2,041$ ve belirlilik katsayısı ise $R^2 = 93,7$ olarak bulunmuştur. Bu model için yapılan F testi anlamlı bulunmuş olup bağımlı ve bağımsız değişkenler arasında doğrusal ilişki olduğuna karar verilmiştir.

Bulunan model için otokorelasyon sorunu olup olmadığını araştırmak için yapılan Durbin-Watson testi sonucu, modelde otokorelasyon sorunu olduğu yani hata terimleri birbirleriyle ardışık bağımlı olduğuna karar verilmiştir. Değişkenlere dönüşüm yaptırılarak oluşturulan yeni modelin anakütle katsayıları bir kez daha tahmin edilmiştir. Yapılan çoklu doğrusal regresyon denklemi ve her parametrenin standart hatası,

$$Y = 16,5 + 0,0101X_1 - 0,149X_2 + 0,0343X_3 + 0,01014X_4$$

$$(1,911) \quad (0,0032) \quad (0,1422) \quad (0,0645) \quad (0,0572)$$

olarak bulunmuştur. Ayrıca modelin standart hatası $S = 1,466$ ve belirleme katsayısı ise $R^2 = 87,6$ olarak bulunmuştur. Bu model için yapılan F testi anlamlı bulunmuş olup bağımlı ve bağımsız değişkenler arasında doğrusal ilişki olduğuna karar verilmiştir.

Bulunan yeni model için otokorelasyon sorunu olup olmadığını araştırmak için yapılan test sonucu, modelde hala otokorelasyon sorunu olduğu yani hata terimleri birbirleriyle ardışık bağımlı olduğuna karar verilmiştir.

Dönüştürülmüş değişkenler esas alınarak bu değişkenlere yeniden dönüşüm yaptırılarak modelin anakütle katsayıları bir kez daha tahmin edilmiştir. Tahmin edilen yeni model,

$$Y = 9,40 + 0,0096X_1 - 0,151X_2 + 0,0471X_3 - 0,0075X_4$$

$$(1,017) \quad (0,0030) \quad (0,1027) \quad (0,0528) \quad (0,0504)$$

olarak bulunmuştur. Modelin standart hatası $S = 1,297$ ve belirleme katsayısı ise $R^2 = 74,4$ olarak bulunmuştur. Bu model için yapılan F testi anlamlı bulunmuş olup bağımlı ve bağımsız değişkenler arasında doğrusal ilişki olduğuna karar verilmiştir.

Bulunan bu model için otokorelasyon sorunu olup olmadığını araştırmak için yapılan test sonucu, modelde otokorelasyon sorunu kalmadığı yani hata terimleri birbirleriyle ardışık bağımlı olmadığına karar verilmiştir. Bu durumda otokorelasyondan arındırılmış olan yeni model tahmin için kullanılabilir.

Kaynaklar

- Akkaya, Ş. ve Pazarlıoğlu, V. (2000), Ekonometri-I, Anadolu Matbaacılık, Dördüncü Baskı, İzmir.
- Draper, N.R. ve H. Smith, (1981), Applied Regression Analysis. 2.ed., New York: John Wiley & Sons, 1981.
- Gujarati, D. (2001), Temel Ekonometri, Çeviren: Ümit Şenesen, Literatür Yayıncılık, İstanbul.
- Gürsaka, N. (2002), Bilgisayar Uygulamalı İstatistik II, Alfa Yayınları, İstanbul.
- Johnston, J. (1981), Ekonometrik Metodlar, Çeviren: Yüksel İşyar ve Ergün Kip, Atatürk Üniversitesi Basımevi.
- Kalaycı Ş. (2006). SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri, Asil Yayınları, Ankara.
- Karagöz, M. (1999), İstatistik Yöntemleri, Dördüncü Baskı, Malatya.
- Koutsoyiannis A. (1989). Ekonometri Kuramı, Verso Yayıncılık, Çevirenler: Ümit Şenesen ve Gülay Göktürk Şenesen, Ankara.

- Kutlar A. (1998). Bilgisayar Uygulamalı Ekonometriye Giriş, Beta Yayıncılık, İstanbul.
- Newbold, P. (2002), İşletme ve İktisat için İstatistik, Çev: Ümit Şenesen ve Gülay Şenesen, Literatür Yayınları, Yayın No: 44, İstanbul.
- Orhunbilge, N. (1996), Uygulamalı Regresyon ve Korelasyon Analizi, İstanbul Üniversitesi İşletme Fakültesi Yayını, Yayın No:267, İstanbul.
- Şahinler, S. (2000), “En Küçük Kareler Yöntemi ile Doğrusal Regresyon Modeli Oluşturmanın Temel Prensipleri”, Mustafa Kemal Üniversitesi Ziraat Fakültesi Dergisi, Sayı:5, ss.57 73.
- Ünver, Ö. ve Gamgam, H. (1996), Uygulamalı İstatistik Yöntemler, İkinci Baskı, Siyasal Kitabevi, Ankara.
- Yıldız, N., Akbulut, Ö. ve Bircan, H. (1999), İstatistiğe Giriş, Aktif Yayınevi, İkinci Baskı, Erzurum.
- http://bilgiteknoloji.net/ekonometri/ekotest/e03_otokorelasyon.asp
- <http://web.sakarya.edu.tr/~adurmus/statistik/acikogretim/unite11.pdf>