Uygulamalı Ekonometri

prof.dr. bülent miran

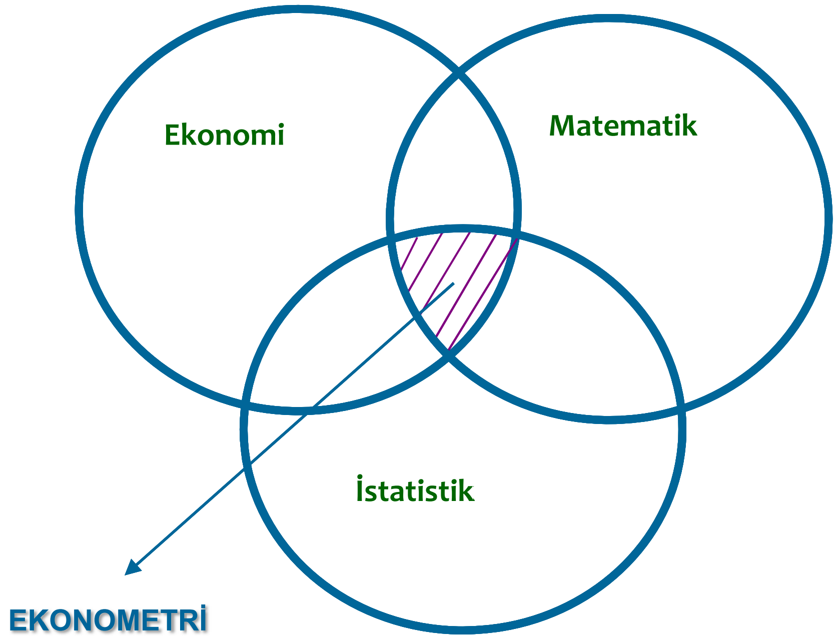
bornova 2011

1. GİRİŞ

1.1. Ekonometrinin Tanımı

Ekonometrinin kelime anlamı, ekonomik ölçümdür. Yaşanan ekonomik olayların ve işletmecilik olgularının, sayısal ölçümü ve analizi, ekonometrinin ilgi alanına girer. Ekonometri, ekonomik gerçekler ve ekonomi teorisi arasında, köprü görevi yapar. Bu bağlamda ekonometri; ekonomi teorisi, matematik ve istatistik bilimlerini biraraya getirir. Bir başka ifadeyle ekonometri, ekonomi teorisinin, matematik ve istatistik yöntemlerle kanıtlanması çabalarıdır.

Ekonometrinin tanımı basit bir Venn şemasıyla aşağıdaki gibi gösterilebilir:



Tipik bir ekonometrik çalışma, şu öncüllerle başlar:

- y ve x bir anakitleyi temsil eden iki değişkendir
- x'i kullanarak y'yi açıklamak istiyoruz
- x'deki değişmeler y'yi nasıl değiştirir

x'in y'yi açıkladığı bir modeli yazarken üç soruyla karşılaşırız:

1. y'yi tek başına x açıklayamayacağına göre, diğer faktörlerin y'ye etkisini ne yapacağız?
2. y ve x arasında nasıl bir fonksiyonel ilişki vardır?
3. y ve x arasındaki ilişkiyi yakalayabildiğimize nasıl emin olabileceğiz?

<u>Ekonometrinin Tarihçesi</u>

İstatistik ve matematik araçların ekonomik verilere ilk uygulanışı 17. yy'ın sonlarına rastlamaktadır. İlk deneysel talep modeli Charles Davenant tarafından 1699 yılında yayınlanmıştır. İlk modern istatistik çalışması ise, İtalyan istatistikçi Rudolfo Enini tarafından 1907 yılında gerçekleştirilmiştir. Ekonometrinin gelişmesinde asıl itici güç, 1930'da ekonometri topluluğunun kurulması ve Ocak 1933'de ekonometri dergisinin yayınlanmaya başlaması olmuştur.

<u>Ekonometrinin Kullanım Alanları</u>

Ekonometrinin üç ana kullanımı söz konusudur. Bunlar:

- Ekonomik gerçeğin tanımlanması;
- ekonomi teorisiyle ilgili hipotezlerin test edilmesi;
- ekonomik değişkenlerin geleceğinin kestirilmesidir;

Ekonomik Gerçeğin Tanımlanması

Ekonometrinin en basit kullanımı, tanımlama amacına dönüktür. Basit sembollerle ifade ettiğimiz denklemleri sayısallaştırmak için ekonometriden yararlanabiliriz. Örneğin bir mala ilişkin tüketici talebi, talep miktarı (Q) ile malın fiyatı (P), rakip mal fiyatı (P_s) ve harcanabilir gelir (Y_d) arasındaki ilişki olarak düşünülebilir. Genel olarak, talep miktarı ile malın fiyatı arasında negatif; talep miktarı ile rakip mal fiyatı ve harcanabilir gelir arasında ise pozitif ilişkiler vardır. Ekonometri; talep ile malın fiyatı, rakip mal fiyatı ve gelir arasındaki ilişkiyi geçmiş verilere dayalı olarak tahmin etmemizi sağlar. Bu ilişkiyi:

Denklem 1-1: $Q = f(P, P_s, Y_d)$

fonksiyonuyla ifade edersek, tahmin sonuçları:

Denklem 1-2: $Q = 20.7 - 0.85P + 0.15P_s + 0.19Y_d$

şeklinde olabilir. Bu tarz, fonksiyonu çok daha anlaşılabilir hale getirmiştir. Denklem 1-1, sadece fonksiyonel ilişki bulunduğunu gösterirken; Denklem 1-2, malın fiyatı bir birim arttığında, talebin 0.85 birim düşeceği ayrıntısını da verebilmektedir. 20.7, –0.85, 0.15 ve 0.19 sayısal değerleri, tahmin edilen regresyon katsayılarıdır. Ekonometriyi değerli kılan da, bu katsayıları tahmin etme gücüdür.

Ekonomi Teorisiyle İlgili Hipotezlerin Test Edilmesi

Ekonometrinin en önemli ve en yaygın kullanımı, ekonomi teorilerinin, sayısal kanıtlarla test edildiği, hipotez testleridir. Ekonomide çoğu kez teorik modeller oluşturulup, kanıtlarla test edilir. Burada hipotez testi, kaçınılmaz ve en uygun bilimsel yaklaşımdır. Örneğin Denklem 1-1'deki harcanabilir gelire ait katsayının, normal bir mal için pozitif olması gerektiği test edilebilir. Y_d'nin işareti pozitif görünmektedir ancak istatistiki açıdan önemli olup olmadığını hipotez testiyle belirlemeden önce böyle bir sonuca gidilmemesi gerekir. Katsayı, beklendiği gibi pozitif olsa bile, yeterince sıfırdan farklı olmayabilir ve bu durumda, katsayı sıfır kabul edilir.

Ekonomik Değişkenlerin Geleceğinin Kestirilmesi

Ekonometrinin en zor kullanımı, geçmişte ne olduğuna bakarak, gelecek 3 ayda, 6 ayda veya yılda muhtemelen ne olacağının kestirilmesi sırasında ortaya çıkar. Örneğin ekonomistler ekonometrik modelleri, satış, kâr, GSMH ve enflasyonu kestirmek amacıyla kullanmak isterler. Bu kestirimlerin doğruluk dereceleri, geçmişin geleceğe ne denli rehberlik edeceğine bağlıdır. İşletme yöneticileri ve politikacılar, ekonometrik modelleri, geleceğe dönük kararlarında kullanırlar. Verecekleri kararların yanlışlığı, ağır sonuçlara yol açacağından, ekonometrinin gücünü kararlarına destek olarak katmak isterler.

1.2. Alternatif Ekonometrik Yaklaşımlar

Sayısal çalışmalarda kullanılabilecek çok sayıda yaklaşım bulunmaktadır. Fizik, biyoloji, psikoloji gibi bilimlerde de, ekonomi ve işletmeciliktekine benzer sayısal problemlerle karşılaşılmaktadır. Ancak bu bilimlerin problem yapıları birbirinden farklıdır. Ekonomik verilerin belli özelliklere sahip olması gerektiği kabul edilir. Bu nedenle, sadece ekonomik verilerin analizine özgü ekonometrik yöntemlere ihtiyaç duyulur.

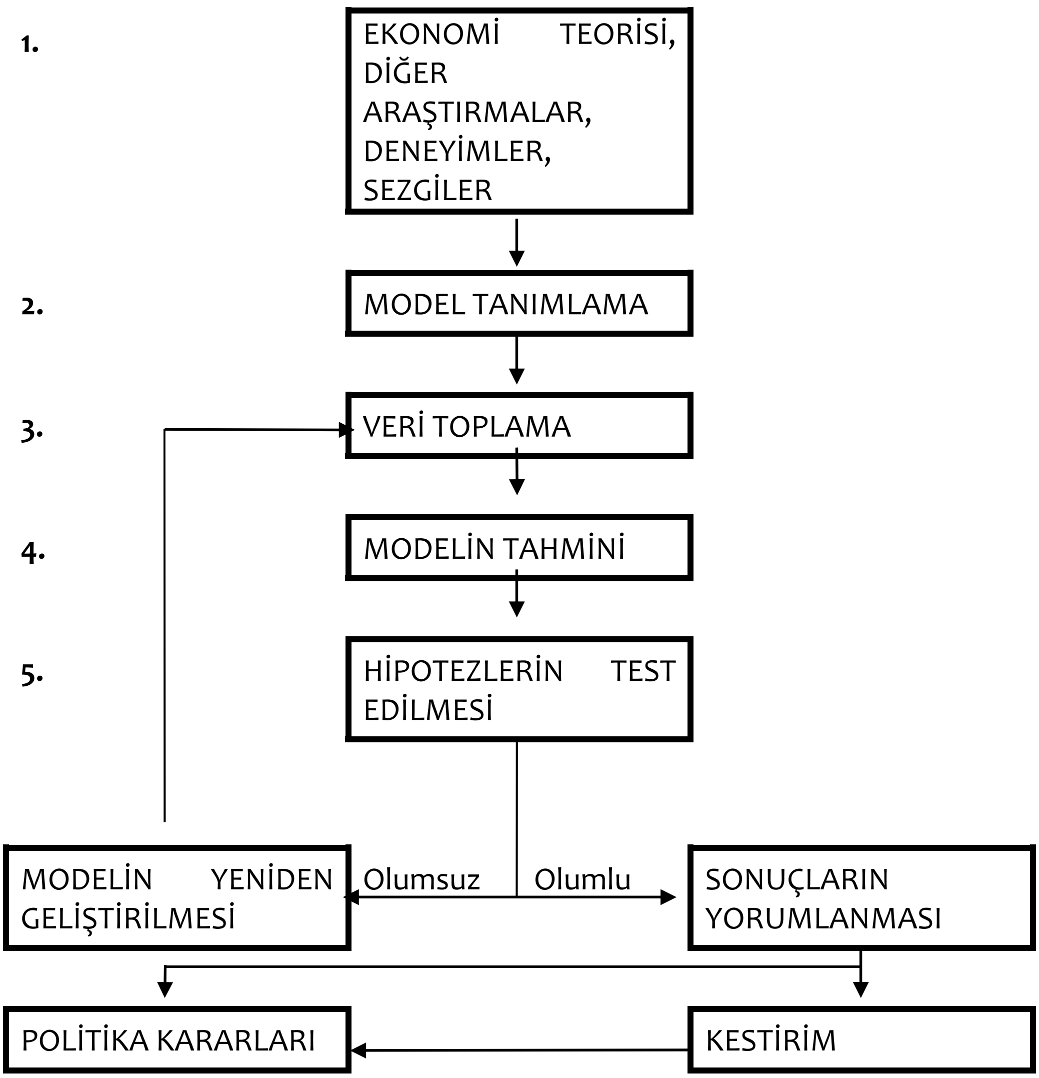
Herhangi bir sayısal araştırmada aşağıdaki aşamalardan geçilir:

1. Araştırılacak model veya ilişkilerin tanımlanması
2. Modelleri sayısallaştırmaya yarayacak verilerin toplanması
3. Modellerin sayısallaştırılması

1. ve 2. aşamalar, tüm araştırmalarda birbirine benzer. Ancak, modellerin sayısallaştırılması, disiplinlere göre büyük ölçüde farklılık gösterir. Ekonomik modellerin sayısallaştırılmasında kullanılan ekonometrik yöntemler de kendi içinde ayrıca farklılıklar arzeder. Teoriye uygun en iyi yöntemin ve değişkenlerin seçilmesi, ekonometri sanatı olarak görülür. Aynı denklemin sayısallaştırılmasında, pek çok yaklaşımdan yararlanılabilir ve her yaklaşımdan belli sonuçlar alınabilir. Yaklaşımın seçimi, ekonometristin kendisine bırakılır. Ancak, ekonometristin doğru yaklaşımı seçtiğini kanıtlama zorunluluğu vardır.

1.3. Ekonometrik Araştırmanın Aşamaları

Tipik bir ekonometri araştırmasında aşağıdaki süreç izlenir:



1. *Ekonomi Teorisi, Deneyim, Diğer Araştirmalar ve Sezgiler*

   Ekonometri, ekonomi teorisinin kanıtlanmasını hedefler. Bu nedenle ilk referans, ekonomi teorisi olmalıdır. Araştırma konusuyla ilgili ekonomi teorisi derinliğine incelendikten sonra, daha önce yapılmış benzeri çalışmalar ve elde edilen sonuçlar irdelenmelidir. Araştırıcının ilgili konudaki deneyim ve birikimleri de teori ve önceki araştırmalara önemli katkılarda bulunabilecektir. Özellikle daha ayrıntılı ve yeni değişken tanımlamaları ve bunlara ilişkin beklentiler, araştırıcının sezgilerini gerektirir. Bu aşama, tahmin edilecek modelin doğruluğu konusunda rehber niteliğindedir.

2. *Model Tanımlama*

   Ekonometrik modelin tanımlanması (spesifikasyonu) aşağıdaki aşamalarla gerçekleştirilir:

   - Model Değişkenlerinin belirlenmesi

     Gerek ekonomi teorisinin gerektirdiği gerek diğer araştırmalarda kullanılan ve gerekse araştırıcının dahil etmek istediği tüm değişkenler listelenir.

   - Bağımlı değişken ve bağımsız değişken ayrımının yapılması

     Ekonometrik modelde hangi değişkenin, diğer değişken ya da değişkenlere bağlı olarak tahmin edileceği açıklığa kavuşturulur.

   - Model katsayılarının işaret ve büyüklüklerinin tartışılması

     Model tahmin edildikten sonra bağımsız değişkenlere ait katsayıların beklenen işaretleri önceden bilinmelidir. Ekonomi teorisi ve önceki araştırmalar bu konuda önemli ölçüde bilgi sağlayacaktır. Diğer değişkenlere ilişkin işaret beklentileri ise deneyim ve sezgiler yoluyla ortaya konulabilir.

   - Modelin matematiksel şeklinin belirlenmesi

     Ekonometrik modellemede kullanılabilecek farklı matematiksel formlar bulunmaktadır. İkinci veya üçüncü dereceden denklem tahminleri yapılabileceği gibi, logaritmik, yarı logaritmik formlar da dikkate alınabilir.

3. *Veri toplama*

   Modelde yer alan değişkenlere ait veriler, amaca uygun olarak; zaman serileri, kesit verileri ya da karma veri olarak toplanır.

4. *Modelin tahminlenmesi*

   Ekonometrik modelin bağımsız değişkenlerine ait katsayıların tahminlenmesinde çeşitli yöntemler kullanılabilir. En basit ve en yaygın tahminleme yöntemi olan, en küçük kareler yönteminden (EKK) yararlanılabileceği gibi amaca ve teoriye uygun olarak dolaylı en küçük kareler yöntemi (DEKK), 2 Aşamalı EKK ve doğrusal olmayan yöntemler de tahminleme de kullanılabilir.

5. *Hipotezlerin test edilmesi*

   Ekonometrik model tahmin edildikten sonra, değişkenlere ait katsayılan istatistiki açıdan anlamlı olup olmadıkları test edilir. Böylece katsayıların belli bir güven düzeyinde bağımsız değişkenlerin bağımlı değişkeni ne derece etkilediği belirlenebilir. Bu amaçla her katsayı için t-testi, modelin bütünü için F-testi yapılır. Bunlara ek olarak çeşitli regresyon hatalarına karşın; otokolerasyon, farklı Varyanslılık, çoklu bağlantı testleri gerçekleştirilebilir. Model testlere olumsuz tepkiler veriyorsa, model tanımlama aşamasına dönülmesi gerekebilir.

Tüm bu aşamalardan sonra ekonometrik modelden politika kararları, işletmeci tercihleri ve geleceğe dönük öngörümleme amacıyla yararlanılabilir.

1.4. Ekonomik ve Ekonometrik Modeller

Ekonomik model, ekonomideki bir sektörü veya ekonomik davranışı yaklaşık olarak tanımlayan varsayımlar dizisidir. Ekonometrik modelde ise:

- Ekonomik modelden çıkarılan davranış denklemleri kümesi vardır. Bu denklemler, gözlenen bazı değişkenleri ve hataları birlikte içerir.
- Gözlenen değişkenlere ait verilerde hata olup olmadığı irdelenir.
- Hataların olasılık dağılımı tanımlanmıştır.

Örneğin ekonometrik bir talep modeli için davranış denklemi:

$$q = \alpha + \beta p + u$$

şeklinde olsun. Burada q ve p gözlenen değişkenler; u ise, hata terimidir. u'nun olasılık dağılımı tanımlanmıştır. Farklı gözlemlerden elde edilen u değerleri, ortalaması sıfır ve varyansı σ^2 olan normal dağılış gösterir.

Bu tanımlamalardan sonra, talep teorisinin ya da diğer bir ifadeyle $\beta>0$ hipotezinin deneysel olarak test edilmesi gerekir. Kuşkusuz, talep fonksiyonunun geleceğe dönük tahminleme veya politika kararları için kullanılması da mümkündür.

Ekonometrik bir çalışmada ilk iş, ekonometrik modeli formüle etmektir. Modelin olabildiğince basit formüle edilmesi istenir. Basit modeli anlamak, anlatmak ve verilerle test etmek daha kolaydır; ancak aşırı basitleştirme ve varsayımların gerçek dışı olması gibi sorunları da beraberinde getirebilir.

Bir modeli, en basit haliyle başlatıp, adım adım kapsamlı bir model haline getirmek mümkündür. Tam tersi bir yol izleyip, genel veya kapsamlı bir model hazırlayıp, basite doğru gitme yolu da izlenebilir.

1.5. Hata teriminin önemi

Hata terimi modele alınmayan, ancak Y'yi etkileyebilecek tüm değişkenlerin temsilcisidir. O halde bu değişkenleri neden modele dahil etmiyoruz? Bunun pek çok nedeni vardır. Ancak biz bunlardan bazılarını açıklayacağız:

Teorideki eksiklik: Y'nin davranışını belirleyen teori genellikle eksiktir. Örneğin haftalık gelirin haftalık tüketim harcamasını etkilediği düşünülüyorsa, diğer değişkenlerin etkisi hakkında bilgi sahibi olamayız. Bu durumda ihmal ettiğimiz değişkenlerin etkisini hata terimi temsil edecektir.

Eksik veri: Tüm değişkenleri modele dahil etsek bile, bu değişkenleri sayısal olarak ölçemeyebiliriz. Örneğin tüketim fonksiyonunda gelirin yanında aile varlığını da dikkate almak gerekir. Ancak çoğu zaman aile varlığını elde edemeyiz. Bu durumda, varlık değişkeni teorik olarak çok önemli olmakla birlikte, model dışında bırakma zorunluluğu doğacaktır.

Temel değişkenler ve önemsiz değişkenler: Tüketim örneğimizde gelir değişkenine ek olarak çocuk sayısı, cinsiyet, eğitim ve coğrafi bölge değişkenlerinin de tüketim harcaması üzerinde etkili olduğunu düşünebiliriz. Ancak bu değişkenlerin ortak etkisinin düşük olduğunun varsayılması ya da bu değişkenlere ait verileri toplamanın çok maliyetli olması, model dışı bırakılmalarını gerektirebilir. Bu durumda bu değişkenlerin ortak etkisinin hata terimi içinde olduğu kabul edilir.

Davranış dışına çıkma: Olası tüm değişkenleri modele alsak bile, her bireyin tüketim harcaması davranışı, her zamanki tipik davranışın dışına çıkabilecektir. Bu sıradışı davranışları hata terimi içinde düşünmemiz gerekecektir.

Ölçüm hataları: Regresyon modeli, Y ve X değişkenlerinin doğru ölçüldüğünü kabul etse de uygulamada %100 doğru ölçümler elde etmek mümkün değildir. Ölçüm hatalarını hata terimi içine dahil etmemiz gerekecektir.

En az değişkenli model ilkesi: Modelimizi olabildiğince az değişkenle kurma eğilimi, diğer değişkenlerin etkisini hata terimine gönderme zorunluluğu doğuracaktır.

Yanlış fonksiyonel form: Teorik olarak gerekli tüm değişkenleri modele alsak bile, ilişkiyi doğru fonksiyonel formla tanımlayamama ihtimali, hata teriminin varlığını gerektirecektir. Örneğin ikinci dereceden terim gerektiren bir model yerine, doğrusal bir modelden yararlanmak, fonksiyonel form hatasına yol açacaktır. Bu yanlışlık, hata terimi içinde yer alacaktır.

1.6. Veri Tipleri

Uygulamalı araştırmalarda üç tip veri söz konusudur:

- Zaman serileri
- Kesit verileri
- Karma (panel) veri

Şimdi veri tiplerini ayrıntılı olarak inceleyebiliriz.

1.6.1. Zaman serileri

Birbirini izleyen periyodik dönemlere ait verilere, zaman serisi denir. Günlük, haftalık, aylık, üç aylık, altı aylık ya da yıllık veriler, zaman serilerine örnek olarak verilebilir. Zaman serilerinde hiç bir döneme ait veri, eksik olmamalıdır. Aşağıda aylık, üç aylık ve yıllık değişkenlere ait örnekler bulunmaktadır:

Aylık zaman serileri

Dönem	Tarımsal işgücü	Dönem	Tarımsal işgücü
1983.01	94341	1984.01	98463
1983.02	94399	1984.02	99104
1983.03	95023	1984.03	99898
1983.04	95655	1984.04	100437
1983.05	96032	1984.05	101567
1983.06	97836	1984.06	102932
1983.07	99144	1984.07	103536
1983.08	99179	1984.08	102982

1983.09	98825	1984.09	102247
1983.10	99252	1984.10	102994
1983.11	99866	1984.11	103019
1983.12	99852	1984.12	103037

Üçer aylık zaman serileri

Dönem	Fiyat	Talep	Gelir	Dönem	Fiyat	Talep	Gelir
1978.1	841	1317	1271	1982.1	480	943	1036
1978.2	957	1615	1295	1982.2	530	1175	1019
1978.3	999	1662	1313	1982.3	557	1269	1047
1978.4	960	1295	1150	1982.4	602	973	918
1979.1	894	1271	1289	1983.1	658	1102	1137
1979.2	851	1555	1245	1983.2	749	1344	1167
1979.3	863	1639	1270	1983.3	827	1641	1230
1979.4	878	1238	1103	1983.4	858	1225	1081
1980.1	792	1277	1273	1984.1	808	1429	1326
1980.2	589	1258	1031	1984.2	840	1699	1228
1980.3	657	1417	1143	1984.3	893	1749	1297
1980.4	699	1185	1101	1984.4	950	1117	1198
1981.1	675	1196	1181	1985.1	838	1242	1292
1981.2	652	1410	1116	1985.2	884	1684	1342
1981.3	628	1417	1190	1985.3	905	1764	1323
1981.4	529	919	1125	1985.4	909	1328	1274

Yıllık zaman serileri

Yıl	GELIR	TÜKETİM	Yıl	GELIR	TÜKETİM
1976	1562.2	1417.2	1991	2710.1	2448.4
1977	1653.5	1497	1992	2733.6	2447.1
1978	1734.3	1573.8	1993	2795.8	2476.9
1979	1811.4	1622.4	1994	2820.4	2503.7
1980	1886.8	1707.5	1995	2893.6	2619.4
1981	1947.4	1771.2	1996	3080.1	2746.1
1982	2025.3	1813.5	1997	3162.1	2865.8
1983	2099.9	1873.7	1998	3261.9	2969.1
1984	2186.2	1978.4	1999	3289.5	3052.2
1985	2334.1	2066.7	2000	3404.3	3162.4
1986	2317	2053.8	2001	3464.9	3223.3
1987	2355.4	2097.5	2002	3524.5	3272.6
1988	2440.9	2207.3	2003	3538.5	3259.4
1989	2512.6	2296.6	2004	3648.1	3349.5
1990	2638.4	2391.8	2005	3704.1	3458.7

1.6.2. Kesit verileri

Zamanın belli bir diliminde veya noktasında, bireylerden, hanehalklarından, firmalardan veya tarım işletmelerinden toplanan veriler, kesit verileridir. Anket yoluyla toplanan veriler, kesit verileridir. Nüfus sayımı buna iyi bir örnektir. İllere, coğrafi bölgelere, ülkelere göre belli bir zaman dilimi için toplanan veriler de kesit verileridir. Aşağıda farklı kesit verilerine ilişkin örnekler verilmektedir:

10 Firmaya ait kesit verileri

Firma no	Üretim	Kapasite
1	70	80
2	65	100
3	90	120
4	95	140
5	110	160
6	115	180
7	120	200
8	140	220
9	155	240
10	150	260

14 Öğrenciye ait kesit verileri

Öğrenci no	Vize	Final
1	10	30
2	30	20
3	70	80
4	100	70
5	90	90
6	50	60
7	40	50
8	80	100
9	20	10
10	60	40
11	70	30
12	75	80
13	85	40
14	15	55

14 Eve ait kesit verileri

Ev no	Fiyat	Alan (m^2)
1	199.9	106.5
2	228	125.4
3	235	130
4	285	157.7
5	239	160
6	293	175
7	285	180
8	365	187
9	295	193.5
10	290	194.8
11	385	225.4

12	505	260
13	425	280
14	415	300

20 İle ait kesit verileri

İl no	**Gini KATSAYISI**	**Gelir**	**İşsizlik oranı**
1	0.4759	2950	6.3
2	0.3939	670	9
3	0.3732	16340	2.9
4	0.4454	11780	2.3
5	0.2885	200	2.5
6	0.5245	720	2.9
7	0.596	2590	3.3
8	0.2069	2830	2.5
9	0.2741	18970	2.4
10	0.5788	2060	9.9
11	0.36	390	4.2
12	0.4607	1660	5.7
13	0.5046	870	12.7
14	0.205	22850	5.7
15	0.3674	400	4.9
16	0.5906	960	3.9
17	0.54	870	3.8
18	0.2334	2780	0.3
19	0.3046	370	6.6
20	0.3274	16050	2.9

1.6.3. Karma (panel) veri

Zaman serisi ve kesit verilerinin bir araya getirilmesiyle, karma veri elde edilir. Örneğin 1999-2004 yılları arasında bölgelere göre buğday verimleri, 1990-2005 yılları arasında firmalara göre süt üretim miktarları ve süt maliyetleri, karma verilere örnek olarak verilebilir. Aşağıda karma veri örnekleri bulunmaktadır:

1997-2001 arasında bölgelere göre süt üretimleri ve reel fiyatları

YIL	**Ege**		**Marmara**		**Akdeniz**		**İç Anadolu**	
	Üretim	*Fiyat*	*Üretim*	*Fiyat*	*Üretim*	*Fiyat*	*Üretim*	*Fiyat*
1997	100	1.0	120	0.90	110	0.85	105	0.95
1998	120	1.1	90	1.15	110	1	125	1.20
1999	110	1.2	80	1.25	120	1.2	95	1.05
2000	130	1.1	95	1.05	350	0.95	120	0.90
2001	140	1.25	120	1.25	150	1.25	110	1.25

1990-1997 arasında 7 ülkenin tüketici fiyatları indeksi

YIL	Kanada	Fransa	Almanya	İtalya	Japonya	İngiltere	ABD
1990	135.5	133	112.2	159.6	111.4	148.2	130.7
1991	143.1	137.2	116.3	169.8	115	156.9	136.2
1992	145.3	140.5	122.1	178.8	116.9	162.7	140.3
1993	147.9	143.5	127.6	186.4	118.4	165.3	144.5
1994	148.2	145.8	131.1	193.7	119.3	169.4	148.2
1995	151.4	148.4	133.5	204.1	119.1	175.1	152.4
1996	153.8	151.4	135.5	212	119.3	179.4	156.9
1997	156.3	153.2	137.8	215.7	121.3	185	160.5

1.7. Ekonometrik Modelleme Uygulamaları

Yaygın olarak kullanılan ekonometrik modeller aşağıda verilmiştir:

1. Mikro Ekonomik Modeller

a) Üretim Fonksiyonları

$Y = f(X_1, X_2, \dots , X_n)$

İkinci ve üçüncü dereceden üretim fonksiyonları:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3$

(Y: Üretim; X: Girdi)

Cobb Douglas üretim fonksiyonu:

$\ln Q = \alpha + \beta_1 \ln K + \beta_2 \ln L$

(Q: Üretim; K: Sermaye; L: İşgücü)

b) Maliyet Fonksiyonları

Toplam Maliyet Fonksiyonu:

$Y = f(X)$

(Y: Maliyet; X: Girdi)

İkinci ve üçüncü dereceden maliyet fonksiyonları:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3$

(Y: Maliyet; X: Girdi)

c) Hane Halkı Tüketim Fonksiyonları

$Y = f(X) = \beta_0 + \beta_1 X$

(Y: Tüketim harcamaları; X: Gelir)

d) Talep Fonksiyonları

$Q_d = f(P_1, P_2, P_3, Y, H)$

$Q_d = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \beta_3 P_3 + \beta_4 Y + \beta_3 H$

(Q_d: Talep; P_1: Malın fiyatı; P_2: İkame mal fiyatı;
P_3:Tamamlayıcı mal fiyatı; Y:Gelir; H: Hanehalkı sayısı

e) Arz Fonksiyonları

$Q_s = f(P_1, P_2, M, K, T)$

$Q_s = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \beta_3 M + \beta_4 K$

(Q_s: Arz; P_1: Malın fiyatı; P_2: İkame mal fiyatı; M: Ürün maliyeti; K:Kapasite; T:Teknoloji düzeyi)

f) Hedonik Modeller

$P_a = f(K, E, Q)$

$P_a = \beta_0 + \beta_1 K + \beta_2 E + \beta_3 Q$

(P_a: a Malının fiyatı; K: Kapasite; E: Nitel özellikleri; Q: Nicel özellikleri)

g) Firma Yatırım Fonksiyonu

$I_t = b_1 + b_2\Pi_t + b_3K_{t-1} + b_4r_t$

(I: Yatırım; Π: Kâr; K: Sermaye stoğu; r: Faiz)

2. Makro Ekonomik Modeller

a) Tüketim Fonksiyonu

$C_t = \beta_0 + \beta_1 Y_t$

C_t: t dönemindeki tüketim harcaması

Y_t: t dönemindeki gelir

b) Modern Tüketim Fonksiyonu

$C_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1}$

C_t: t dönemindeki tüketim harcaması

Y_t: t dönemindeki gelir

Y_{t-1}: Bir önceki döneme ait gelir

c) Toplam Yatırım Fonksiyonu

$I = f(r,Y)$

$I = \beta_0 + \beta_1 r + \beta_3 Y$

I: Yatırım

r: Faiz

Y: Gelir

d) Philips Eğrisi

$Y_t = b_1 + b_2X_2$

Y_t : enflasyon oranı

X_2: İşsizlik oranı

Model tanımlama uygulamaları

Aşağıda çeşitli araştırmalarda uygulanmış model denemeleri verilmiştir. Bağımsız değişkenlerin beklenen katsayılarına ait işaretler, köşeli parantez içindedir. Değişkenler ve işaretleri, sadece fikir verme amacına dönüktür.

1. Problem: Hanehalkı tüketim harcaması

Veri tipi: Kesit

Model 1

Bağımlı değişken: Aylık tüketim harcamaları (TL/ay)

Bağımsız değişken: Aylık hane geliri (TL/ay) **[+]**

Model 2

Bağımlı değişken: Aylık tüketim harcamaları (TL/ay)
Bağımsız değişken: Aylık hane geliri (TL/ay) [+]
Varlık [+]

2. Problem: Yıllık otomobil tamir-bakım masrafları
Veri tipi: Kesit
Model 1
Bağımlı değişken: Yıllık otomobil tamir-bakım masrafları (TL/yıl)
Bağımsız değişken: Otomobilin yaşı [+]
Model 2
Bağımlı değişken: Yıllık otomobil tamir-bakım masrafları (TL/yıl)
Bağımsız değişken: Otomobilin o ana kadar yaptığı yol (km) [+]
Model 3
Bağımlı değişken: Yıllık otomobil tamir-bakım masrafları (TL/yıl)
Bağımsız değişken: Otomobilin o ana kadar yaptığı yol (km) [+]
Bağımsız değişken: Otomobilin yaşı [+]

3. Problem: İllere göre bina sayısı
Veri tipi: Kesit
Model
Bağımlı değişken: İldeki bina sayısı
Bağımsız değişken: İldeki ortalama ev fiyatı (TL) [+]
Nüfus yoğunluğu (Kişi/km^2)[+]
Ortalama hanehalkı geliri (TL) [+]
İlin nüfus artış hızı (%) [+]
İlin işsizlik oranı (%) [-]

4. Problem: İllere göre göç oranı
Veri tipi: Kesit
Model
Bağımlı değişken: İle göç oranı
Bağımsız değişken: Hayat standardı indeksi [+]
İlde kişi başına gelir [+]
İl geliri / Ülke geliri [+]
İl istihdam oranı [+]
İlin işsizlik oranı [-]
İl eğitim indeksi [+]
İl sağlık indeksi [+]

5. Problem: İllere göre aylık toplu taşım aracıyla seyahat süresi
Veri tipi: Kesit

Model

Bağımlı değişken: İllere göre aylık toplu taşım aracıyla seyahat süresi (Saat/ay)

Bağımsız değişken: İlde ortalama bilet fiyatı (TL) [-]
Kişi başına gelir (TL) [-]
İl nüfusu [+]
İl yerleşim alanı (km^2) [+]

6. Problem: İllere göre çalışan kadın oranı

   Veri tipi: Kesit

   Model

   Bağımlı değişken: İllere göre çalışan kadın oranı (%)

   Bağımsız değişken: İlde ortalama kadın maaşı (TL/ay) [+]
   İlde ortalama erkek maaşı (TL/ay) [-]
   İlde üniversite mezunu kadın oranı (%) [+]
   İlde işsizlik oranı (%) [-]
   İlde boşanma oranı (%) [+]
   Şehirleşme oranı (%) [+]

7. Problem: Ücret

   Veri tipi: Kesit

   Model

   Bağımlı değişken: Ücret (TL/ay)

   Bağımsız değişken: Eğitim (yıl) [+]
   Deneyim (yıl) [+]
   Deneyim^2 [-]

8. Problem: Dayanıklı süt talebi

   Veri tipi: Kesit

   Model 1

   Bağımlı değişken: Kişi başına dayanıklı süt talebi (kg/ay)

   Bağımsız değişken: Dayanıklı süt fiyatı (TL/kg) [-]

   Model 2

   Bağımlı değişken: Dayanıklı süt talebi (kg/ay)

   Bağımsız değişken: Dayanıklı süt fiyatı (TL/kg) [-]
   Hanehalkı geliri (TL/ay) [+]

   Model 3

   Bağımlı değişken: Dayanıklı süt talebi (kg/ay)

   Bağımsız değişken: Dayanıklı süt fiyatı (TL/kg) [-]
   Hanehalkı geliri (TL/ay) [+]
   Sokak sütü fiyatı (TL/kg) [+]

9. Problem: Ev fiyatı (Hedonik model)
   Veri tipi: Kesit
   Model

Bağımlı değişken:	Ev fiyatı (TL)
Bağımsız değişken:	Evin kullanım alanı (m^2) [+]
	Oda sayısı (m^2) [+]
	Banyo sayısı [+]
	Kent merkezine uzaklık (km) [-]

10. Problem: Otomobil fiyatı (Hedonik model)
    Veri tipi: Kesit
    Model

Bağımlı değişken:	Otomobil fiyatı (TL)
Bağımsız değişken:	Motor hacmi (cm^3) [+]
	Beygir gücü [+]
	Tork [+]
	0-100 km hızlanma (sn) [-]
	Dizel olma durumu [+]
	Ek güvenlik sistemlerinin varlığı [+]

11. Problem: Süt arzı
    Veri tipi: Zaman serisi
    Model 1

Bağımlı değişken:	Süt arzı (ton/yıl)
Bağımsız değişken:	Süt fiyatı [+]
	Süt sığırı sayısı [+]
	Süt maliyeti [-]

12. Problem: Yıllık pamuk ekiliş alanı (dekar)
    Veri tipi: Zaman serisi
    Model 1

Bağımlı değişken:	Pamuk ekiliş alanı (dekar)
Bağımsız değişken:	Bir yıl önceki pamuk fiyatı [+]
	Bir yıl önceki pamuk ekiliş alanı [+]
	Bir yıl önceki rakip ürün fiyatı [-]

2. TEK AÇIKLAYICILI DOĞRUSAL REGRESYON

Ekonometride, daha önce tamamen teorik olarak tanımlanmış ekonomik ilişkilerin sayısal karşılıklarını tahmin etmek üzere regresyon analizi kullanılır. Örnek olarak bir ekonomist GSMH ile işsizlik düzeyi arasındaki ilişkiyi ya da arazi değeriyle arazi özellikleri arasındaki ilişkiyi modelleyebilir. Regresyon analizleri iki değişken arasında veya çoklu ilişkilerin modellenmesinde kullanılan istatistik yöntemlerden biridir.

2.1. Kesin (Deterministik) ve Olasılıklı Model

Regresyon analizinin ayrıntılarına inmeden önce, kesin ve olasılıklı modeller üzerinde duralım. Örneğin, arz teorisinde, fiyatın arzı ne şekilde etkilediği büyük önem taşır. Teoriye göre, arz, fiyatın bir fonksiyonudur. Böyle durumda ilk soru şu olmalıdır: "Bu iki değişken arasında *kesin* bir ilişkinin var olduğunu düşünebilir miyiz?". Bir başka ifadeyle "eğer belli bir fiyat düzeyi için arz düzeyini kesinlikle ifade etmenin mümkün olduğunu düşünebilir miyiz?". Çeşitli nedenlerle bunun mümkün olmadığını söyleyebiliriz. Arz miktarı, fiyat dışında pek çok değişkene bağlı olup, örneğin diğer malların fiyatları, girdi fiyatları, geleceğe ilişkin görüşler, teknoloji düzeyi gibi değişkenler de arz miktarını etkileyecektir. Modele çok sayıda değişken dahil edilse bile, yine de arz miktarını *kesinlikle* kestirmemiz mümkün değildir. Arz miktarını modellerken, açıklanamayan tesadüfi olgulara bağlı bir değişkenlik mutlaka bulunacaktır. Değişkenler arasında kesin bir ilişki olduğunu varsayan modeller, kesin (deterministic) modeller olarak adlandırılmaktadır. Örneğin arz miktarı y'nin, fiyat düzeyi *x*'in tam bir buçuk katı olduğuna inanıyorsak:

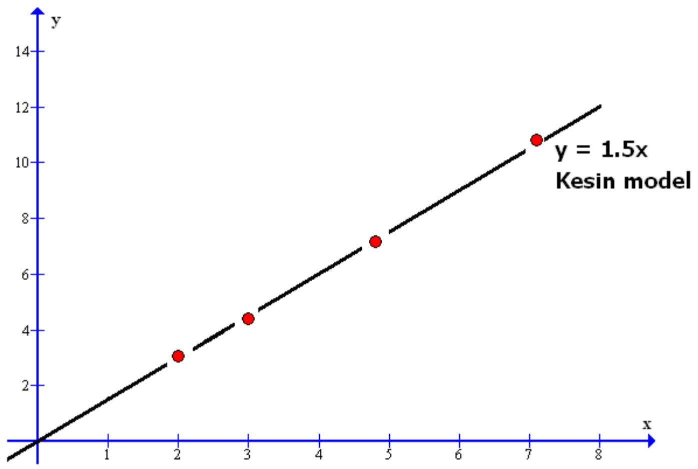
$$y=1.5x$$

yazabiliriz. Bu denklem, *x* ve *y* değişkenleri arasındaki kesin bir ilişkiyi temsil etmektedir. *x*'in değeri bilindiği zaman y'nin değerinin kesinlikle belirlenebileceğini ifade etmektedir. Bu kestirimde hata payı yoktur.

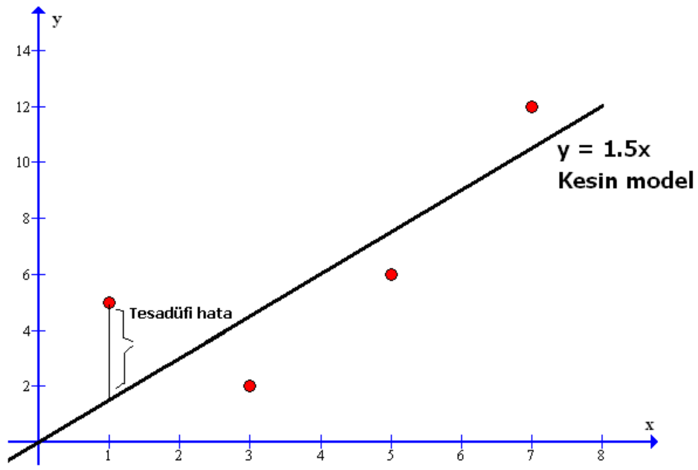
Diğer yandan, eğer arz miktarında –belki de önemli fakat ele alınmayan değişkenlerin veya tesadüfi olguların yol açtığı- açıklanmayan değişimlerin olacağına inanıyorsak kesin model yerine, tesadüfi hataya yer veren modeli kullanabiliriz. Olasılıklı model hem kesin ögeyi hem de tesadüfi hata ögesini içerir. Örneğin eğer arz miktarı y'nin, fiyat düzeyi *x* ile:

$$y = 1.5x + \text{Tesadüfi Hata}$$

şeklinde bir ilişkisi olduğunu düşünüyorsak, *x* ile *y* arasında *olasılıklı* bir ilişki olduğunu anlarız. Görüldüğü gibi, olasılıklı modelin kesin ögesi 1.5x'tir.



Kesin Model: *y=1.5x*



Olasılıklı Model: *y=1.5x + Tesadüfi hata*

Yukarıdaki grafikte, kesin model, fiyatı temsil eden *x*'in dört farklı değeri için olası arz düzeyleri göstermektedir. Kesin model hataya yer vermediğinden, tüm tepkiler kesinlikle doğru üzerine düşmek zorundadır. Olasılıklı modelde ise, *x*'in aynı değerlerine karşılık gelen olası tepkiler Y görülmektedir. Modelin kesin kısmının (doğrunun kendisi) aynı olduğuna dikkat edelim. Ancak şimdi tesadüfi hata teriminin dahil edilmesiyle arz miktarları bu doğrunun dışında bir noktaya düşebilmektedir. *x*'in belirli bir değeri için tepki süresinin tesadüfi olarak değişebileceğini bildiğimizden, olasılıklı modelin, kesin modelden daha gerçekçi bir model olduğunu söyleyebiliriz.

Olasılıklı Modelin Genel İfadesi:

y= Kesin öge + Tesadüfi hata

şeklindedir. Burada *y* kestirilecek değişkendir. Tesadüfi hatanın ortalama değeri her zaman sıfıra eşit olduğundan y'nin ortalama değeri EY'nin, modelin kesin ögesine eşit olduğunu kabul edebiliriz:

EY = Kesin öge

Olasılıklı modellerin en basiti, doğrusal modeldir. Bir veri kümesine doğrusal model uydurulması, regresyon analizlerine veya regresyon modellerine bir örnektir. Şimdi doğrusal regresyon modelini irdeleyelim.

Birinci dereceden doğrusal olasılıklı model:

$y = b_0 + b_1x + e$

olup, burada:

y= Bağımlı değişken (Modellenecek değişken)

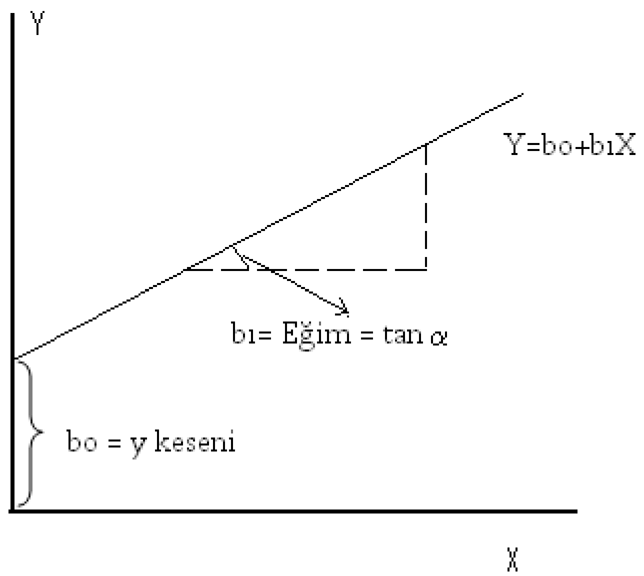
x=Bağımsız ya da açıklayıcı değişken.

e=Tesadüfi hata ögesi

b_0 = Tahminci; doğrunun y eksenini kesim noktası

b_1 = Tahminci; doğrunun eğimidir. Yani x'de meydana gelecek her 1 birimlik değişime karşılık, y'de meydana gelecek değişim miktarıdır.

Doğrusal modelin grafiği aşağıda sunulmuştur.



Modele ait doğrunun y eksenini kesim noktasını b_0 ile, eğimini b_1 ile göstereceğiz. b_0 ve b_1 örneğe ait tüm (*x*,*y*) gözlemlerini elde ettikten sonra bulabileceğimiz tahmincilerdir. b_0 bazı doğrusal modellerde, daha modellemenin başında sıfır olarak kabul edilebilmektedir. Ancak b_0 böylesi bir uygulamanın regresyon analizi açısından bazı sorunları da beraberinde getirebileceğini belirtmemiz gerekiyor.

2.2. Regresyon Modellerinin Tahminlenmesi

Regresyon modelinin tahminlenmesini, bir örnek üzerinden ele alabiliriz. Bir malın arzı ile o malın fiyatı arasındaki ilişkiyi belirlemek üzere beş yıllık bir veri setine sahip olduğumuzu varsayalım. Bu ilk örnekte aritmetik işlem karmaşasından kaçınmak için, gözlem sayısı az ve ölçümler basit tutulmuştur.

Yıl	Arz Miktarı (10000 ton/yıl) Y	Fiyat (TL/kg) X
1	1	1
2	1	2
3	2	3

4	2	4
5	4	5

Şimdi adım adım modelleme sürecini gerçekleştirelim:

1. Adım: Olasılıklı modelin kesin ögesine ilişkin varsayımda bulunma

Modelimiz, doğrusaldır. Arz miktarı, fiyatın doğrusal fonksiyonudur. Modelin ifadesi:

$$EY = b_0 + b_1 x$$

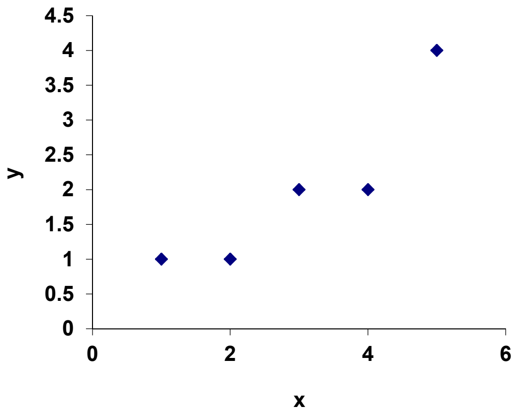
şeklindedir. Olasılıklı modelimiz ise:

$$y = b_0 + b_1 x + e$$

olarak ifade edilir.

2. Adım: Örnek verilerini kullanarak, modeldeki bilinmeyen tahmincilerin hesaplanması

y ile x arasında doğrusal bir ilişkinin mantıklı olup olmadığını belirlemek için örnek verilerinin serpilme çizelgesini hazırlamak yararlı olacaktır.



Serpilme çizelgesi, x arttıkça y'nin arttığı şeklindeki genel bir eğilimi ortaya koymaktadır. Bir doğrunun bir veri setine ne denli iyi uyduğunu anlamanın yollarından biri, veri noktalarının doğrudan sapma miktarını (tesadüfi hata) dikkate almaktır. Kuşkusuz, grafik üzerindeki noktalara en yakın geçen doğru, en az toplam tesadüfi hataya yolaçacaktır. Bir başka deyişle, verilerin tam ortasından geçen bir doğru belirlersek, en iyi doğruya ulaşmış oluruz. Hatırlanacağı üzere, bir veri setindeki değerlerin, veri setinin tam ortasında yer alan aritmetik ortalamadan farklarının toplamı sıfırdı. Bu nedenle, grafik üzerinde alabileceğimiz pek çok doğru için, toplam hata hep sıfır olacaktır.

Hataların negatif ve pozitif değerler almasıyla, toplam hata sıfırı verecektir. Olasılıklı modelimiz:

$$y = b_0 + b_1 x + e$$

şeklindeydi. Denklemi hata terimi için çözersek:

$$e = y - b_0 - b_1 x$$

olur. Gözlem sayısı kadar hata terimi olacağından:

$e_i = y_i - b_0 - b_1 x_i$

ve

$$\sum_{i=1}^{n} e_i = 0$$

sonucuna ulaşırız.

Hata kareleri toplamını aldığımızda ise, olasılıklı bir model için çok büyük bir ihtimalle sıfırdan büyük bir sonuca ulaşacağımız açıktır. Sıfır olabilmesi için, tüm gözlemlerden geçen bir doğru belirlememiz gerekir. Kuşkusuz bu, olasılıklı değil, kesin bir modeldir.

$$\sum_{i=1}^{n} e_i^2 >= 0$$

Gerçekten de, serpilme çizelgesinde gözlemlere en yakın noktalardan geçen doğruyu belirlemek için, hata kareleri toplamının en küçük olması istenir. Bu nedenle, regresyon modeline ait doğrunun tesbitinde kullanılan yönteme, *en küçük kareler (EKK)* yöntemi denmektedir.

Hata kareleri toplamını en küçük kılan tek bir doğru vardır. Bu doğruya en küçük kareler doğrusu, regresyon doğrusu veya en küçük kareler kestirim denklemi denir.

(x_1, y_1), (x_2, y_2), ...,(x_n, y_n) gibi veri çiftlerinin oluşturduğu n gözlem noktası için en küçük kareler doğrusunu bulmak istediğimizi varsayalım. Yukarıdaki örneğimizde, n=5 gözlem noktası (1,1), (2,1), (3,2), (4,2) ve (5,4) şeklindeydi. y'nin x'e bağlı fonksiyonu:

$y = b_0 + b_1x + e$

ise, x'in herhangi bir değeri için y'nin ortalama (veya beklenen) değeri:

$EY= b_0 + b_1x$

ve uydurulacak doğru:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x$$

$\hat{y}$, y şapka şeklinde okunur ve y'nin tahmincisidir. Nitekim $\hat{y}$, y'nin ortalama değerinin $[E_Y]$ tahmincisi ve y'nin gelecekteki bir değerinin kestirimcisidir.

Sözgelimi (x_i, y_i) gibi bir gözlem noktası için, y'nin gözlenen değeri y_i'dir. $\hat{y}$ ise:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$$

denkleminde x_i konulmak suretiyle hesaplanır. y'nin i'inci değerinin, tahmin edilen y değerinden sapması:

$$(y_i - \hat{y}_i) = [y_i - (\hat{b}_0 + \hat{b}_1 x_i)]$$

ise, n gözlem için gerçek gözlem değerleriyle tahmin değerleri arasındaki farkların karelerinin toplamı, hata kareleri toplamını (HKT) verecektir:

$$HKT = \sum [y_i - (\hat{b}_0 + \hat{b}_1 x_i)]^2$$

Hata kareler toplamını en küçük kılan $\hat{b}_0$ ve $\hat{b}_1$ değerleri, anakitle parametrelerinin en küçük kareler tahmincileridir. Bunları hesaplamanın en kolay yolu, iki bilinmeyenli iki denklemden yararlanmaktır:

$\Sigma y = nb_0 + b_1\Sigma x$

$\Sigma xy = b_0\Sigma x + b_1\Sigma x^2$

Yukarıdaki denklem çifti, en küçük kareler yönteminin matematiksel bir süreç sonunda elde edilmiş halidir.

Verilerimizi kullanarak denklemleri çözebiliriz. Gerekli ön hesaplamalar:

y	x	y^2	x^2	xy
1	1	1	1	1
1	2	1	4	2
2	3	4	9	6
2	4	4	16	8
4	5	16	25	20
10	**15**	**26**	**55**	**37**

$\Sigma y=10$, $\Sigma x=15$, $\Sigma y^2=26$, $\Sigma x^2=55$, $\Sigma xy=37$

$\bar{y} = 2$, $\bar{x} = 3$

Şimdi denklemlerimizde yerine koyalım:

$10 = 5b_0 + 15b_1$

$37 = 15b_0 + 55b_1$

$-30 = -15b_0 - 45b_1$

$37 = 15b_0 + 55b_1$

$7 = 10\, b_1$

b_1=0.7

$10 = 5b_0 + 15b_1$

$10 = 5b_0 + 15(0.7)$

b_0 = -0.1

İki bilinmeyenli iki denklemi çözmek kolay olmakla birlikte, regresyon analizinin devamında daha fazla istatistik ve hesaplamaya ihtiyaç duyacağız. Bu nedenle hata kareler toplamını en küçük kılan $\hat{b}_0$ ve $\hat{b}_1$ değerlerini aşağıdaki formüllerle hesaplayacağız.

Eğim : $\hat{b}_1 = \dfrac{SS_{xy}}{SS_{xx}}$

y-keseni: $\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x}$

Formüllerde:

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$SS_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^{\,2} - \frac{(\sum x_i)^2}{n}$$

$\bar{x}$ = x'e ait gözlemlerin aritmetik ortalaması

$\bar{y}$ = y'ye ait gözlemlerin aritmetik ortalaması

n = Gözlem sayısı

Şimdi arz örneği için, doğrusal regresyon modelini yeni formüllerle çözelim.

$$SS_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 37 - \frac{(15)(10)}{5} = 7$$

$$SS_{xx} = \sum x_i^{\,2} - \frac{(\sum x_i)^2}{n} = 55 - \frac{15^2}{5} = 10$$

$$\hat{b}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{7}{10} = 0.7$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 2 - (0.7)(3) = -0.1$$

En küçük kareler doğrusu:

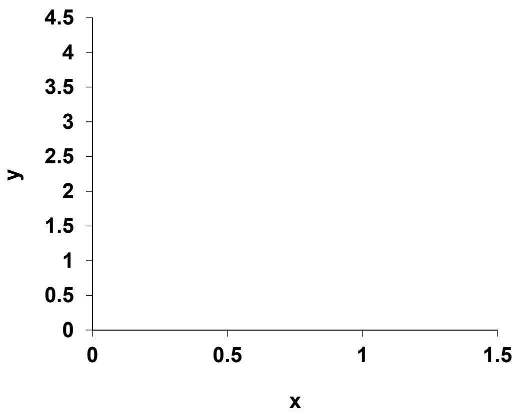
$$\hat{y} = -0.1 + 0.7x$$

Denklemdeki eğimin anlamı ($\hat{b}_1 = 0.7$): fiyatın bir birim artması, arz miktarını 0.7 birim artırmaktadır.

Bu doğrunun grafiğini çizebilmek için iki nokta yeterlidir. x'in en küçük ve en büyük değerleri için, tahmin denklemimizden elde edeceğimiz y tahmin değerlerini grafikte işaretleyip, bu noktaları birleştirdiğimizde, model doğrumuzu elde ederiz:

x=1, $\hat{y} = -0.1 + 0.7(1) = 0.6$ => (1, 0.6)

x=5, $\hat{y} = -0.1 + 0.7(5) = 3.4$ => (5, 3.4)



2.3. Değişkenliğin Bileşenlerine Ayrıştırılması

Doğru denklemimizi elde ettiğimize göre, hataları hesaplayabiliriz. Bunun için, her x değeri için y tahmin değerini elde edip, bunu gerçek y gözlem değerinden çıkarırız.

x	y	$\hat{y}$	e = y- $\hat{y}$	e² = (y- $\hat{y}$)²

1	1	0.6	0.4	0.16
2	1	1.3	-0.3	0.09
3	2	2.0	0	0.00
4	2	2.7	-0.7	0.49
5	4	3.4	0.6	0.36
15	**10**	**10**	**0.0**	**1.10**

Görüldüğü gibi, gerçek y gözlemleri ve y tahmin gözlemlerinin toplamı aynıdır. Yukarıda belirttiğimiz gibi, hata terimlerinin toplamı, sıfırdır. Hata kareleri toplamı ise:

HKT= Σe^2=1.10'dur.

HKT'nı, daha kolay hesaplayabilmek için:

$$HKT = \sum (y_i - \hat{y}_i)^2 = SS_{yy} - \hat{b}_1 SS_{xy}$$

formülünü kullanabiliriz. Burada:

$$SS_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

Problem değerlerimizi yerine koyarsak:

$$SS_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 26 - \frac{10^2}{5} = 6$$

$$HKT = SS_{yy} - \hat{b}_1 SS_{xy} = 6 - 0.7(7) = 1.10$$

2.4. Doğrusal Regresyon Modelinin Varsayımları

EKK ile tahmin edilen regresyon modeli, doğrusaldır. Doğrusal regresyon modelinin aşağıdaki varsayımlara sahip olduğu düşünülür:

1. Hata teriminin olasılık dağılımının ortalaması sıfırdır. Yani sonsuz sayıda denemeler yapıldığında her bir *x* bağımsız değişkeni için hataların ortalaması sıfırdır.
2. Bağımsız değişken *x*'in tüm setleri için hata teriminin olasılık dağılımının varyansı sabittir. Bizim doğrusal modelimiz için bu varsayım *x*'in her bir değeri için e'nin varyansının bir sabite, sözgelimi σ^2'ye eşit olduğu anlamına gelir.
3. e'nin olasılık dağılımı, normal dağılımdır.
4. Herhangi 2 farklı gözleme ait hatalar birbirinden.Yani y'nin herhangi bir değerine ilişkin hatanın, y'nin diğer değerlerine ilişkin hatalar üzerine etkisi yoktur.

σ^2'nin Tahmincisi

e'nin değişkenliği σ^2 ile ölçülür. σ^2 ne kadar büyükse model parametreleri b_0 ve b_1'in hataları daha büyük olacak ve herhangi bir *x* değeri için y'yi tahminde kullanıldığında, y'nın hatası da büyük olacaktır.

Doğrusal regresyon modelinin varyansı:

$$s^2 = \frac{HKT}{n-k}$$

formülüyle hesaplanır. Burada *n* gözlem sayısı; k tahminci sayısıdır. $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$ doğrusal denkleminde $\hat{b}_0$ ve $\hat{b}_1$ tahmincilerimiz olduğuna göre, tek açıklayıcılı doğrusal bir modelde, k=2'dir.

$$s^2 = \frac{HKT}{n-k} = \frac{1.10}{5-2} = 0.3667$$

Modelin standart hatası ise:

$$s = \sqrt{\frac{HKT}{n-k}} = 0.6055\text{'tir.}$$

2.5. Tahmincilerin Yarayışlılığı

Y-Keseninin Yarayışlılığı**:**

Doğrusal modeldeki b_0 sabiti, bazı durumlarda önemli ipuçları sunabilmektedir. Bu nedenle yarayışlılığının test edilmesinde fayda bulunmaktadır.

Çift yanlı test hipotezlerimiz:

H_0: $\hat{b}_0$ =0

H_1: $\hat{b}_0 \neq$0

Tek yanlı test hipotezlerimiz:

H_0: $\hat{b}_0$ = 0 H_0: $\hat{b}_0$ = 0

H_1: $\hat{b}_0$ < 0 H_1: $\hat{b}_0$ > 0

şeklindedir.

$\hat{b}_0$'ın standart hatası:

$$s_{\hat{b}_0} = \sqrt{\frac{s^2 \sum x^2}{n.SS_{xx}}} = \sqrt{\frac{0.3667(55)}{5(10)}} = 0.6351$$

$\hat{b}_0$'ın t:

$$t_{\hat{b}_0} = \frac{\hat{b}_0}{s_{\hat{b}_0}} = -0.1 / 0.6351 = 0.157$$

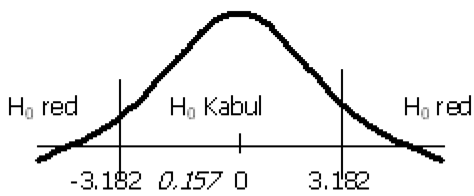
$\hat{b}_0$'ın istatistiki açıdan anlamlılığını belirlemek için, t tablosuna ihtiyaç duyacağız. Bunun için serbestlik derecesini:

sd=n-k

formülüyle hesaplayacağız. Burada n gözlem sayısı, k tahminci sayısıdır.

sd=5-2=3

α=0.05 için çiftyanlı testimizde $t_{0.025,\,3} = \pm 3.182$'dir.



$t_{\hat{b}_1}$, kabul bölgesinde olduğundan sıfır hipotezini kabul ederiz. Sabit ($\hat{b}_0$), sıfırdır.

*Eğimin Yarayışlılığı***:**

Arz örneğimizde, arzın fiyat düzeyiyle hiçbir ilgisinin bulunmadığını varsayalım. Doğrusal modelde bunun anlamı; gerçek eğim olarak ifade edilen b_1'in "0" olması demektir. Bu nedenle sıfır hipotezi, "doğrusal model *y*'nin tahmininde bir katkı sağlamıyor" şeklinde düzenlenmelidir. Alternatif hipotez ise "doğrusal modelin y'yi tahminde yararlı olduğu"dur. Buna göre çift yanlı testte hipotezlerimiz:

H_0: $\hat{b}_1 = 0$

H_1: $\hat{b}_1 \neq 0$

olarak hazırlanır. Tek yanlı testte ise:

H_0: $\hat{b}_1 = 0$ H_0: $\hat{b}_1 = 0$

H_1: $\hat{b}_1 < 0$ H_1: $\hat{b}_1 > 0$

şeklindedir.

Eğer veriler alternatif hipotezi doğrularsa *x*'in, y'nin tahmininde katkısı bulunduğu sonucuna varırız.

$\hat{b}_1$'in standart hatası

Doğrunun eğimine ait standart hata:

$$s_{\hat{b}_1} = \frac{s}{\sqrt{SS_{xx}}} = 0.6055 / 3.1623 = 0.1915$$

t:

$$t_{\hat{b}_1} = \frac{\hat{b}_1}{s_{\hat{b}_1}} = 0.7 / 0.1915 = 3.65$$

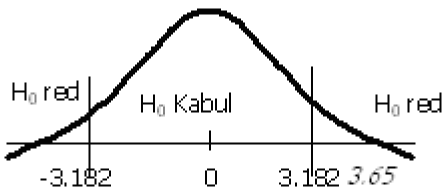
Tahmin ettiğimiz $\hat{b}_1$'nın istatistiki açıdan anlamlılığını test edebilmek için, yine t tablosuna, dolayısıyla serbestlik derecesine ve α'ya ihtiyaç duyacağız. Serbestlik derecesini:

sd=n-k

formülüyle hesaplayacağız. Burada n gözlem sayısı, k tahminci sayısıdır.

sd=5-2=3

α=0.05 için çiftyanlı testimizde $t_{0.025,3} = \pm 3.182$'dir.



$t_{\hat{b}_1}$, red bölgesine düşmektedir. Sıfır hipotezini reddederiz. Eğim ($\hat{b}_1$), sıfırdan farklı bir değerdir. Bulduğumuz değeri, anakitle tahmincisi olarak kullanabiliriz.

2.6. Modelin Yarayışlılığı

Son olarak, eğimin yarayışlılığını test ettik. Şimdi, modelin bütün olarak istatistiki yarayışlılığını irdeleyeceğiz. Bu amaçla F testinden yararlanacağız. Çift yanlı testte hipotezlerimiz:

H_0: $\hat{b}_1 = 0$

H_1: $\hat{b}_1 \neq 0$

Test istatistiğinin hesaplanması:

F testini daha önce varyans analizinde görmüştük. Regresyon analizinde de varyans analizine benzer şekilde değişkenliği bileşenlerine ayıracağız. Buna göre, toplam değişkenliğin (TD), regresyona bağlı değişkenlik (RBD) ve hataya bağlı değişkenlik (HBD) bileşenlerinden oluştuğunu kabul edeceğiz:

TD = RBD + HBD

Toplam değişkenlik:

$$TD = \sum (y_i - \bar{y})^2 = SS_{yy}$$

Regresyona bağlı değişkenlik:

$$RBD = \sum (\hat{y}_i - \bar{y})^2$$

Hataya bağlı değişkenlik:

$$HBD = \sum (y_i - \hat{y}_i)^2 = HKT$$

Tüm bunları Anova tablosu olarak gösterebiliriz:

Değişkenlik Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F
RBD	$\sum (y_i - \bar{y})^2$	k-1	a= $\sum (y_i - \bar{y})^2 / (k-1)$	a/b
HBD	$\sum (y_i - \hat{y}_i)^2$	n-k	b= $\sum (y_i - \hat{y}_i)^2 / (n-k)$	
TD	$\sum (y_i - \bar{y})^2$	n-1		

Tablodan anlaşılabileceği gibi, F istatistiği:

F=(Açıklanan Değişkenlik) / (Açıklanamayan Değişkenlik)

$$F_{hes} = \frac{RBD/(k-1)}{HBD/(n-k)}$$

olarak tanımlanmaktadır.

Test Kararı:

Eğer hesaplanan F değeri, tablo F değerinden büyükse, sıfır hipotezi reddedilir. Bu, regresyon modeli anakitleyi temsil etmektedir, güvenle kullanılabilir anlamına gelmektedir.

Şimdi arz örneğimiz için Anova tablosunu hazırlayıp, F testini gerçekleştirelim. Anova tablosunda, toplam değişkenliği SS_{yy} olarak, hataya bağlı değişkenliği *HKT* olarak hesaplamıştık. Bu değerleri eşitlikte yerine koyduğumuzda, *RBD*'yi de kolayca elde edebiliriz.

TD = RBD + HBD

6=RBD+1.10

RBD=4.9

Değişkenlik Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F
RBD	4.9	1	a=4.900	13.36
HBD	1.1	3	b=0.3667	
TD	6.0	4		

Test istatistiğimiz, F_{hes}=13.36'dır.

α=0.05

v_1=k-1=2-1=1

v_2=n-k=5-2=3

$F_{0.05,1,3}$=10.13

2.7. Korelasyon Katsayısı

İki değişken arasındaki ilişki, korelasyon olarak adlandırılır. Korelasyonun yön ve derecesi, korelasyon katsayısıyla belirlenir. Örneğin, içilen sigara sayısıyla akçiğer kanseri olgusunun yüksek korelasyon gösterdiği iddia edilir. Diğer bir yaygın kanı da suçluluk oranıyla işsizlik oranının ilişkili olduğudur.

Herhangi iki değişken arasındaki korelasyon katsayısını Pearson'un korelasyon katsayısı ile hesaplayacağız:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

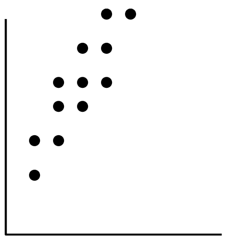
Burada r, korelasyon katsayısıdır ve *x* ve *y* gibi iki tesadüfi değişken arasındaki doğrusal ilişkinin gücünün bir ölçüsüdür. Dikkat edilirse, korelasyon katsayısı, doğru denklemini hesaplarken kullandığımız istatistiklerle hesaplanmaktadır. Korelasyon katsayısı r birimsizdir ve *x* ve *y*'nin birimi ne olursa olsun -1 ile + arasında bir değer verir.

$-1 \leq r \leq 1$

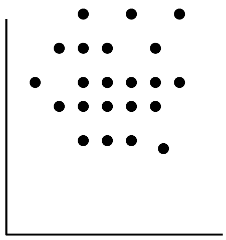
Sıfıra yakın ya da eşit bir r değeri, *x* ile *y* arasında çok az ya da hiç doğrusal ilişki olmadığını gösterir. Aksine r, +1 ve -1'e yaklaştıkça *x* ve *y* arasındaki doğrusal ilişki daha güçlenir. Eğer r = +1 ya da r = -1 ise tüm örnek gözlem noktaları, en küçük kareler doğrusunun tam üzerine düşer. r'nin pozitif değerleri y ile x arasında pozitif (aynı yönlü) doğru bir doğrusal ilişkiyi ifade eder. Yani *x* arttıkça *y* artar.

r'nin negatif değerleri de *y* ile *x* arasında bir doğrusal ilişkiyi ifade eder ancak bu kez *x* azaldıkça *y* azalır .

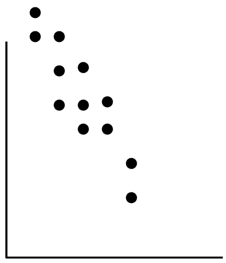
İki değişken arasındaki korelasyonla ilgili ön bilgiyi serpilme çizelgesinden elde edebiliriz. Aşağıdaki grafiklerde, bazı örneklere yer verilmiştir.



r, 1'e yakın



r, sıfıra yakın



r, -1'e yakın

r =1

r =-1

Şimdi arz örneğimiz için korelasyon katsayını hesaplayalım:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{7}{\sqrt{(10)(6)}} = 0.904$$

r pozitif ve 1'e yakın olduğundan, 5 gözlemden oluşan örnek için kandaki fiyat arttıkça, tepki süresinin artma eğilimi gösterdiğini ifade eder. En küçük kareler doğrusunun eğimiyle, korelasyon katsayısının işareti aynıdır.

Korelasyon Katsayısının Yarayışlılığı Testi:

Hesapladığımız doğrusal korelasyon katsayısının istatistiki açıdan anlamlı olma durumunu test edebiliriz. Bunun için t testinden yararlanacağız. Hipotezlerimiz:

H_0: $\rho = 0$

H_1: $\rho \neq 0$

şeklinde hazırlanır.

Test istatistiği

$$t_{hes} = \frac{r}{s_r}$$

Burada: $s_r = \sqrt{\frac{1-r^2}{n-k}}$

Daha sonra, test istatistiği, n-k serbestlik derecesi ve $\alpha/2$ anlamlılık düzeyinde test edilir.

Şimdi, arz örneğimizin korelasyon katsayısını test edelim. Korelasyn katsayımız, 0.904 idi:

Hipotezler:

H_0: $\rho = 0$

H_1: $\rho \neq 0$

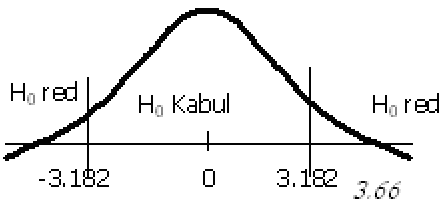
Test istatistiği

$$s_r = \sqrt{\frac{1-0.904^2}{5-2}} = 0.247$$

$$t_{hes} = \frac{0.904}{0.247} = 3.66$$

α=0.05 ve sd=5-2=3 için:

$t_{0.025,3} = \pm 3.182$



Test kararımız:

Test istatistiğimiz, red bölgesindedir. Sıfır hipotezini reddederiz. Korelasyon katsayısı, sıfırdan farklıdır ve güvenle kullanılabilir.

2.8. Model Seçim Kriterleri

2.8.1. Belirleme Katsayısı (R^2)

y'nin değerini kestirirken x'in katkısını ölçmenin bir başka yolu da, y'nin tahminindeki hataların ne kadarının x tarafından sağlanan bilgilerin yardımıyla azaltıldığıdır. Basit olarak:

$$R^2 = \frac{\text{Regresyon Tarafından Açıklanan Değişkenlik}}{\text{Toplam Değişkenlik}}$$

anlamına gelir.

Determinasyon katsayısı, korelasyon katsayısının karesidir. *x* ile *y* arasındaki doğrusal ilişkinin açıklayabildiği *y* etrafındaki toplam örnek değişkenliği oranını temsil eder ve:

$$R^2 = \frac{SS_{yy} - HKT}{SS_{yy}}$$

formülüyle hesaplanır. R^2 determinasyon katsayısıdır ve korelasyon katsayısının karesine eşittir. R^2'nin her zaman 0 ile 1 arasında değer alır.

Arz örneğimizin determinasyon katsayısı:

$$R^2 = \frac{SS_{yy} - HKT}{SS_{yy}} = \frac{6-1.1}{6} \cong 0.8167 \cong \%81.67$$

r'yi, 0.904 olarak hesaplamıştık. $r^2=(0.904)^2=0.82$'dir. Bunun anlamı, fiyat, arzı %82 oranında açıklamaktadır. Ya da, arzdaki değişkenliğin %82'si fiyatta değişmelerden kaynaklanmaktadır.

2.8.2. Düzeltilmiş R^2

Serbestlik derecesindeki değişmelerden etkileneceği için, hesaplanan determinasyon katsayısında düzeltme yapılır:

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

R_d^2, serbestlik derecesine göre düzeltilmiş belirleme katsayısıdır ve belirleme katsayısından biraz daha küçük bir değerdir. Yorumu, determinasyon katsayısıyla aynıdır. Arz örneği için R_d^2:

$$R_d^2 = 1-(1-R^2)\frac{n-1}{n-k} = 1-(1-0.8167)\frac{4}{3} = 0.7556 = \%75.56$$

2.9. Bilgi Kriterleri

Sol tarafı aynı formda olan regresyon modellerinin birbiriyle karşılaştırılmasında yaygın olarak kullanılan iki kriter bulunmaktadır. Bunlar:

1. Akaike bilgi kriteri (AIC)
2. Schwarz Bayesgil kriteri (SCHWARZ)

AIC hesaplamasında kullanılan formül:

2.10.Regresyon Analizlerinin Genel Sunumu

Regresyon analizi sürecini, arz örneğiyle gerçekleştirdik. Ancak analiz sonuçları dağınık haldedir. Elde ettiğimiz tüm istatistik ve hesaplamaları iki basit tabloda toplayabiliriz: (1) Tahminciler tablosu; (2) Varyans analizi tablosu.

Tahminciler tablosu:

	Tahminci	StHata	t	p
Sabit	$\hat{b}_0$	$s_{\hat{b}_0}$	$t_{\hat{b}_0}$	
x	$\hat{b}_1$	$s_{\hat{b}_1}$	$t_{\hat{b}_1}$	

s= R^2= R_d^2=

Anova tablosu

Değişkenlik Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
RBD	$\sum(\hat{y}_i-\bar{y})^2$	k-1	a= $\sum(y_i-\bar{y})^2/(k-1)$	a/b	
HBD	$\sum(y_i-\hat{y}_i)^2$	n-k	b= $\sum(y_i-\hat{y}_i)^2/(n-k)$		
TD	$\sum(y_i-\bar{y})^2$	n-1			

Gerek tahminciler gerekse Anova tablosundaki p olasılık değerleri, hipotez testleri bölümünde değindiğimiz istatistiklerdir. Bilgisayar çözümlerinde, doğrudan bu olasılık değerlerini kullanarak, tahmincilerin ve modelin hipotez testleri kolayca gerçekleştirilebilir. Burada α anlamlık düzeyi ve p olasılık değeri karşılaştırılır. Eğer $p<\alpha$ ise sıfır hipotezi reddedilir.

Şimdi, tahminciler ve Anova tablosuna, arz örneğimizin sonuçlarını yerleştirelim:

Tahminciler tablosu:

	Katsayı	*StHata*	*t*	*p*
Sabit	-0.1	0.6351	-0.16	0.885
X	0.7	0.1915	3.66	0.035

S = 0.6055 R^2 = % 81.67 R_d^2= % 75.56

Anova tablosu

Değişkenlik Kaynağı	*Kareler Toplamı*	*Sd*	*Kareler Ortalaması*	*F*	*p*
RBD	4.9	1	4.900	13.36	0.035
HBD	1.10	3	0.3667		
TD	6.0	4			

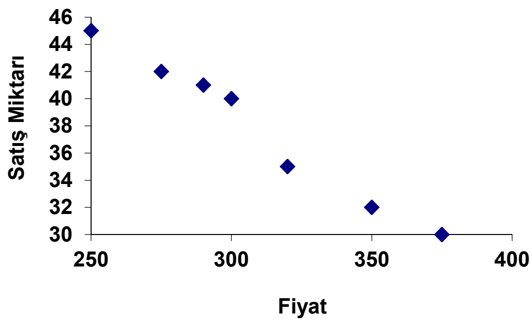
Örnek 2-1

Bir ürünün fiyatı ile satılan miktarına ilişkin veriler aşağıda sunulmuştur:

Fiyat (TL/kg) X	Satılan Miktar (100 adet) Y
250	45
275	42
290	41
300	40
320	35
350	32
375	30
400	25
410	23
430	20

Bu verilere uygun regresyon çözümlemesini yapalım:

Serpilme çizelgesi



Serpilme çizelgesi, satış miktarı ve fiyat arasında ters yönlü doğrusal bir ilişki olduğunu göstermektedir.

Malın satış miktarının, fiyatın doğrusal bir fonksiyonu olduğunu, ekonomi derslerinden hatırlıyoruz. Buna göre, satış miktarı bağımlı değişken Y, fiyatı ise bağımsız veya açıklayıcı değişkendir X.

x	y	x^2	y^2	xy
250	45	62500	2025	11250
275	42	75625	1764	11550
290	41	84100	1681	11890
300	40	90000	1600	12000
320	35	102400	1225	11200
350	32	122500	1024	11200
375	30	140625	900	11250
400	25	160000	625	10000
410	23	168100	529	9430
430	20	184900	400	8600
Σ 3400	333	1190750	11773	108370

$\Sigma y=333$, $\Sigma x=3400$, $\Sigma y^2=11773$, $\Sigma x^2=1190750$, $\Sigma xy=108370$

$\bar{y} = 33.3$, $\bar{x} = 340$

$$SS_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 108370 - \frac{(3400)(333)}{10} = -4850$$

$$SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 1190750 - \frac{3400^2}{10} = 34750$$

$$SS_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 11773 - \frac{333^2}{10} = 684.1$$

$$\hat{b}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-4850}{34750} = -0.139568$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 33.3 - (-0.139568)(340) = 80.753$$

Doğrusal modelimiz:

$$\hat{y} = 80.753 - 0.139568x$$

$$HKT = SS_{yy} - \hat{b}_1 SS_{xy} = 684.1 - (-0.139568)(-4850) = 7.1952$$

Varyans:

$$s^2 = \frac{HKT}{n-k} = \frac{7.1952}{10-2} = 0.90$$

Standart hata:

$$s = \sqrt{s^2} = \sqrt{0.90} = 0.948$$

Y-Keseninin Yarayışlılığı**:**

H₀: $\hat{b}_0$ =0

H₁: $\hat{b}_0 \neq$0

$\hat{b}_0$’ın standart hatası:

$$s_{\hat{b}_0} = \sqrt{\frac{s^2 \sum x^2}{n.SS_{xx}}} = \sqrt{\frac{0.90(1190750)}{10(34750)}} = 1.755$$

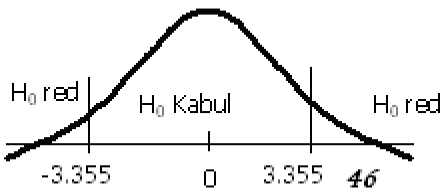
$\hat{b}_0$’ın t:

$$t_{\hat{b}_0} = \frac{\hat{b}_0}{s_{\hat{b}_0}}$$ =80.753 /1.755 = 46.0

sd=n-k

sd=5-2=3

α=0.01 için çiftyanlı testimizde $t_{0.005,\,8} = \pm3.355$’dir.



$t_{\hat{b}_1}$, red bölgesinde olduğundan sıfır hipotezini reddederiz. Sabit ($\hat{b}_0$), sıfırdan farklıdır. Anakitle tahmincisi olarak kullanılabilir.

Eğimin Yarayışlılığı**:**

H₀: $\hat{b}_1$ =0

H₁: $\hat{b}_1 \neq$0

$\hat{b}_1$’in standart hatası

$$s_{\hat{b}_1} = \frac{s}{\sqrt{SS_{xx}}}$$ = 0.948 /186.41= 0.00508

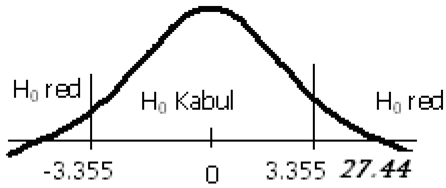
t:

$$t_{\hat{b}_1} = \frac{\hat{b}_1}{s_{\hat{b}_1}}$$ =-0.139568 / 0.00508 = 27.44

sd=n-k

sd=5-2=3

α=0.01 için çiftyanlı testimizde $t_{0.005,\,3} = \pm3.355$’dir.



$t_{\hat{b}_1}$, red bölgesine düşmektedir. Sıfır hipotezini reddederiz. Eğim ($\hat{b}_1$), sıfırdan farklı bir değerdir. Bulduğumuz değeri, anakitle tahmincisi olarak kullanılabilir.

Modelin Yarayışlılığı:

H_0: $\hat{b}_1$=0

H_1: $\hat{b}_1 \neq$0

Test istatistiğimiz:

Toplam değişkenlik:

$$TD = \sum (y_i - \bar{y})^2 = SS_{yy} = 684.1$$

Hataya bağlı değişkenlik:

$$HBD = \sum (y_i - \hat{y}_i)^2 = 7.19$$

RBD=TD-HBD=684.1-7.19=676.91

$$F_{hes} = \frac{RBD/(k-1)}{HBD/(n-k)} = \frac{676.91/1}{7.19/8} = 752.8$$

Kritik Değer:

Test istatistiğimiz, F_{hes}=752.8'dir.

α=0.01

v_1=k-1=2-1=1

v_2=n-k=10-2=8

$F_{0.01,1,8}$=11.26

F_{hes}, tablo değerinden büyük olduğu için sıfır hipotezi reddedilir. Model, anakitleyi temsil etmektedir, güvenle kullanılabilir.

Korelasyon Katsayısı:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-4850}{\sqrt{(34750)(684.1)}} = -0.9947$$

r negatif ve 1'e çok yakındır. Malın satış miktarı ile fiyatı arasında ters yönde çok güçlü bir ilişki vardır.

Korelasyon Katsayısının Yarayışlılığı Testi:

H_0: ρ = 0

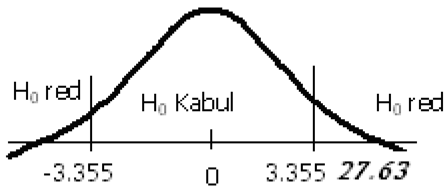
H_1: $\rho \neq$ 0

$$s_r = \sqrt{\frac{1-(-0.9947)^2}{10-2}} = 0.036$$

$$t_{hes} = \frac{0.9947}{0.036} = 27.63$$

α=0.01 ve sd=10-2=8 için:

$t_{0.005,8}$= $\pm$3.355



Test kararımız:

Test istatistiğimiz, red bölgesindedir. Sıfır hipotezini reddederiz. Korelasyon katsayısı, sıfırdan farklıdır ve güvenle kullanılabilir.

Determinasyon Katsayısı:

$$R^2 = \frac{684.1-7.19}{684.1} = \frac{6-1.1}{6} \cong 0.989 \cong \%98.9$$

Malın satış miktarındaki değişmelerin %98.9'u fiyatından kaynaklanmaktadır.

Düzeltilmiş R^2

$$R_d^2 = 1-(1-0.989)\frac{10-1}{10-2} = 0.9876 = \%98.76$$

Hesaplamalarımızı, tahminciler tablosu ve Anova tablosu olarak toplu halde sunalım:

Tahminciler tablosu:

	Katsayı	*StHata*	*t*	*p*
Sabit	80.753	1.755	46.0	0.000
X	-0.139568	0.00508	27.44	0.000

S = 0.948 R^2 = % 98.9 R_d^2= % 98.8

Değişkenlik Kaynağı	*Kareler Toplamı*	*Sd*	*Kareler Ortalaması*	*F*	*p*
RBD	676.91	1	676.91	752.8	0.000
HBD	7.19	8	0.90		
TD	684.1	9			

Son olarak, elde ettiğimiz modeli yorumlayalım. Doğrusal modelimiz:

$$\hat{y} = 80.753 - 0.139568x$$

şeklindeydi.

x'in katsayısına baktığımızda; ele aldığımız malın satış miktarıyla satış fiyatı arasında ters yönlü bir ilişki olduğunu anlıyoruz. Malın kg fiyatı bir birim arttığınıda, satış miktarı 0.139568 birim (=0.139568x100=13.9568 adet) azalacaktır.

Örnek 2-2

Aşağıdaki veriler, aynı marka 16 otomobilin yaşları (yıl) ve yıllık tamir bakım masraflarına (TL / yıl) aittir.

Oto No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T.B.Masr.	109	75	21	135	67	125	71	52	25	70	126	58	30	47	120	105
Yaş	8	3	1	9	5	7	5	2	1	3	6	2	1	2	6	8

Tamir bakım masrafları ve yaş arasındaki doğrusal ilişkiyi; tamir bakım masrafları, yaşın doğrusal bir fonksiyonudur şeklinde belirleyebiliriz. Buna göre, bağımlı değişken tamir bakım masrafları, bağımsız değişken ise yaştır.

Regresyon çözümlerini bu kez doğrudan, tahminciler tablosu ve Anova tablosu olarak vereceğiz:

Tahminciler tablosu:

	Katsayı	*StHata*	*t*	*P*
Sabit	22.605	7.275	3.11	0.008
X	12.671	1.432	8.85	0.000

S = 15.39 R^2 = % 84.8 R_d^2= % 83.7

Anova:

Değişkenlik Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
RBD	18535	1	18535	78.3	0.000
HBD	3314	14	237		
TD	21849	15			

Doğrusal modelimiz:

$$\hat{y} = 22.605 + 12.671x$$

- t testi için p değerlerine baktığımızda, α=0.01 için gerek y keseni, gerekse eğim katsayıları anlamlıdır ($p < \alpha$).
- F testi için p değerine baktığımızda, α=0.01 için doğrusal model anakitleyi temsil eden bir modeldir.
- Eğim katsayısı, otomobil yaşlandıkça, tamir bakım masraflarının arttığını göstermektedir. Otomobil yaşı 1 yıl arttığında, yıllık tamir bakım masrafları, yıllık yaklaşık 12.7 TL artmaktadır.
- y keseni katsayısına göre, yeni bir otomobilin tamir bakım masrafları ise 22 civarındadır.
- R^2, tamir bakım masraflarının %84.8'inin otomobilin yaşına bağlı olduğunu göstermektedir.

Örnek 2-3

12 öğrencinin matematik ve fizik notları aşağıdaki gibidir:

Öğrenci	1	2	3	4	5	6	7	8	9	10	11	12
Fizik	60	70	50	45	75	80	90	30	85	60	70	55
Matematik	70	65	45	40	60	90	100	35	80	60	75	60

Şimdi bu verilerle, doğrusal regresyon analizini gerçekleştirelim.

Fizik, matematik temelli bir ders olduğundan, fizik notlarının matematik notlarına bağlı bir fonksiyon olmasını bekleriz. Buna göre, bağımlı değişken fizik notları, bağımsız değişken ise matematik notlarıdır.

Tahminciler tablosu:

	Katsayı	*StHata*	*t*	*p*
Sabit	10.774	7.794	1.38	0.197
x	0.8214	0.1152	7.13	0.000

S = 7.468 R^2 = % 83.6 R_d^2= % 81.9

Anova:

Değişkenlik Kaynağı	*Kareler Toplamı*	*Sd*	*Kareler Ortalaması*	*F*	*p*
RBD	2833.9	1	2833.9	50.81	0.000
HBD	557.7	10	55.8		
TD	3391.7	11			

Doğrusal modelimiz:

$$\hat{y} = 10.774 + 0.8214x$$

- p değerlerine baktığımızda, α=0.05 için y keseni istatistiki olarak anlamlı değildir; ancak eğim katsayısı anlamlıdır ($p < \alpha$).
- Yine p değerine göre, α=0.01 için doğrusal model anakitleyi temsil eden bir modeldir.
- Eğim katsayısı, matematik notu ile fizik notu arasında doğru yönlü bir ilişkiye işaret etmektedir. Matematik notu 1 puan arttığında, fizik notu 0.82 artmaktadır.
- R^2'ye göre, matematik notu, fizik notunun % 83.6'sını açıklamaktadır.

2.11. Uygulamalar

Örnek 2-4

Y=Süt Arzı (ton/ay) – X=Fiyat (kr/kg)

Y	X	Y^2	X^2	XY	$\hat{Y}$	e
90	100	8100	10000	9000	90.06738	-0.067375887
95	110	9025	12100	10450	99.3227	-4.322695035
120	130	14400	16900	15600	117.8333	2.166666667
130	135	16900	18225	17550	122.461	7.539007092
135	150	18225	22500	20250	136.344	-1.343971631
137	155	18769	24025	21235	140.9716	-3.971631206

$\sum y$	$\sum x$	$\sum y^2$	$\sum x^2$	$\sum xy$	$\sum e$	Y-ort	X-ort	SSyy	SSxx	SSxy
707	780	85419	103750	94085	0	117.8333	130	2110.833	2350	2175

Tahmin sonuçları:

	Katsayı	*St.Hata*	*t*	*P*
Sabit	-2.485815603	13.41303561	-0.18533	0.861989
X	0.925531915	0.102001991	9.073665	0.000818

ANOVA

Değişkenlik Kaynağı	Sd	HKT	HKO	F	P
Regresyon	1	2013.031915	2013.032	82.3314	0.000818
Hata	4	97.80141844	24.45035		
Toplam	5	2110.833333			

St Hata=4.9447 R kare=0.95366 Düz.R kare=0.94208

Doğrusal modelimiz:

$$\hat{y} = -2.485815603 + 0.925531915x$$

- p değerlerine göre, α=0.05 için y keseni istatistiki olarak anlamlı değildir; ancak eğim katsayısı sıfırdan farklıdır ($p < \alpha$).
- Yine p değerine göre, α=0.01 için doğrusal model anakitleyi temsil eden bir modeldir.
- Eğim katsayısı, süt fiyatının süt arzını doğru yönlü etkilediğine işaret etmektedir. Daha açık bir ifadeyle, süt fiyatı 1 birim arttığında süt arzı 0.92 birim artmaktadır.
- R^2'ye göre, süt arzındaki değişmenin %95'I süt fiyatı tarafından açıklanmaktadır.

Örnek 2-5

Y=Tereyağ talebi (10 ton/ay) – X=Fiyat (TL/kg)

Y	X	Y^2	X^2	XY	$\hat{Y}$	e
170	15	28900	225	2550	123.6922	46.30784708
80	16	6400	256	1280	115.3421	-35.34205231
75	18	5625	324	1350	98.64185	-23.64185111
70	22	4900	484	1540	65.24145	4.758551308
60	23	3600	529	1380	56.89135	3.108651911
45	25	2025	625	1125	40.19115	4.808853119

∑y	∑x	$∑y^2$	$∑x^2$	∑xy	∑e	Y-ort	X-ort	SSyy	SSxx	SSxy
500	119	51450	2443	9225	0	83.3333	19.8333	9783.333	82.833	-691.667

	Katsayı	*Standart Hata*	*t*	*P*
Sabit	248.943662	70.17931339	3.547251	0.024
X	-8.350100604	3.477947377	-2.40087	0.074

St Hata=31.65377972 R kare=0.590 Düz.R kare=0.488

Anova

Değişkenlik Kaynağı	Sd	HKT	HKO	F	p
Regresyon	1	5775.486251	5775.486	5.764178	0.074
Hata	4	4007.847082	1001.962		
Toplam	5	9783.333333			

Örnek 2-6

Y=Tüketim Harcamaları (TL/Ay) – X=Gelir (TL/Ay)

Y	X	Y^2	X^2	XY	$\hat{Y}$	e
75	65	5625	4225	4875	83.0682	-8.068198776
60	55	3600	3025	3300	81.3886	-21.38859872
80	80	6400	6400	6400	85.5876	-5.587598866
100	110	10000	12100	11000	90.6264	9.373600955
150	350	22500	122500	52500	130.9368	19.06319952
175	400	30625	160000	70000	139.3348	35.66519922
180	500	32400	250000	90000	156.1308	23.86919863
120	600	14400	360000	72000	172.9268	-52.92680197

∑y	∑x	$∑y^2$	$∑x^2$	∑xy	∑e	Y-ort	X-ort	SSyy	SSxx	SSxy
940	2160	125550	918250	310075	0	117.5	270	15100	335050	56275

	Katsayı	*St.Hata*	*t*	*P*
Sabit	72.15079839	17.95784963	4.0177861	0.007
X	0.167960006	0.053005246	3.1687431	0.019

St Hata=30.68 R kare=0.626 Düz.R kare=0.564

ANOVA

Değişkenlik Kaynağı	Sd	HKT	HKO	F	P
Regresyon	1	9451.949336	9451.9493	10.04093	0.01935
Hata	6	5648.050664	941.34178		
Toplam	7	15100			

Örnek 2-7

Y=Birim Maliyet (TL/kg) – X=Toplam Üretim (ton/ay)

Y	X	Y^2	X^2	XY	$\hat{Y}$	e
650	10	422500	100	6500	671.2222	-21.2222
640	15	409600	225	9600	624.3889	15.6111
600	20	360000	400	12000	577.5556	22.4444
550	25	302500	625	13750	530.7222	19.2777
490	30	240100	900	14700	483.8889	6.1111
400	35	160000	1225	14000	437.0556	-37.0555
350	40	122500	1600	14000	390.2222	-40.2222
340	45	115600	2025	15300	343.3889	-3.38888
335	50	112225	2500	16750	296.5556	38.44444

∑y	∑x	$\sum y^2$	$\sum x^2$	∑xy	∑e	Y-ort	X-ort	SSyy	SSxx	SSxy
4355	270	2245025	9600	116600	0	483.889	30	137688.8889	1500	-14050

	Katsayı	*St.Hata*	*t*	*P*
Sabit	764.8888	24.86733	30.7587	0.000
X	-9.36666	0.761403	-12.3018	0.000

St hata=29.489 R kare=0.956 Düz.R kare=0.955

ANOVA

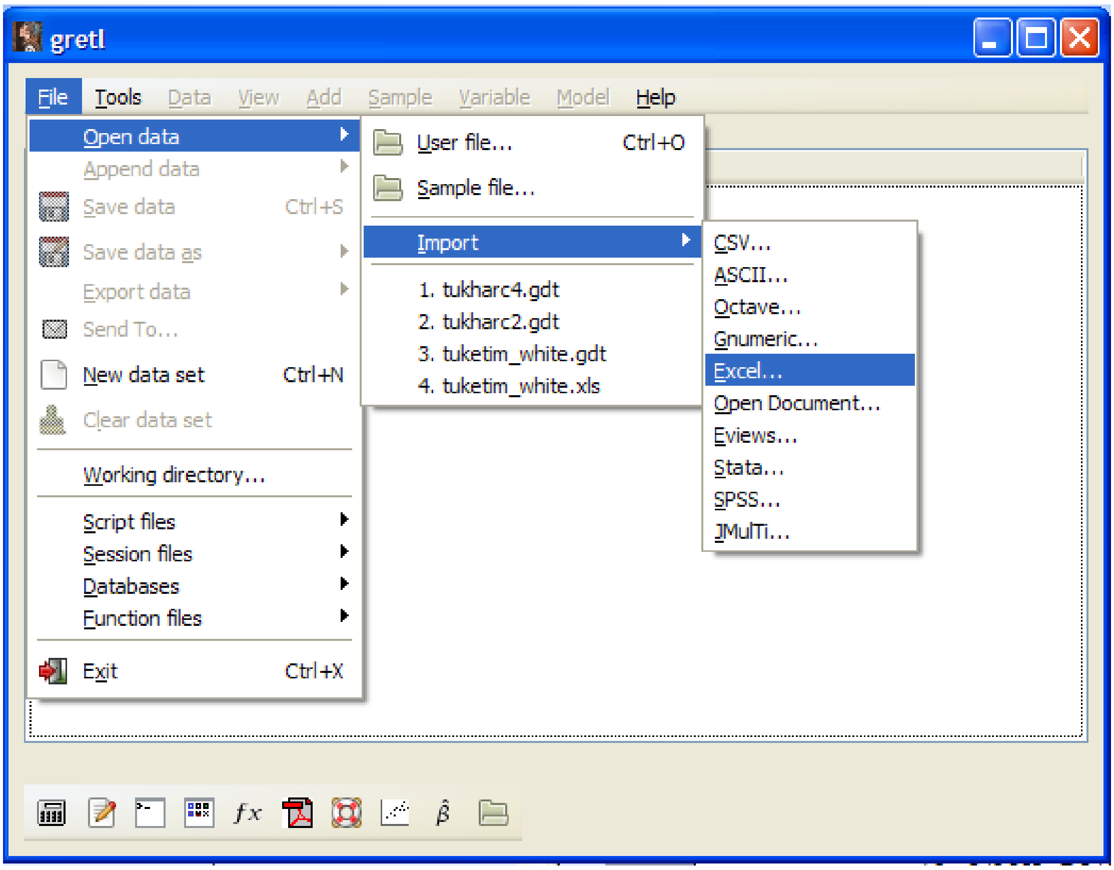
Değişkenlik Kaynağı	Sd	HKT	HKO	F	P
1	1	131601.6667	131601.67	151.3353	0.000
6	7	6087.222222	869.60317		
7	8	137688.8889			

2.12.Gretl Kullanarak Ekonometrik Modelleme

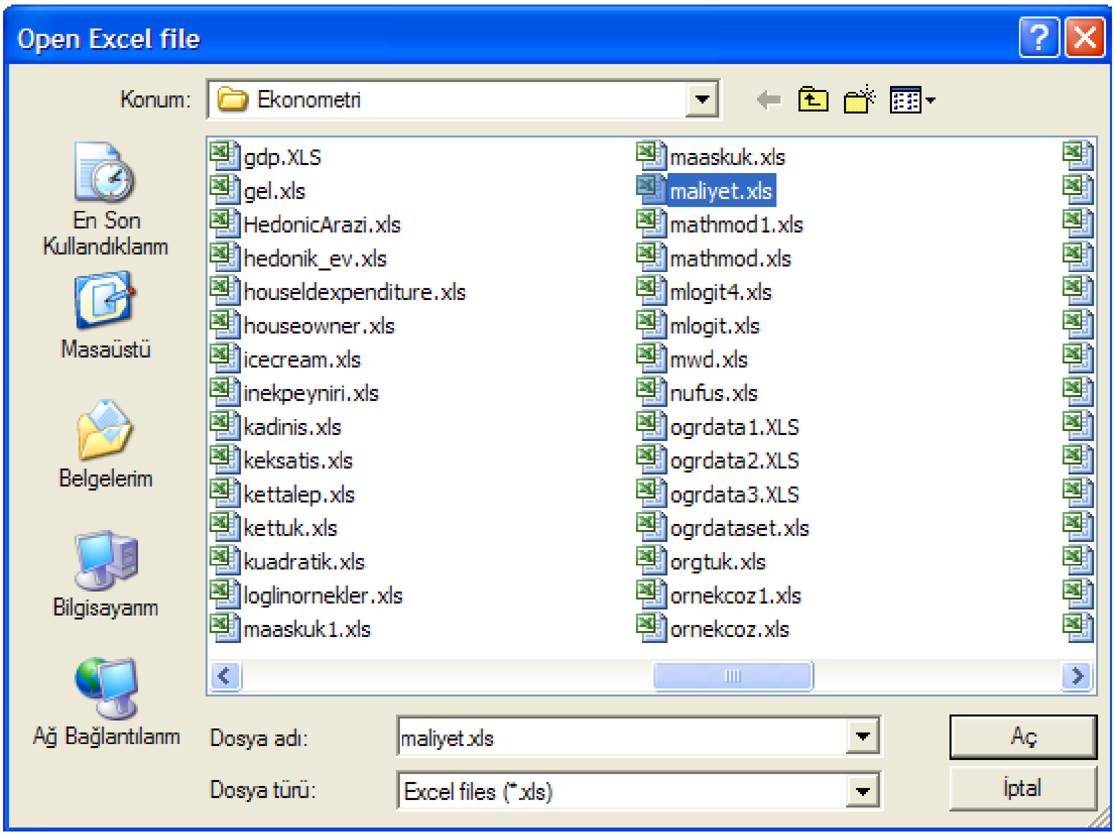
Gretl ismi, (Gnu Regression, Econometrics and Time-series Library) kelimelerinin baş harflerinden oluşmaktadır. Açık kod ekonometri yazılımıdır. http://gretl.sourceforge.net/ adresinden ücretsiz olarak indirilebilmektedir. Ekonometristler tarafından geliştirilmiş, kullanımı son derece kolay bir yazılımdır. Bu kitapta Gretl'dan sıkça yararlanacağız.

2.12.1.Veri Girişi

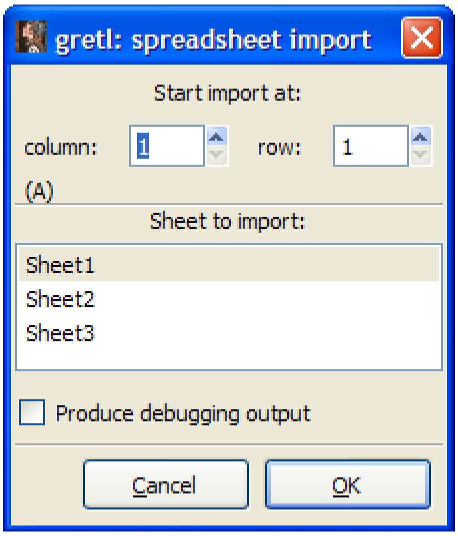
Gretl'in kendine özel veri giriş editörü bulunmaktadır. Ancak, verilerin Excel'de girilip daha sonra Gretl'a aktarılması önerilir. Yalnızca verilerdeki küçük değişiklikler için Gretl'ın kullanılması uygundur. Excel'de hazırlanan bir veri dosyasını Gretl'a almak için aşağıdaki süreç izlenir:



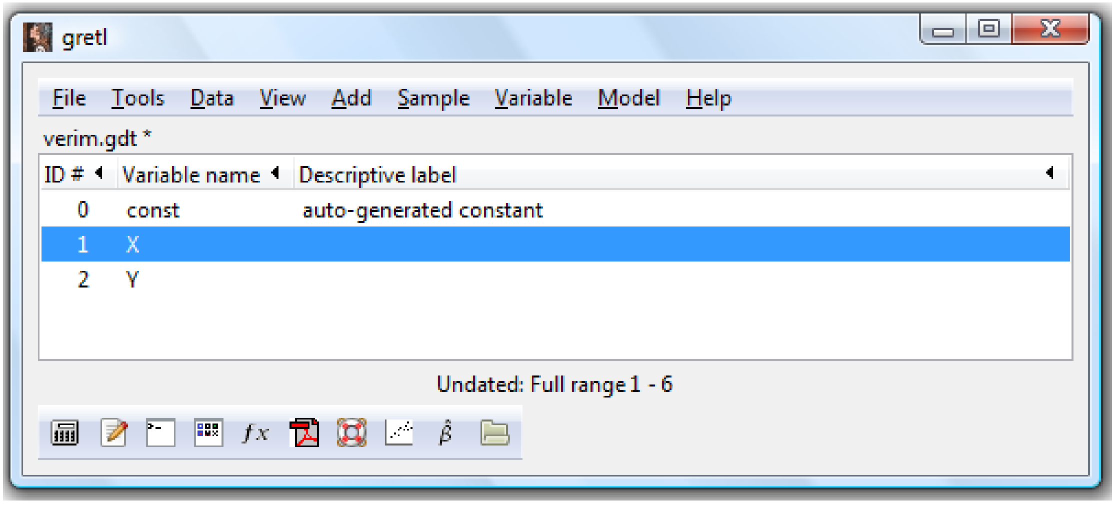
Excel formatında hazırlanan dosyanın seçilmesi gerekiyor:



Verilerin bulunduğu Excel dosyasının hangi sayfasında ve hangi satır/sütun aralığında olduğu soruluyor:

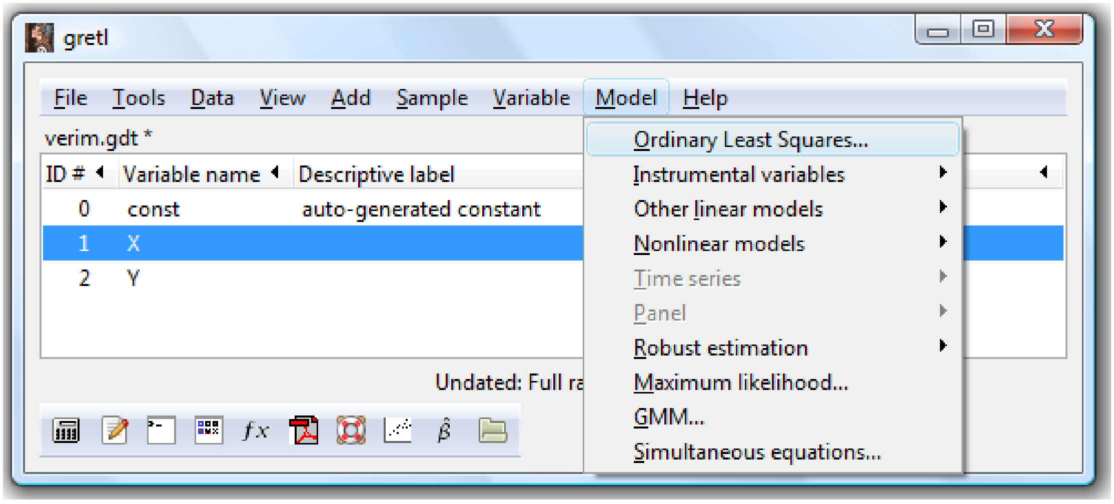


Artık verileriniz Gretl'a alınmış durumdadır.

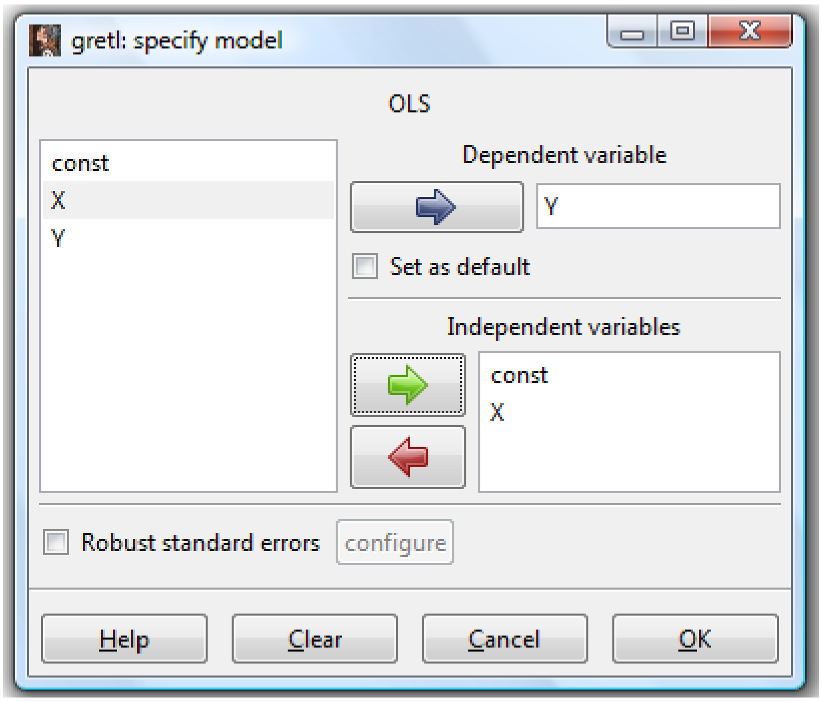


2.12.2. *EKK Tahminlemesi*

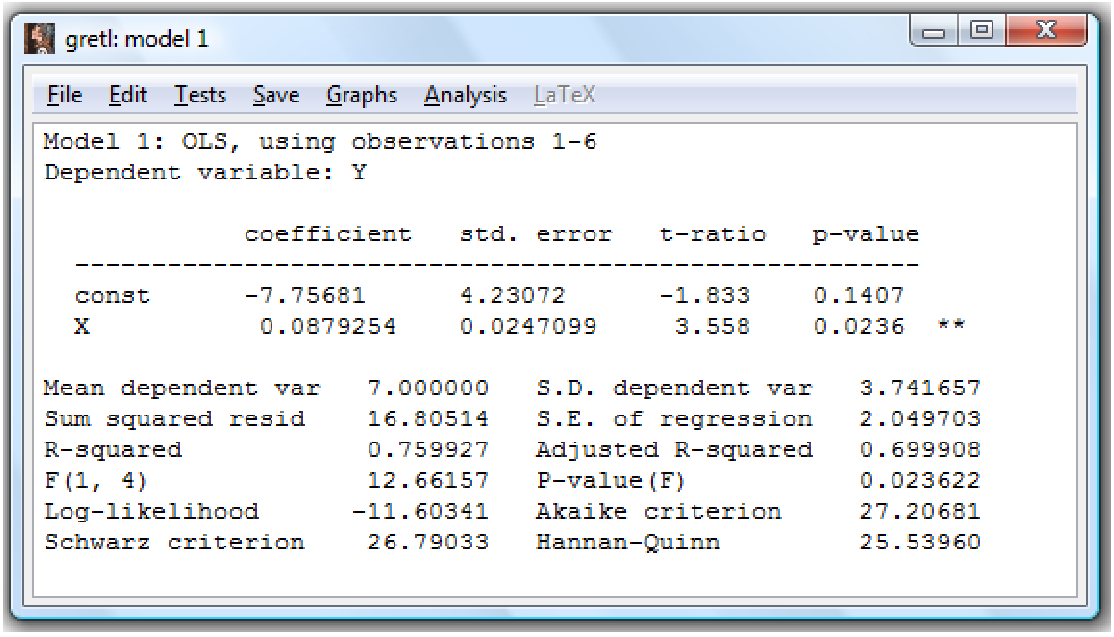
Gretl'da EKK modellemesi yapmak için menüden Model'i seçip, Ordinary Least Squares'i tıklamanız gerekiyor.



OLS (en küçük kareler) ekranında, sol tarafta değişkenlerimiz bulunmaktadır. Bağımlı değişkeni, **dependent variable** başlığı altındaki -> butonuyla metin kutusuna aktarılır. Bağımsız değişkenler sol taraftaki pencerede işaretlendikten sonra, -> butonu ile **Bağımsız variables** bölümüne aktarılır. Son olarak OK butonuna tıklanarak, EKK tahmin sonuçları alınır.

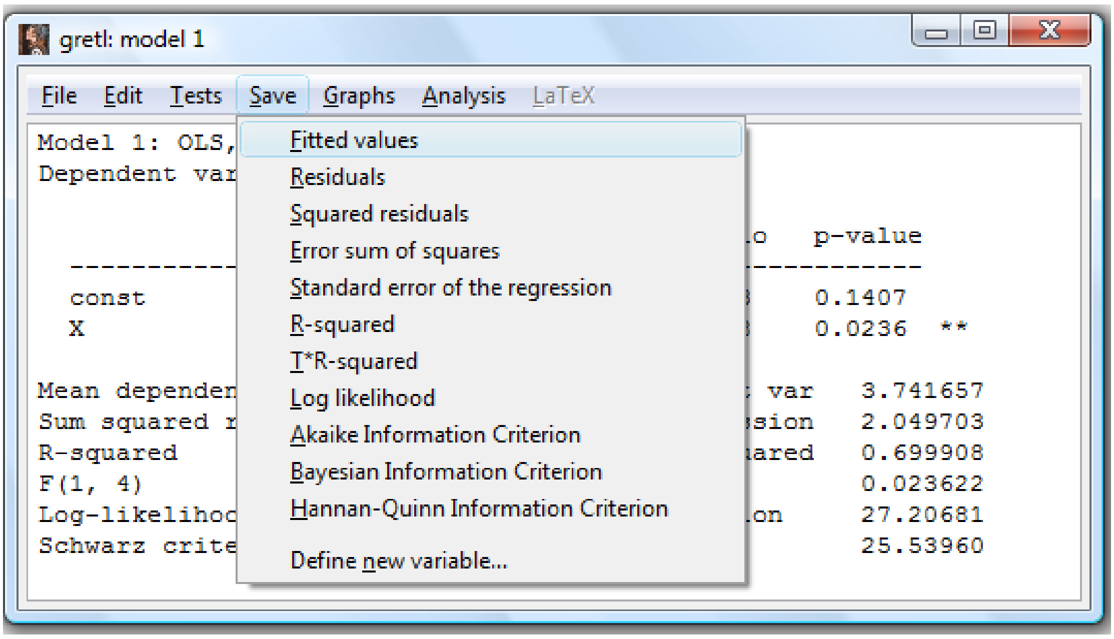


EKK tahmin sonuçları, model seçim kriterleri eşliğinde sunulur.

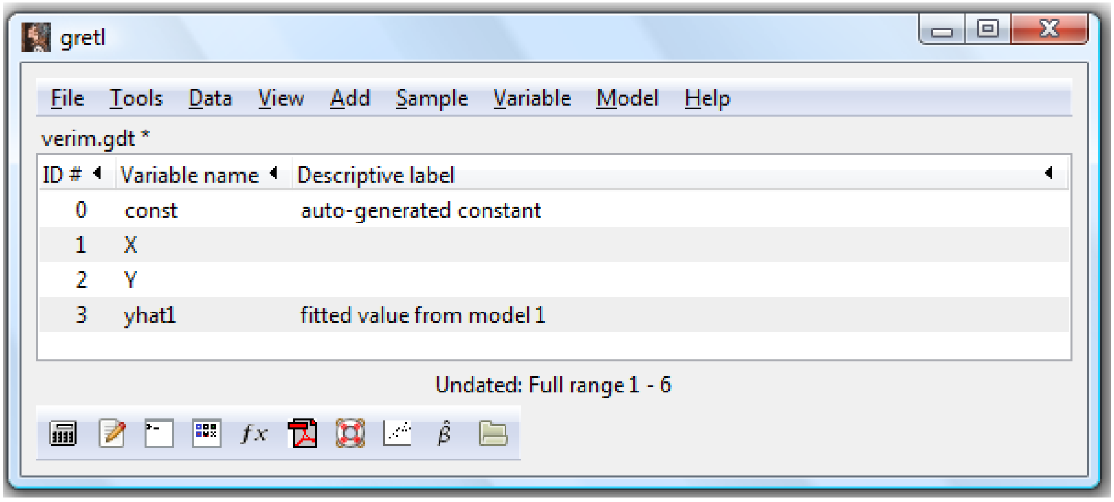



```
Model 1: OLS, using observations 1-6
Dependent variable: Y

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const      -7.75681      4.23072      -1.833    0.1407
  X           0.0879254    0.0247099     3.558    0.0236   **

Mean dependent var    7.000000   S.D. dependent var    3.741657
Sum squared resid    16.80514    S.E. of regression    2.049703
R-squared             0.759927   Adjusted R-squared    0.699908
F(1, 4)              12.66157    P-value(F)            0.023622
Log-likelihood      -11.60341    Akaike criterion     27.20681
Schwarz criterion    26.79033    Hannan-Quinn         25.53960
```

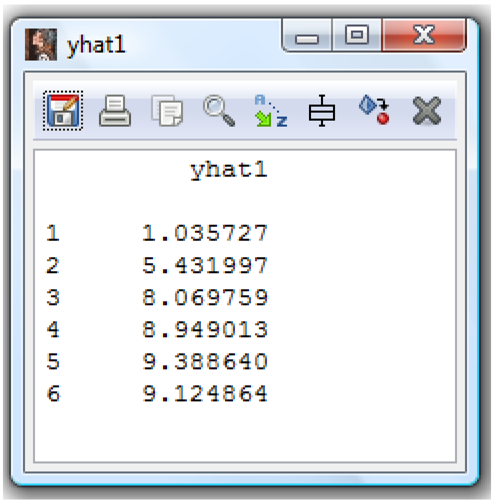
EKK modeliyle ilgili çeşitli istatistikler, model seçim kriterleri, hata terimi ve tahmin değerleri Save menüsünde saklanmaktadır.



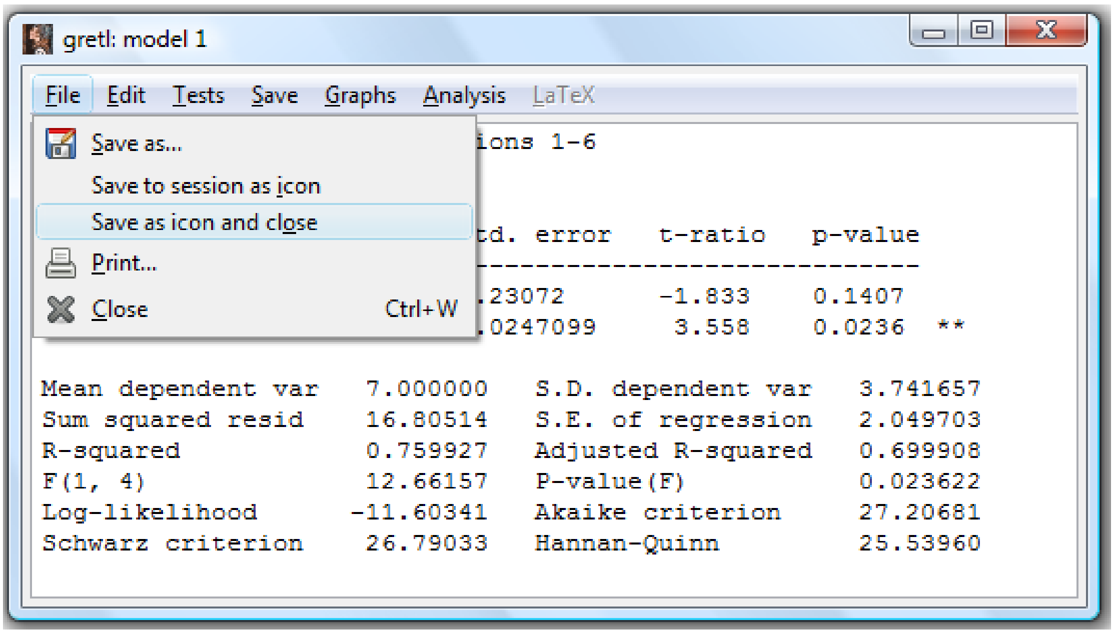
Örneğin tahmin değerlerine ulaşmak için **Fitted values** tıklanır. Tahmin değerleriniz **yhat** olarak adlandırılır ve değişken listesine eklenir.



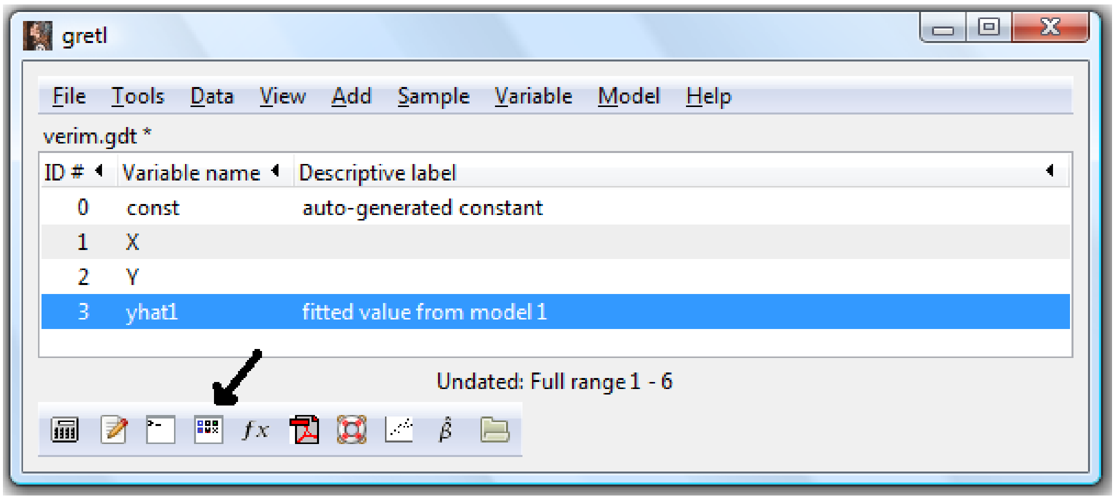
yhat1 değişkeninin üzerine tıklandığında, EKK modelinden hesaplanan tahmin değerleri sunulur.



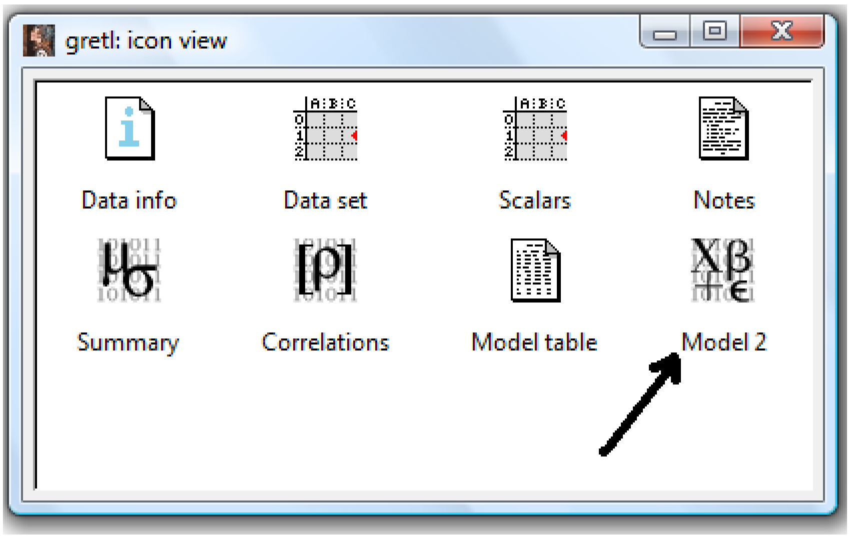
Tahmin edilen EKK modeli, daha sonra ulaşmak üzere saklanabilir. Bunun için **File/Save as icon and close** menüsünden yararlanılabilir.



Tahmin edilen modellerin her birini bu şekilde saklayıp, daha sonra **session icon view'**i tıklayarak ulaşmak mümkündür.



Tahmin ettiğimiz EKK modeli, Model 2 (tahmin edilen model numarası) olarak saklanmıştır. Tıklandığında sakladığınız EKK modelini tekrar sunacaktır.


gretl: icon view
Data info
Data set
Scalars
Notes
Summary
Correlations
Model table
Model 2

3. ÇOK AÇIKLAYICILI DOĞRUSAL REGRESYON

Regresyon modellerinde, bir değişkeni tek bir değişkenin açıkladığı varsayımına dayalı tek açıklayıcılı regresyon modelleri üzerinde durmuştuk. Gerçek hayat koşullarında hiç bir olay, tek bir nedene bağlı olarak gerçekleşmez. Basit bir olay bile, çok sayıda faktörün bir bileşkesi olarak ortaya çıkar. Buğday verimi; gübre, ilaç, su, nem, aylara göre ortalama sıcaklıklar, toprağın yapısı, çiftçinin deneyimi gibi çok sayıda etkenin etkisi altında gerçekleşir. Bir malın talebi; o malın fiyatı, rakip malların fiyatları, tamamlayıcı malların fiyatları, nüfus, gelir, alışkanlıklar, moda gibi sayısız faktöre bağlıdır. Regresyon modelleri, basitleştirme ilkesiyle, gerçek koşullardaki etkili faktörleri makûl sayıya indirmeye çalışır. Çok sayıda faktörün modele alınması hem kontrolü zorlaştırır, hem de hepsine ilişkin verilerin toplanması zaman alıcı ve masraflıdır. O halde, kaç faktör alınmalı ki en iyi tahminelemede bulunabilelim. Regresyon modelindeki değişkenlerin sayısı ve niteliği, araştırmacının amaçlarına, deneyimine ve becerisine bağlı olacaktır. Aynı konuda, farklı araştırıcılar farklı modeller kurabilir. Ancak dikkate alınması gereken hususun, çoğu azla açıklamak olduğunu belirtmeliyiz. Bunu yaparken, modelde olmazsa olmaz değişkenleri asla gözardı etmememiz gerekir.

Çok açıklayıcılı regresyon modelleri, tek bir değişkenle kısıtlı kalmaktan kurtaracak bir araçtır. Böylece, diğer faktörlerin değişmediğini veya sabit kaldığını varsaymanın getireceği bilgi kaybını en aza indirebileceğiz. Çok açıklayıcılı regresyon modelleri de yine kesin ve olasılıklı olmak üzere ele alınabilir. Biz, olasılıklı regresyon modelleriyle ilgileneceğiz. Olasılıklı regresyon modelinin genel ifadesi:

y= Kesin öge + Tesadüfi hata

şeklindedir.

Birinci dereceden doğrusal çok açıklayıcılı olasılıklı model:

$$y = b_0 + b_1x_1 + b_2x_2 +...+ b_kx_k + e$$

olup, burada:

y= Bağımlı değişken (Modellenecek değişken)

x_i=Bağımsız ya da açıklayıcı değişken (i=1..k).

k = Açıklayıcı değişken sayısı

e=Tesadüfi hata ögesi

b_0 = Doğrunun y eksenini kesim noktası

b_1 = Doğrunun x_1 değişkenine göre eğimidir. Diğerleri sabitken x_1'de meydana gelecek her 1 birimlik değişime karşılık, y'de meydana gelecek değişim miktarıdır.

b_k = Doğrunun x_k değişkenine göre eğimidir. Diğerleri sabitken x_k'da meydana gelecek her 1 birimlik değişime karşılık, y'de meydana gelecek değişim miktarıdır.

Çok değişkenli modellerin tahminlenmesi, esas olarak, tek değişkenli modellerinkinden hiç farklı değildir. Daha önce olduğu gibi, yine en küçük kareler yöntemi kullanılmaktadır. Tek açıklayıcılı bir regresyon modelinin tahmincilerini hesaplamak için, 2 bilinmeyenli 2 denklem çözmüştük. İki açıklayıcılı bir modeli çözebilmek için ise 3 bilinmeyenli 3 denklem çözmemiz gerekir. İster tek ister çok açıklayıcılı olsun, regresyon modelleri, matris yöntemlerle çözülebilmektedir. Matris çözüm, makûl sayıda değişken için elle çözüme de imkan vermektedir.

3.1. Çok Açıklayıcılı Regresyon Modellerinin Çözümü

Tek açıklayıcılı regresyon modellerinin çözümünde izlediğimiz yolu, çok açıklayıcılı modeller için de kullanacağız. Aşağıdaki 2 açıklayıcılı modeli ele alalım:

$$y = b_0 + b_1x_1 + b_2x_2$$

2 açıklayıcılı modeli 3 bilinmeyenli 3 denklem halinde ifade edip, tahmincileri çözmemiz gerekiyor. Bu denklemler:

$$\Sigma y = nb_0 + b_1\Sigma x_1 + b_2\Sigma x_2$$

$$\Sigma x_1y = b_0\Sigma x_1 + b_1\Sigma x_1^2 + b_2\Sigma x_1x_2$$

$$\Sigma x_2y = b_0\Sigma x_2 + b_1\Sigma x_1x_2 + b_2\Sigma x^2_2$$

şeklindedir.

Bir malın arzının, o malın kendi fiyatı ve diğer malın fiyatına bağlı olduğu çok açıklayıcılı regresyon modelini hazırlayalım:

Arz Y	Fiyat (x_1)	Diğer Malın Fiyatı (x_2)
32	157	8
35	159	10
27	149	6
34	162	11
28	151	8
29	150	7
38	155	10
28	148	9
27	152	10
25	142	6
38	161	12
34	157	9

$\Sigma y=375$, $\Sigma x_1=1843$, $\Sigma x_2=106$, $\Sigma x_1y=57842$, $\Sigma x_2y=3383$, $\Sigma x_1^2=283443$ $\Sigma x^2_2=976$, $\Sigma x_1x_2=16379$

Denklemlerde hesaplama sonuçlarımızı yerlerine koyarsak:

$$375 = 12b_0 + 1843b_1 + 106b_2$$

$$57842 = 1842b_0 + 283443b_1 + 16379b_2$$

$$3383 = 106b_0 + 16379b_1 + 976b_2$$

Elde ettiğimiz 3 bilinmeyenli 3 denklemi çözdüğümüzde:

$$\hat{b}_0 = -51.6 \qquad \hat{b}_1 = 0.511 \qquad \hat{b}_2 = 0.501$$

buluruz. b_0 ve b_1'in tahmincileri, sırasıyla fiyat ve diğer malın fiyatı değişkenlerine göre kısmi eğim değerleridir. Hangi değişkene göre eğim için yorum yapılıyorsa, "diğer değişkenler sabit tutulduğu takdirde" koşulunu dikkate almalıyız. Buna göre, fiyat için eğim, diğer mal fiyatı sabitken o malın fiyatı 1 birim arttığında, arzın 0.511 birim artacağını; diğer malın fiyatı için eğim ise, malın fiyatı sabitken diğer malın fiyatı 1 birim arttığında arzın 0.501 birim artacağını ifade etmektedir. Ancak bu yorumlamalar, istatistiki açıdan güvenli olmaları durumunda yapılabilecektir. Bu nedenle tahmincilerin yarayışlılığını test edeceğiz.

3.2. Tahmincilerin Yarayışlılığı

Hesaplanan tahmincilerin tek tek bağımlı değişken üzerinde etkili olup olmadıkları, t testiyle belirlenecektir. Bu testte, tahmincinin bağımlı değişken üzerinde etkisi yoktur (veya sıfırdır) hipotezi, tahminci bağımlı değişken üzerinde etkilidir hipotezine karşı test edilir.

b_0 tahmincisi için çift yanlı test hipotezleri:

H_0: $\hat{b}_0 = 0$

H_1: $\hat{b}_0 \neq 0$

b_0 tahmincisi için tek yanlı test hipotezleri:

H_0: $\hat{b}_0 = 0$ H_0: $\hat{b}_0 = 0$

H_1: $\hat{b}_0 < 0$ H_1: $\hat{b}_0 > 0$

şeklindedir.

b_1 tahmincisi için çift yanlı test hipotezleri:

H_0: $\hat{b}_1 = 0$

H_1: $\hat{b}_1 \neq 0$

b_1 tahmincisi için tek yanlı test hipotezleri:

H_0: $\hat{b}_1 = 0$ H_0: $\hat{b}_1 = 0$

H_1: $\hat{b}_1 < 0$ H_1: $\hat{b}_1 > 0$

şeklindedir.

b_2 tahmincisi için çift yanlı test hipotezleri:

H_0: $\hat{b}_2 = 0$

H_1: $\hat{b}_2 \neq 0$

b_2 tahmincisi için tek yanlı test hipotezleri:

H_0: $\hat{b}_2 = 0$ H_0: $\hat{b}_2 = 0$

H_1: $\hat{b}_2 < 0$ H_1: $\hat{b}_2 > 0$

şeklindedir.

Bir modelde k adet tahminci olabileceğine göre:

b_k tahmincisi için çift yanlı test hipotezleri:

H_0: $\hat{b}_k = 0$

H_1: $\hat{b}_k \neq 0$

b_k tahmincisi için tek yanlı test hipotezleri:

H₀: $\hat{b}_k = 0$ H₀: $\hat{b}_k = 0$

H₁: $\hat{b}_k < 0$ H₁: $\hat{b}_k > 0$

3.3. Modelin Yarayışlılığı

Kısmi eğimlerin yarayışlılığını test ettikten sonra, modelin bütün olarak istatistiki yarayışlılığı test edilir. Yine F testinden yararlanacağız. Hipotezlerimiz:

H_0: $\hat{b}_1 = 0$ ve $\hat{b}_2 = 0$ ve $\hat{b}_3 = 0$ ve .... ve $\hat{b}_k = 0$

H_1: $\hat{b}_1 \neq 0$ veya $\hat{b}_2 \neq 0$ veya $\hat{b}_3 \neq 0$ veya .... veya $\hat{b}_k \neq 0$

H_0 hipotezi, b_0 tahmincisini içermez. Zira, b_0 bir sabittir. Tek başına istatistiki olarak geçerli bir b_0, değişkenlik taşımadığından, bir anlam taşımaz. Sıfır hipotezinde, tahmincilerin tamamının eksiksiz olarak sıfıra eşit olduğu iddia edilir. O yüzden "ve" bağlacı kullanılmıştır. Alternatif hipotezde ise, en az bir tahmincinin sıfırdan farklı olacağı iddiası vardır. Buna göre, alternatif hipotezinin kabul edilmesi, tahmincilerin tamamının veya en az birinin istatistiki olarak anlamlı olduğu sonucunu verir. Bir başka deyişle, modelin geçerli olması için, tek bir tahmincinin dahi sıfırdan farklı olması yeterlidir.

Test istatistiğinin hesaplanması:

Tek açıklayıcılı regresyon modelinde olduğu gibi, değişkenlik bileşenlerine ayrılır. Buna göre, toplam değişkenlik (TD), regresyona bağlı değişkenlik (RBD) ve hataya bağlı değişkenlik (HBD) bileşenlerinin toplamından oluşur:

TD = RBD + HBD

Toplam değişkenlik:

$$TD = \sum (y_i - \bar{y})^2 = SS_{yy}$$

Regresyona bağlı değişkenlik:

$$RBD = \sum (\hat{y}_i - \bar{y})^2$$

Hataya bağlı değişkenlik:

$$HBD = \sum (y_i - \hat{y}_i)^2 = HKT$$

Değişkenliğin bileşenlerini Anova tablosu olarak gösterebiliriz:

Değişkenlik Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F
RBD	$\sum (y_i - \bar{y})^2$	k-1	a= $\sum (y_i - \bar{y})^2 /(k-1)$	a/b
HBD	$\sum (y_i - \hat{y}_i)^2$	n-k	b= $\sum (y_i - \hat{y}_i)^2 /(n-k)$	
TD	$\sum (y_i - \bar{y})^2$	n-1		

Tabloda F istatistiği:

$$F_{hes} = \frac{RBD/(k-1)}{HBD/(n-k)}$$

olarak hesaplanmaktadır.

Test Kararı:

Eğer hesaplanan F değeri, tablo F değerinden büyükse, sıfır hipotezi reddedilir. Bu, regresyon modeli anakitleyi temsil etmektedir, güvenle kullanılabilir anlamına gelmektedir.

Tek açıklayıcılı regresyon modellerinin analizi, test ve yorum açısından tek açıklayıcı modellerle hemen hemen aynıdır. Bu nedenle, burada sadece çözülmüş regresyon analizleri üzerinde durulacaktır.

Örnek 3-1

Bir malın talep miktarı, fiyatı ve gelir düzeyine ilişkin veriler aşağıdadır:

Talep Y	Fiyat (x_1)	Gelir (x_2)
20	2	60
17	4	130
16	5	170
11	7	160
10	9	250
6	10	250
4	12	350
3	13	350

Talep miktarını, malın fiyatı ve gelir düzeyinin bir fonksiyonu olarak ifade edebiliriz. Buna göre talep, bağımlı değişken; fiyat ve gelir, açıklayıcı değişkenlerdir.

Regresyon analizi sonuçları:

Tahminciler tablosu:

	Katsayı	*StHata*	*t*	*P*
Sabit	22.8987	0.606	37.79	0.000
x_1	-2.3017	0.3222	-7.14	0.001
x_2	0.02705	0.01218	2.22	0.077

S = 0.7007 R^2 = % 99.1 R_d^2= % 98.8

Anova:

Değişkenlik Kaynağı	*Kareler Toplamı*	*Sd*	*Kareler Ortalaması*	*F*	*p*
RBD	278.42	2	139.21	283.55	0.000
HBD	2.45	5	0.49		
TD	280.87	7			

Doğrusal modelimiz:

$$\hat{y} = 22.8987 - 2.3017x_1 + 0.02705x_2$$

- p değerlerine baktığımızda, α=0.05 için; *y* keseni istatistiki olarak anlamlıdır (p < α); x_1 değişkeni talebi etkilemektedir (p < α); x_2 değişkeni ise talep üzerinde etkili değildir. (p > α).
- Yine p değerine göre, α=0.05 için doğrusal model anakitleyi temsil eden bir modeldir.

- x_1'in katsayısı, talep ile fiyat arasında ters yönlü bir ilişkiyi göstermektedir. Fiyat 1 birim arttığında, talep 2.3 birim düşmektedir.
- x_2'de meydana gelen değişmelerin, talep üzerinde herhangi bir etkisi yoktur.
- Talebin %98.8'i, fiyat ve gelir tarafından açıklanmaktadır.
- Bu örnekte, eğer α=0.10 olarak alınsaydı, gelir de talep üzerine etkili bir değişken olarak kabul edilebilirdi. Burada tercih, araştırmacınındır.

Örnek 3-2

Bu örnekte, hedonik modelleme yapacağız. Otomobillerin fiyatıyla, motor gücü ve 0'dan 100 km'ye ulaşma süresi arasındaki ilişkiyi incelemek üzere, aşağıdaki veriler elde edilmiştir:

Marka	Güç (HP) (x_1)	0-100 km Süresi (sn) (x_2)	Fiyat (1000 $) (Y)
BMW M3	240	6.0	38.4
Corvette	300	5.7	41.4
Dodge Viper	400	4.8	54.8
Mustang	240	6.9	25.8
Honda Prelude	190	7.1	25.6
Mitsubishi 3000 GT	320	5.7	43.7
Toyota Supra	320	5.3	48.2
Nissan 300 ZX	300	6.0	40.8
Alfa Romeo	320	7.6	38.1
Mazda RX-7	255	5.5	35.0

Otomobil fiyatının, motor gücü ve 0'dan 100 km'ye ulaşma süresine bağlı olarak oluştuğunu varsayabiliriz. Bu durumda, otomobil fiyatı bağımlı değişken, motor gücü ve 0'dan 100 km'ye ulaşma süresi bağımsız değişkenler olarak alınmalıdır.

Regresyon analizi sonuçları:

Tahminciler tablosu:

	Katsayı	*StHata*	*t*	*p*
Sabit	30.96	12.37	2.5	0.041
x_1	0.10829	0.02006	5.4	0.000
x_2	-3.8	1.349	-2.82	0.026
x_3	0.10829	0.02006	5.4	0.000

S = 2.968 R^2 = % 91.6 Düz.R^2= % 89.2

Anova:

Değişkenlik Kaynağı	*Kareler Toplamı*	*Sd*	*Kareler Ortalaması*	*F*	*p*
RBD	674.34	2	337.17	38.26	0.000
HBD	61.68	7	8.81		
TD	736.02	9			

Doğrusal modelimiz:

$\hat{y} = 30.96 - 3.8x_1 + 0.10829x_2$

- p değerlerine baktığımızda, α=0.05 için; *y* keseni istatistiki olarak anlamlıdır (p<α); x_1 değişkeni fiyatı etkilemektedir (p<α); x_2 değişkeni de fiyat üzerinde etkilidir (p<α).
- Yine p değerine göre, α=0.05 için doğrusal model anakitleyi temsil eden bir modeldir.
- x_1'nin katsayısı, fiyatla 100 km'ye ulaşma süresi arasında ters yönlü bir ilişkiyi göstermektedir. 100 km'ye ulaşma süresi 1 sn düştüğünde, fiyat 3800 $ yükselmektedir.
- X_2'in katsayısı, motor gücü arttıkça, fiyatın artacağına işaret etmektedir. Motor gücünün 1 beygir artması, fiyatın yaklaşık 108 $ artmasına neden olmaktadır.
- Fiyatın %89.2'si, hızlanma süresi ve motorgücüyle açıklanmaktadır.

3.4. Matrisleri Kullanarak EKK

En küçük kareler yöntemini, matris yöntemden yararlanarak çözmek mümkündür. Tahmincileri hesaplamak için:

$b=(X'X)^{-1}X'Y$

formülü kullanılır. Burada X, bağımsız değişkenler matrisi; Y, bağımlı değişken matrisi; X', X matrisinin devriğidir.

Örnek 3-3

Tek açıklayıcılı doğrusal bir modeli ele alalım:

Y	X
1	1
3	1
7	3
7	3
8	5
9	6
12	9

Amacımız doğrusal modeli tahmin etmek:

$Y=b_0 + b_1X$

İlk olarak X matrisini hazırlayalım. Denklemde sabit terim olduğundan, bağımsız değişkenler matrisine, b_0'ı temsil eden bir sütun ekleyeceğiz. Bu sütunun tamamı 1 olacaktır.

$$X=\begin{vmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 3 \\ 1 & 3 \\ 1 & 5 \\ 1 & 6 \\ 1 & 9 \end{vmatrix}$$

Y matrisimiz, bağımlı değişken değerleri içerecektir:

$$Y=\begin{vmatrix} 1 \\ 3 \\ 7 \\ 7 \\ 8 \\ 9 \\ 12 \end{vmatrix}$$

$$X'=\begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 3 & 5 & 6 & 9 \end{vmatrix}$$

$$X'X=\begin{vmatrix} 7 & 28 \\ 28 & 162 \end{vmatrix}$$

$$X'X^{-1}=\begin{vmatrix} 81/175 & -2/25 \\ -2/25 & 1/50 \end{vmatrix}$$

$$X'Y=\begin{vmatrix} 47 \\ 248 \end{vmatrix}$$

$$b=\begin{vmatrix} 1.914286 \\ 1.2 \end{vmatrix}$$

Örnek 3-4

Çok açıklayıcılı bir modeli, matris yöntemle tahmin edelim. Maaş, eğitim ve deneyim.

y	x1	x2
30	4	10
20	3	8
36	6	11
24	4	9
40	8	12
35	7	7
45	13	6

$$X=\begin{vmatrix} 1 & 4 & 10 \\ 1 & 3 & 8 \\ 1 & 6 & 11 \\ 1 & 4 & 9 \end{vmatrix} \qquad X'=\begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 3 & 6 & 4 & 8 & 7 & 13 \\ 10 & 8 & 11 & 9 & 12 & 7 & 6 \end{vmatrix}$$

$$\begin{vmatrix} 1 & 8 & 12 \\ 1 & 7 & 7 \\ 1 & 13 & 6 \end{vmatrix}$$

$$X'X=\begin{vmatrix} 7 & 45 & 63 \\ 45 & 359 & 389 \\ 63 & 389 & 595 \end{vmatrix} \qquad (X'X)^{-1}=\begin{vmatrix} 5\ 216/877 & -\ 81/424 & -\ 152/353 \\ -\ 81/424 & 7/424 & 1/106 \\ -\ 152/353 & 1/106 & 3/73 \end{vmatrix}$$

$$X'Y=\begin{vmatrix} 230 \\ 1642 \\ 2067 \end{vmatrix} \qquad b=\begin{vmatrix} 2\ \ 296/319 \\ 2\ \ 71/106 \\ 1\ \ 136/325 \end{vmatrix} \qquad b=\begin{vmatrix} 2.927898 \\ 2.669811 \\ 1.418464 \end{vmatrix}$$

HKT=Y'Y-b'X'Y

$$Y'Y= 8022 \qquad b'=\begin{vmatrix} 2.927898 & 2.669811 & 1.418464 \end{vmatrix}$$

b'X'Y= 7989.211

HKT= 8022 –7989.211 = 32.78908

s^2 =HKT/(n-k) =32.78908/(7-3) = 8.197271

var-cov(b) = s^2(X'X)-1

$$(X'X)^{-1}= 8.197271 \begin{vmatrix} 5\ 216/877 & -81/424 & -152/353 \\ -\ 81/424 & 7/424 & 1/106 \\ -\ 152/353 & 1/106 & 3/73 \end{vmatrix}$$

$$\text{var-cov(b)}=\begin{vmatrix} 43\ \ 1/189 & -1\ 223/394 & -3\ 339/640 \\ -1\ 223/394 & 18/133 & 29/375 \\ -3\ 339/640 & 29/375 & 31/92 \end{vmatrix}$$

$$\text{var-cov (b)}=\begin{vmatrix} 43.0053 & -1.5660 & -3.5297 \\ -1.5660 & 0.1353 & 0.0773 \\ -3.5297 & 0.0773 & 0.3369 \end{vmatrix}$$

$$s_b=\begin{vmatrix} 6.5578 & & \\ & 0.3679 & \\ & & 0.5805 \end{vmatrix}$$

s_{b0}=6.5578 s_{b1}=0.3679 s_{b2}=0.5805

n=7 $\bar{Y}$ =32.8571 $\bar{Y}^2$=1080.25

Anova

Değişkenlik Kaynağı	*Kareler Toplamı*	*Serbestlik Derecesi*	*Kareler Ortalaması*	*F*
RBD	$\hat{b}'X'Y - n\overline{Y}^2$	k-1	a=($\hat{b}'X'Y - n\overline{Y}^2$)/(k-1)	a/b
HBD	$Y'Y - \hat{b}'X'Y$	n-k	b=($Y'Y - \hat{b}'X'Y$)/(n-k)	
TD	$Y'Y - n\overline{Y}^2$	n-1		

Anova

Değişkenlik Kaynağı	*Kareler Toplamı*	*Serbestlik Derecesi*	*Kareler Ortalaması*	*F*
RBD	7989.211-7(1080.25) =427.461	3-1=2	a=213.7305	26.07
HBD	8022 –7989.211 =32.78908	7-3=4	b=8.19727	
TD	8022-7(1080.25) =460.25	7-1=6		

4. MODEL TANIMLAMA

4.1. Nasıl Bir Modelleme

Bir ekonometristin birinci görevinin, ekonometrik modeli formüle etmek olduğunu belirtmiştik. Model, gerçek olayın yalınlaştırılmış veya basitleştirilmiş bir gösterimi olarak tanımlanabilir. Örneğin; portakal talebini, portakal fiyatlarına bağlı olarak modelleyebiliriz. Aslında portakal talebini etkileyen çok sayıda başka değişkenler de vardır. Örneğin, tüketicilerin geliri, diyet yapma bilinci, elma fiyatı vb. Kuşkusuz diğer değişkenlerin sonu yoktur. Öyle ki benzin fiyatlarındaki değişme bile, portakal talebini etkileyebilir.

Ekonometristler basitliğin derecesini tartışmaktadır. Basit modellerin anlaşılması, iletişimi ve verilerin deneysel olarak test edilmesi daha kolaydır. Karl Popper ve Milton Fredman, modellerin olabildiğince basit olması gerektiğini savunmaktadır. Ancak gerçek dünyanın karmaşık olaylarını açıklamak için basit bir modelden yararlanması, iki noktada eleştiriye maruz kalmaktadır:

- Model, olması gerektiğinden fazla basitleştirilebilir
- Basitleştirme varsayımları gerçekdışı olabilir

Verilen örnekteki portakal talebi için, talebin yalnızca portakal fiyatına bağlı olduğunu söylemek fazla basitleştirmek olduğu kadar, gerçekçi olmayan bir varsayımdır.

Basitleştirme yanlıları, basitleştirilmiş bir modelle başlayıp, giderek daha karmaşık modellere ulaşılmasını daha doğru bulmaktadır. Koopmans bunun savunucularındandır. Diğer yandan, bazı ekonometristler ise çok genel bir modelle başlayıp, modelin giderek basitleştirilmesini savunmaktadır. Örneğin, L.J. Savage "bir model, bir fil kadar büyük olmalıdır" demektedir.

Fredman'a göre, varsayımların gerçekçi olup olmadığını belirlemenin kesin bir yolu yoktur. Ona göre, bir teori hakkındaki varsayımlar tam olarak gerçekçi olamasa bile, amaca başarılı yaklaşımlar sağlaması yeterlidir. Bunu anlamak için teorinin işleyip işlemediğine bakılabilir. Portakal örneğine geri dönecek olursak: Portakal talebinin yalnızca portakal fiyatlarına bağlı olması gerçekçi bir varsayım olmayabilir. Bununla birlikte modele elma fiyatı, gelir gibi değişkenlerin eklenmesi de onu daha gerçekçi hale getirmeyebilir. Zira böyle bir durumda da, çok sayıda değişkenin dışarıda kalabileceği açıktır. Sonuçta hangi modelin portakal talebini tahmin için daha yararlı olacağını belirlemek için, sahip olduğumuz ya da olabileceğimiz değişkenlerden yola çıkarak karar vermek daha doğru olabilir.

Biz bu kitapta, basit modellerle başlayıp, giderek daha karmaşık modeller kurmayı deneyeceğiz.

4.2. Model Tanımlama Esasları

Bir denklem tahmin edilmeden önce, kuşkusuz model tanımlamasının doğru bir şekilde yapılması gerekir. Ekonometrik model tanımlamada üç temel seçim söz konusudur:

- Doğru bağımsız değişkenlerin seçilmesi
- Doğru fonksiyonel formun belirlenmesi
- Hata teriminin formunun doğru seçilmesi

Bunlardan birindeki yanlışlık, model tanımlamanın da yanlış olması sonucunu doğurur. Araştırıcının doğru değişkenleri seçmesi, ekonometrinin hem zayıf hem de güçlü yönünü teşkil eder. Güçlü yönü, araştırıcının denklemleri bireysel ihtiyaçlarını giderecek şekilde formüle edilebilmesinde; zayıf yönü ise, aksini kanıtlayan pek çok sonuç olsa bile araştırıcının düşüncesini kanıtlayan çok sayıda modeli tahmin edebilmesinde yatmaktadır.

Bir değişken modele alınırken dikkat edilecek en önemli husus, ekonomi teorisi açısından o değişkenin temel nitelikte olmasıdır. Eğer ekonomi teorisinin kaçınılmaz bir unsuruysa, tahminleme sonucunda istatistiki açıdan sıfır bulunsa bile, o değişken mutlaka modelde tutulmalıdır. Ekonomi teorisinin açık bir görüş belirtmediği değişkenlerde ise kararsızlık yaşanabilir.

Model tanımlaması sırasında değişken seçimi yaparken, aşağıdaki dört soru dikkate alınmalıdır:

1. *Teori:* Değişkenin denklemdeki yeri kaçınılmaz ve teorik olarak güçlü müdür?
2. *t-testi:* Değişkenin tahmincisi, beklenen işarete sahip ve istatistiki olarak önemli mi?
3. *Düzeltilmiş R^2 :* Değişkenin modele girmesi, düzeltilmiş R^2'yi yükseltiyor mu?
4. *Sapma:* Değişken modele alındığında, diğer değişkenlerin katsayıları önemli ölçüde değişiyor mu?

Eğer tüm bu soruların cevabı, evet ise o değişken modele aittir. Soruların hepsine, hayır cevabı veriliyorsa, o değişken modele alınmaz.

4.3. Bağımsız Değişkenlerin Seçimi

4.3.1. İhmal Edilen Değişkenler

Bir araştırıcı, kurduğu modele, olması gereken tüm değişkenleri dahil etmeyebilir. Bazan da, önemli değişkenlerden birine ait verileri bulamayabilir. Hangi nedenle olursa olsun, modelde bulunması gereken bir değişken model dışı bırakılmışsa, bu değişkene *ihmal edilmiş* değişken denir.

İhmal edilmiş bir değişken, modeli kuşkulu hale getirir. Örneğin arz veya talep fonksiyonlarında fiyatın olmaması düşünülemez. Israrla fiyatsız bir arz veya talep modeli üzerinde çalışmak, modeldeki diğer değişkenlerin tahmin edilen katsayılarında sapmaya yol açar. Buna, ihmal edilmiş değişken sapması veya tanımlama sapması denir.

Herhangi bir ihmal edilmiş değişkenin bulunmadığı *doğru* modelin:

$$Y = b_0 + b_1X_1 + b_2X_2$$

olduğunu varsayalım. Ancak tahmin edilen model:

$$Y = b_0 + b_1X_1$$

olsun. Bir başka ifadeyle, b_2 sıfırdan farklı olmayan bir değişkenin tahmincisi olmasına rağmen, bu değişkeni tahmin ettiğimiz modele almadık. Bu durumda:

- Modele dahil edilen değişkenlerin tahmincileri sapmalıdır. İhmal edilen değişken, modeldeki değişkenlerden en az biriyle korelasyon halinde olmadığı sürece, sapma söz konusu olacaktır.

- İhmal edilmiş değişkenle, modeldeki değişken(ler) arasında bir korelasyon olsa bile, sabit terim sapmalı olacağından, kestirimler sapmalı olacaktır.
- Modele alınmış bir değişkene ait tahmincinin varyansı sapmalı olacağından, hipotez testleri de geçersiz olabilecektir.

İhmal edilmiş değişken probleminin çözümü zor değildir. İlgili ekonomi teorisine göre olması zorunlu değişkenler incelendiğinde, bu değişken kolayca belirlenebilir. Böylesi ihmaller, tahmincilerin beklenenin tersine bir işarete sahip olmasına bile neden olabilir. Ancak, küçük sapmalara yol açan ihmal edilmiş değişkenlerin bulunması oldukça zordur.

Bazı araştırmacılar, ihmal edilmiş değişken problemini farkettiklerinde, ilk çözüm olarak, olası tüm değişkenleri modele dahil etmektedirler. Ancak bu kez de gereksiz değişkenlerin modele dahil edilmesi problemi ortaya çıkabilecektir. Araştırmacıların bir başka çözüm yolu da, önemli gördükleri değişkenleri modelde sabit tutup, doğru işaretleri bulana kadar diğer değişkenleri, modele alıp çıkarmaktır. Ancak bu yolla beklentiye uygun bir model yakalamak, büyük ihtimalle rastlantı olacaktır.

İhmal edilmiş değişken probleminin en önemli nedenleri:

- Teorik yapının yeterince iyi bir şekilde ortaya konulmaması
- Örneklemenin zayıf olmasıdır.

İhmal edilmiş bir değişken modele alındığında:

- Düzeltilmiş R^2 yükselecek
- Diğer değişkenlerin tahmincileri, önemli ölçüde değişecektir.

Örnek 4-1

Bu örnekte tavuk eti talebini modelleyeceğiz. Model değişkenlerimiz:

KBTET: Kişi Başına Tavuk Eti Tüketimi
Fiyat: Tavuk Eti Fiyatı
Gelir: Kişi Başına Gelir
SEF: Sığır eti fiyatı

Çizelge 4-1: Kişi Başına Tavuk Eti Tüketimi

Yıl	Kişi BaşınaTavuk Eti Tüketimi Y	Tavuk Eti Fiyatı X1	Kişi Başına Gelir X2	Sığır Eti Fiyatı X3
1977	27.8	42.2	397.5	78.3
1978	29.9	38.1	413.3	79.2
1979	29.8	40.3	439.2	79.2
1980	30.8	39.5	459.7	79.2
1981	31.2	37.3	492.9	77.4
1982	33.3	38.1	528.6	80.2
1983	35.6	39.3	560.3	80.4
1984	36.4	37.8	624.6	83.9
1985	36.7	38.4	666.4	85.5
1986	38.4	40.1	717.8	93.7
1987	40.4	38.6	768.2	106.1
1988	40.3	39.8	843.3	104.8
1989	41.8	39.7	911.6	114
1990	40.4	52.1	931.1	124.1
1991	40.7	48.9	1021.5	127.6
1992	40.1	58.3	1165.9	142.9
1993	42.7	57.9	1349.6	143.6
1994	44.1	56.5	1449.4	139.2
1995	46.7	63.7	1575.5	165.5
1996	50.6	61.6	1759.1	203.3
1997	50.1	58.9	1994.2	219.6
1998	51.7	66.4	2258.1	221.6
1999	52.9	70.4	2478.7	232.6

Burada doğru modelin:

$$KBTET = b_0 + b_1 Fiyat + b_2 Gelir$$

olduğu açıktır. Ancak biz talep teorisinin can damarı olan gelir değişkenini gözardı edip, modeli:

$$KBTET = b_0 + b_1 Fiyat$$

şeklinde yanlış tanımladığımızı varsayalım. Bu modelin tahmin sonuçları:

	Katsayı	St.Hata	t	p
Sabit	12.933	3.865	3.35	0.003
Fiyat	0.55706	0.07853	7.09	0.000
S = 4.095		R^2 =0.706		

Görüldüğü gibi, b_1 tahmincisi, istatistiki açıdan anlamlıdır. Ancak gerek b_0, gerekse b_1 tahmincisi, sapmalıdır. Gerçekten de, tavuk eti tüketimiyle fiyat arasında ters yönlü bir ilişki beklenirken, pozitif bir tahminci elde edilmiştir. Bu nedenle, b_1 tahmincisinin test sonucuna güvenmemiz de beklenemez.

Kuşkusuz, talep teorisinin önemli bir ögesi olan gelir gibi olmazsa olmaz bir değişken ihmal edilmiştir. Şimdi gelir değişkenini de dahil ederek, modelimizi yeniden tahmin edelim:

$KBTET = b_0 + b_1 Fiyat + b_2 Gelir$

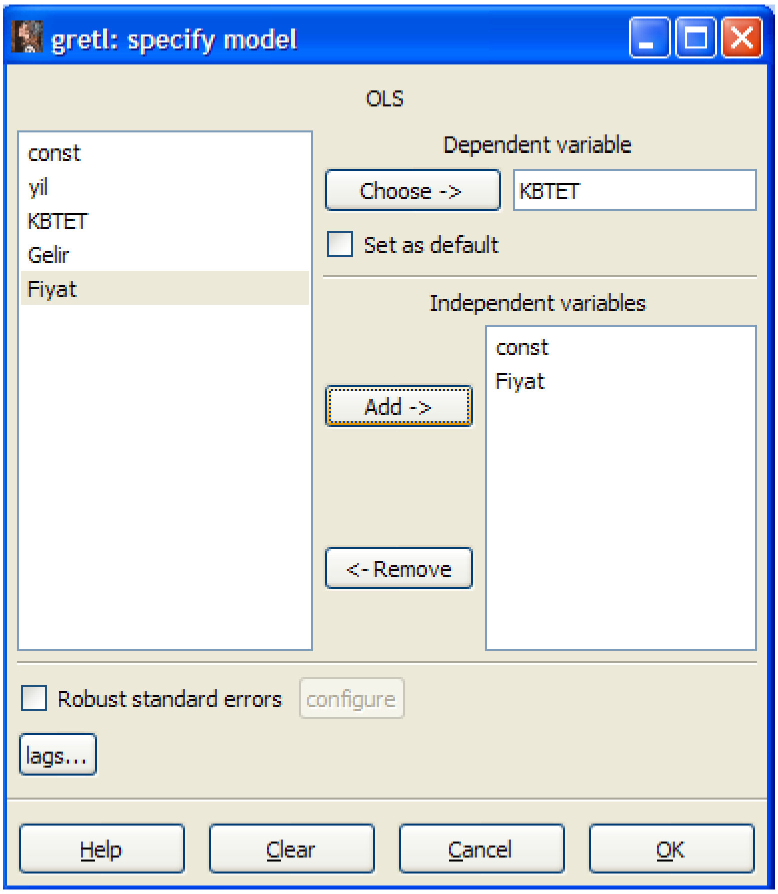
	Katsayı	St.Hata	t	p
Sabit	34.516	3.856	8.95	0.000
Fiyat	-0.2136	0.1219	-1.75	0.095
Gelir	0.014884	0.002193	6.79	0.000
$S = 2.309$		$R^2 = 0.845$		

Doğru modelimizin sonuçları, beklendiği gibidir; gerçekten de bu modelde fiyatın işareti negatif, gelirin işareti pozitiftir. Hem b_0, hem de b_1, %10 düzeyinde istatistiki açıdan sıfırdan farklıdır.

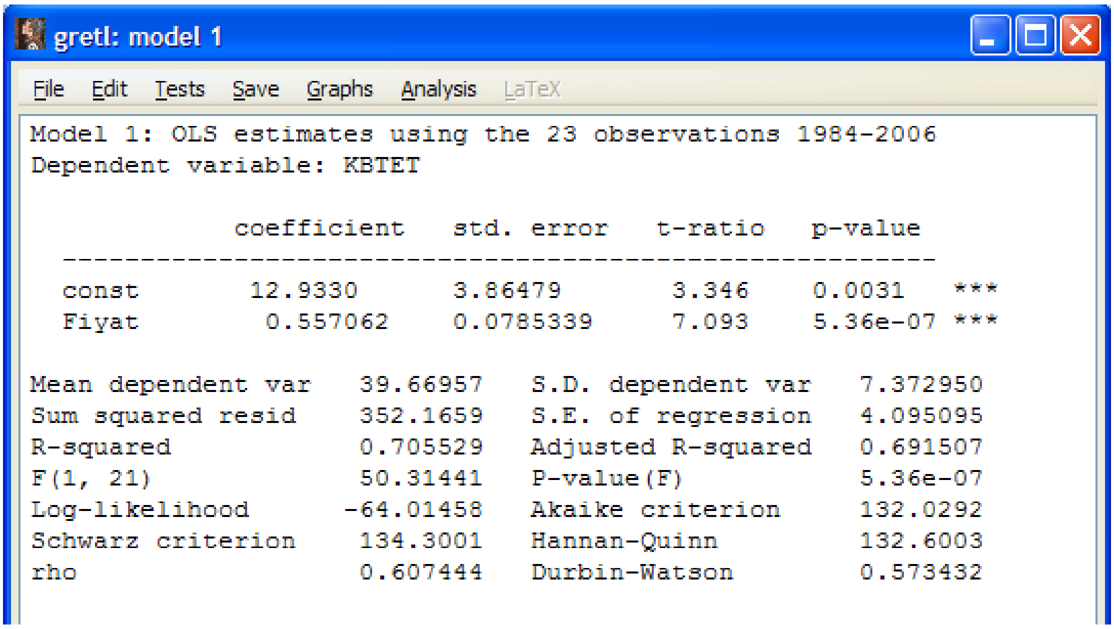
Şimdi bu süreci Gretl'da deneyelim:

Başlangıç modelimiz: $KBTET = b_0 + b_1 Fiyat$

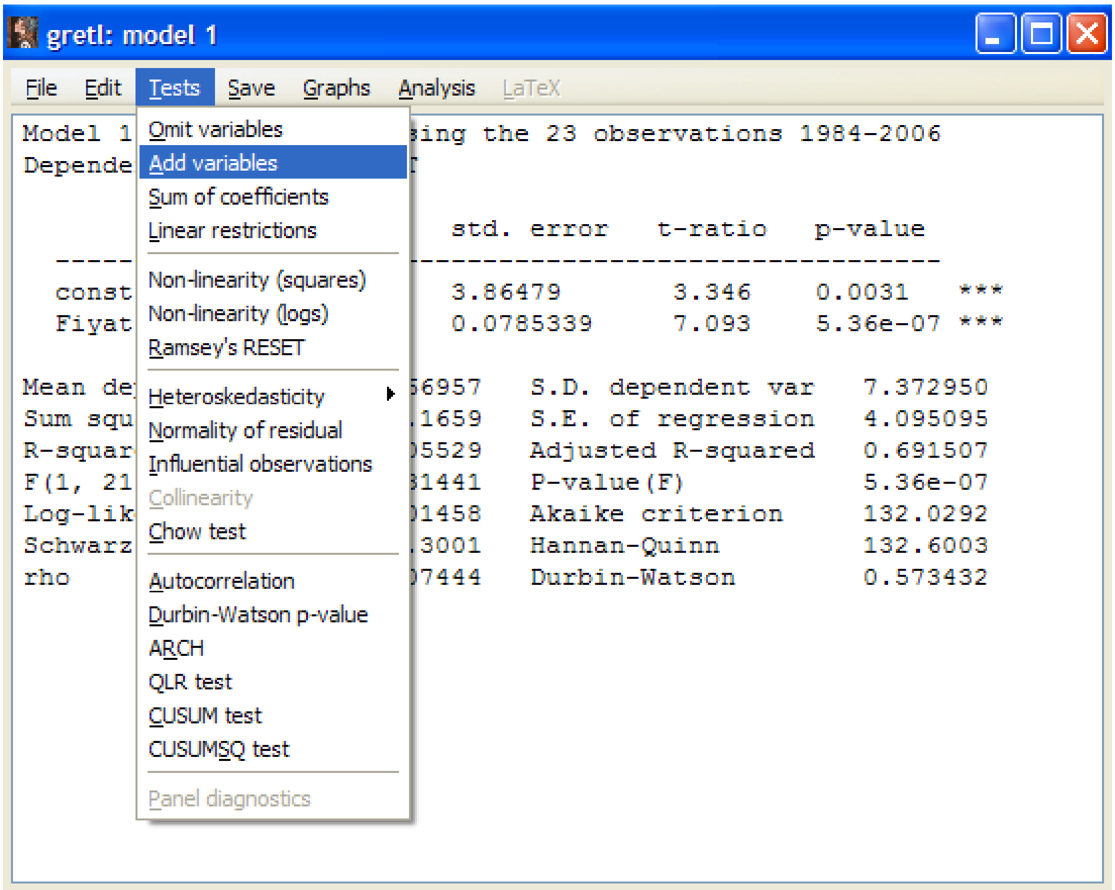
İhmal ettiğimiz değişken Gelir'dir.



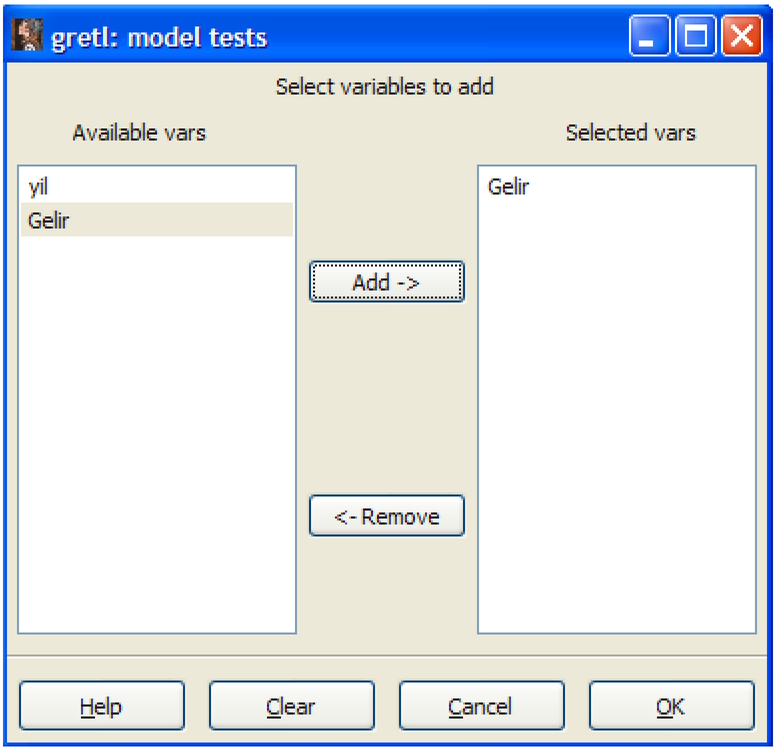
Başlangıç modelimizin EKK tahmini:



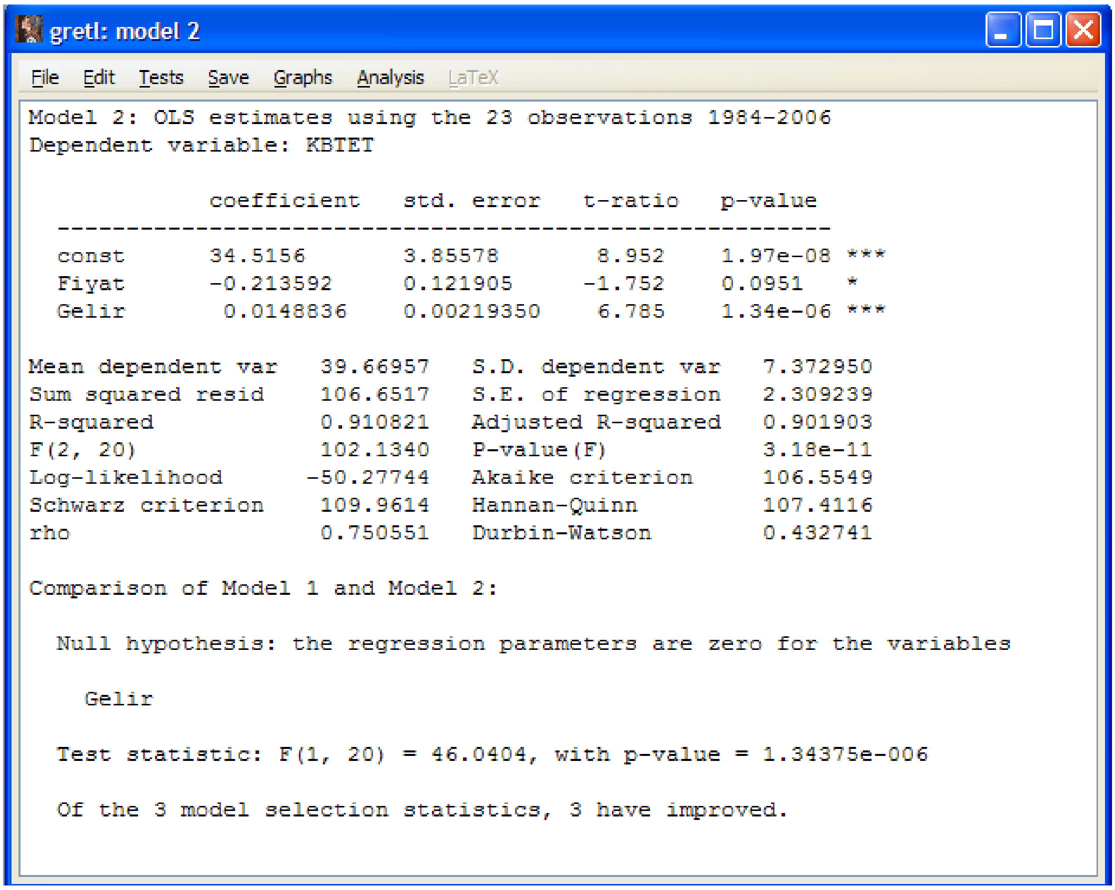
İhmal edilen değişken testi, **Tests/Add variables** menüsünden yapılır.



İhmal edilen değişkenimizi işaretleyip Add -> düğmesini tıklayarak seçelim.



İhmal edilen değişkenin EKK tahminindeki katsayısı istatistiki açıdan anlamlıdır. Başlangıç modeli (Model 1) ve gelir değişkeninin eklendiği Model 2 karşılaştırması, F testi ile yapılmaktadır. Sıfır hipotezi, eklenen değişken(ler)in katsayı(sı/ları)nın sıfır olduğunu kabul eder. Örneğimizde üç model seçim kriterinden üçü iyileşmiştir mesajıylala birlikte, F testi, sıfır hipotezini reddetmektedir. O halde Gelir ihmal edilen değişkendir. Bir başka ifadeyle, Gelir mutlaka modelde yer almalıdır.

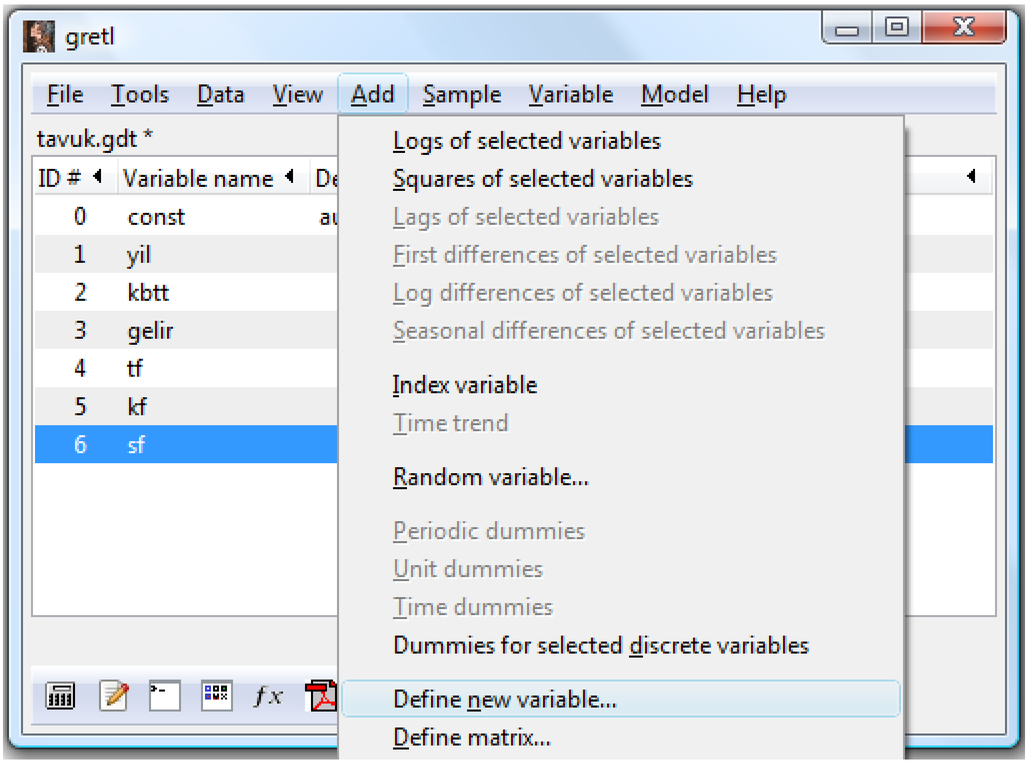


gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

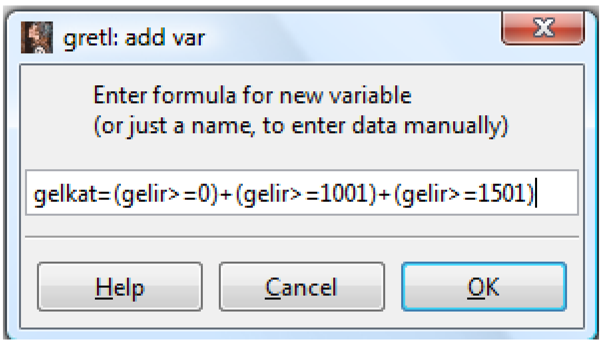
```
Model 2: OLS estimates using the 23 observations 1984-2006
Dependent variable: KBTET

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const      34.5156       3.85578       8.952    1.97e-08 ***
  Fiyat      -0.213592     0.121905     -1.752    0.0951   *
  Gelir       0.0148836    0.00219350    6.785    1.34e-06 ***

Mean dependent var    39.66957   S.D. dependent var    7.372950
Sum squared resid     106.6517   S.E. of regression    2.309239
R-squared             0.910821   Adjusted R-squared    0.901903
F(2, 20)              102.1340   P-value(F)            3.18e-11
Log-likelihood       -50.27744   Akaike criterion      106.5549
Schwarz criterion     109.9614   Hannan-Quinn          107.4116
rho                   0.750551   Durbin-Watson         0.432741

Comparison of Model 1 and Model 2:

  Null hypothesis: the regression parameters are zero for the variables

    Gelir

  Test statistic: F(1, 20) = 46.0404, with p-value = 1.34375e-006

  Of the 3 model selection statistics, 3 have improved.
```

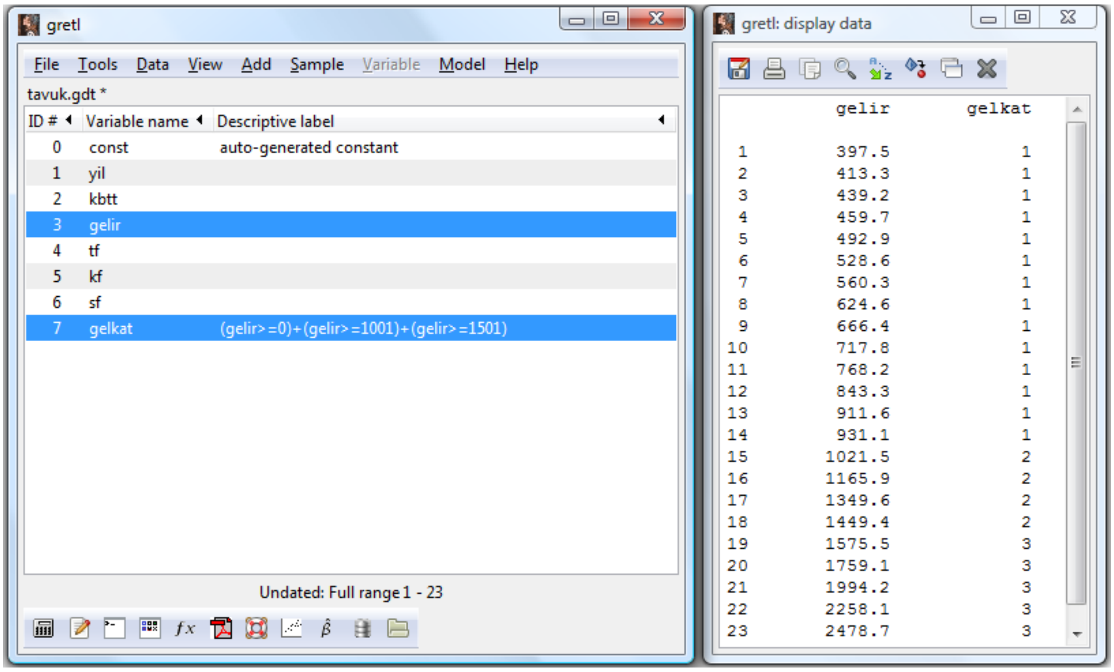
Gelir değişkeni bazen kategorik olarak modele girebilir. Özellikle bağımlı değişkenin gelir kategorilerine göre değişimi incelemek gerektiğinde, gelir değişkeni belli kıstaslar kullanılarak kategorize edilebilir. Biz burada gelir kategorilerinin daha önceden belirlendiğini varsayacağız. Buna göre gelir 0 ile 1000 arasında ise 1. kategori, 1001 ile 1500 arasında ise 2. kategori, 1500'den büyükse 3. kategori olarak tanımlanmış olsun. Gretl bu işlemi basit bir komutla gerçekleştirebilir. Bunun için önce *Add/Define new variable…* seçeği tıklanır:



Ardından gelir kategorisi değişkenini temsil eden **gelkat** değişkeni aşağıdaki formülle tanımlanır:

$$\text{gelkat} = \underbrace{(\text{gelir}>=0)}_{\text{Kategori 1}} + \underbrace{(\text{gelir} >= 1001)}_{\text{Kategori 2}} + \underbrace{(\text{gelir} >= 1501)}_{\text{Kategori 3}}$$





4.3.2. *Gereksiz Değişkenler*

İhmal edilmiş değişkenlerin aksine, modelde olmaması gereken değişken değişkenler, modele dahil edilmiş olabilir. Bu değişkenlere, gereksiz değişkenler denmektedir.

Tahminleme sonucunda, modeldeki bir değişkenin mutlak değer t'si 1'den küçükse, bu değişkenin modelden çıkarılması, düzeltilmiş R^2'yi yükseltir; mutlak değer t'si 1'den büyükse, bu değişkenin modelden çıkarılması, düzeltilmiş R^2'yi düşürür.

Herhangi bir ihmal edilmiş değişkenin bulunmadığı *doğru* modelin:

$$Y = b_0 + b_1X_1$$

olduğunu varsayalım. Ancak tahmin edilen model:

$$Y = b_0 + b_1X_1 + b_2X_2$$

olsun. X_2 modelde olmaması gereken bir değişkendir. Böyle bir değişkenin tahmincisi, istatistiki açıdan sıfırdır. Herhangi bir modelde gereksiz değişken bulunması halinde:

- Diğer değişkenler yansız ve tutarlıdır
- Modelin varyansı, gereksiz değişkenlerin olmadığı duruma göre, daha yüksek olacak; dolayısıyla tahminciler etkin olmayacaktır.
- Tahmincilerin varyansları yansız olduğundan, hipotez testleri geçerlidir.

Eğer gereksiz bir değişken modele dahil edilirse:

- Tahmincisi istatistiki açıdan önemsiz olacak
- Düzeltilmiş R^2 düşecek
- Diğer değişkenlerin tahmincileri, çok az değişecektir.

Örnek 4-2

Çizelge 4-1'deki veri setinden yararlanarak, gereksiz değişken irdelemesi yapalım.

Burada doğru modelin:

$$KBTET = b_0 + b_1Fiyat + b_2Gelir$$

olduğunu kabul edelim. Ancak biz sığır eti fiyatı değişkenini de (SEF) modele alıp, yanlış model olarak kabul edelim:

$$KBTET = b_0 + b_1Fiyat + b_2Gelir + b_3SEF$$

Bu modelin tahmin sonuçları:

	Katsayı	St.Hata	t	p
Sabit	33.799	4.287	7.88	0.000
TF	-0.2223	0.1262	-1.76	0.094
Gelir	0.01298	0.0050	2.58	0.018
SEF	0.02499	0.0591	0.42	0.677
S = 2.358		R^2= 0.912	Düz. R^2= 0.898	

Sığır eti fiyatı tahmincisi (b_3), istatistiki açıdan sıfırdır. Buna göre, modelde yer almasa da olabilir. O halde sığır eti fiyatının modelde olmadığı tahmin sonuçlarını hatırlayalım:

$KBTET = b_0 + b_1Fiyat + b_2Gelir$

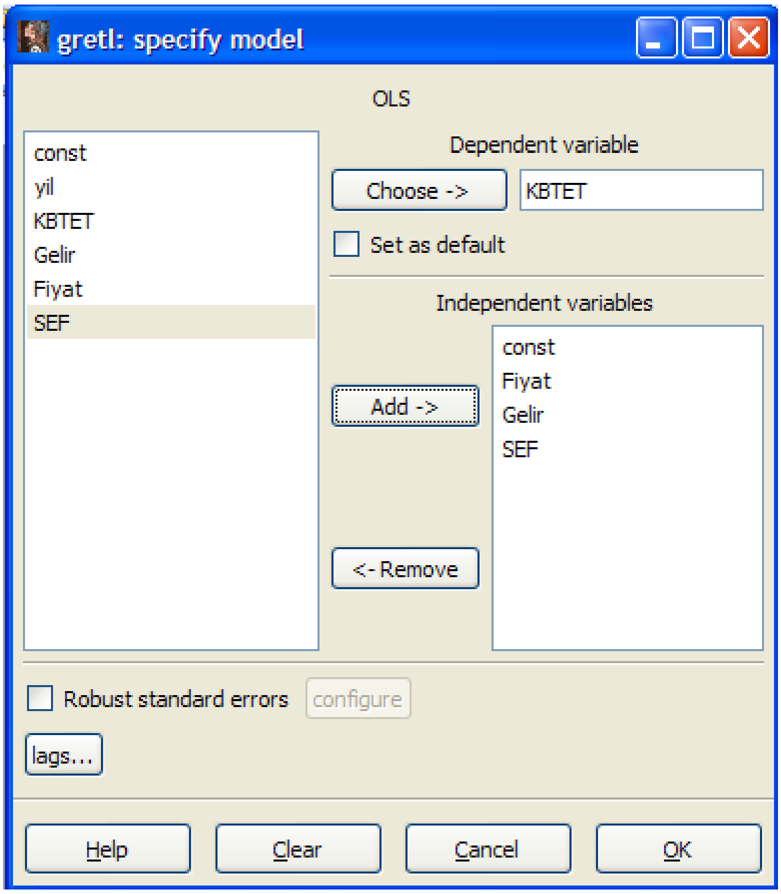
	Katsayı	St.Hata	t	p
Sabit	34.516	3.856	8.95	0.000
Fiyat	-0.2136	0.1219	-1.75	0.095
Gelir	0.0149	0.0022	6.79	0.000
S = 2.309		R^2 = 0.845	Düz. R^2=0.902	

Dikkat edilirse, yanlış modelin varyansı daha yüksek, düzeltilmiş R^2 değeri daha düşüktür. Yanısıra, yanlış modelin hipotez testleri hala geçerlidir.

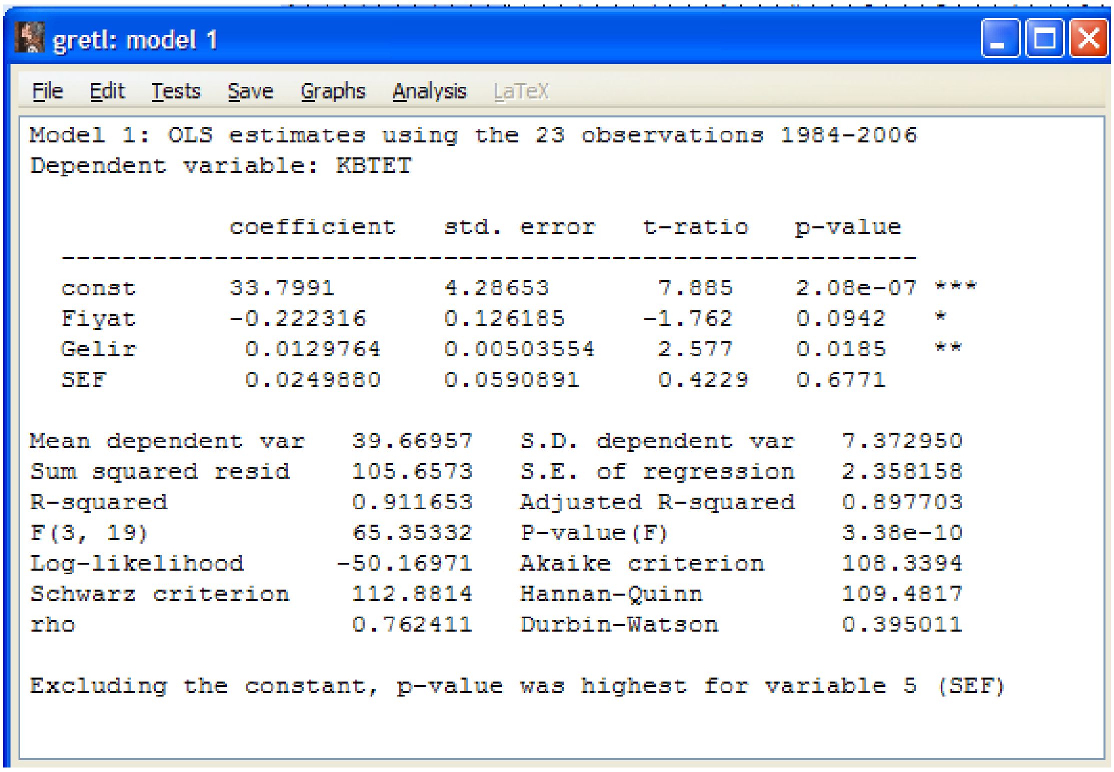
Bu süreci Gretl yardımıyla yapmayı deneyelim:

Başlangıç modelimiz: $KBTET = b_0 + b_1Fiyat + b_2Gelir + b_3SEF$

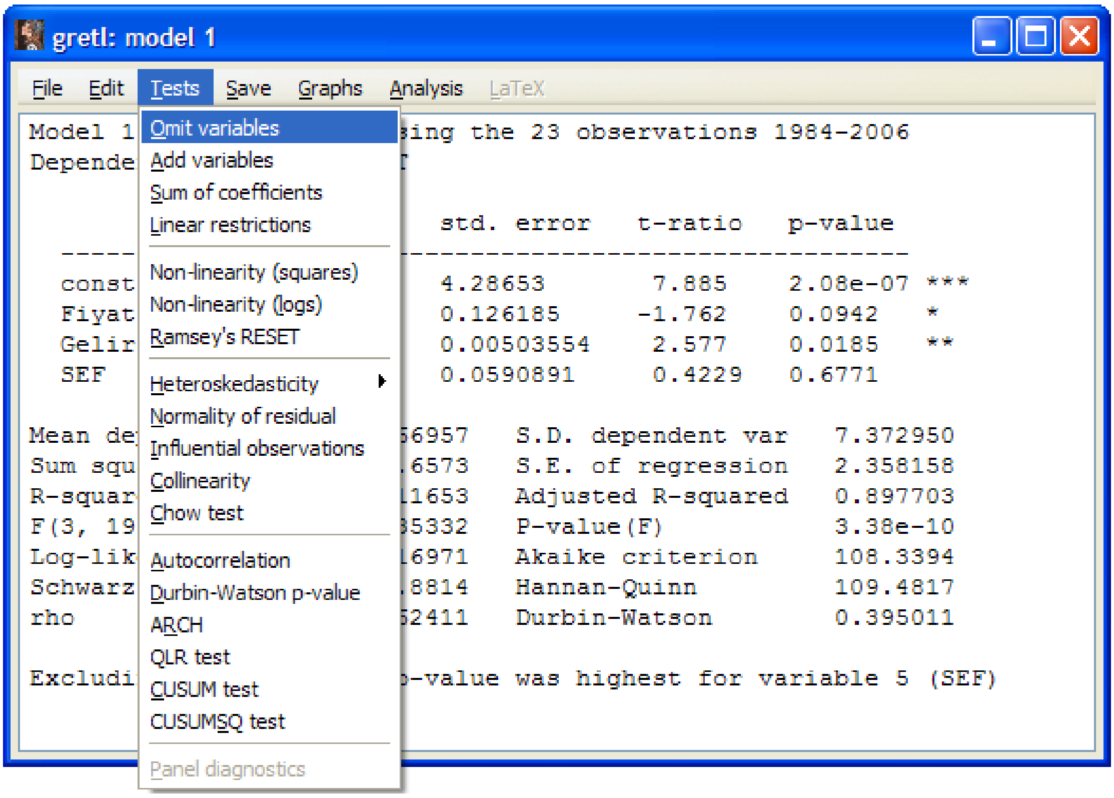
SEF gereksiz değişken olarak belirlenmiştir.



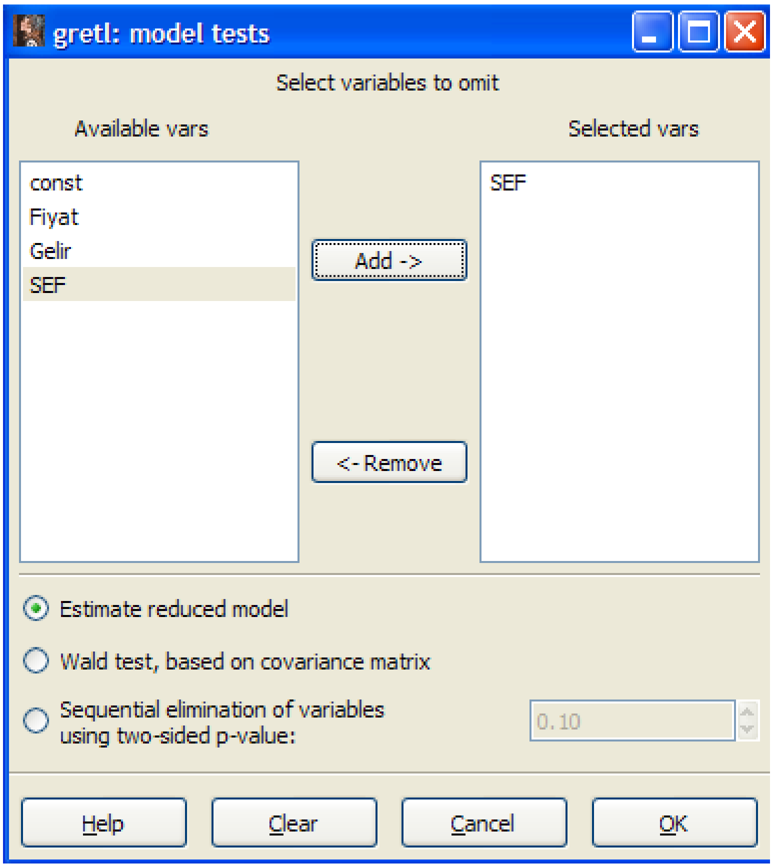
EKK tahminimiz SEF değişkeninin gereksiz olduğunu doğrulamaktadır.



SEF değişkeninin gereksiz olup olmadığı **Tests/Omit variables** menüsünden test edilir.



SEF değişkeni gereksiz değişken olarak seçilir.



Başlangıç modeli (Model 1) ve SEF değişkeninin çıkarıldığı Model 2 karşılaştırması, F testi ile yapılmaktadır. Sıfır hipotezi, çıkarılan değişken(ler)in katsayı(sı/ları)nın sıfır olduğunu kabul eder. Örneğimizde üç model seçim kriterinden üçü iyileşmiştir mesajıylala birlikte, F testi, sıfır hipotezini reddetmemektedir. O halde SEF sıfırdır; yani gereksiz değişkendir. Bir başka ifadeyle, SEF modelden çıkarılmalıdır.

```
gretl: model 2
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 2: OLS estimates using the 23 observations 1984-2006
Dependent variable: KBTET

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const      34.5156       3.85578       8.952    1.97e-08 ***
  Fiyat      -0.213592     0.121905     -1.752    0.0951   *
  Gelir       0.0148836    0.00219350    6.785    1.34e-06 ***

Mean dependent var   39.66957   S.D. dependent var   7.372950
Sum squared resid    106.6517   S.E. of regression   2.309239
R-squared            0.910821   Adjusted R-squared   0.901903
F(2, 20)             102.1340   P-value(F)           3.18e-11
Log-likelihood      -50.27744   Akaike criterion     106.5549
Schwarz criterion    109.9614   Hannan-Quinn         107.4116
rho                  0.750551   Durbin-Watson        0.432741

Comparison of Model 1 and Model 2:

  Null hypothesis: the regression parameters are zero for the variables

    SEF

  Test statistic: F(1, 19) = 0.178833, with p-value = 0.677126

  Of the 3 model selection statistics, 3 have improved.
```

4.3.3. Uygulamalar

Örnek 4-3

Aşağıdaki verileri kullanarak talep modellemesi yapacağız:

Talep	13	14	13	10	9	9	8	7	9	5	5	3	2	2	5
Fiyat	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20

$\text{Talep} = b_0 + b_1\text{Fiyat}$

Modeli için EKK sonuçları:

	Katsayı	St.Hata	t	p
Sabit	17.32137	1.041005	16.63908	0.000
Fiyat	-0.76346	0.076758	-9.94633	0.000
S = 1.3878		$R^2 = 0.88$	Düz. R^2=0.87	

Model sonuçları, beklenen işaretlere sahip ve istatistiki açıdan anlamlıdır.

Örnek 4-4

Aşağıdaki verileri kullanarak yeni bir talep modellemesi yapacağız:

Fiyat	7.8	7	5.7	5.2	6.3	5.3	6.4	7.9	6.5	6.9	4.9	5.1	4.7	6.4	6.9

Talep	18	19	20	22	20	20	21	18	20	20	21	22	20	21	20

Talep = b_0 + b_1Fiyat

modeli için EKK sonuçları:

	Katsayı	St.Hata	t	p
Sabit	25.49295	1.367991	18.63532	0.000
Fiyat	-0.86445	0.217928	-3.96669	0.000
S = 0.8287		R^2 = 0.55	Düz. R^2=0.51	

Model, beklenen işaretlere sahip ve istatistiki açıdan anlamlıdır.

Örnek 4-5

Aşağıdaki verileri kullanarak bir başka talep modellemesi yapacağız:

Talep	75	79	58	51	63	74	51	71	76	70	64	62	64	60	40
Fiyat	71	54	89	93	76	61	92	60	57	41	73	69	75	80	97

Talep = $b_0 + b_1$Fiyat

Modeli için EKK sonuçları:

	Katsayı	St.Hata	t	p
Sabit	105.9762	7.033228	15.06793	0.000
Fiyat	-0.58055	0.094812	-6.12322	0.000
S = 5.709		R^2 = 0.74	Düz. R^2=0.72	

Model, beklenen işaretlere sahip ve istatistiki açıdan anlamlıdır.

Örnek 4-6

Aşağıdaki verileri kullanarak yeni bir talep modellemesi yapacağız:

Fiyat	24	37	41	21	39	41	37	35	23	39	21	31	41	26	18
Talep	25	34	28	23	32	29	30	30	41	29	44	36	31	41	43

Talep = $b_0 + b_1$Fiyat

Modeli için EKK sonuçları:

	Katsayı	StHata	t	p
Sabit	44.32665	6.179661	7.172991	0.000
Fiyat	-0.35633	0.189236	-1.88298	0.082
S = 6.03661		R^2 = 0.21	Düz. R^2=0.15	

Modelde fiyat değişkeni beklenen işaretli ve istatistiki açıdan anlamlıdır.

Örnek 4-7

Aşağıdaki verileri kullanarak üç açıklayıcılı yeni bir talep modellemesi yapacağız:

Talep	Fiyat	Gelir	İkame Fiyat	Talep	Fiyat	Gelir	İkame Fiyat
28.4	5	5	9	8.68	4	4	2
21.9	4	5	8	5.02	5	6	1
23.4	3	5	7	11.44	6	12	2
19.92	6	6	8	7.18	5	4	3
8.72	7	6	6	8.5	6	5	4
7.4	6	5	5	9.9	5	5	4
18.5	3	5	6	11.12	4	6	4
15.78	3	4	5	9.8	5	5	5
19.1	2	5	4	10.48	6	4	6
4.02	7	6	3	8.8	7	5	7

Talep = $b_0 + b_1$Fiyat

modeli için EKK sonuçları:

	Katsayı	StHata	t	p
Sabit	25.068	4.713	5.319	0.000

Fiyat	-2.458	0.915	-2.687	0.015
S = 5.853	R^2 = 0.29	Düz. R^2=0.25		

Model, beklenen işaretlere sahip ve istatistiki açıdan anlamlıdır. Ancak R^2 çok düşüktür.

Talep = $b_0 + b_1$Fiyat + b_2Gelir

modeli için EKK sonuçları:

	Katsayı	StHata	t	p
Sabit	23.833	5.789	4.117	0.001
Fiyat	-2.564	0.977	-2.625	0.018
Gelir	0.327	0.845	0.387	0.704
S = 5.853	R^2 = 0.29	Düz. R^2=0.21		

Model, beklenen işaretlere sahiptir. Ancak fiyatın tahmincisi istatistiki açıdan anlamlı olmakla birlikte gelirin tahmincisi anlamsızdır. R^2'nin düşük olması da dikkati çekmektedir.

Talep = $b_0 + b_1$Fiyat + b_2Gelir+ b_3IkameFiyat

modeli için EKK sonuçları:

	Katsayı	St.Hata	t	p
Sabit	7.732	2.799	2.763	0.014
Fiyat	-2.880	0.383	-7.518	0.000
Gelir	1.300	0.345	3.772	0.002
IkameFiyat	2.506	0.257	9.766	0.000
S = 2.343	R^2 = 0.90	Düz. R^2=0.88		

Model, beklenen işaretlere sahiptir. Aynı zamanda tüm değişkenlerin tahmincileri istatistiki açıdan anlamlıdır. R^2'nin önceki iki modele göre oldukça yükselmiş olması da dikkati çekmektedir.

Örnek 4-8

Şimdi beyaz inek peyniri arzının modelle denemelerini yapacağız. Model değişkenlerimiz:

ARZ : Beyaz inek peyniri arzı
FIYAT : Beyaz inek peyniri fiyatı
IKAMEFIYAT : Koyun peyniri fiyatı
SUTFIYAT: İnek sütü fiyatı (Maliyet)

Beyaz inek peyniri arzı	Fiyat	Koyun peyniri fiyatı	İnek sütü fiyatı
18	9	12	7
22	8	10	7
18	7	8	7
18	8	7	9
21	7	8	6
27	11	8	9
27	12	9	8
10	6	9	8
18	8	9	7
31	11	8	7
16	7	8	9

25	12	11	8
29	9	9	7
13	7	8	8
26	9	10	7
20	9	11	8
29	10	8	7
24	9	8	7
25	10	10	9

Önce:

$$Arz = b_0 + b_1 Fiyat$$

modelini tahmin edelim:

	Katsayı	StHata	t	p
Sabit	-1.313	4.317	-0.304	0.765
FIYAT	2.615	0.477	5.487	0.000
S = 3.559		R^2 = 0.64	Düz. R^2=0.62	

Model, beklenen işaretlere sahiptir ve fiyatın tahmincisi istatistiki açıdan anlamlıdır.

Şimdi:

$$Arz = b_0 + b_1 FIYAT + b_2 IKAMEFIYAT$$

modelini tahmin edelim:

	Katsayı	StHata	T	P
Sabit	5.188	6.217	0.835	0.416
FIYAT	2.792	0.480	5.822	0.000
IKAMEFIYAT	-0.897	0.633	-1.417	0.176
S = 3.456		R^2 = 0.68	Düz. R^2=0.64	

Modelde fiyatın tahmincisi beklenen işaretli ve anlamlı olmakla birlikte, ikame mal fiyatı anlamlı değildir.

Modelimizi biraz daha geliştirelim ve:

$$Arz = b_0 + b_1 FIYAT + b_2 IKAMEFIYAT + b_3 SUTFIYAT$$

modelini tahmin edelim:

	Katsayı	StHata	t	p
Sabit	22.361	7.774	2.876	0.012
FIYAT	3.006	0.401	7.487	0.000
IKAMEFIYAT	-1.111	0.526	-2.111	0.052
SUTFIYAT	-2.247	0.766	-2.935	0.010
S = 2.847		R^2 = 0.80	Düz. R^2=0.76	

Modeldeki tüm tahminciler, beklenen işaretlere sahip ve istatistiki açıdan anlamlıdır. R^2 de önceki modellerden daha yüksektir.

Örnek 4-9

Burada zeytinyağı arzının ekonometrik tahminlerini yapacağız. Modelimizde yer alan değişkenler ve anlamları:

ZyagFiy:	Zeytinyağı fiyatı
CicYFiy:	Çiçek yağı fiyatı
ZYMaliy:	Zeytinyağı maliyeti
ZyagArz:	Zeytinyağı arzı

ZYagFiy	CicYFiy	ZYMaliy	ZyagArz
37	36	7	33.29056
32	37	5	33.10796
30	61	6	35.97284
28	64	6	35.56312
35	52	8	35.29082
39	41	6	35.61058
37	58	8	36.95927
44	51	5	39.97583
32	16	8	25.49522
33	26	8	29.08814
25	49	8	31.01176
44	40	6	36.87313
41	32	6	34.04514
30	25	5	29.36549
36	38	6	34.00242
42	28	7	32.64008
43	19	8	29.42697
41	23	5	31.98229
44	65	8	40.33482
39	43	5	36.76716
44	21	8	30.40931

Önce zeytinyağı arzını zeytinyağı fiyatı ile ilişkilendirelim:

Bağımlı değişken: ZYagArz

	Katsayı	*St.Hata*	*t*	*p*	
sabit	25.5518	5.09181	5.0182	0.000076	***
ZYagFiy	0.219879	0.136163	1.6148	0.122834	

R^2 = 0.120681
Düzeltilmiş R^2 = 0.0744015
Akaike bilgi kriteri = 115.049
Schwarz Bayesian kriteri = 117.138

Modelde zeytinyağı fiyatı istatistiki açıdan anlamlı değildir. O halde bu model bizim kullanabileceğimiz nitelikte değildir.

Zeytinyağı arzını zeytinyağı fiyatı ve rakip ürün olan çiçek yağının fiyatı ile ilişkilendirelim:

Bağımlı değişken: ZYagArz

	Katsayı	*St.Hata*	*t*	*p*	
sabit	12.5629	1.88312	6.6713	< 0.00001	***
ZYagFiy	0.334775	0.0437762	7.6474	< 0.00001	***
CicYFiy	0.222553	0.0168996	13.1691	< 0.00001	***

R^2 = 0.917316
Düzeltilmiş R^2 = 0.908129
F-istatistiği (2, 18) = 99.8487 (p < 0.00001)
Akaike bilgi kriteri = 67.4026
Schwarz Bayesian kriteri = 70.5362

F testi, bu modelin istatistiki açıdan anlamlı olduğunu göstermektedir. Gerek zeytinyağı fiyatı gerekse çiçek yağı fiyatı hem sıfırdan farklı hem de teorik beklentiye uygundur. O halde bu modeli kullanabiliriz.

Bu kez yukarıdaki modele, arz fonksiyonunun önemli bir değişkeni olan zeytinyağı maliyetini dahil edeceğiz.

Bağımlı değişken: ZYagArz

	Katsayı	*St.Hata*	*t*	*P*	
sabit	17.8007	1.12595	15.8095	< 0.00001	***
ZYagFiy	0.33198	0.02103	15.7860	< 0.00001	***
CicYFiy	0.220086	0.00812352	27.0925	< 0.00001	***
ZYMaliy	-0.761077	0.0974318	-7.8114	< 0.00001	***

R^2 = 0.981983
Düzeltilmiş R^2 = 0.978804
F-istatistiği (3, 17) = 308.857 (p < 0.00001)
Akaike bilgi kriteri = 37.4044
Schwarz Bayesian kriteri = 41.5825

F testi bu modeli güvenle kullanabileceğimizi göstermektedir. Modeldeki tüm değişkenlere ait katsayılar istatistik açıdan anlamlı ve beklenen işaretlere sahiptir. Başta R^2 olmak üzere, Akaike ve Schwarz kriterleri bu modelin en iyi model olduğunu kanıtlamaktadır.

Örnek 4-10

Gözlem	ZYagArz	ZYagFiy	CicYFiy	ZYMaliy	l_ZYagAr	l_ZYagFi	l_CicYFi	l_ZYMali
1	34.3	37	36	5	3.535145	3.610918	3.583519	1.609438
2	30.1	32	37	9	3.404525	3.465736	3.610918	2.197225
3	34	30	61	6	3.526361	3.401197	4.110874	1.791759
4	33.6	28	64	6	3.514526	3.332205	4.158883	1.791759
5	31.3	35	52	6	3.443618	3.555348	3.951244	1.791759
6	34.6	39	41	5	3.543854	3.663562	3.713572	1.609438
7	41	37	58	5	3.713572	3.610918	4.060443	1.609438
8	37	44	51	5	3.610918	3.78419	3.931826	1.609438
9	27.5	32	16	9	3.314186	3.465736	2.772589	2.197225
10	33.1	33	26	8	3.499533	3.496508	3.258097	2.079442
11	28	25	49	10	3.332205	3.218876	3.89182	2.302585
12	39.9	44	40	4	3.686376	3.78419	3.688879	1.386294
13	30	41	32	8	3.401197	3.713572	3.465736	2.079442
14	30.4	30	25	8	3.414443	3.401197	3.218876	2.079442
15	38	36	38	6	3.637586	3.583519	3.637586	1.791759
16	28.6	42	28	9	3.353407	3.73767	3.332205	2.197225
17	29.4	43	19	9	3.380995	3.7612	2.944439	2.197225
18	35	41	23	5	3.555348	3.713572	3.135494	1.609438
19	39.3	44	65	8	3.671225	3.78419	4.174387	2.079442
20	33.8	39	43	5	3.520461	3.663562	3.7612	1.609438
21	31.4	44	21	8	3.446808	3.78419	3.044522	2.079442

Bağımlı değişken: ZYagArz

	Katsayı	*St.Hata*	*t*	*p*	
sabit	24.1197	5.40362	4.4636	0.000266	***
ZYagFiy	0.249724	0.144501	1.7282	0.100174	

R^2 = 0.135838
Düzeltilmiş R^2 = 0.0903553
Akaike bilgi kriteri = 117.546
Schwarz Bayesian kriteri = 119.635

Bağımlı değişken: ZYagArz

	Katsayı	*St.Hata*	*t*	*p*	
sabit	14.6066	4.91377	2.9726	0.008157	***
ZYagFiy	0.333875	0.114228	2.9229	0.009085	***
CicYFiy	0.163	0.0440974	3.6964	0.001652	***

R^2 = 0.508735
Düzeltilmiş R^2 = 0.454151
F-istatistiği (2, 18) = 9.32007 (p = 0.00167)
Akaike bilgi kriteri = 107.685
Schwarz Bayesian kriteri = 110.819

Bağımlı değişken: ZYagArz

	Katsayı	St.Hata	t	p	
sabit	29.4111	5.87741	5.0041	0.000109	***
ZYagFiy	0.206351	0.0984848	2.0953	0.051430	*
CicYFiy	0.10322	0.0393247	2.6248	0.017744	**
ZYMaliy	-1.12928	0.334564	-3.3754	0.003594	***

R^2 = 0.705864
Düzeltilmiş R^2 = 0.653957
F-istatistiği (3, 17) = 13.5988 (p = 8.95e-005)
Akaike bilgi kriteri = 98.9135
Schwarz Bayesian kriteri = 103.092

Bağımlı değişken: ZYagArz

	Katsayı	St.Hata	t	p	
sabit	1.04627	18.2137	0.0574	0.954791	
l_ZYagFi	8.98067	5.05871	1.7753	0.091872	*

R^2 = 0.142276
Düzeltilmiş R^2 = 0.0971326
Akaike bilgi kriteri = 117.389
Schwarz Bayesian kriteri = 119.478

Bağımlı değişken: l_ZYagAr

	Katsayı	St.Hata	t	p	
sabit	2.54312	0.540219	4.7076	0.000153	***
l_ZYagFi	0.266122	0.150042	1.7737	0.092150	*

Düzeltilmiş R^2 = 0.142051
Düzeltilmiş R^2 = 0.0968961
Akaike bilgi kriteri = -30.3656
Schwarz Bayesian kriteri = -28.2765

Bağımlı değişken: l_ZYagAr

	Katsayı	St.Hata	t	p	
sabit	2.76443	0.531486	5.2013	0.000072	***
l_ZYagFi	0.208875	0.103486	2.0184	0.059614	*
l_CicYFi	0.111573	0.0432846	2.5777	0.019562	**
l_ZYMali	-0.220194	0.0681018	-3.2333	0.004884	***

R^2 = 0.712647
Düzeltilmiş R^2 = 0.661938

F-istatistiği (3, 17) = 14.0536 (p = 7.37e-005)
Akaike bilgi kriteri = -49.3361
Schwarz Bayesian kriteri = -45.158

4.3.4. *Sabit Terimin Kullanımı ve Yorumu*

En basit regresyon denklemi:

$$Y=b_0 + b_1X + e$$

şeklindeki doğrusal regresyon denklemidir. Burada b_0, sabit terimdir. Anlamı, diğer tüm bağımsız değişkenler ve hata terimi (e) sıfır iken, Y'nin beklenen değeridir. Çoğu kez, b_0'ın teorik önemi de vardır. Örneğin toplam masraf fonksiyonu (TC), çıktının (Q) bir fonksiyonu olarak tanımlanabilir:

$$TC = b_0 + b_1Q$$

Burada b_1, değişken masrafları temsil ederken; b_0, sabit masrafları göstermektedir. İstendiği takdirde, model tanımlamasında, b_0 denkleme alınmayabilir. Böyle bir tercih, bazı durumlarda teoriye ters düşecek sonuçlara yol açabileceği gibi, doğrusal regresyonun varsayımlarından olan, hata terimlerinin toplamının sıfır olması varsayımına da zarar verebilir.

Sabit terimsiz modellerle çalışıldığı takdirde:

- Eğim katsayısı muhtemelen sapmalı olacaktır.
- Tahmincilerin t değerleri, aşırı ölçüde büyüyecektir.
- Sabit terimi olmayan bir modelin tahmin edilmesi, kuşkusuz mümkündür. Ancak, ilgili teori gerektirse bile, sabit terimsiz bir modelle çalışılması önerilmemektedir.

Sabit terim, analiz ve yorumlama açısından önemli olmakla birlikte, iki açıdan dikkatli olunmasında fayda vardır:

Birincisi; hata terimi, kısmen, modele alınmayan değişkenlerden dolayı ortaya çıkar; sabit terim ise ortalama etkiyi temsil eder. Sabit terim, çöp toplama görevini üstlenerek, miktarı bilinmeyen ortalama etkiyi bünyesinde toplar. Denklemin bir bütün olmasının sonucu olarak, böyle bir görevi üstlenmiş olması, hata terimi tahmincisinin, olması gerektiğinden çok daha farklı değerler almasına yolaçabilir. Bu nedenle, sabit terim için hipotez testi yapmak, anlamlı bir çaba değildir.

İkincisi; sabit terim, tüm bağımsız değişkenler ve hata terimi sıfırken, bağımlı değişkenin aldığı değerdir. Ancak, ekonomik değişkenlerin aldığı değerler, genellikle sıfırdan büyüktür. O nedenle, orijin, gözlem aralığının dışında kalır.

4.3.5. *Yaygın Matematiksel Formlarda Tahminleme*

Modelde yer alacak bağımsız değişkenler doğru bir şekilde belirlenmiş olsa bile, süreç tamamlanmış değildir. Bundan sonraki adım, doğru fonksiyonel formu seçmektir. Orijinden geçen bir doğru mu, yoksa doğru yerine bir eğri mi tercih edilecektir? Modelin fonksiyonel formu yanlışsa, doğru seçilmiş bir bağımsız değişkenin işareti ters veya istatistiki açıdan önemsiz bulunması ihtimali vardır. Kuşkusuz bu durumda, yorumlama ve kestirim hatalarına düşülebilecektir.

Teorik yapı, genellikle fonksiyonel form konusunda yol gösterici olacaktır. Ekonomi ve işletmecilik ilkelerine en uygun şekli veren matematiksel formun seçilmesi gerekir.

Ekonometrik çalışmalarda yaygın olarak kullanılan matematiksel formlar:

1. Doğrusal fonksiyon: **$Y = b_0 + b_1X$**
2. Ters fonksiyon: **$Y = b_0 + b_1(1/X)$**
3. Logaritmik fonksiyon
   a. Yarı logaritmik:
      i. Log-lin: **$\ln Y = b_0 + b_1X$**
      ii. Lin-log: **$Y = b_0 + b_1\ln X$**
   b. Tam logaritmik: **$\ln Y = b_0 + b_1\ln X$**

4.3.5.1. Doğrusal Form

Akla ilk gelen ve kullanımı en yaygın olan form, doğrusal formdur. Tahmin edilmesi, anlaşılması ve yorumlanması en kolay olan formdur.

Tek açıklayıcılı doğrusal model: $Y=b_0 + b_1X$

Doğrusal form, bağımlı ve bağımsız değişken(ler) arasındaki ilişkiyi en yalın şekilde ifade eder. b_0, denklemin y eksenini kestiği nokta, b_1 ise eğimdir. Tek bağımsız değişkenli doğrusal fonksiyonlarda Y'nin X'e göre türevi, eğimi verir:

$dy/dX = b_1$

Bağımsız değişkendeki bir birim değişmenin, bağımlı değişkende meydana getireceği değişme, eğimdir ve doğrudan b_1 katsayısına bakarak, kolayca bulmak mümkündür. Doğrusal formda eğim sabittir ve bağımsız değişkeninin tüm değerleri için aynıdır.

Doğrusal formdaki modellerdee eğim, doğrudan doğrusal denklemdeki değişkenin tahmincisi iken; esneklik için ayrıca hesaplama yapmaya ihtiyaç duyarız. Esneklik, bağımsız değişkendeki oransal (veya yüzde) değişmenin, bağımlı değişkende meydana getireceği oransal (veya yüzde) değişmeyi ölçer. Örneğin arzı, fiyatın doğrusal bir fonksiyonu olarak tanımlayalım. Bu durumda, arzın fiyat esnekliğini:

E = Miktardaki yüzde değişme / Fiyattaki yüzde değişme

formülüyle hesaplarız. Şimdi bu tanımlamayı matematiksel sembollerle gerçekleştirelim: Y miktarı, X fiyatı, ΔY miktardaki değişmeyi, ΔX fiyattaki değişmeyi temsil etsin:

$$E = \frac{\frac{\Delta Y}{Y}.100}{\frac{\Delta X}{X}.100}$$

Yeniden düzenlediğimizde:

$$E = \frac{\Delta Y / \Delta X}{Y / X} = \frac{dY / dX}{Y / X}$$

olur. Daha anlaşılabilir bir ifadeyle, esneklik, marjinal fonksiyonun, ortalama fonksiyona oranıdır:

E = MF / OF

Buna göre, esneklik eğimin bir fonksiyonudur. Bu nedenle, esneklik hiçbir zaman eğime *eşit değil*dir.

Formülden anlaşılacağı üzere, esneklik tüm X değerleri için farklıdır. Her X değeri için esneklik hesabı yapmak mümkün olmakla birlikte, pratik bir yol değildir. Bunun yerine X değişkenini temsil eden ortalama X değerini kullanmak daha uygundur.

Ortalama fonksiyonun hesaplanmasında Y değerine ihtiyaç duyulmaktadır. Bunun için ya ortalama Y değerini, ya da ortalama X değerine göre hesaplanmış Y tahmin değerini kullanabiliriz. Biz, Y tahmin değerini tercih edeceğiz.

Esnekliği yorumlarken, %1 veya %10 gibi çok küçük yüzde değişmeler kullanılmalıdır. Her fiyat için farklı bir esneklik olduğu düşünülürse, çok büyük oransal değişmeler seçmek, doğru bir yaklaşım olmayacaktır.

Modelin matematiksel formuna göre eğim ve esneklik hesaplamada kullanılan formüller farklılık gösterir. Aşağıdaki tabloda bu formüller sunulmuştur:

Çizelge 4-2: Fonksiyon Tiplerine Göre Marjinal Etki ve Esneklik Hesaplamaları

Fonksiyon tipi	*Matematiksel gösterim*	*Eğim (Marjinal Etki)*	*Esneklik*
Linear	$Y = b_0 + b_1 X$	b_1	$b_1 \frac{\overline{X}}{\overline{Y}}$
Ters	$Y = b_0 + b_1 \frac{1}{X}$	$-\frac{b_1}{\overline{X}^2}$	$-b_1 \frac{1}{\overline{X}\overline{Y}}$
Lin-log	$Y = b_0 + b_1 \ln X$	$\frac{b_1}{\overline{X}}$	$\frac{b_1}{\overline{Y}}$
Log-lin	$\ln Y = b_0 + b_1 X$	$b_1 \overline{Y}$	$b_1 \overline{X}$
Log-log	$\ln Y = b_0 + b_1 \ln X$	$b_1 \frac{\overline{Y}}{\overline{X}}$	b_1
Kuadratik	$Y = b_0 + b_1 X + b_2 X^2$	$b_1 + 2b_2 X$	$\frac{(b_1 + 2b_2 X)X}{Y}$

Örnek 4-11

Bir malın arzıyla, fiyatı arasındaki ilişkiyi doğrusal bir fonksiyonla irdeleyelim. Doğrusal fonksiyonumuz:

$$Y=b_0 + b_1X$$

Gözlemlerimiz:

Arz (Y)	Fiyat (X)
75	65
60	55
80	80
100	110
150	350
175	400
180	500
120	600

olsun.

Arzın fiyat fonksiyonunu doğrusal formda tahmin edelim:

	Katsayı	StHata	t	p
Sabit	72.15	17.96	4.02	0.007
X	0.16796	0.05301	3.17	0.019
S = 30.68	R^2 = 0.626	Düz. R^2=0.564	F=10.04	p=0.019

Görüldüğü gibi, denklem istatistiki açıdan geçerlidir. Denklemdeki tahminciler %5 önem düzeyinde sıfırdan farklı ve b_1 beklendiği gibi pozitiftir. Şimdi modelimizi denklem formunda yazalım:

Y=72.15 + 0.16796 X

Eğim:

Doğrusal formda eğim, b_1 tahmincisidir ve 0.16796'ya eşittir. Fiyat 1 birim arttığında, arzın 0.16796 birim artış göstereceği veya fiyat 1 birim azaldığında arzın 0.16796 birim azalacağı anlamına gelir. Daha genel bir ifadeyle, fiyat ne olursa olsun, 1 birimlik değişme, arzda her zaman 0.16796 birimik değişmeye neden olacaktır.

Esneklik:

Doğusal formda arzın fiyat esnekliği, her fiyat düzeyi için farklıdır. Örneğin fiyat 110 iken, telep esnekliğini hesaplayalım:

$$E = \frac{dY / dX}{Y / X}$$

dY/dX = 0.16796

Y(110) = 72.15 + 0.16796 (110) = 90.6256

E = 0.16796 / (90.6256/110)

E = 0.204

Esneklik, 0.204'tür. Buna göre, fiyat 110 iken, %10'luk bir artış, arzı % 2.04 artıracaktır:

0.10 x 0.204 = 0.0204 = % 2.04

Fiyat 110 iken, %5 azalırsa, arz % 1.02 azalacaktır:

0.05 x 0.204 = 0.0102 = % 1.02

Şimdi fiyatı temsil etmek üzere ortalama X değerini alıp, esnekliği yeniden hesaplayalım. Ortalama X, 270'tir. Buna göre:

dY/dX = 0.16796

Y(270) = 72.15 + 0.16796 (110) = 117.4992

E = 0.16796 / (117.4992/270)

E = 0.157

Esneklik bu kez, 0.157'dir. Buna göre, fiyat 270 iken, %10'luk bir artış, arzı % 1.57 artıracaktır:

0.10 x 0.157 = 0.0157 = % 1.57

Fiyat 270 iken, %5 azalırsa, arz % 0.785 azalacaktır:

0.05 x 0.157 = 0.0102 = % 0.785

Çok açıklayıcılı doğrusal model:

k adet açıklayıcı bulunan bir doğrusal model:

$$Y=b_o + b_1X1 + b_2X2+\ldots+b_rXr \qquad (r=1, 2,..,k)$$

şeklinde gösterilir. Çok açıklayıcı doğrusal modellerde de eğim kolayca bulunabilir: Değişkenlere ait tahmincilerin her biri, o değişkene ait eğimi gösterir. Ancak birden fazla değişken olduğundan, içlerinden sadece birine ait eğimi hesaplayabilmek için, diğer değişkenlerin sabit tutulması gerekecektir. Bu nedenle, eğim değil, kısmi eğimden söz edilir. Örneğin Y'nin X1 değişkenine göre kısmi eğimi, b_1; Y'nin X2 değişkenine göre kısmi eğimi ise, b_2'dir. Genel olarak gösterecek olursak, Y'nin herhangi bir r değişkenine göre kısmi eğimi, Y'nin r değişkenine göre kısmi türevidir:

$$\partial y/\partial Xr = b_r$$

Çok açıklayıcı doğrusal modellerde kısmi esneklik hesaplayabiliriz. Örneğin, bir malın talebi (*Q*), kendi fiyatına (*P*), rakip ürün fiyatına (P_A) ve gelire Y bağlı olsun:

$$Q = f(P, P_A, Y)$$

Talebin kısmi fiyat esnekliği:

$$E_P = (\partial Q/\partial P) / (Q/P)$$

formülüyle; talebin çapraz fiyat esnekliği:

$$E_C = (\partial Q/\partial P_A) / (Q/P_A)$$

formülüyle; talebin gelir esnekliği:

$$E_Y = (\partial Q/\partial Y) / (Q/Y)$$

formülüyle hesaplanır.

Doğrusal modellerde de kısmi esneklikler sabit değildir. İlgili değişkenin her bir değeri için ayrı bir esneklik söz konusudur. Ancak bunun yerine, değişkenin ortalama değeri için kısmi esneklik hesaplamak daha uygundur. Kısmi esneklik yorumlamalarında, daha önce belirtildiği gibi, %1 veya %10 gibi küçük yüzde değişimler kullanılmalıdır.

Örnek 4-12

1973-2000 yılları arasındaki kişi başına kırmızı et tüketimi Y ile reel kırmızı et fiyatı (X1), ve kişi başına gelir (X2) arasındaki ilişkiyi, çok değişkenli doğrusal bir fonksiyonla irdeleyelim.

Çizelge 4-3: 1974-98 Yılları Kişi Başına Kırmızı Et Tüketimi

Yıl	Kişi Başına K.Et Tüketimi (Y)	Reel K.Et Fiyatı (X1)	Kişi Başına Gelir (X2)
1973	85.1	20.4	6.036
1974	87.8	20.2	6.113
1975	88.9	21.3	6.271
1976	94.5	19.9	6.378
1977	99.9	18	6.727
1978	99.5	19.9	7.027
1979	104.2	22.2	7.28
1980	106.5	22.3	7.513
1981	109.7	23.4	7.728
1982	110.8	26.2	7.891
1983	113.7	27.1	8.134
1984	113	29	8.322
1985	116	33.5	8.562
1986	108.7	42.8	9.042
1987	115.4	35.6	8.867
1988	118.9	32.2	8.944
1989	127.4	33.7	9.175
1990	123.5	34.4	9.381
1991	117.9	48.5	9.735
1992	105.4	66.1	9.829
1993	103.2	62.4	9.722
1994	104.2	58.6	9.769
1995	103.7	56.7	9.725
1996	105.7	55.5	9.93
1997	105.5	57.3	10.419
1998	106.5	53.7	10.625
1999	107.3	52.6	10.905
2000	103.3	61.1	10.97

Çok açıklayıcılı doğrusal modelimizi:

$$Y = f(X1, X2)$$

olarak tanımlayalım. Bu modelin tahmin sonuçları:

	Katsayı	StHata	t	p
Sabit	37.54	10.04	3.74	0.001
X1	-0.8826	0.1647	-5.36	0.000
X2	11.891	1.762	6.75	0.000
S = 6.081	R^2 = 0.658		Düz. R^2=0.631	F=24.05

Modelimizi denklem formunda yazalım:

$Y = 37.54 - 0.8826X1 + 11.891X2$

Modelimiz, % 1 düzeyinde bile istatistiki açıdan anlamlıdır. Tahminciler de %1 için istatistiki olarak anlamlı ve işaretleri, beklendiği gibidir: fiyatın işareti negatif, gelirin işareti pozitiftir.

Kısmi eğimler:

Reel kırmızı et fiyatına göre kısmi eğim –0.8826'dır. Kırmızı et fiyatı 1 birim arttığında, kişi başına kırmızı et tüketimi 0.8826 birim düşecektir.

Gelire göre kısmi eğim, 11.891'dir. Gelir 1 birim arttığında, kişi başına kırmızı et tüketimi 11.891 birim artacaktır.

Kısmi esneklikler:

Kısmi esneklikleri hesaplayabilmek için, ortalama X1'in, 37.664; ortalama X2'nin, 8.6079 olduğunu belirtelim.

Talebin kısmi fiyat esnekliği:

$$E_P = (\partial Y/\partial X1) / (Y/X1)$$

$$\partial Y/\partial X1 = -0.8826$$

$$Y(37.664, 8.6079) = 37.54 - 0.8826(37.664) + 11.891(8.6079)$$

$$Y(37.664, 8.6079) = 106.65$$

$$E_P = -0.8826 / (106.65/37.664)$$

$$E_P = -0.31$$

Buna göre, kırmızı et fiyatı 37.664 iken, %10 artarsa, kişi başına kırmızı et tüketimi:

$$0.10 \times -0.31 = 0.031 = \%3.1$$

%3.1 azalacaktır.

Talebin kısmi gelir esnekliği:

$$E_Y = (\partial Y/\partial X2) / (Y/X2)$$

$$\partial Y/\partial X1 = 11.891$$

$$Y(37.664, 8.6079) = 37.54 - 0.8826(37.664) + 11.891(8.6079)$$

$$Y(37.664, 8.6079) = 106.65$$

$$E_Y = 11.891 / (106.65/8.6079)$$

$$E_P = 0.96$$

Eğer gelir 8.6079 iken, %10 artarsa:

$$0.10 \times 0.96 = 0.096 = \%9.6$$

kişi başına kırmızı et tüketimi, %9.6 artacaktır.

4.3.5.2. Ters Fonksiyon

Ters fonskiyon:

$$Y = b_0 + b_1 (1/x)$$

şeklinde ifade edilir. Bu fonksiyonu tahmin edebilmek için önce 1/x dönüşümünün yapılması gerekir. Buna göre artık Y bağımlı değişken, 1/x bağımsız değişkendir.

Y	X	1/x
75	65	0.015385
60	55	0.018182
80	80	0.0125
100	110	0.009091
150	350	0.002857
175	400	0.0025
180	500	0.002
120	600	0.001667

Tahmin sonuçları:

	Katsayı	St.Hata	t	p
Sabit	167.86	12.4	13.54	0.002
1/x	-6277	1219	-5.15	0.002
S = 21.55	R^2 = 0.815	Düz. R^2=0.785	F=26.5	p=0.002

Denklemimiz:

Y = 167.86 + -6277 (1/x)

Denklemdeki tahmincilerin yorumu, doğrusal fonksiyondan farklıdır. x'deki bir birim değişmenin y üzerine etkisini ölçebilmek için eğimi hesaplamamız gerekir. Bu, y'nin x'e göre türevi demektir:

$$\text{Eğim} = -\frac{b_1}{\overline{X}^2}$$

Eğim= -(-6277 / 270^2)= 0.0861

Bunu, x bir birim arttığında y 0.0861 birim artar şeklinde yorumlarız.

Ters fonksiyonda esneklik:

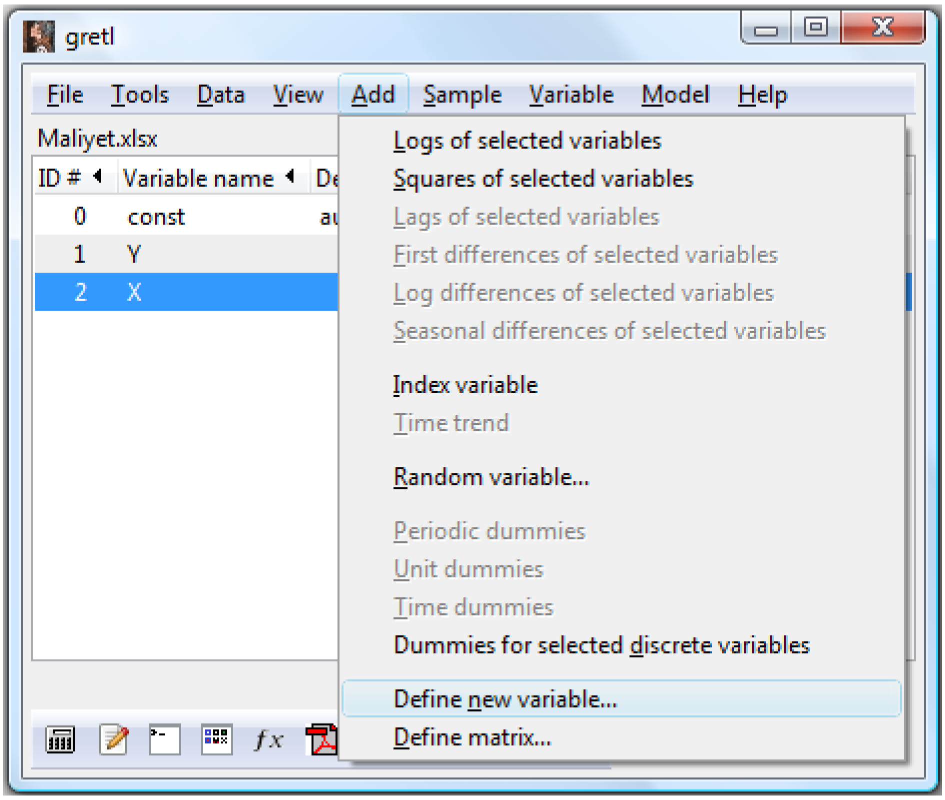
$$E = -b_1 \frac{1}{\overline{X}\overline{Y}}$$

E= -6277 / (270*117.5)

E= 0.1979

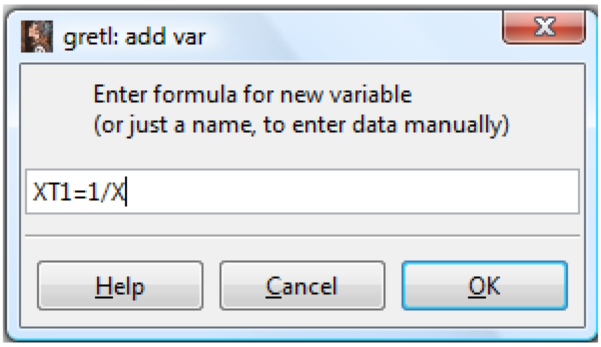
şeklinde hesaplanır. Buna göre, x'teki %10'luk artış, y'yi %1.979 artacaktır.

Şimdi ters fonksiyonu Gretl yardımıyla tahmin edelim. Önce X değişkeninin tersini almamız gerekir. Bunun için *Add* menüsü açılır ve *Define new variable* seçeneği tıklanır:

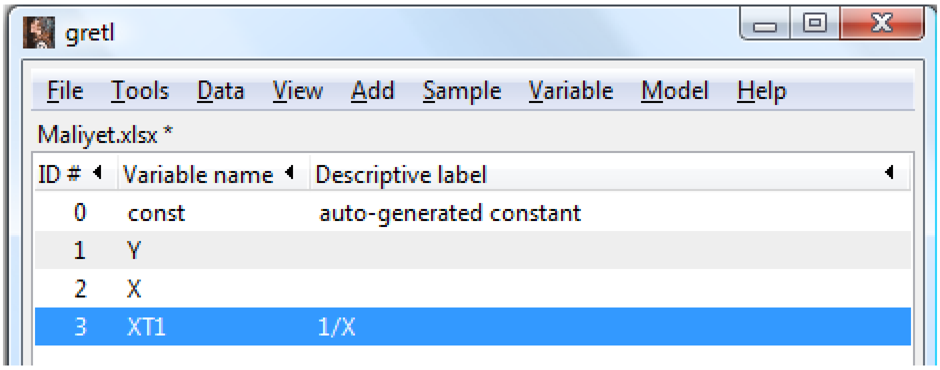


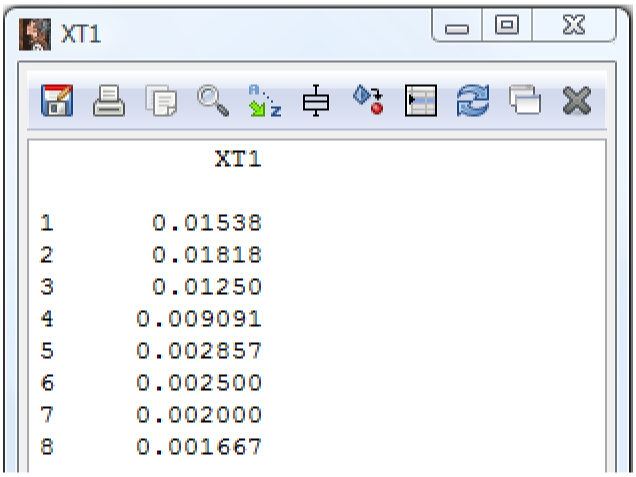
Açılan metin kutusuna aşağıdaki formül yazılır:

XT1= 1 / X



XT1, X değişkeninin çarpmaya göre tersini temsil eden değişkendir. Son duruma göre Gretl değişken listemize XT1 değişkenini eklemiş olduk:





	XT1
1	0.01538
2	0.01818
3	0.01250
4	0.009091
5	0.002857
6	0.002500
7	0.002000
8	0.001667

4.3.5.3. Yarı Logaritmik Fonksiyonlar

lin-log fonksiyon

Yarı logaritmik fonksiyonlardan lin-log formu:

$$Y = b_0 + b_1 \ln X$$

şeklindedir. Bu modeli tahmin edebilmek için önce x'in logaritmasını (ln) almamız gerekir. Tahminlemede artık y bağımlı değişken, lnX bağımsız değişkendir.

Y	X	lnX
75	65	4.17439
60	55	4.00733
80	80	4.38203
100	110	4.70048
150	350	5.85793
175	400	5.99146
180	500	6.21461
120	600	6.39693

$Y = b_0 + b_1 \ln X$ fonksiyonunun tahmin sonuçları:

	Katsayı	StHata	t	p
Sabit	-97.45	47.68	-2.04	0.087
X	41.213	9	4.58	0.004
S = 23.66	R^2 = 0.778	Düz. R^2=0.740	F=20.97	p=0.004

lin-log fonksiyonda eğim:

$$\text{Eğim}=\frac{b_1}{\overline{X}}$$

Eğim=41.213/270 = 0.15

Anlamı, x 1 birim arttığında, y 0.15 birim artar.

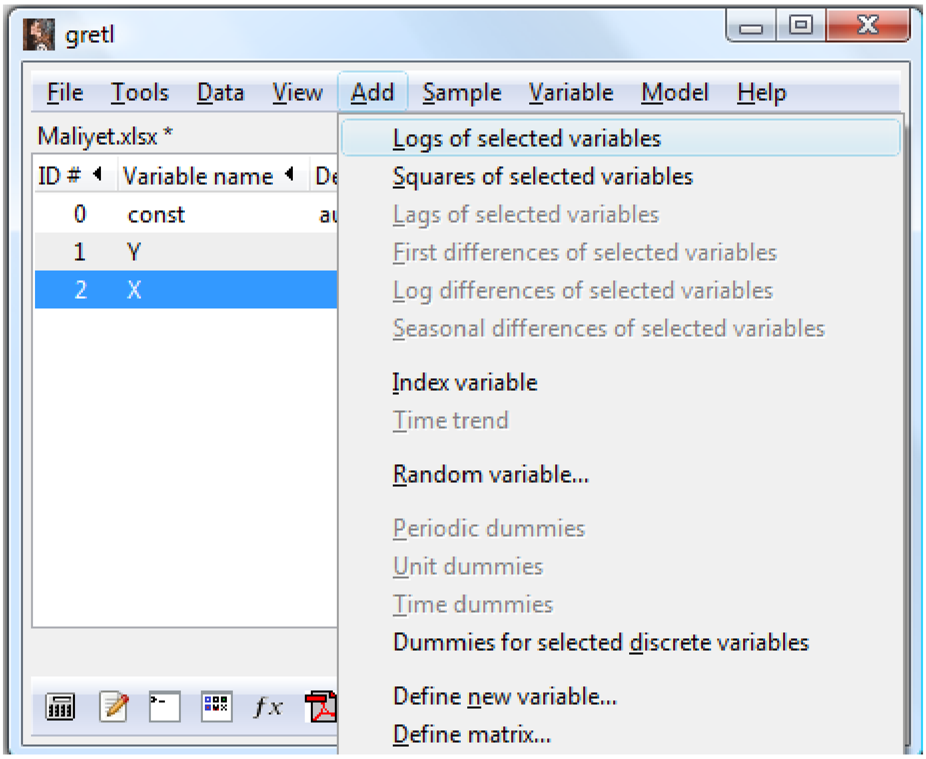
lin-log fonksiyonda esneklik:

$$E=\frac{b_1}{\overline{Y}}$$

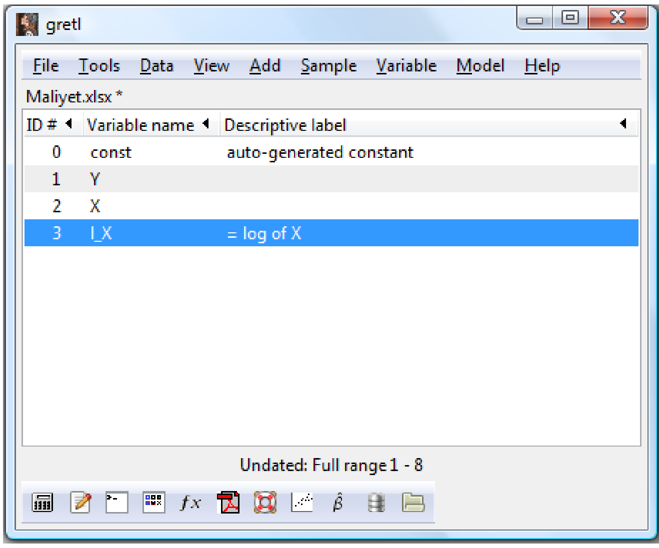
E= 41.213 / 117.5 = 0.35

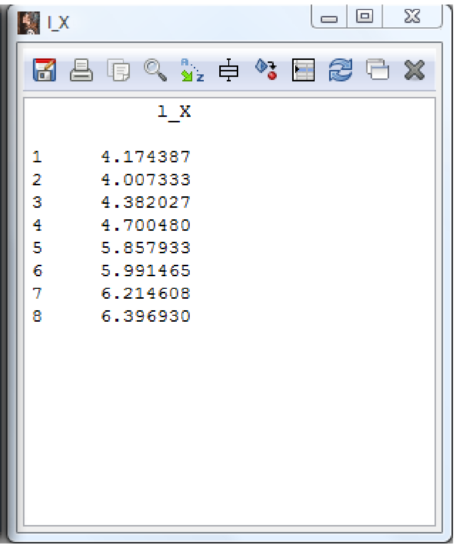
Anlamı, x %10 arttığında, y %3.5 artar.

Lin-log fonksiyonu Gretl yardımıyla tahmin etmek için X değişkeninin logaritmasını almamız gerekir. Bunun için, X değişkeni işaretliylen *Add* menüsü açılır ve *Logs of selected variables* seçeneği tıklanır.

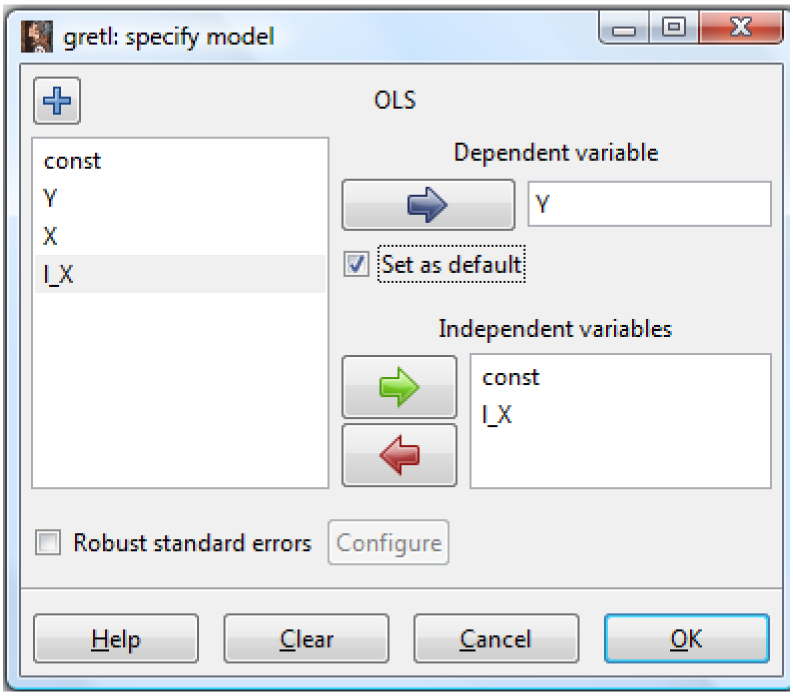


Bu işlemde, X değişkeninin e tabanına göre logaritması alınır ve l_X değişkenine atanır.





Elde edilen l_X değişkeni lin-log modelin tahmin edilmesi sırasında bağımsız değişken olarak kullanılır.



log-lin Fonksiyon

Diğer yarı logaritmik fonksiyonumuz olan log-lin formu:

$\ln Y = b_0 + b_1 X$

denklemiyle ifade edilir. Bu denklemi tahmin edebilmek için önce y'nin logaritmasını (ln) almamız gerekir. Modelin tahmin edilmesi sırasında, lnY bağımlı değişkeni, X bağımsız değişkeni temsil edecektir.

Y	X	lnY
75	65	4.31749
60	55	4.09434
80	80	4.38203
100	110	4.60517
150	350	5.01064
175	400	5.16479
180	500	5.19296
120	600	4.78749

$\ln Y = b_0 + b_1 X$ modelinin tahmin sonuçları:

	Katsayı	StHata	t	p
Sabit	4.282	0.152	28.17	0.000
X	0.001527	0.000449	3.4	0.014
S = 0.2597	R^2 = 0.659	Düz. R^2=0.602	F=11.59	p=0.014

log-lin modelin eğim hesabı:

Eğim= $b_1 \bar{Y}$

Eğim= 0.001527 * 117.5

Eğim= 0.179

Yorumu: x bir birim arttığında, y 0.179 birim artar.

log-lin modelin esneklik hesabı:

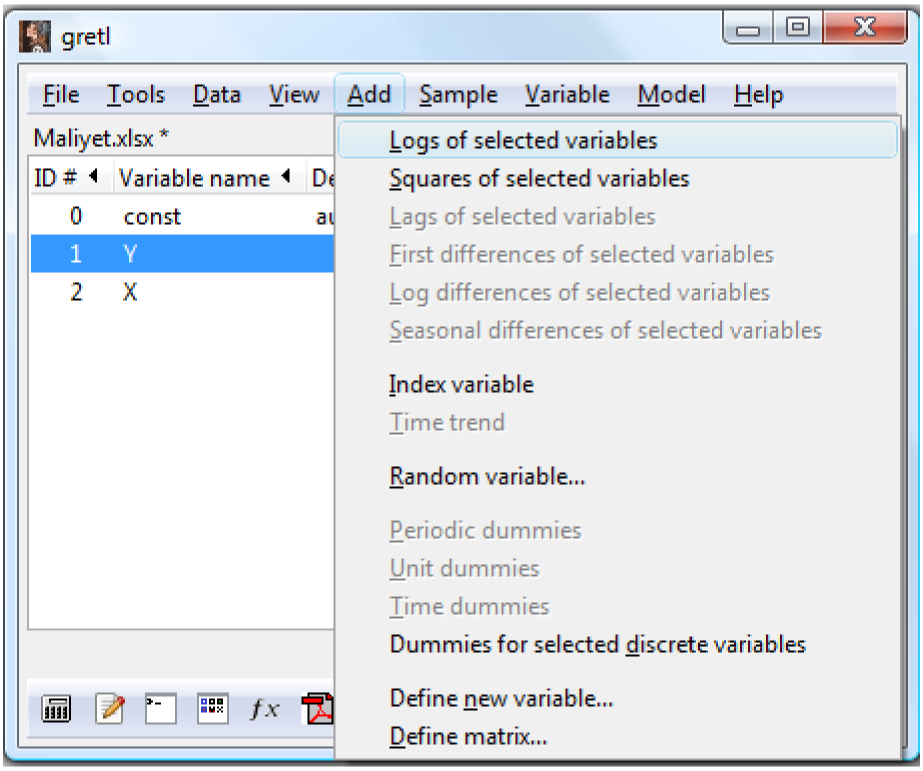
$E = b_1 \overline{X}$

E= 0.001527 * 270

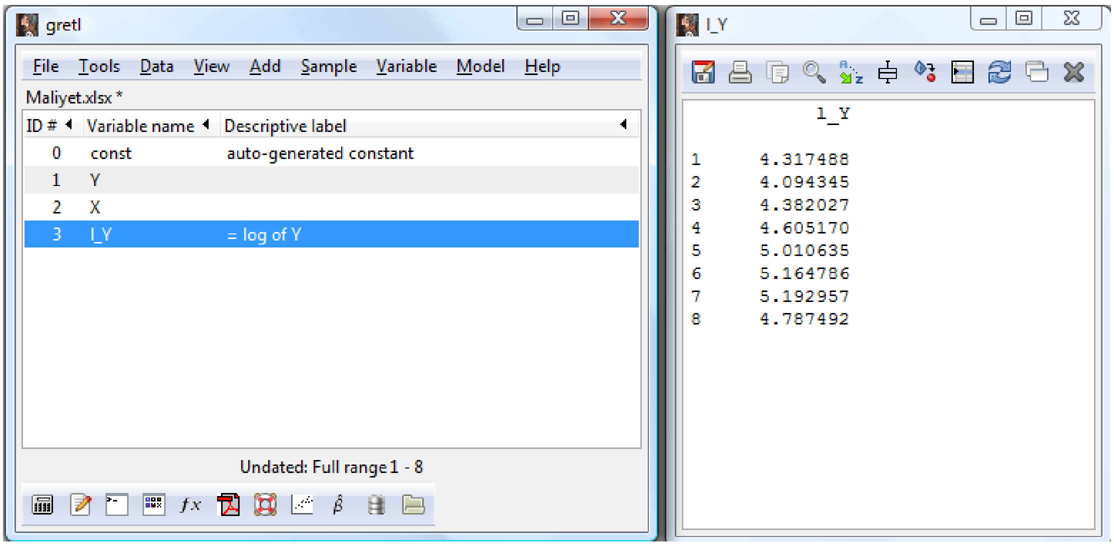
E= 0.4122

Yorumu: x %10 arttığında, y %4.122 artar.

Log-lin fonksiyonu Gretl yardımıyla tahmin edebilmek için bu kez Y değişkeninin logaritmasını almamız gerekir. Bunun için, Y değişkeni işaretliyken *Add* menüsü açılır ve *Logs of selected variables* seçeneği tıklanır.



Bu işlemde, Y değişkeninin e tabanına göre logaritması alınır ve l_Y değişkenine atanır.



4.3.5.4. Çift Logaritmik Fonksiyon

Çift logaritmik form:

$\ln Y = b_0 + b_1 \ln X$

denklemiyle ifade edilir. Bu modeli tahmin edebilmek için hem y'nin hem de x'in logaritmasını (ln) almamız gerekir. Artık lny bağımlı değişken, lnx bağımsız değişkendir.

Y	X	Lny	lnx
75	65	4.31749	4.17439
60	55	4.09434	4.00733
80	80	4.38203	4.38203
100	110	4.60517	4.70048
150	350	5.01064	5.85793
175	400	5.16479	5.99146
180	500	5.19296	6.21461
120	600	4.78749	6.39693

$\ln Y = b_0 + b_1 \ln X$ modelinin tahmin sonuçları:

	Katsayı	StHata	t	p
Sabit	2.7326	0.3754	7.28	0.000
X	0.37613	0.07086	5.31	0.002
S = 0.1863	R^2 = 0.824	Düz. R^2=0.795	F=28.18	p=0.002

log-log formda eğim:

Eğim= $b_1 \frac{\bar{Y}}{\bar{X}}$

Eğim= 0.37613*(117.5 / 270)

Eğim=0.16369

Yorumu: x birim artarsa, y 0.16369 birim artar.

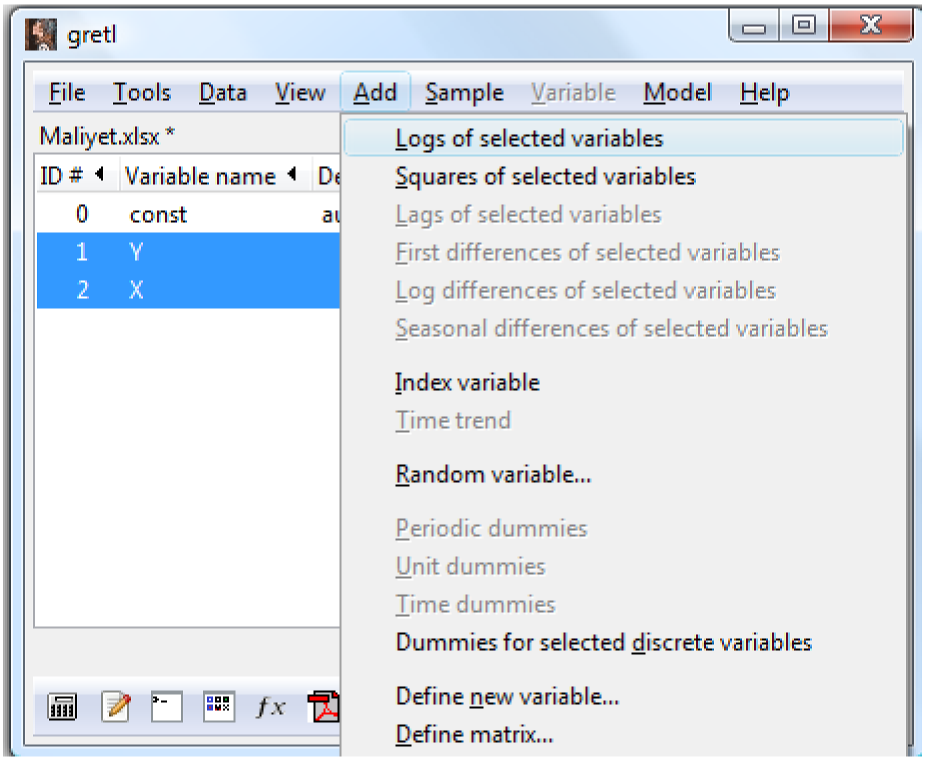
log-log formda esneklik:

E= b_1

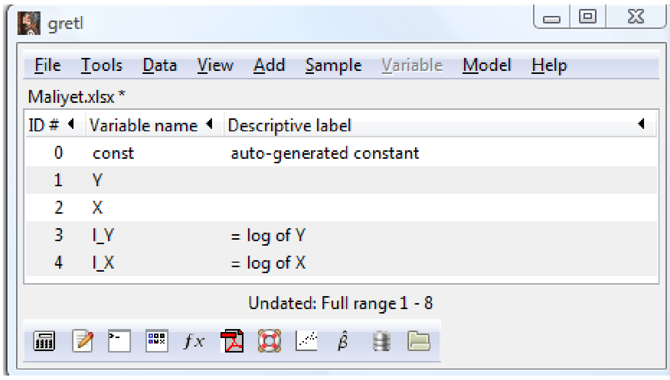
E= 0.37613

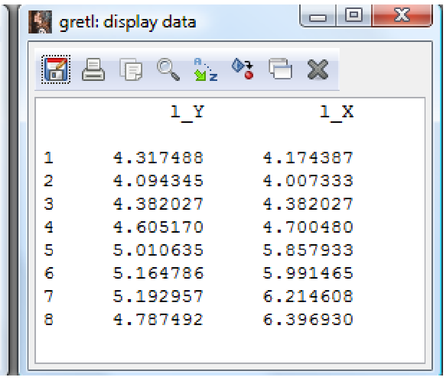
Yorumu: x %10 artarsa, y %3.16 artar.

Log-log fonksiyonu Gretl yardımıyla tahmin edebilmek için bu kez hem Y hem de X değişkeninin logaritmasını almamız gerekir. Bunun için, Y ve X değişkenleri aynı anda işaretliyken *Add* menüsü açılır ve *Logs of selected variables* seçeneği tıklanır.



Bu işlemde, gerek Y gerekse X değişkeninin e tabanına göre logaritması alınır ve sırasıyla l_Y ve l_X değişkenlerine atanır.





4.3.5.5.Polinomial Fonksiyonlar

Masraf ve üretim fonksiyonlarında, üretim miktarı değiştikçe, masraf ve üretimin eğimi değişir. Gerçekten de azalan verimler kanununa göre, kullanılan girdi arttıkça, daha az üretim elde edilir. Böyle durumlarda, ikinci veya daha yüksek dereceden terimlerin modele alınması zorunluluğu doğar. Örneğin ikinci dereceden bir model:

$$Y = b_0 + b_1X + b_2X^2$$

ya da üçüncü dereceden bir model:

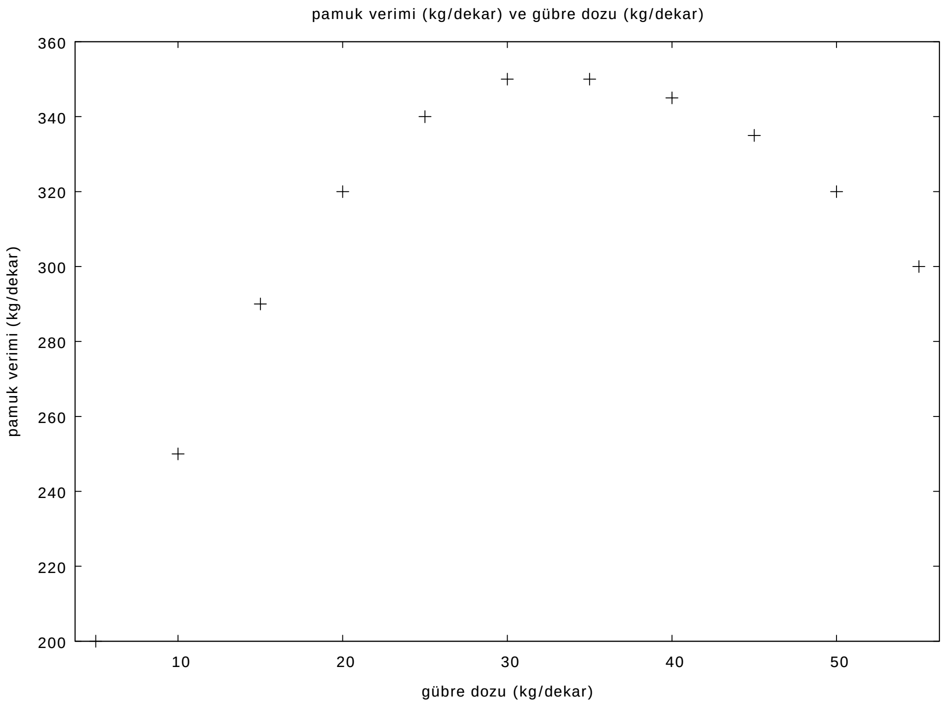
$$Y = b_0 + b_1X + b_2X^2 + b_3X^3$$

şeklinde tanımlanabilir. Aslında bu modellerde X gibi tek bir bağımsız değişken vardır. Bu değişkenin karesi alınıp modele dahil edilmesiyle, model, iki açıklayıcılı model haline gelmektedir. Aynı şekilde, X değişkeninin karesini ve küpünü alıp, modele dahil ettiğimizde, üç açıklayıcılı bir model ortaya çıkar.

Aşağıdaki tabloda belli gübre dozları (kg/dekar) için elde edilen pamuk verimi (kg/dekar) sunulmuştur:

Pamuk verimi (kg/dekar) y	Gübre dozu (kg/dekar) x
200	5
250	10
290	15
320	20
340	25
350	30
350	35
345	40
335	45
320	50
300	55

Önce pamuk verimi ve gübre verimi arasındaki ilişkiye dair ipucu aramak üzere grafiğini çizelim:



Grafik, pamuk verimi ve gübre dozu arasında doğrusal bir ilişki aramanın hatalı olacağını göstermektedir. Hiperbolik veya ikinci dereceden fonksiyonu tercih etmenin doğru bir karar olacağı anlaşılmaktadır.
İkinci dereceden fonksiyonu tahmin etmeden önce, doğrusal fonksiyonu tahmin edelim:

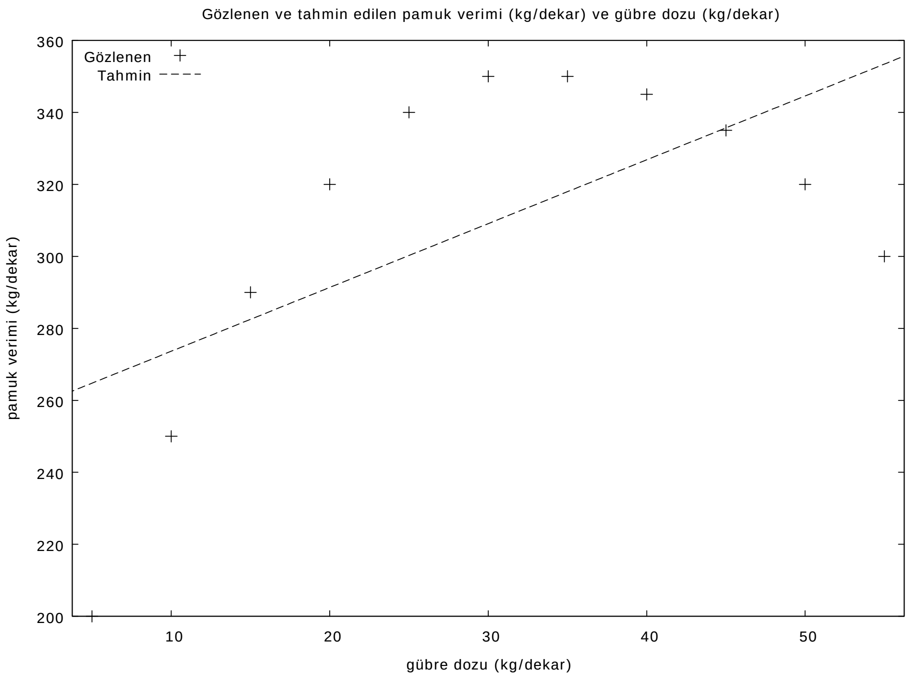
$$y = b_0 + b_1 x$$

Bağımlı değişken: y

	Katsayı	St.Hata	t	p	
sabit	255.909	25.2292	10.1434	<0.00001	***
x	1.77273	0.743969	2.3828	0.04104	**

R-kare	0.386826	Düzeltilmiş R-kare	0.318696
F(1, 9)	5.677727	P(F)	0.041039
Log-likelihood	-54.80777	Akaike kriteri	113.6155
Schwarz kriteri	114.4113	Hannan-Quinn	113.1139

Görüldüğü gibi doğrusal fonksiyon istatistiki açıdan anlamlıdır. Ancak R^2 düşüktür. Tahmin ettiğimiz denklemi grafik üzerinde görebiliriz:



Şimdi ikinci dereceden fonksiyon tahminini yapalım. Bunun için önce x^2 terimini hesaplamamız gerekir. Bir başka ifadeyle gübre dozunu temsil eden x değişkeninin karesini alıp, x_kare olarak adlandıralım. Yeni tablomuz:

Pamuk verimi (kg/dekar) y	Gübre dozu (kg/dekar) x	x_kare
200	5	25
250	10	100
290	15	225
320	20	400
340	25	625
350	30	900
350	35	1225

345	40	1600
335	45	2025
320	50	2500
300	55	3025

İkinci dereceden fonksiyonumuz:

$$y = b_0 + b_1 x + b_2 x^2$$

veya

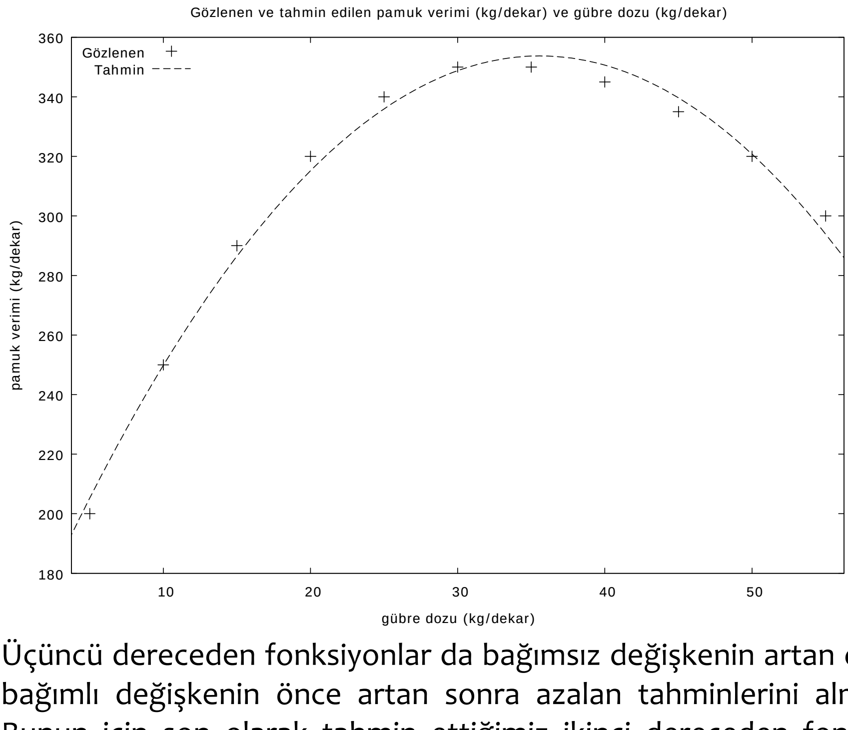
$$y = b_0 + b_1 x + b_2 x\_kare$$

Bağımlı değişken: y

	Katsayı	*St.Hata*	*t*	*p*	
sabit	152.727	5.29557	28.8406	<0.00001	***
X	11.2972	0.405649	27.8497	<0.00001	***
x_kare	-0.158741	0.00658482	-24.1071	<0.00001	***

R-kare	0.991674	Düzeltilmiş R-kare	0.989592
F(2, 8)	476.4135	P(F)	4.81e-09
Log-likelihood	-31.16192	Akaike kriteri	68.32384
Schwarz kriteri	69.51753	Hannan-Quinn	67.57139

İkinci dereceden fonksiyon da istatistiki açıdan anlamlıdır. Modeldeki tüm tahminciler %1 düzeyinde anlamlıdır. R^2, AIC (68.32) ve SC (69.52), ikinci derece denklemin doğrusala göre daha iyi olduğunu açıkça göstermektedir. Grafiğimiz:



Üçüncü dereceden fonksiyonlar da bağımsız değişkenin artan değerlerine karşılık, bağımlı değişkenin önce artan sonra azalan tahminlerini almamızı sağlayabilir. Bunun için son olarak tahmin ettiğimiz ikinci dereceden fonksiyona x^3 terimini eklememiz yeterli olacaktır. O halde x değişkeninin küpünü alıp, x_kup adını verelim. Yeni veri tablomuz:

Pamuk verimi (kg/dekar) y	Gübre dozu (kg/dekar) x	x_kare	x_kup
200	5	25	125
250	10	100	1000
290	15	225	3375
320	20	400	8000
340	25	625	15625
350	30	900	27000
350	35	1225	42875
345	40	1600	64000
335	45	2025	91125
320	50	2500	125000
300	55	3025	166375

Üçüncü dereceden fonksiyonumuz:

$$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3$$

veya

$$y = b_0 + b_1 x + b_2 x\_kare + b_3 x\_kup$$

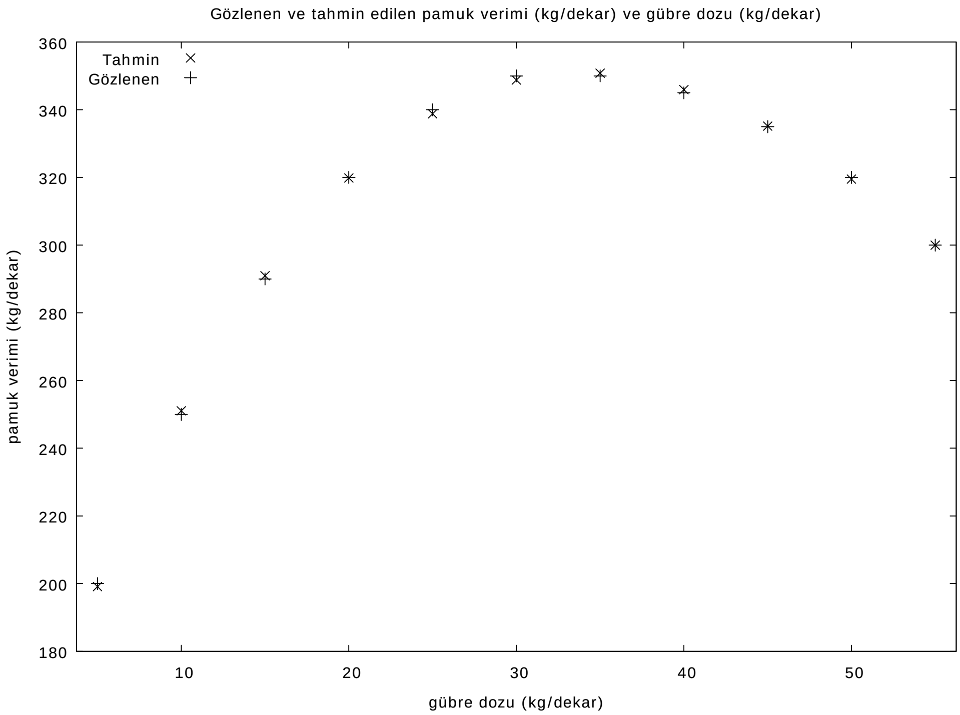
olur. Bu denklemi tahmin ettiğimizde:

Model 5: EKK tahminleri 11 gözlem 1-11
Bağımlı değişken: y

	Katsayı	*St.Hata*	*t*	*p*	
sabit	134.167	1.83883	72.9633	<0.00001	***
x	14.3634	0.253819	56.5892	<0.00001	***
x_kare	-0.281119	0.00961662	-29.2326	<0.00001	***
x_kup	0.00135975	0.000105684	12.8663	<0.00001	***

R-kare	0.999662	Düzeltilmiş R-kare	0.999517
F(3, 7)	6905.234	P(F)	1.65e-12
Log-likelihood	-13.53595	Akaike kriteri	35.07190
Schwarz kriteri	36.66348	Hannan-Quinn	34.06863

Üçüncü dereceden fonksiyonun gerek doğrusal gerekse ikinci derece fonksiyondan daha iyi olduğu açıkça görülmektedir.



Tahmin ettiğimiz denklemleri aynı tabloda görebiliriz:

Bağımlı değişken: y

	Doğrusal	İkinci derece	Üçüncü derece
Sabit	255.9**	152.7**	134.2**
	(25.23)	(5.296)	(1.839)
x	1.773**	11.30**	14.36**
	(0.7440)	(0.4056)	(0.2538)
x_kare		-0.1587**	-0.2811**
		(0.006585)	(0.009617)
x_kup			0.001360**
			(0.0001057)
n	11	11	11
Adj. R^2	0.3187	0.9896	0.9995
Akaike I C	113.6155	68.32384	35.07190

Standart hatalar parantez içindedir
* %10'da önemli
** %5'te önemli

Şimdi, üçüncü dereceden denklemi kullanarak, ortalama ürün ve marjinal ürün fonksiyonlarını elde edeceğiz.

Ortalama ürün (AP):

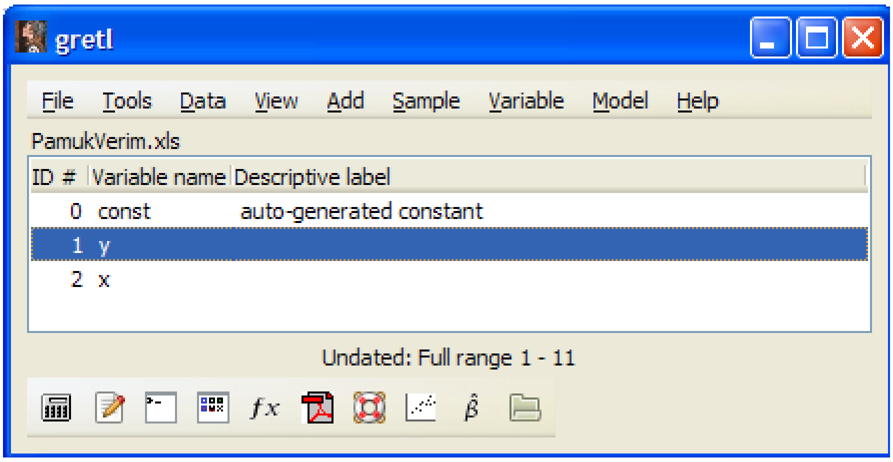
AP: $y/x = (134.2 + 14.36\ x - 0.2811\ x^2 + 0.001360\ x^3)\ /x$

$y/x = 134.2/x + 14.36 - 0.2811\ x + 0.001360\ x^2$

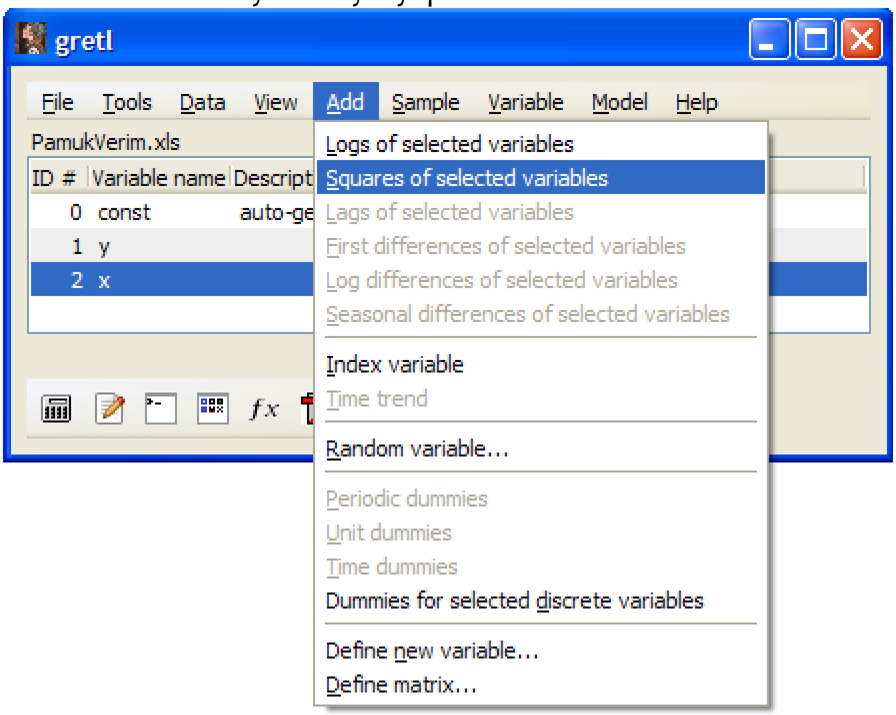
Marjinal ürün (AP):

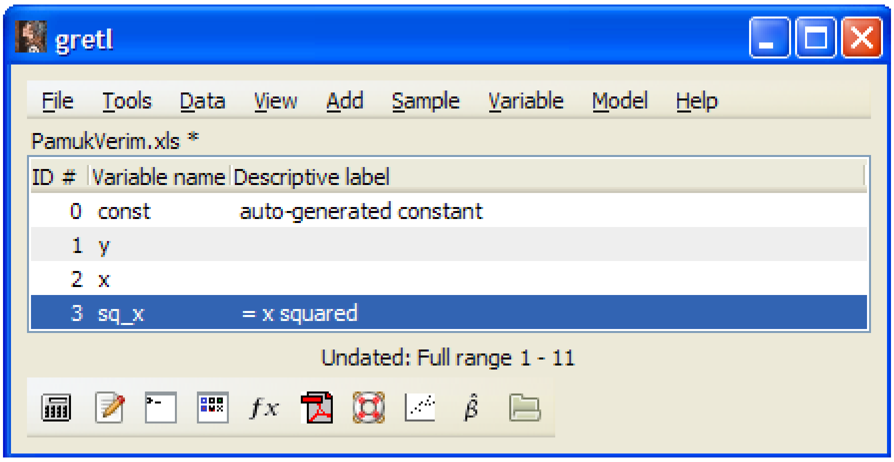
MP: dy/dx = 14.36 - 0.5622 x + 0.00408 x^2

Tüm bunları Gretl yardımıyla yapmamız mümkündür. Değişkenlerimiz sadece y ve x'tir.

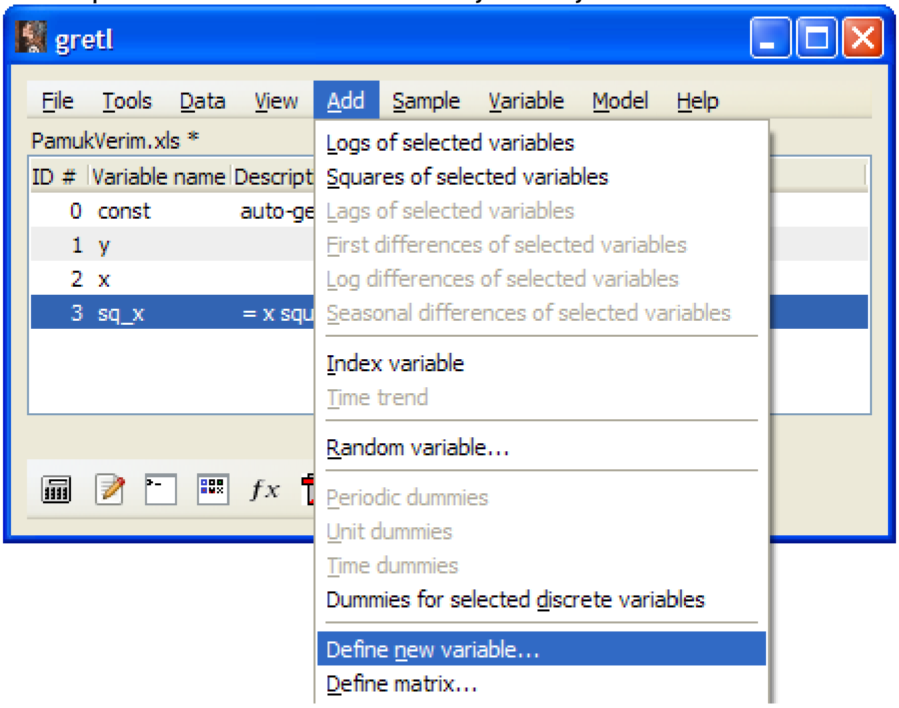


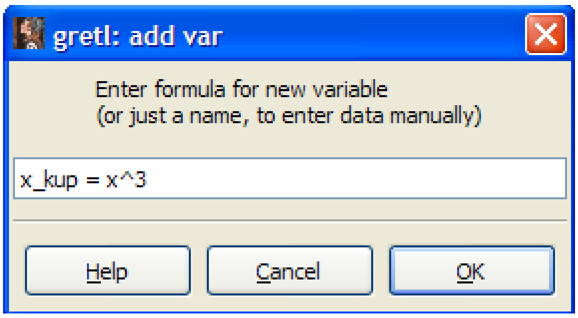
x'in karesini menü yardımıyla yapabiliriz:

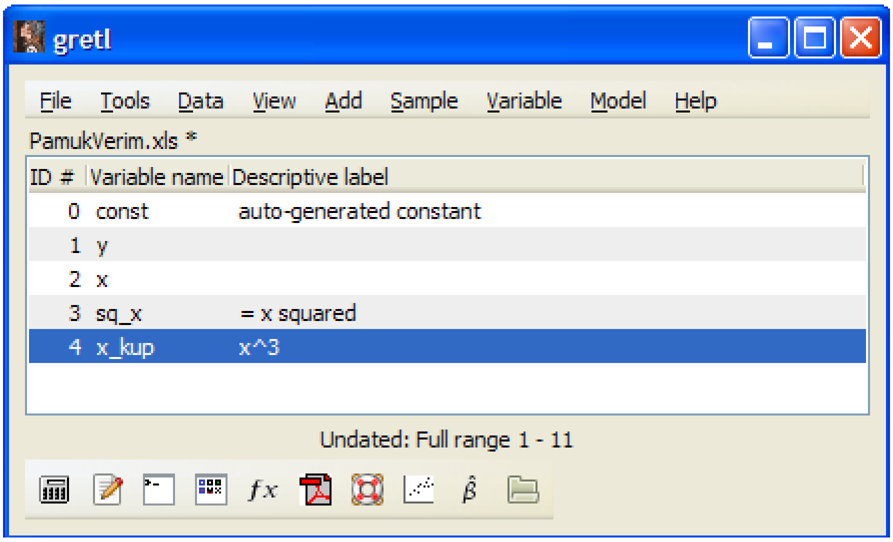




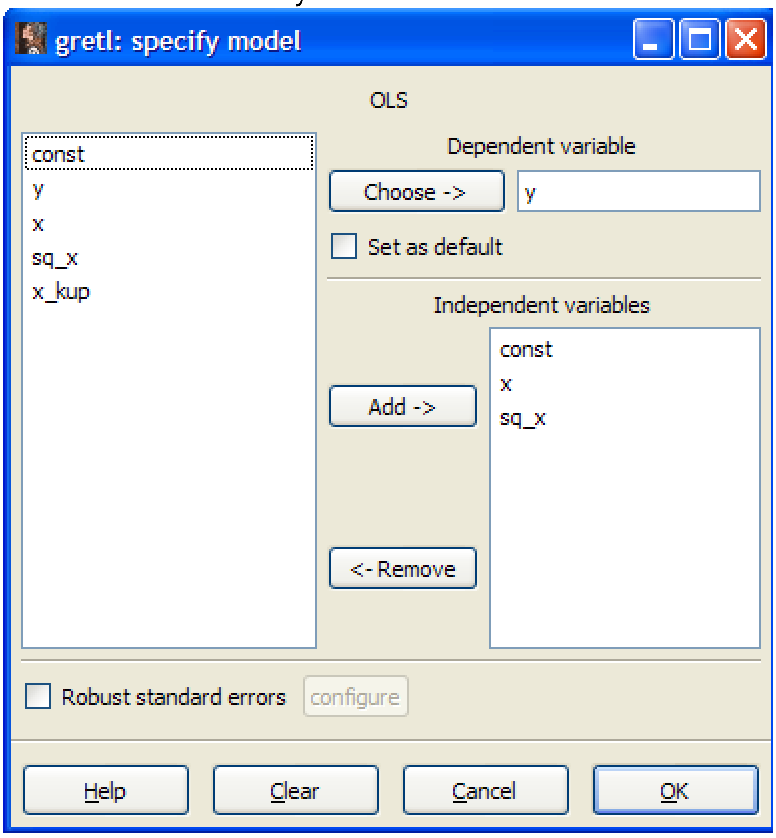
x'in küpünü “Define new variable” yardımıyla alabiliriz:

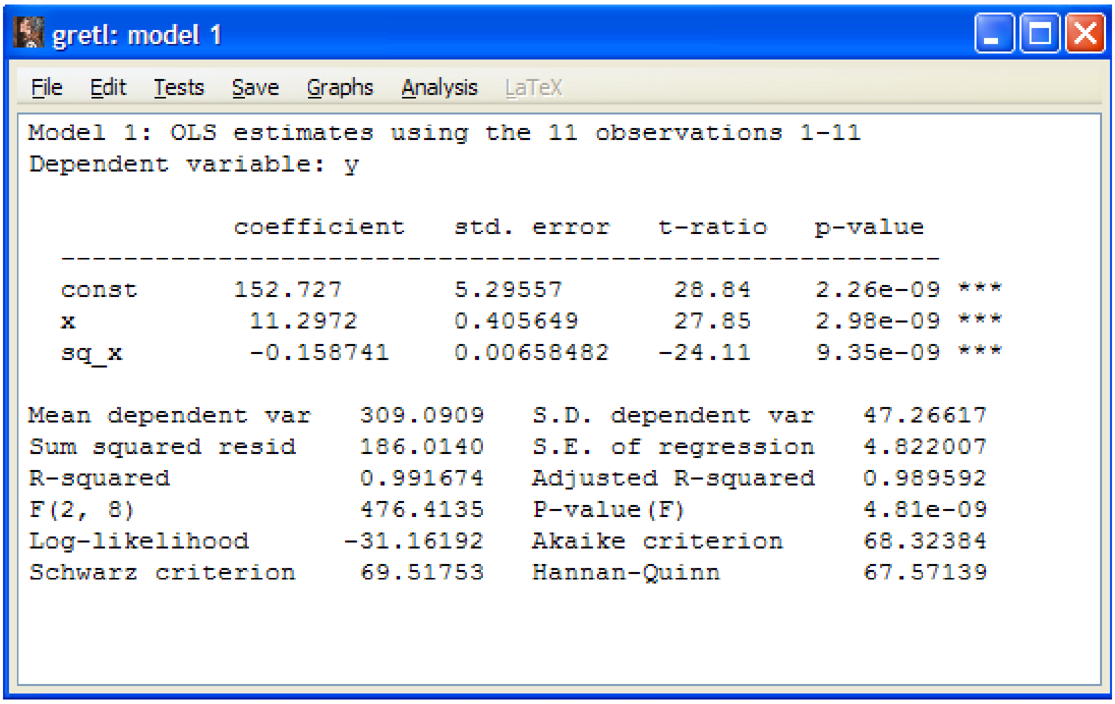




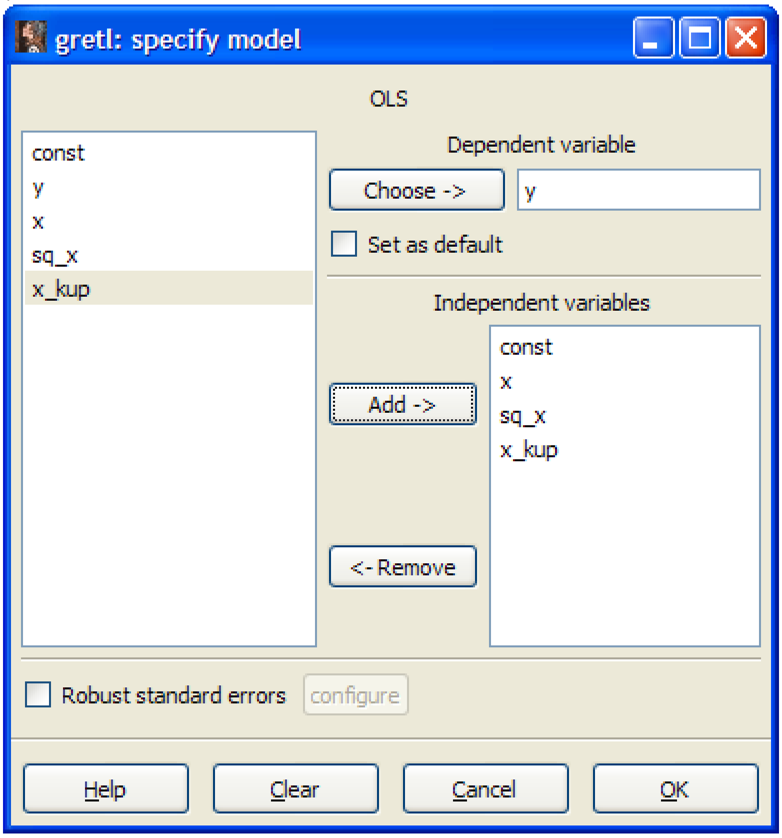


Önce kuadratik fonksiyonu tahmin edelim:





Şimdi de kübik fonksiyonu tahmin edelim:



```
gretl: model 2
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 2: OLS estimates using the 11 observations 1-11
Dependent variable: y

             coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const      134.167        1.83883        72.96    2.39e-011 ***
  x           14.3634       0.253819       56.59    1.41e-010 ***
  sq_x        -0.281119     0.00961662    -29.23    1.41e-08  ***
  x_kup        0.00135975   0.000105684    12.87    3.98e-06  ***

Mean dependent var   309.0909   S.D. dependent var   47.26617
Sum squared resid    7.546620   S.E. of regression   1.038310
R-squared            0.999662   Adjusted R-squared   0.999517
F(3, 7)              6905.234   P-value(F)           1.65e-12
Log-likelihood      -13.53595   Akaike criterion     35.07190
Schwarz criterion    36.66348   Hannan-Quinn         34.06863
```

4.3.5.6. Gretl ile Uygulamalar

Örnek 4-13

Y=Tereyağ talebi (10 ton) – X=Fiyat (100000 TL/kg)

Y	X	1/X	lnY	lnX
170	15	0.0667	5.135798	2.70805
80	16	0.0625	4.382027	2.772589
75	18	0.0556	4.317488	2.890372
70	22	0.0455	4.248495	3.091042
60	23	0.0435	4.094345	3.135494
45	25	0.0400	3.806662	3.218876

Doğrusal: $Y = b_0 + b_1 X$

	Katsayı	StHata	t	p
Sabit	248.944	70.1793	3.5473	0.023857
X	-8.3501	3.47795	-2.4009	0.074286
S = 31.6538	R^2 = 0.59	Düz. R^2=0.48		

Ters: $Y = b_0 + b_1 (1/X)$

	Katsayı	StHata	t	p
Sabit	-87.6414	63.4749	-1.3807	0.239496
1/X	3270.63	1192.71	2.7422	0.051789
S = 29.1424	R^2 = 0.65	Düz. R^2=0.56		

Lin-log: $Y = b_0 + b_1 \ln X$

	Katsayı	StHata	t	p
Sabit	579.306	193.972	2.9865	0.040476
lnX	-167.028	65.1896	-2.5622	0.062493
S = 30.4308	R^2 = 0.62	Düz. R^2=0.52		

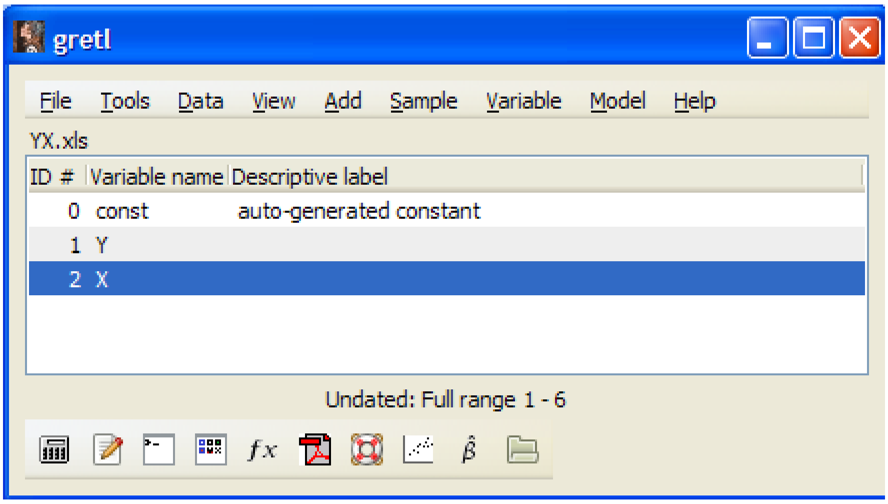
Log-lin: ln Y = b0 + b1 X

	Katsayı	StHata	t	p
Sabit	6.17397	0.579037	10.6625	0.000438
X	-0.0929328	0.0286959	-3.2385	0.031718
S = 0.26117	R^2 = 0.72	Düz. R^2=0.65		

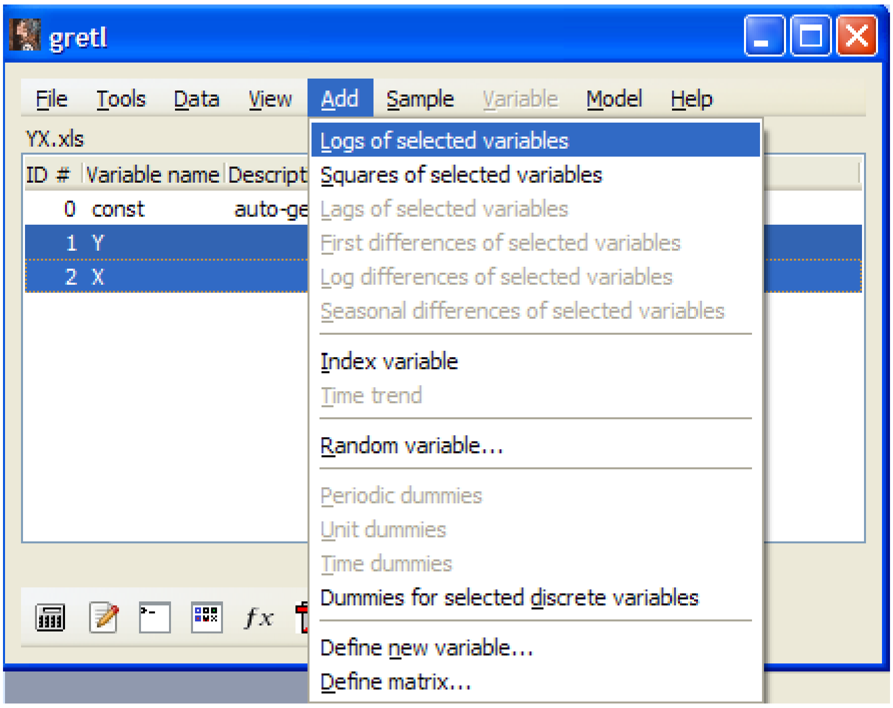
Log-log: ln Y = b0 + b1 lnX

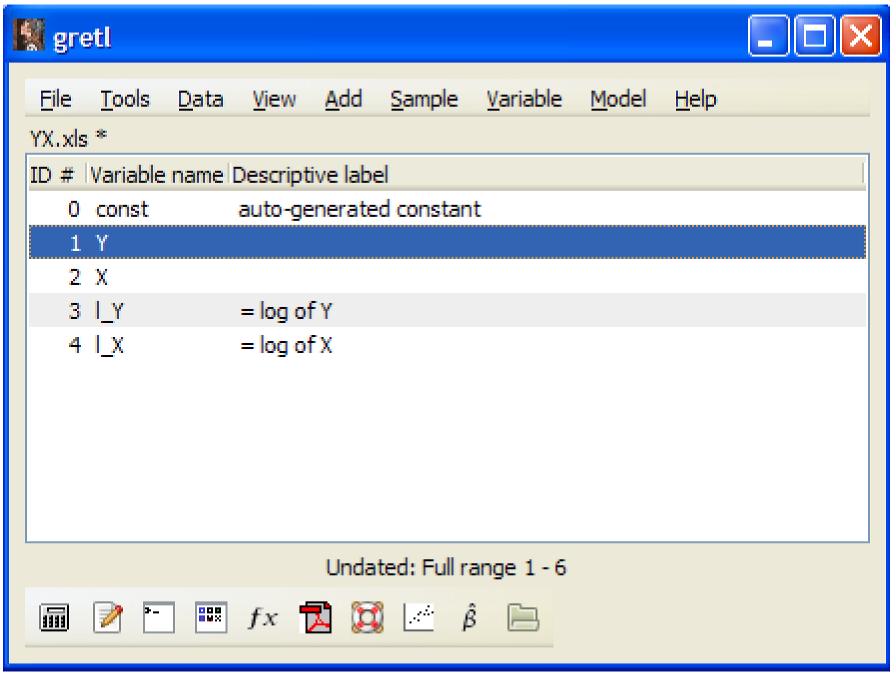
	Katsayı	StHata	t	p
Sabit	9.77426	1.61239	6.0620	0.003739
lnX	-1.83318	0.541886	-3.3830	0.027709
S = 0.25295	R^2 = 0.74	Düz. R^2=0.67		

Şimdi farklı formlarda modellerin tahminini Gretl yardımıyla yapacağız. Değişkenlerimiz başta sadece Y ve X olacak.

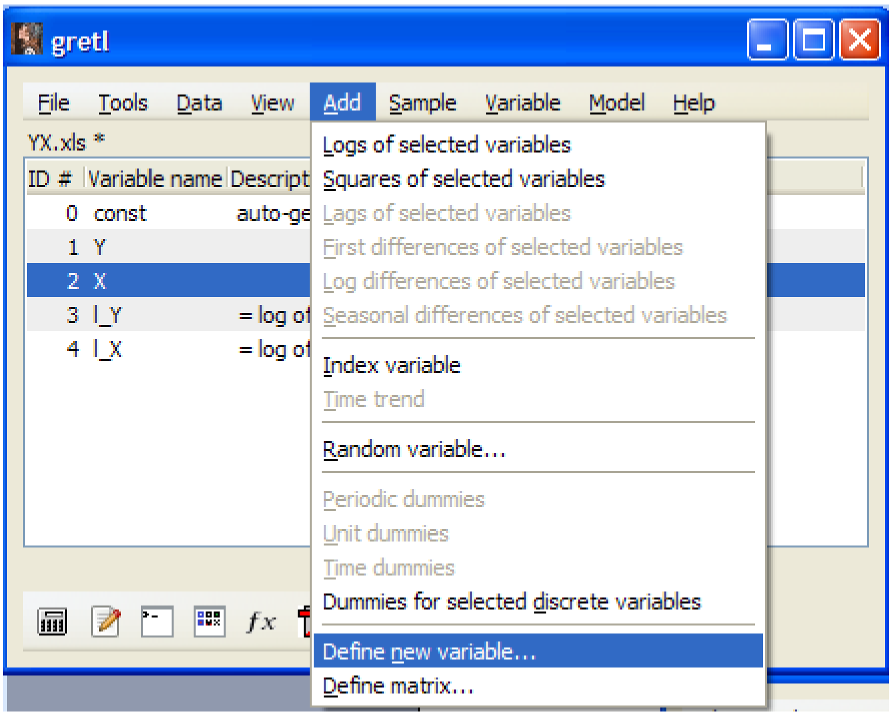


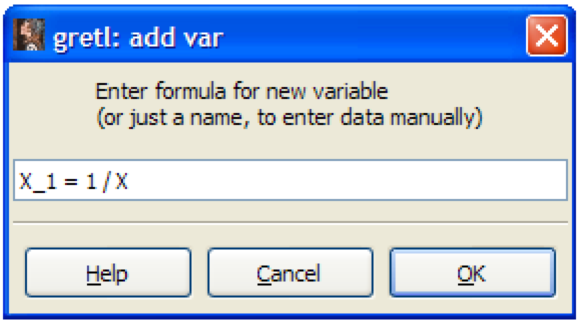
Logaritmik dönüştürme işlemlerinbir çırpıda yapabiliriz. Bunun için önce Y ve X değişkenlerini işaretleyip, **Add/Logs of selected variables** menüsünü seçmemiz gerekir:



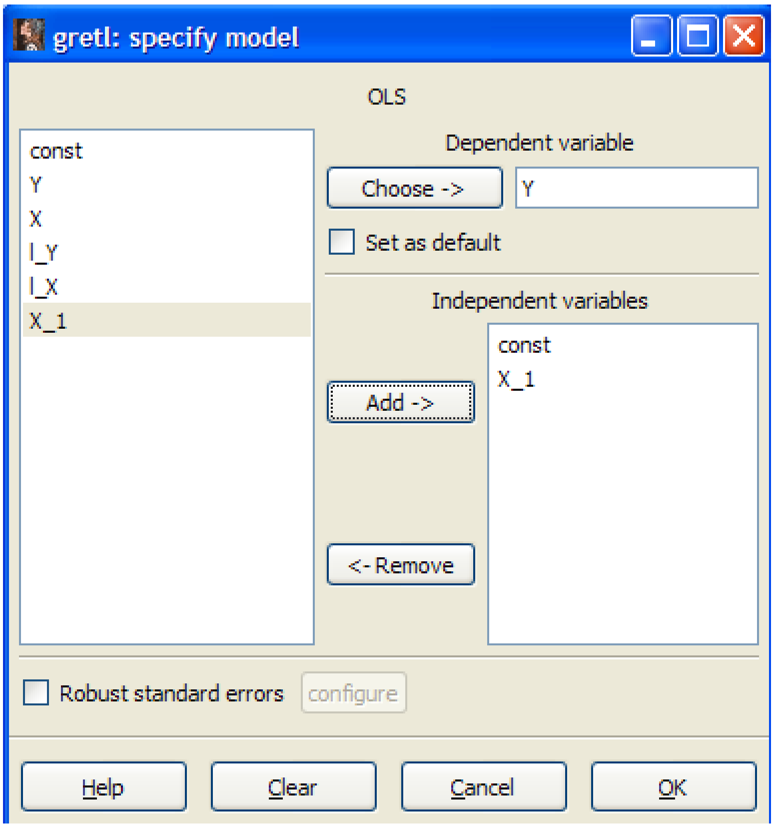


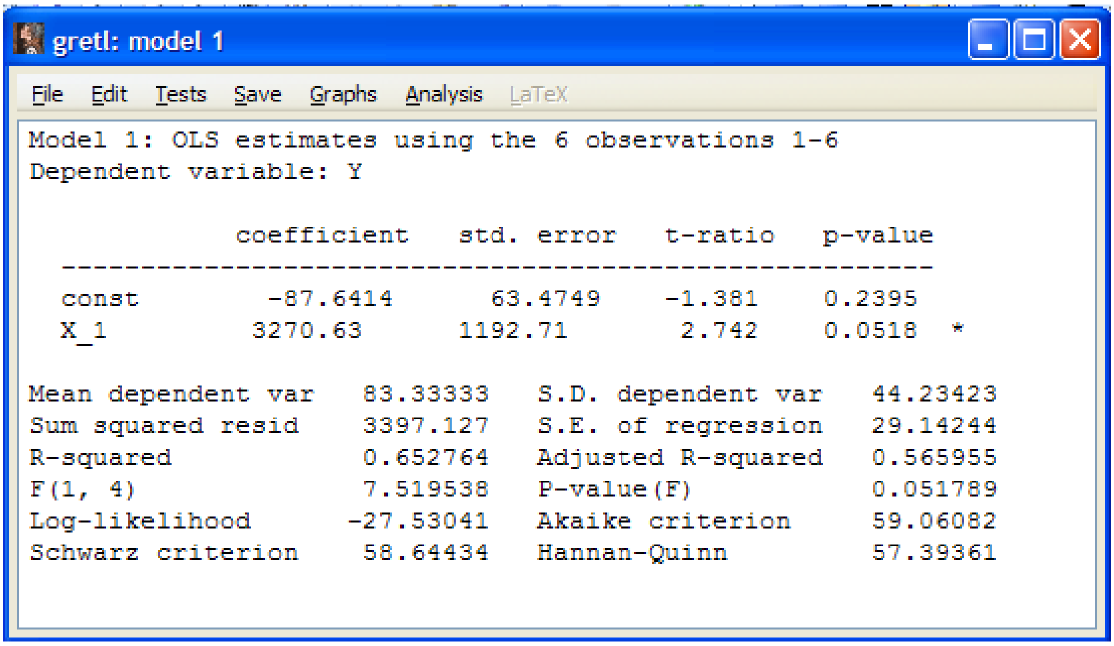
1/X dönüşümünü yapalım:





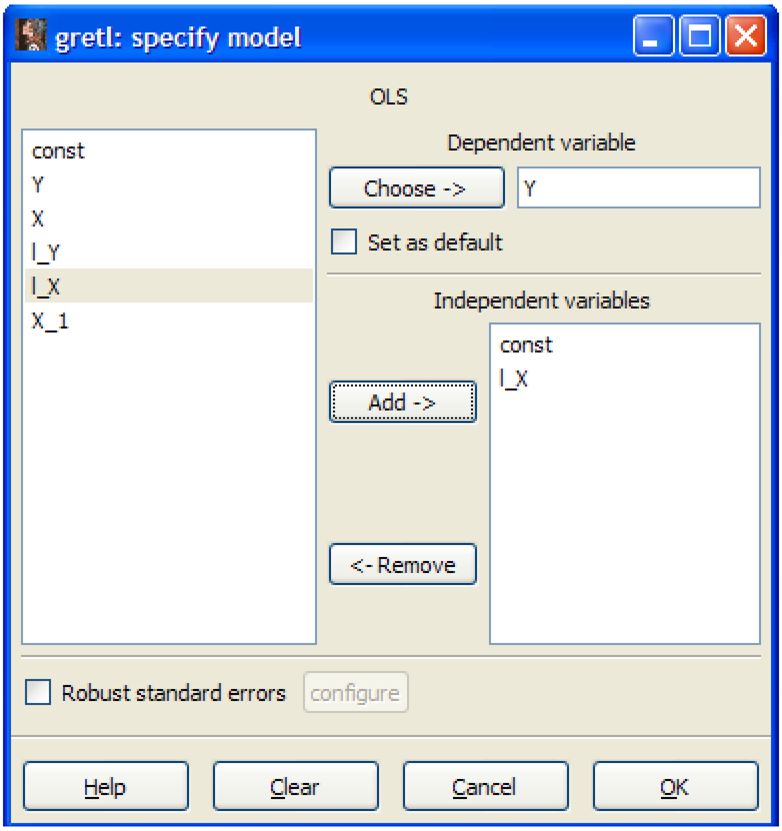
Ters fonksiyonu tahmin ediyoruz:

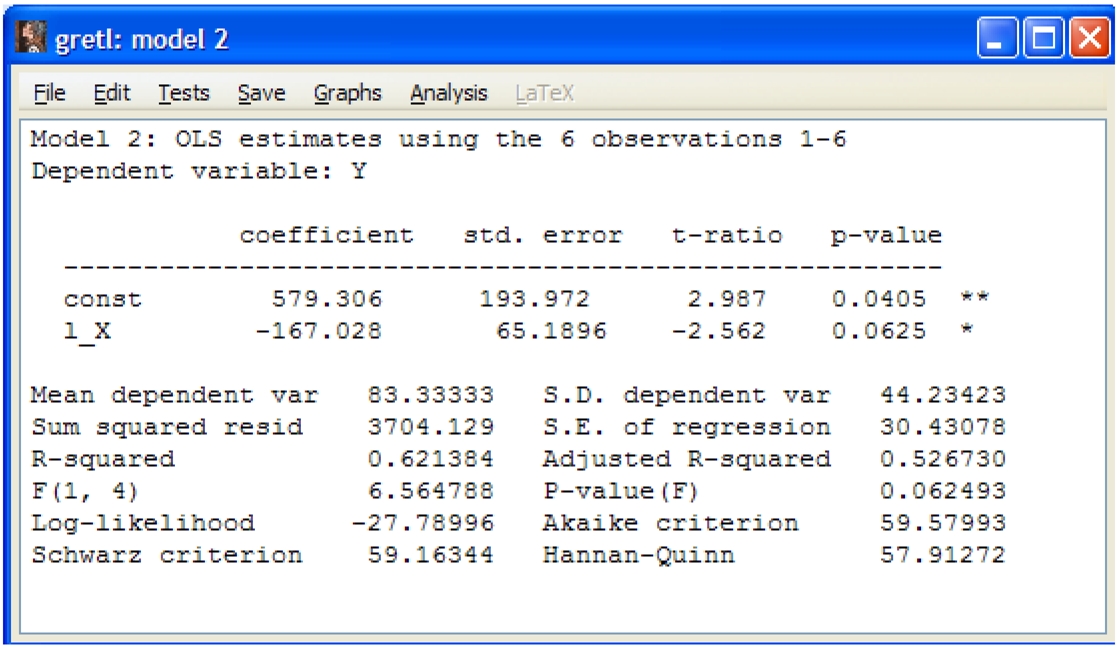





```
Model 1: OLS estimates using the 6 observations 1-6
Dependent variable: Y

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const        -87.6414      63.4749    -1.381    0.2395
  X_1         3270.63      1192.71       2.742    0.0518  *

Mean dependent var    83.33333   S.D. dependent var    44.23423
Sum squared resid     3397.127   S.E. of regression    29.14244
R-squared             0.652764   Adjusted R-squared    0.565955
F(1, 4)               7.519538   P-value(F)            0.051789
Log-likelihood       -27.53041   Akaike criterion      59.06082
Schwarz criterion     58.64434   Hannan-Quinn          57.39361
```

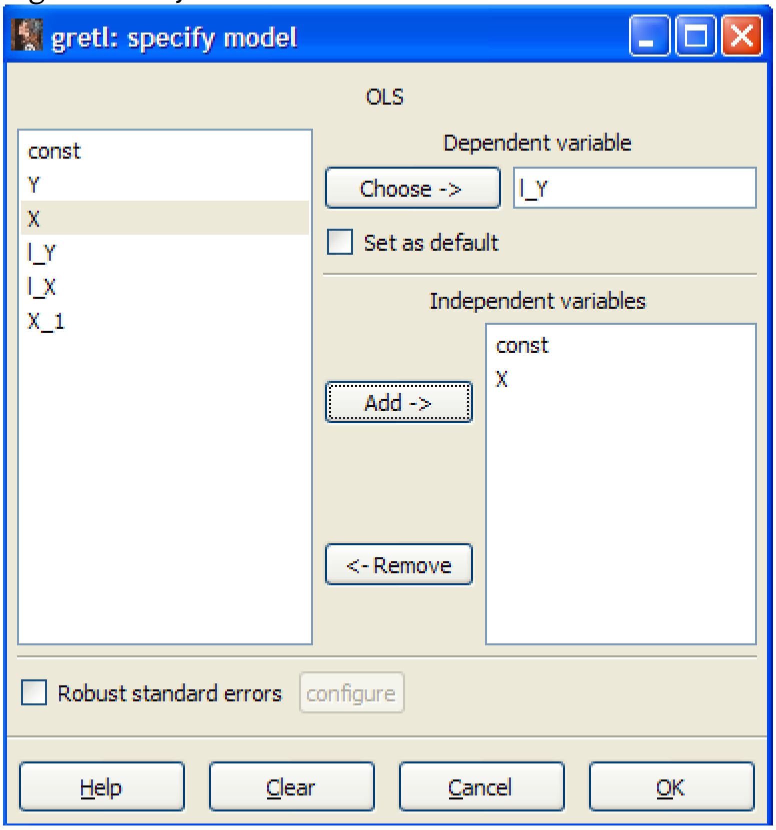
Lin-log fonksiyonu tahmin edelim:

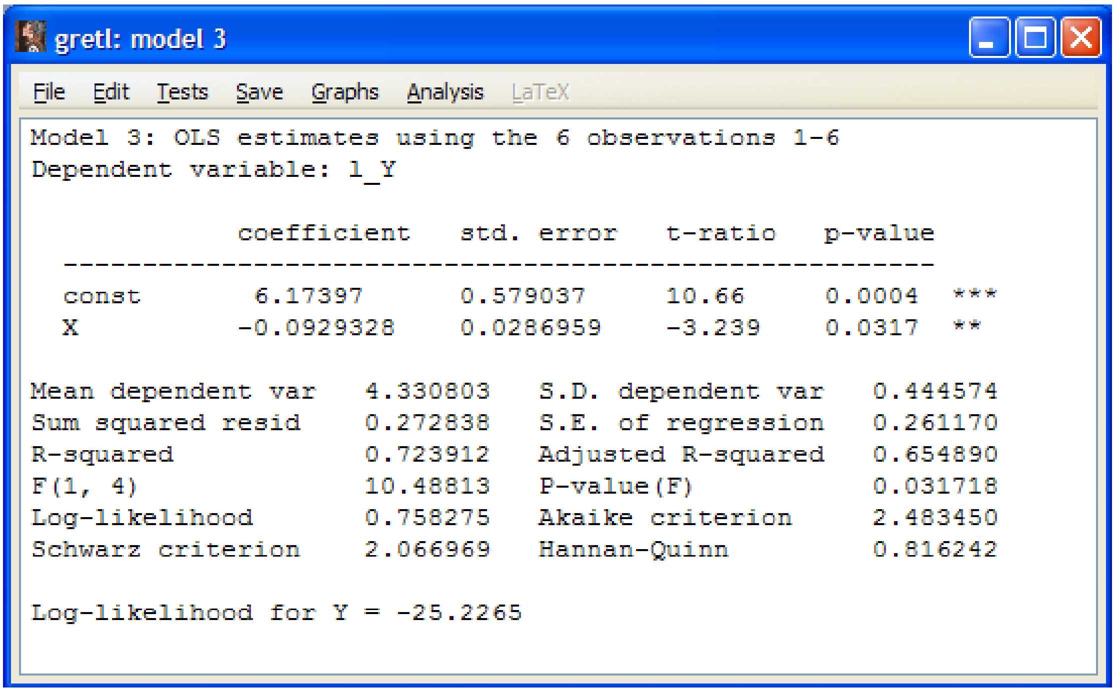




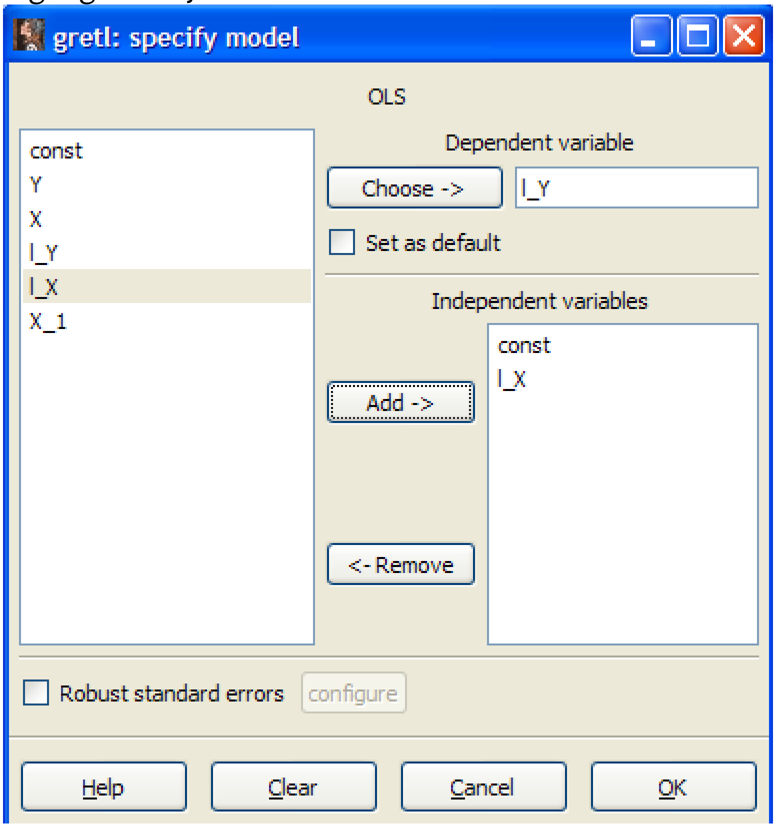

```
Model 2: OLS estimates using the 6 observations 1-6
Dependent variable: Y

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------------
  const        579.306      193.972      2.987    0.0405   **
  l_X         -167.028       65.1896    -2.562    0.0625   *

Mean dependent var    83.33333   S.D. dependent var    44.23423
Sum squared resid     3704.129   S.E. of regression    30.43078
R-squared             0.621384   Adjusted R-squared    0.526730
F(1, 4)               6.564788   P-value(F)            0.062493
Log-likelihood       -27.78996   Akaike criterion      59.57993
Schwarz criterion     59.16344   Hannan-Quinn          57.91272
```

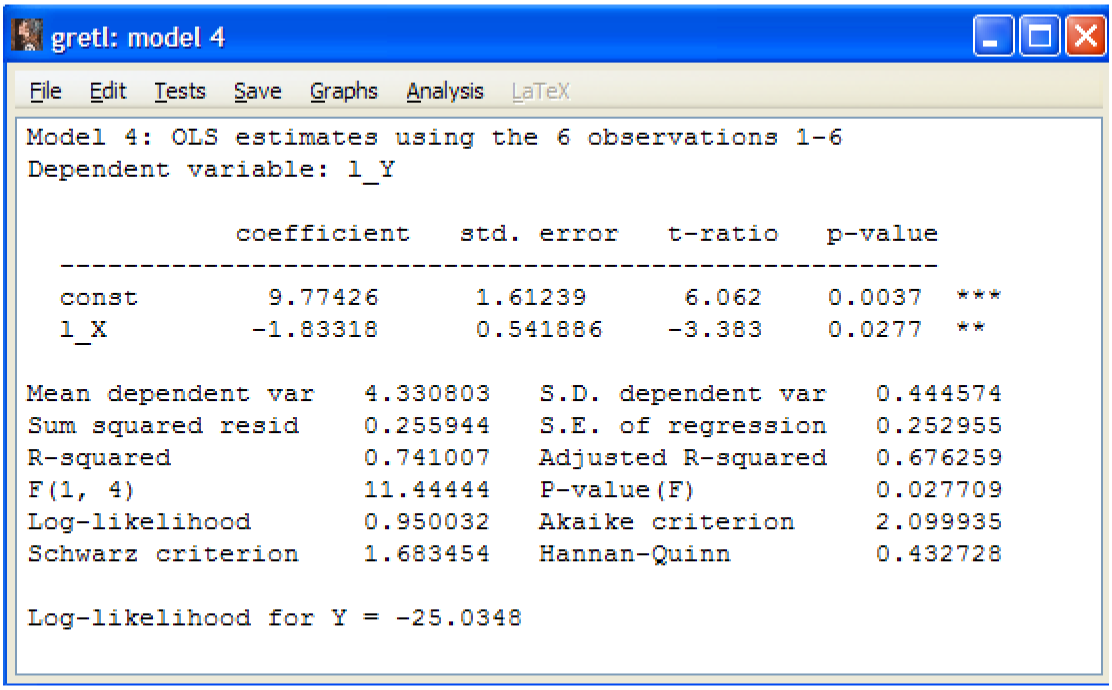
log-lin fonksiyonu tahmin edelim:





log-log fonksiyonu tahmin edelim:






```
Model 4: OLS estimates using the 6 observations 1-6
Dependent variable: l_Y

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const        9.77426      1.61239      6.062    0.0037  ***
  l_X         -1.83318      0.541886    -3.383    0.0277  **

Mean dependent var    4.330803   S.D. dependent var    0.444574
Sum squared resid     0.255944   S.E. of regression    0.252955
R-squared             0.741007   Adjusted R-squared    0.676259
F(1, 4)               11.44444   P-value(F)            0.027709
Log-likelihood        0.950032   Akaike criterion      2.099935
Schwarz criterion     1.683454   Hannan-Quinn          0.432728

Log-likelihood for Y = -25.0348
```

Örnek 4-14

Y= Gelir (TL) – X=Eğitim (yıl)

Y	X	1/X	lnY	lnX
250	8	0.1250000000	5.521461	2.079442
350	15	0.0666666667	5.857933	2.70805
100	5	0.2000000000	4.60517	1.609438
800	20	0.0500000000	6.684612	2.995732
80	3	0.3333333333	4.382027	1.098612
300	11	0.0909090909	5.703782	2.397895

Doğrusal: Y = bo + bıX

	Katsayı	StHata	t	p
Sabit	-86.2787	81.3259	-1.0609	0.348544
lX	38.6721	6.85698	5.6398	0.004866
S = 38241.5	R^2 = 0.89	Düz. R^2=0.86		

Ters: Y = bo + bı (1/X)

	Katsayı	StHata	t	p
Sabit	573.699	144.475	3.9709	0.016526
(1/X)	-1804.11	829.654	-2.1745	0.095326
S = 198.04	R^2 = 0.54	Düz. R^2=0.43		

lin-log: Y = bo + bı lnX

	Katsayı	StHata	t	p
Sabit	-371.354	212.341	-1.7489	0.155223
lnX	318.727	94.6795	3.3664	0.028138
S = 149.423	R^2 = 0.74	Düz. R^2=0.67		

Log-lin: ln Y = bo + bı X

	Katsayı	StHata	t	p
Sabit	4.12879	0.205279	20.1131	0.000036
X	0.128746	0.0173081	7.4385	0.001744
S = 0.246804	R^2 = 0.93	Düz. R^2=0.92		

Log-log: ln Y = bo + bı lnX

	Katsayı	StHata	t	p
Sabit	2.94995	0.330037	8.9382	0.000866
lnX	1.16806	0.147158	7.9374	0.001364
S = 0.232245	R^2 = 0.94	Düz. R^2=0.93		

Örnek 4-15

Farklı matematiksel formlarda talep modelleri:

Fiyat	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20
Talep	13	14	13	10	9	9	8	7	9	5	5	3	2	2	5

Doğrusal: Talep = $b_o + b_l$Fiyat

	Katsayı	StHata	t	p
Sabit	17.3214	1.04101	16.6391	0.00001
Fiyat	-0.763458	0.0767578	-9.9463	0.00001
S = 1.38788	R^2 = 0.88	Düz. R^2=0.87		

lin-log: Talep = $b_o + b_l$lnFiyat

	Katsayı	StHata	t	p
Sabit	28.338	2.22726	12.7232	0.00001
lnFiyat	-8.41631	0.891012	-9.4458	0.00001
S = 1.45228	R^2 = 0.87	Düz. R^2=0.86		

Log-lin: lnTalep = $b_o + b_l$Fiyat

	Katsayı	StHata	t	p
Sabit	3.34542	0.23493	14.2401	0.00001
lnFiyat	-0.116123	0.0173224	-6.7036	0.000015
S = 0.31321	R^2 = 0.77	Düz. R^2=0.76		

Lin-log: lnTalep = $b_o + b_l$ lnFiyat

	Katsayı	StHata	t	p
Sabit	4.90612	0.545022	9.0017	0.00001
lnFiyat	-1.23348	0.218035	-5.6573	0.000078
S = 1.64183	R^2 = 0.71	Düz. R^2=0.69		

Örnek 4-16

Farklı matematiksel formlarda talep modelleri:

Talep	Fiyat	lnTalep	lnFiyat
18	7.8	2.890372	2.054124
19	7	2.944439	1.94591
20	5.7	2.995732	1.740466
22	5.2	3.091042	1.648659
20	6.3	2.995732	1.84055
20	5.3	2.995732	1.667707
21	6.4	3.044522	1.856298
18	7.9	2.890372	2.066863
20	6.5	2.995732	1.871802
20	6.9	2.995732	1.931521
21	4.9	3.044522	1.589235

22	5.1	3.091042	1.629241
20	4.7	2.995732	1.547563
21	6.4	3.044522	1.856298
20	6.9	2.995732	1.931521

Doğrusal: $Y = b_o + b_l X$

	Katsayı	StHata	t	p
Sabit	25.4929	1.36799	18.6353	0.00001
Fiyat	-0.864454	0.217928	-3.9667	0.001610
S = 0.8287	R^2 = 0.55	Düz. R^2=0.51		

Lin-log: $Y = b_o + b_l \ln X$

	Katsayı	StHata	t	p
Sabit	29.3536	2.5473	11.5234	0.00001
lnFiyat	-5.08885	1.40046	-3.6337	0.003030
S = 0.86779	R^2 = 0.50	Düz. R^2=0.47		

Lin-log: $\ln Y = b_o + b_l X$

	Katsayı	StHata	t	p
Sabit	3.27205	0.0680258	48.1001	0.000
Fiyat	-0.0437612	0.0108369	-4.0382	0.001
S = 0.04120	R^2 = 0.56	Düz. R^2=0.52		

Log-log: $\ln Y = b_o + b_l \ln X$

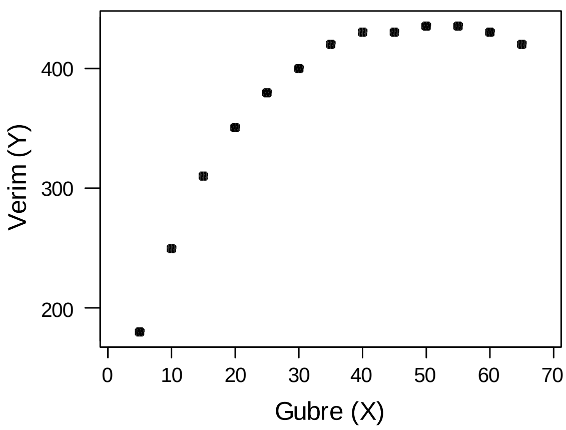
	Katsayı	StHata	t	p
Sabit	3.46672	0.127089	27.2779	0.000
lnFiyat	-0.257192	0.0698713	-3.6809	0.003
S = 0.04329	R^2 = 0.51	Düz. R^2=0.47		

Örnek 4-17

Aşağıdaki veriler, belli gübre dozları X karşısında elde edilen buğday verimlerini Y göstermektedir.

Y (kg/daa)	**X** (kg/daa)
180	5
250	10
310	15
350	20
380	25
400	30
420	35
430	40
430	45
435	50
435	55
430	60
420	65

Önce verilerimizin grafiğini çizelim:



Grafikten anlaşılacağı üzere, verilerimize uygun olan fonksiyonel form, ikinci dereceden denklem gibi görünmektedir. Ancak biz önce birinci dereceden doğrusal denklem tahminini yapalım:

$$Y=b_0 + b_1X$$

Modelin tahmin sonuçları:

	Katsayı	StHata	t	P
Sabit	250.58	26.51	9.45	0.000
X	3.544	0.6679	5.31	0.000
S = 45.05	R^2 = 0.719	Düz.R^2 = 0.694	F=28.15	p=0.000

Model, istatistiki açıdan geçerlidir. Denklem formunda yazarsak:

$Y = 250.58 + 3.544X$

Şimdi ikinci dereceden formu tahmin edelim:

$Y = b_0 + b_1X + b_2X^2$

Denklemdeki X^2'yi bizim hesaplamamız gerekir. Tahmin sonuçları:

	Katsayı	StHata	t	p
Sabit	136.5	10.08	13.54	0.000
X	12.6698	0.6622	19.13	0.000
X^2	-0.13037	0.009205	-14.16	0.000
S = 10.30	R^2 = 0.987	Düz.R^2 = 0.984	F=369.79	p=0.000

ikinci dereceden model %1 için önemlidir. Tahminciler de aynı şekilde % 1 için sıfırdan farklıdır. Bu model, birinci dereceden modele göre daha yüksek bir R^2'ye sahiptir. İkinci dereceden terim, modeli daha da iyileştirmiştir.

Modelimizi, denklem formunda yazalım:

$Y = 136.5 + 12.6698X - 0.13037X^2$

Son olarak, üçüncü dereceden modeli deneyelim:

	Katsayı	St Hata	t	p
Sabit	100.28	5.16	19.43	0.000
X	17.9309	0.6147	29.17	0.000
X^2	-0.31149	0.02002	-15.56	0.000
X^3	0.001725	0.000189	9.15	0.000
S = 3.382	R^2 = 0.999	Düz.R^2 = 0.998	F=2314.1	p=0.000

Üçüncü dereceden model, 1. ve 2. dereceden modellerden daha yüksek açıklayıcılığa sahiptir.

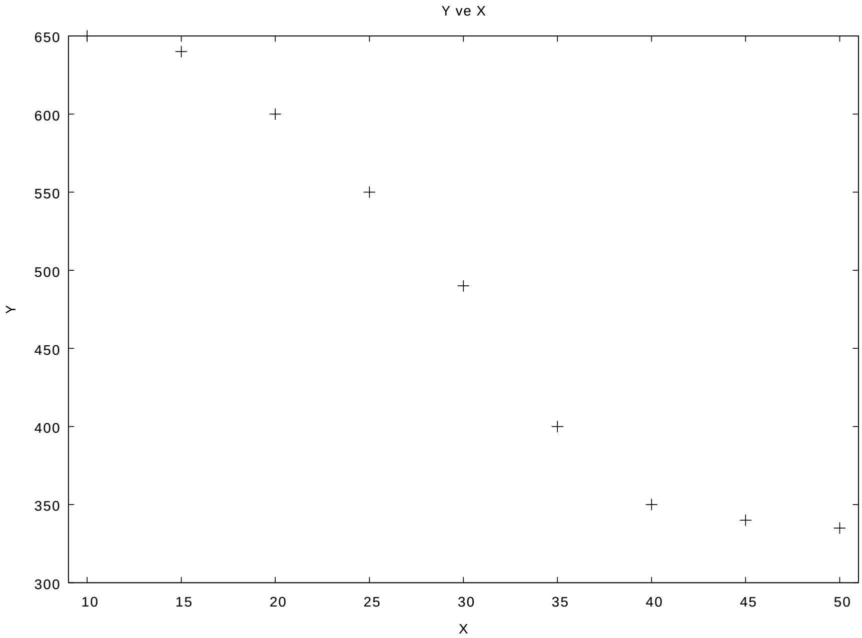
Örnek 4-18

Birim maliyet ile üretim miktarı arasındaki ilişkinin doğrusal fonksiyonla tahmin edilmesi teorik beklentiye cevap vermeyecektir. Bu nedenle ikinci veya üçüncü dereceden bir fonksiyonun daha uygun olacağını düşünebiliriz.

Aşağıdaki veriler Birim Maliyet (Y: TL/kg) ile toplam üretime (X: ton/ay) aittir.

Maliyet (TL/kg) Y	Üretim (100 Ton/ay) X
650	10
640	15
600	20
550	25
490	30
400	35
350	40
340	45
335	50

Önce verilerimizi grafik üzerinde görelim. Grafik doğrusal fonksiyonun uygun olmadığını doğrular niteliktedir.



Önce doğrusal fonksiyonu tahmin edeceğiz:

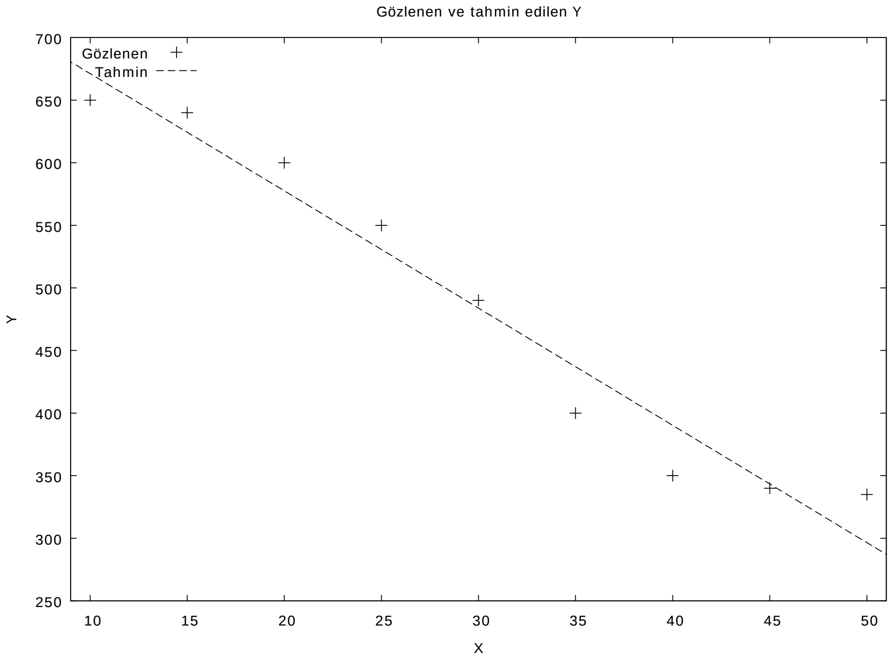
$Y = b_0 + b_1X$

Bağımlı değişken: Y

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	764.889	24.8673	30.7588	<0.00001	***
X	-9.36667	0.761404	-12.3018	<0.00001	***

R-kare	0.955790	Düzeltilmiş R-kare	0.949474
F(1, 7)	151.3353	P(F)	5.38e-06
Log-likelihood	-42.09570	Akaike kriteri	88.19140
Schwarz kriteri	88.58585	Hannan-Quinn	87.34018

Doğrusal fonksiyon istatistiki açıdan anlamlıdır. Ancak teoriye göre üretim miktarı arttıkça birim maliyetin önce azalması, daha sonra artması gerekmektedir. Doğrusal fonksiyonun grafik görüntüsü de bunu sağlamadığını açıkça göstermektedir:



Şimdi ikinci dereceden modeli tahmin edebiliriz:

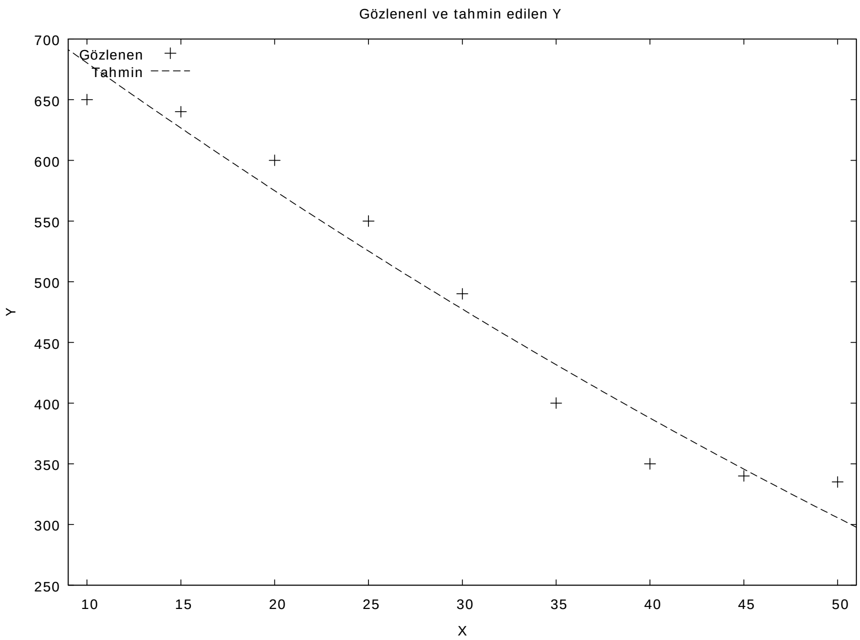
$Y = b_0 + b_1X + b_2X^2$

Bağımlı değişken: Y

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	793.143	58.2134	13.6248	<0.00001	***
X	-11.6784	4.32748	-2.6986	0.03564	**
X_kare	0.0385281	0.0708726	0.5436	0.60629	

R-kare	0.957865	Düzeltilmiş R-kare	0.943820
F(2, 6)	68.20032	P(F)	0.000075
Log-likelihood	-41.87934	Akaike kriteri	89.75867
Schwarz kriteri	90.35035	Hannan-Quinn	88.48184

İkinci dereceden fonksiyon istatistiki açıdan anlamlıdır. Ancak kareli terim anlamlı değildir. Bir başka ifadeyle ikinci dereceden terim modeli iyileştirmemiştir. Model seçim kriterleri de kuadratik fonksiyonun doğrusal fonksiyondan daha iyi bi model olmadığını göstermektedir. Bu görüşü, modelin grafiği de doğrulamaktadır.



Son olarak üçüncü dereceden modelimizi tahmin edebiliriz:

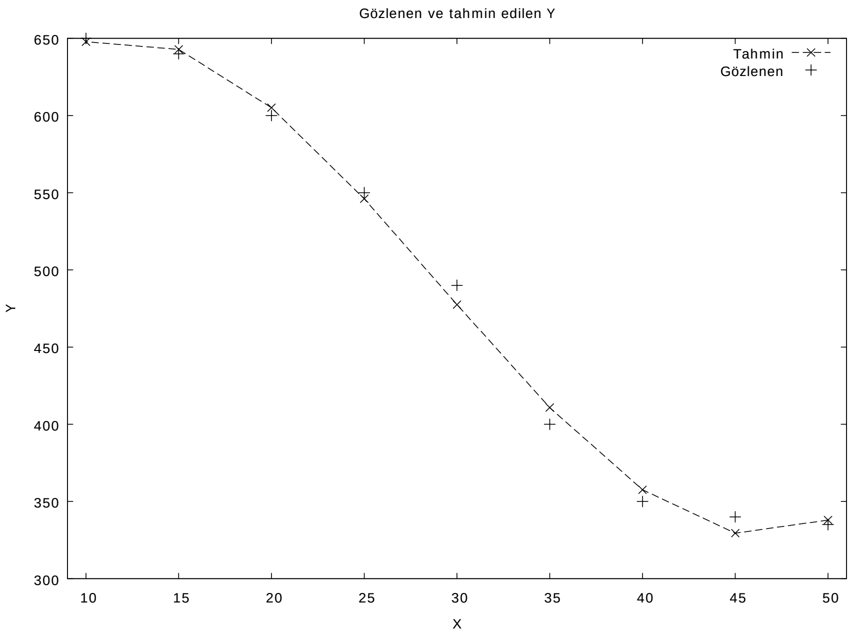
$$Y = b_0 + b_1X + b_2X^2 + b_3X^3$$

Bağımlı değişken: Y

	Katsayı	*St.Hata*	*t*	*p*	
sabit	513.254	42.9589	11.9476	0.00007	***
X	25.4089	5.30555	4.7891	0.00493	***
X_kare	-1.34935	0.192893	-6.9953	0.00092	***
X_kup	0.015421	0.002128	7.2463	0.00078	***

R-kare	0.996337	Düzeltilmiş R-kare	0.994139
F(3, 5)	453.2916	P(F)	1.65e-06
Log-likelihood	-30.88811	Akaike kriteri	69.77622
Schwarz kriteri	70.56512	Hannan-Quinn	68.07378

Üçüncü dereceden fonksiyon istatistiki olarak anlamlıdır. Modeldeki tüm terimler sıfırdann farklıdır. Model seçim kriterleri de üçüncü dereceden fonksiyonun en iyi model olduğunu kanıtlamaktadır. Modelin grafiği de bu başarıyı açıkça göstermektedir.



Bağımlı değişken : Y

	Doğrusal	İkinci derece	Üçüncü derece
sabit	764.9**	793.1**	513.3**
	(24.87)	(58.21)	(42.96)
X	-9.367**	-11.68**	25.41**
	(0.7614)	(4.327)	(5.306)
X_kare		0.03853	-1.349**
		(0.07087)	(0.1929)
X_kup			0.01542**
			(0.002128)
n	9	9	9
Düz. R^2	0.9495	0.9438	0.9941

Standart hatalar parantez içindedir
* %10'da önemli
** %5'te önemli

Ramsey RESET Testi

İkinci veya daha yüksek dereceden değişken içeren modellerin geliştirilmesi sürecini, deneme yanılma yoluyla gerçekleştirdik. Bu süreci, Ramsey RESET testiyle gerçekleştireceğiz. Örnek 4-17'deki verim-gübre örneğinde, doğrusal modele ikinci ve üçüncü dereceden terimleri eklemenin gerekli olup, olmadığını belirlemek üzere, testin aşamaları:

1. Adım

   Hipotezleri hazırla:

   H_0: İkinci ve üçüncü terimler gereksizdir

   H_1: İkinci ve üçüncü terimler gereklidir.

2. Adım

   Birinci dereceden modeli tahmin et ve $\hat{Y}$ değerlerini hesapla

3. Adım

   $\hat{Y}^2$ ve $\hat{Y}^3$, birinci dereceden doğrusal modele ekle ve modeli tahmin et.

4. Adım

   $$F = [(HKT_B - HKT_Y)/M] / [HKT_Y/(n\text{-}k)]$$

   değerini hesapla.

   Burada:

   HKT_B : Birinci dereceden modelin hata kareleri toplamı

   HKT_Y : İkinci ve üçüncü dereceden terimler eklendikten sonraki modelin hata kareleri toplamı

   M: Eklenecek terim sayısı

   n–k: İkinci ve üçüncü dereceden terimlerin eklendiği denklemin serbestlik derecesidir.

5. Adım

   F testini, M ve n–k serbestlik derecelerine göre yap.

Şimdi Ramsey RESET testini, verilerimize uygulayalım.

1. Adım

   Hipotezleri hazırlayalım:

   H_0: İkinci ve üçüncü terimler gereksizdir

   H_1: İkinci ve üçüncü terimler gereklidir.

2. Adım

   Birinci dereceden modeli tahmin edelim:

	Tahminci	StHata	t	p
Sabit	250.58	26.51	9.45	0.000
X	3.544	0.6679	5.31	0.000
S = 45.05	R^2 = 0.719	Düz.R^2 = 0.694	F=28.15	HKT_B=22327

$\hat{Y}$ değerleri

268.297	286.016	303.736	321.456	339.176	356.896	374.615	392.335	410.055	427.775	445.495	463.214	480.934

3. Adım

$\hat{Y}^2$ ve $\hat{Y}^3$, birinci dereceden doğrusal modele ekle ve modeli tahmin edelim.

	Tahminci	StHata	t	p
Sabit	2876.9	228.6	12.58	0.000
X	87.849	6.234	14.09	0.000
$\hat{Y}^2$	-0.05393	0.004766	-11.32	0.000
$\hat{Y}^3$	0.0000388	0.00000424	9.15	0.000
S = 3.382	R^2 = 0.999	Düz.R^2 = 0.998	F=2314.1	HKT_Y=103

4. Adım

2. ve 3. dereceden terimleri ekleyeceğimiz için, M=2'dir.

$F = [(HKT_B - HKT_Y)/M] / [HKT_Y/(n\text{-}k)]$

$F = [(22327 - 103)/2] / [103/9]$

$F = 970.1$

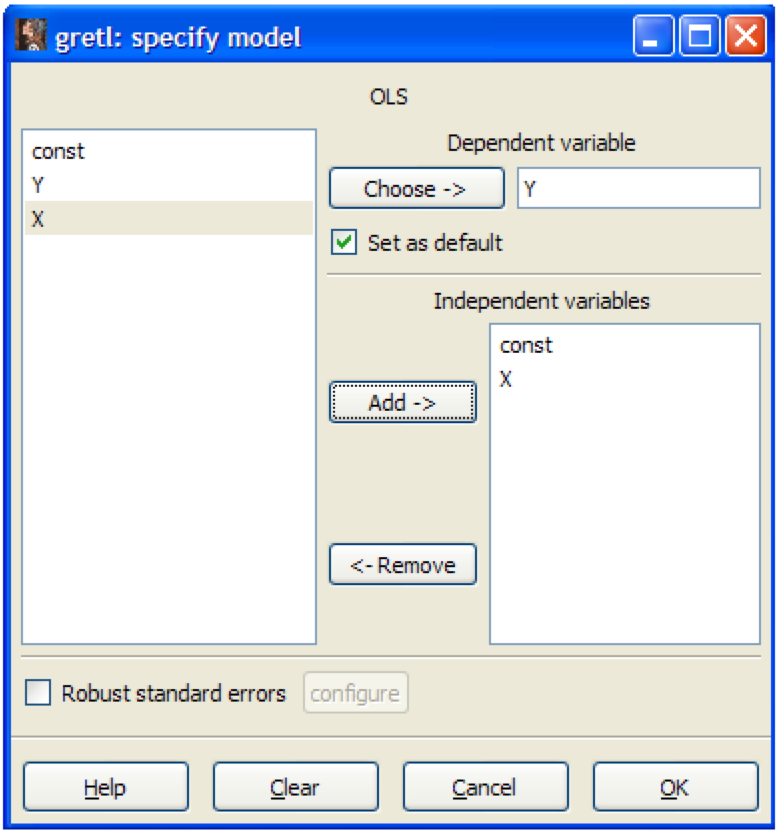
5. Adım

$F_{2,\,9,\,0.05} = 4.26$

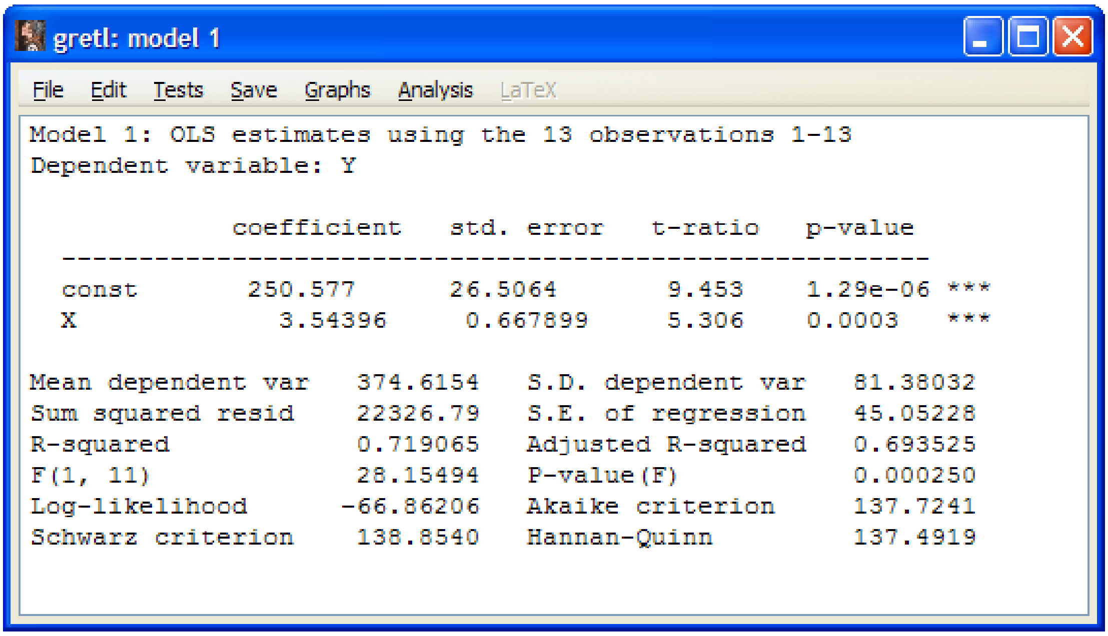
Hesapladığımız F değeri (970.1), tablo değeri olan 4.26'dan büyük olduğundan, H_0 reddedilir. Buna göre, eklenen değişkenler gereklidir.

Ramsey RESET testini Gretl ile yapabiliriz:

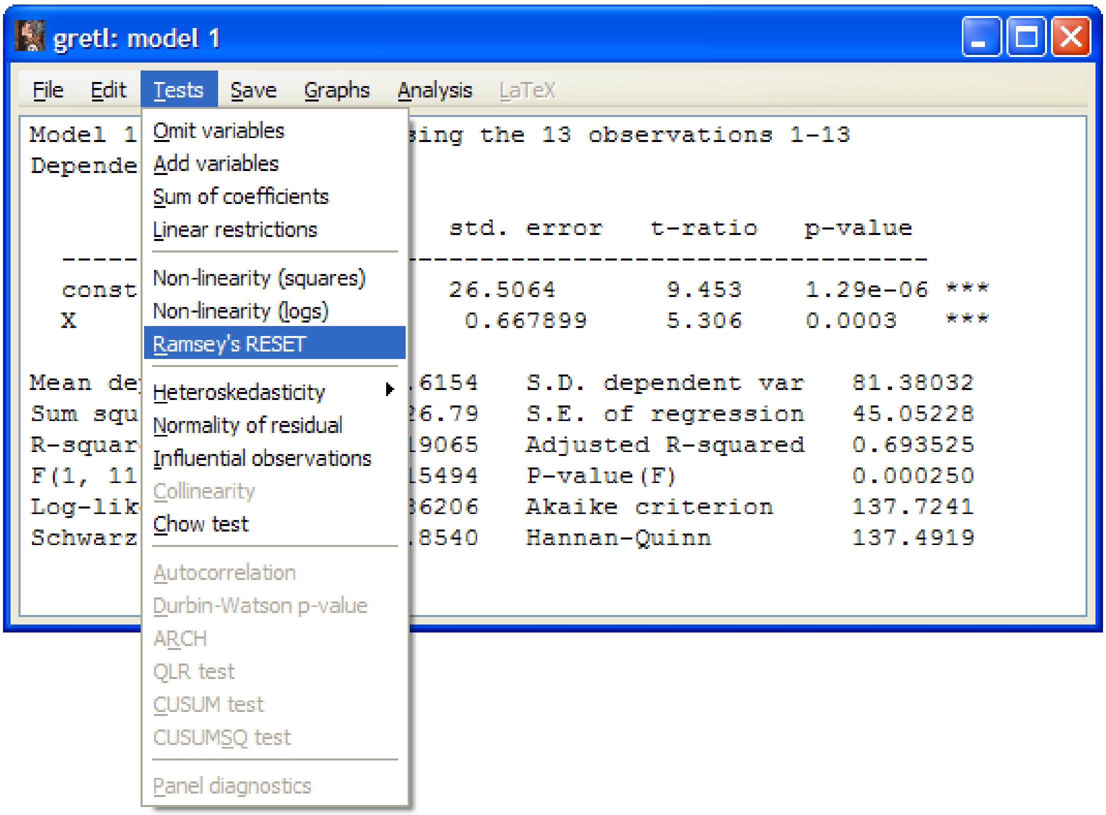
Model değişkenlerimiz sadece Y ve X'tir. Bağımlı değişken olarak Y'yi bağımsız değişken olarak X'i atıyoruz.



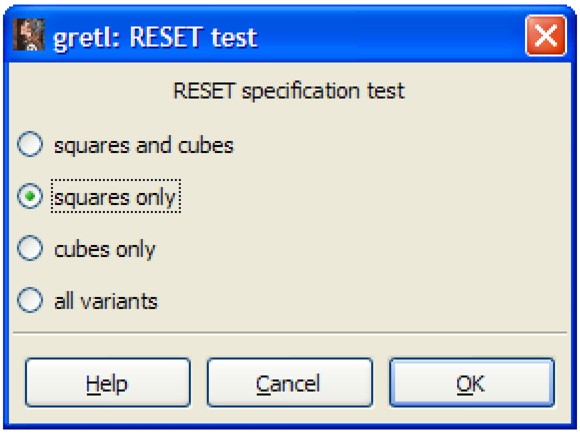
Doğrusal modelimizin EKK tahmin sonuçları:



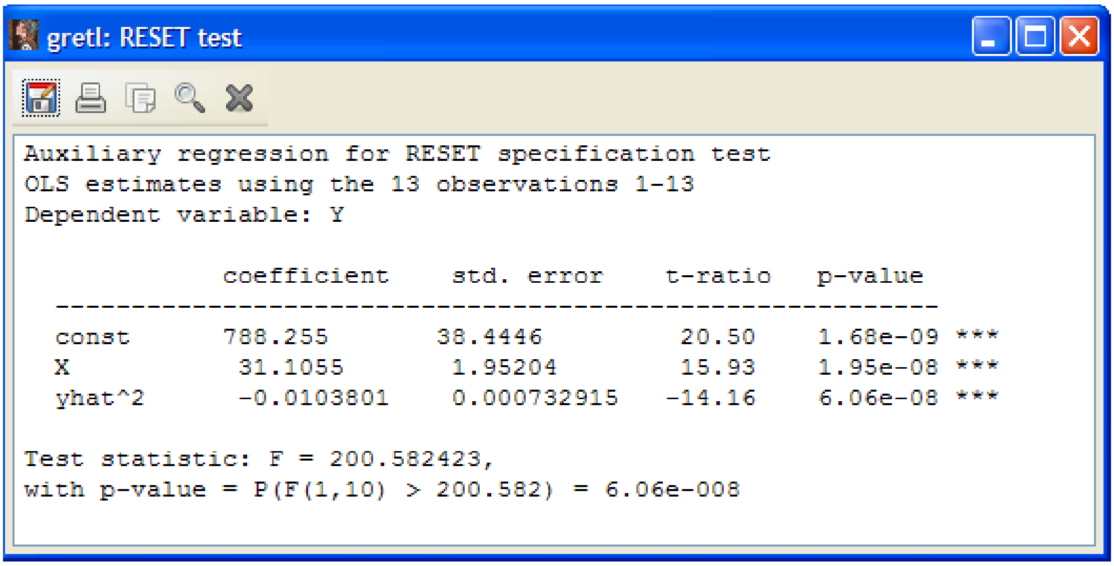
Ramsey RESET testi için, Tests/Ramsey's RESET menü dizisini kullanmamız gerekiyor.



Ramsey RESET testi; kareli ve küplü (squares and cubes), sadece kareli terim (squares only), sadece küplü terim (cubes only) ve tüm kombinasyonlar (all variants) seçenekleri için yapılabilir. Biz şimdi sadece kareli terimi seçeceğiz:

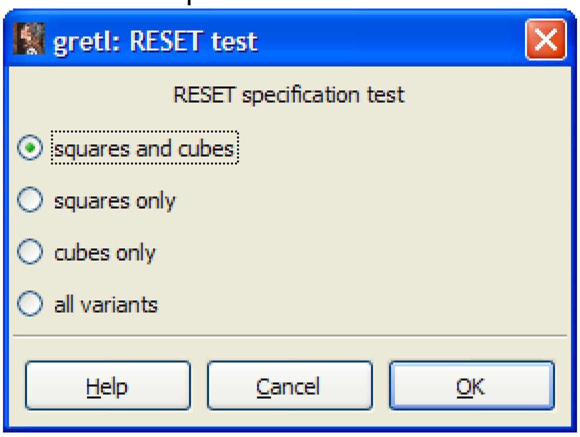


Sıfır hipotezi, kareli terimin sıfır olduğunu, yani kareli terimin modele dahil edilmemesi gerektiğini savunur. Buna göre F testi, sıfır hipotezinin reddedilmesini önerir. Bir başkka ifadeyle, kareli terim modele dahil edilmelidir.

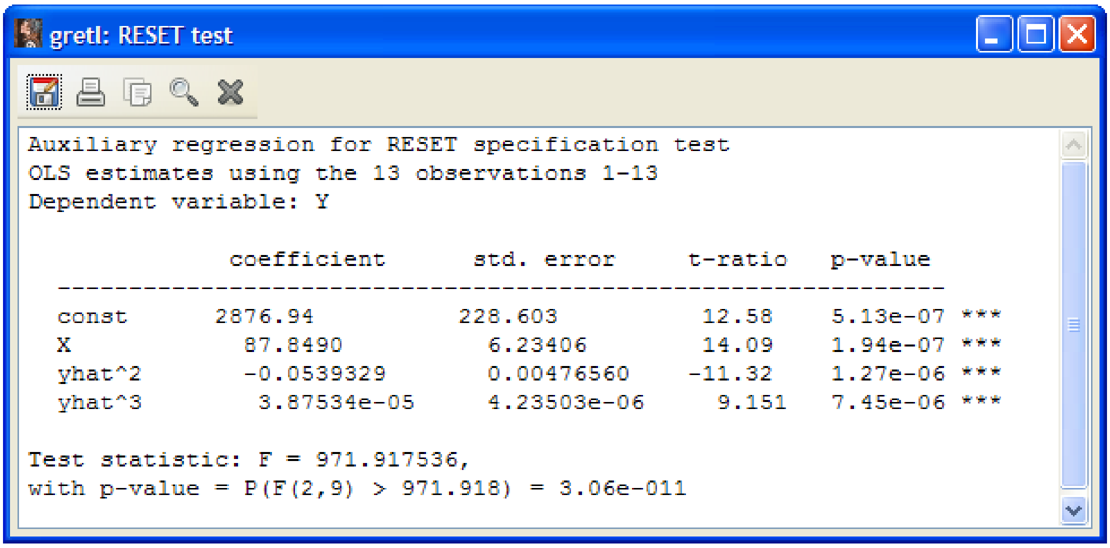



```
Auxiliary regression for RESET specification test
OLS estimates using the 13 observations 1-13
Dependent variable: Y

             coefficient    std. error    t-ratio    p-value
  ---------------------------------------------------------
  const       788.255       38.4446        20.50     1.68e-09 ***
  X            31.1055       1.95204       15.93     1.95e-08 ***
  yhat^2       -0.0103801    0.000732915  -14.16     6.06e-08 ***

Test statistic: F = 200.582423,
with p-value = P(F(1,10) > 200.582) = 6.06e-008
```

Şimdi de kareli ve küplü terimlerin birlikte modele alınması durumunu test edeceğiz.



Sıfır hiptezi, kareli ve küplü terimin her ikisinin birden sıfır olduğunu savunur. F testi, sıfır hipotezinin reddedilmesini önerir. O halde, ikinci ve üçüncü dereceden terimlerden en az birinin modele alınması gerekmektedir. Örneğimizde her ikisinin birden modele alınması gerektiğini, yhat^2 ve yhat^3 terimlerinin istatistiki açıdan anlamlı olmasından anlıyoruz.




```
Auxiliary regression for RESET specification test
OLS estimates using the 13 observations 1-13
Dependent variable: Y

               coefficient      std. error     t-ratio    p-value
  ---------------------------------------------------------------
  const        2876.94          228.603         12.58     5.13e-07 ***
  X              87.8490          6.23406       14.09     1.94e-07 ***
  yhat^2         -0.0539329       0.00476560   -11.32     1.27e-06 ***
  yhat^3          3.87534e-05     4.23503e-06    9.151    7.45e-06 ***

Test statistic: F = 971.917536,
with p-value = P(F(2,9) > 971.918) = 3.06e-011
```

4.3.6. Tahmincilerin Ortak Testi

t testi, tek bir tahmincinin istatistiki açıdan anlamlılığını belirler. Birden fazla değişkene ait tahmincilerin sıfırdan farklı olup olmadığını belirlemek de mümkündür. Böylece modeldeki değişkenlerin modelde tutulmasının gerekliliği irdelenir.

Aşaşıdaki modelleri ele alalım:

U: $Y = b_0 + b_1X1 + b_2X2 + b_3X3 + e$

R: $Y = a_0 + a_1X1 + v$

U, kısıtsız modeli; R, kısıtlı modeli temsil etmektedir. R modelinde b_2 ve b_3 sıfır kabul edildiğinden, kısıtlı model olarak adlandırılmıştır. Bu tahmincilerin gerçekten de sıfır olduğuna karar vermek için Wald testinden yararlanabiliriz.

Wald testinin aşamaları:

1. Adım

   Hipotezleri belirle:

   $H_0: b_1=b_2=0$

   H_1: En az biri sıfırdan farklıdır.

2. Adım

   $Y = b_0 + b_1X1 + b_2X2 + b_3X3 + e$

   modelini tahmin et ve HKT_U'yu hesapla.

   $Y = a_0 + a_1X1 + v$

   modelini tahmin et ve HKT_R'yi hesapla.

3. Adım

   $F=(\text{HKT Farkı} / \text{Kısıt sayısı}) / (HKT_U / Sd_U)$

   $F=[(HKT_R-HKT_U)/(Sd_R-Sd_U)] / (HKT_U/Sd_U)$

hesaplamasını yap.

4. Adım

Belli bir α düzeyi için f_1=k-m için f_2=n-k için F tablosu değerini bul. Burada k, kısıtsız model tahminci sayısı; m, kısıtlı model tahminci sayısıdır.

5. Hesaplanan F değeri, tablo F değerinden büyükse H0 hipotezini reddet.

Örnek 4-19

Burada Wald testi uygulaması yapacağız. Modelimiz, Ev fiyatları ve bunu etkileyen özelliklerden oluşmaktadır:

Fiyat (TL) Y	Alan (m^2) X1	Yatak Odası (m^2) X2	Banyo sayısı X3
20.0	106.5	15	1
22.8	125.4	15	2
23.5	130.0	15	2
28.5	157.7	20	3
23.9	160.0	15	2
29.3	175.0	20	2
28.5	180.0	20	3
36.5	187.0	20	2
29.5	193.5	20	3
29.0	194.8	20	2
38.5	225.4	20	3
50.5	260.0	15	3
42.5	280.0	20	3
41.5	300.0	20	3

Kısıtsız Model: $Y = b_0 + b_1X1 + b_2X2 + b_3X3 + e$

Kısıtlı Model: $Y = a_0 + a_1X1 + v$

1. Adım

   Hipotezleri belirle:

   H_0: $b_1=b_2=0$

   H_1: En az biri sıfırdan farklıdır.

2. Adım

   $Y = b_0 + b_1X1 + b_2X2 + b_3X3 + e$

   modelini tahmin edelim:

	Katsayı	St.Hata	t	p
Sabit	12.077	8.307	1.45	0.177
X1	0.13969	0.02793	5	0.000
X2	-0.5568	0.5342	-1.04	0.322
X3	1.285	2.576	0.5	0.629
S = 27.83	R^2 = 0.972	Düz.R^2 = 0.969	F=282.2	p=0.000

HKT_U=164.23, Sd_U=10

$Y = a_0 + a_1X1 + v$

modelini tahmin edelim:

	Tahminci	StHata	t	p
Sabit	5.24	3.729	1.41	0.185
X1	0.13873	0.01873	7.41	0.000
S = 28.54	R^2 = 0.969	Düz.R^2 = 0.967	F=802.2	p=0.000

HKT_R=,182.74 Sd_R=12

3. Adım

   F=(HKT Farkı / Kısıt sayısı) / (HKT_U / Sd_U)

   F=[(HKT_R-HKT_U)/(Sd_R-Sd_U)] / (HKT_U/Sd_U)

   F=[(182.74-164.23)/(12-10)] /(164.23/10)

   F=0.56

4. Adım

   α = 0.05, f_1=k-m=4-2=2, f_2=14-4=10

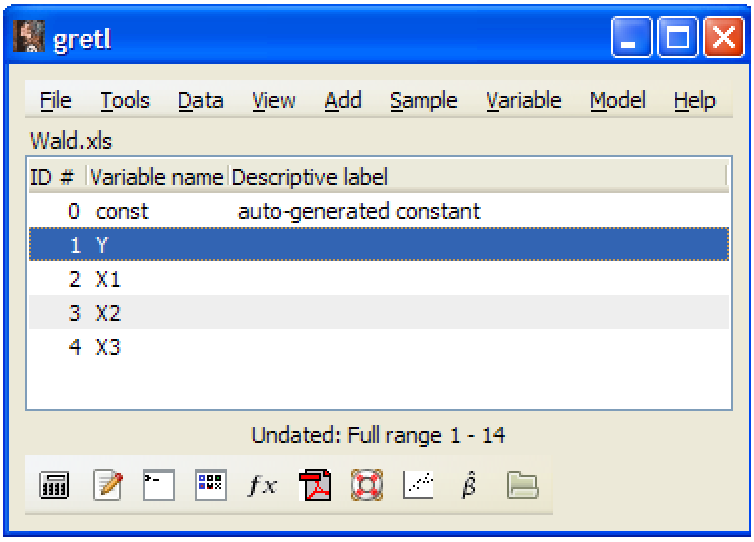
   $F_{0.05,\,2,\,10} = 4.1$

5. Adım

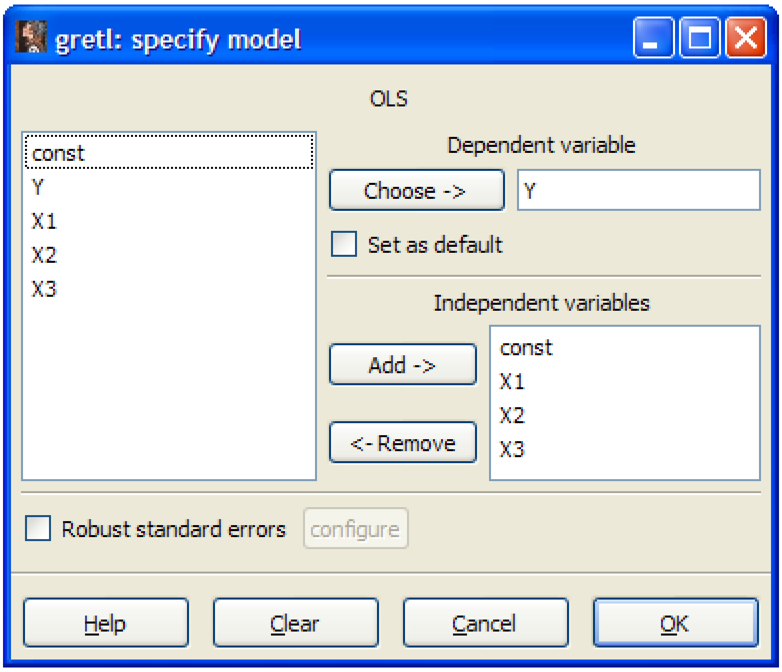
   Hesaplanan F değeri (0.56), tablo F değerinden (4.1) küçük, H_0 hipotezini reddetmeyiz.

Buna göre, X2 ve X3 değişkenlerini modelden çıkarmak yanlış olmaz sonucuna ulaşabiliriz.

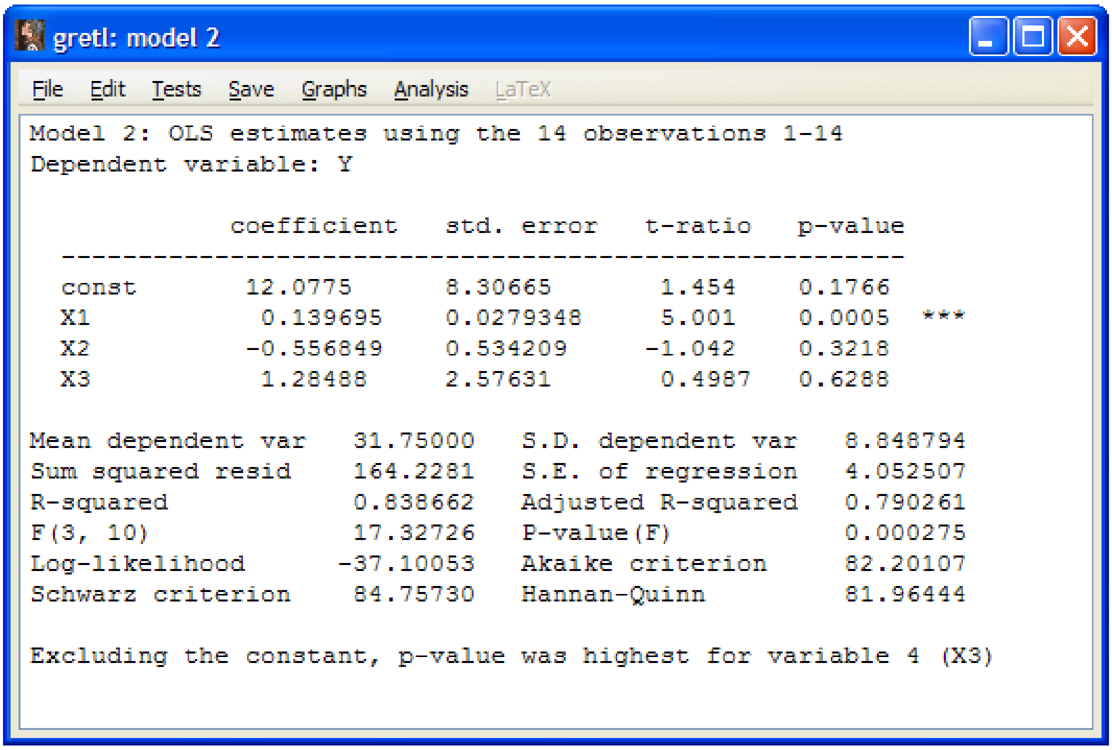
Bu süreci Gretl ile nasıl gerçekleştirebileceğimizi görelim: Değişkenlerimiz, Y, X1, X2 ve X3'tür.



Y'nin bağımlı, X1, X2 ve X3'ün bağımsız değişken olduğu EKK tanımlamamızı yapalım:

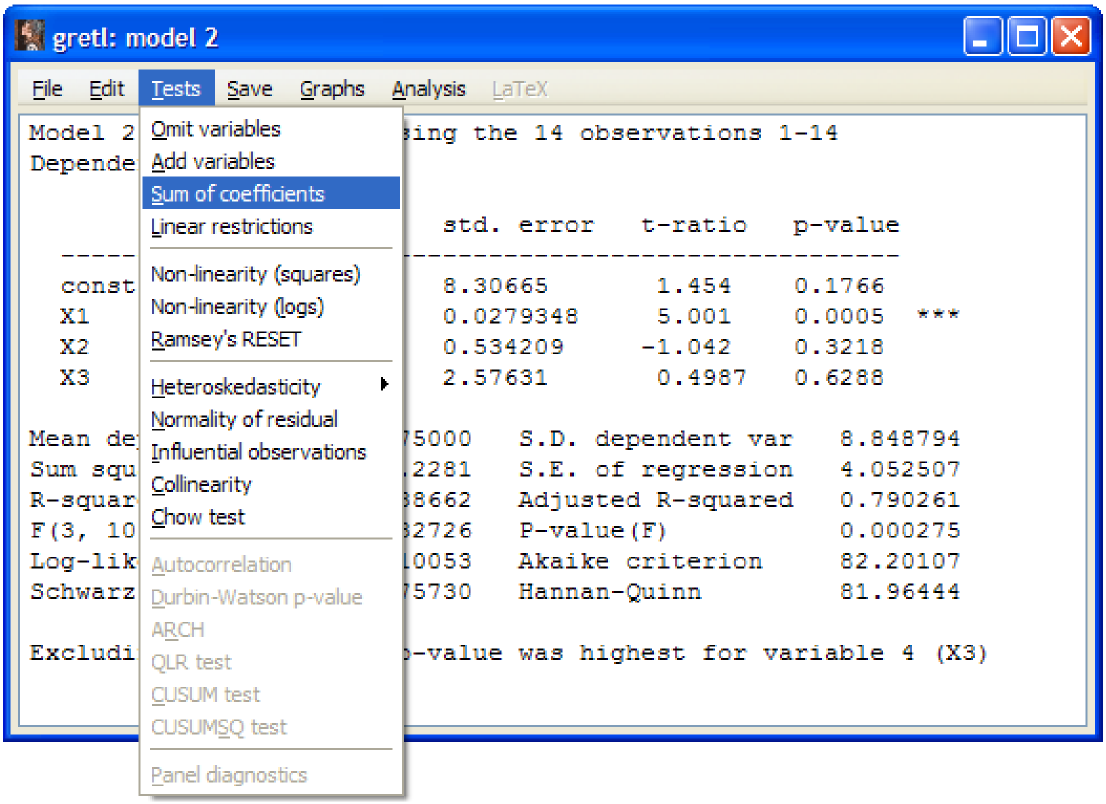


EKK tahmin sonuçları:

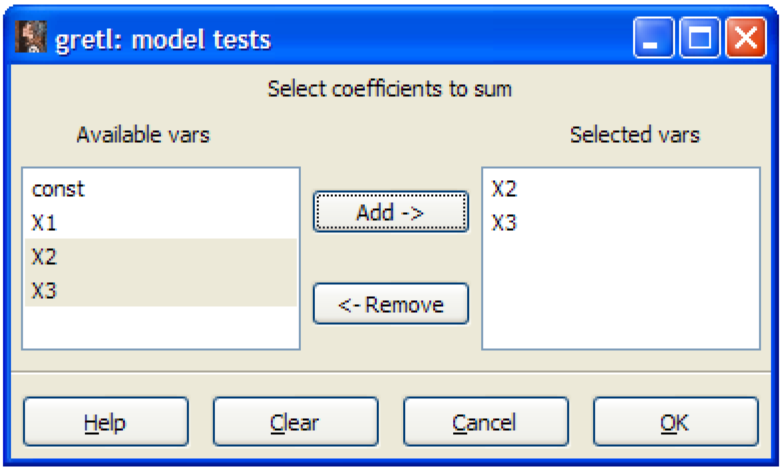



```
Model 2: OLS estimates using the 14 observations 1-14
Dependent variable: Y

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const        12.0775     8.30665       1.454    0.1766
  X1            0.139695   0.0279348     5.001    0.0005   ***
  X2           -0.556849   0.534209     -1.042    0.3218
  X3            1.28488    2.57631       0.4987   0.6288

Mean dependent var     31.75000   S.D. dependent var    8.848794
Sum squared resid      164.2281   S.E. of regression    4.052507
R-squared              0.838662   Adjusted R-squared    0.790261
F(3, 10)               17.32726   P-value(F)            0.000275
Log-likelihood        -37.10053   Akaike criterion      82.20107
Schwarz criterion      84.75730   Hannan-Quinn          81.96444

Excluding the constant, p-value was highest for variable 4 (X3)
```

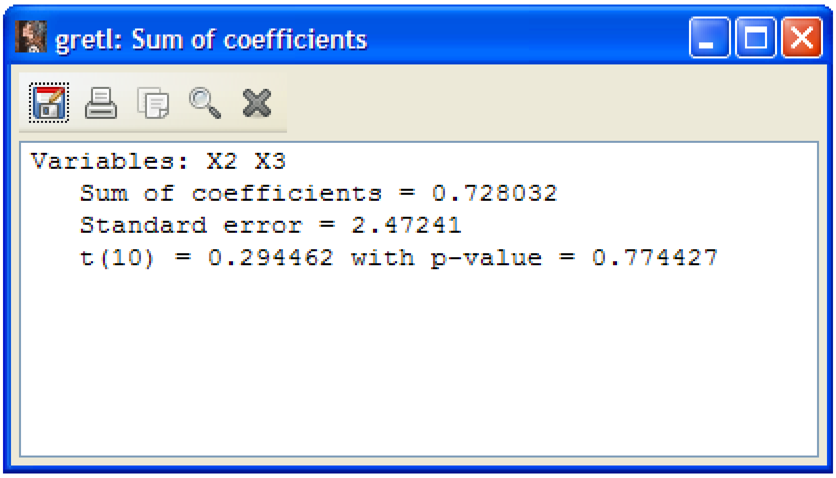
Tahmincilerin ortak testini **Tests/Sum of coefficients** menü dizisiyle yapabiliriz.



Seçtiğimiz değişkenlerin toplamını test edeceğiz. Amacımız: X2+X3=0 testini yapmaktır.



Sıfır hipotezi, X2+X3=0 şeklindedir. t testi, sıfır hipotezini kabul etmektedir.



4.3.7. *Tahmincilerin Doğrusal Kombinasyonlarının Testi*

Tüketim fonksiyonunun:

$C = b_0 + b_1W + b_2P$

olduğunu düşünelim. C, tüketim harcamaları; W, ortalama gelir; P, diğer alanlardan elde edilecek gelirdir. b_1, marjinal tüketim eğilimi; b_2, gelir dışı marjinal tüketim eğilimidir. Amacımız, b_1 ile b_2'nin birbirine eşit olma durumunu test etmek ise, yine Wald testinden yararlanabiliriz.

Şimdi Wald testini

U: $Y = b_0 + b_1X1 + b_2X2$

modeli için uygulayalım. Bu model, kısıtsız modeldir.

1. Adım

   Hipotezleri hazırla:

   H_0: $b_1=b_2$ (veya $b_1-b_2=0$)

   H_1: $b_1 \neq b_2$

2. Adım

   Kısıtsız modelde, $b_1=b_2$ hipotezinden yola çıkarak, b_2 yerine b_1 koyalım:

   $Y = b_0 + b_1X1 + b_1X2$

   $Y = b_0 + b_1(X1 + X2)$

   $Z= X1 + X2$ ise

   R: $Y = b_0 + b_1Z$

   olur. Bu model, kısıtlı modeldir (R).

3. Adım

   Kısıtlı ve kısıtsız modelleri tahmin edip, HKT_R ve HKT_U'yu hesapla.

4. Adım

   $F=(\text{HKT Farkı} / \text{Kısıt sayısı}) / (HKT_U / Sd_U)$

   $F=[(HKT_R-HKT_U)/(Sd_R-Sd_U)] / (HKT_U/Sd_U)$

   hesaplamasını yap.

5. Adım

Belli bir α düzeyi için f1=k-m için f2=n-k için F tablosu değerini bul. Burada k, kısıtsız model tahminci sayısı; m, kısıtlı model tahminci sayısıdır.

6. Adım

   Hesaplanan F değeri, tablo F değerinden büyükse H_0 hipotezini reddet.

Örnek 4-20

Yıl	Tüketim Harcamaları **C**	Reel Gelir **W**	Varlıklar **P**	Reel Gsmh **Z**
1972	2415.9	2278.37	1411.93	3690.3
1973	2532.6	2419.94	1482.36	3902.3
1974	2514.7	2411.89	1476.31	3888.2
1975	2570	2378.25	1486.85	3865.1
1976	2714.3	2509.46	1571.64	4081.1
1977	2829.8	2622.84	1656.46	4279.3
1978	2951.6	2765.50	1728.20	4493.7
1979	3020.2	2847.16	1776.84	4624
1980	3009.7	2827.18	1784.72	4611.9
1981	3046.4	2869.39	1855.51	4724.9
1982	3081.5	2859.94	1763.66	4623.6
1983	3240.6	2899.57	1910.43	4810
1984	3407.6	3087.55	2050.65	5138.2
1985	3565.5	3200.13	2129.37	5329.5
1986	3708.7	3297.95	2191.95	5489.9
1987	3822.3	3404.57	2243.83	5648.4
1988	3972.7	3527.76	2335.14	5862.9
1989	4064.6	3565.16	2495.24	6060.4
1990	4132.2	3609.04	2529.66	6138.7
1991	4105.8	3572.21	2506.79	6079
1992	4219.8	3644.90	2599.50	6244.4
1993	4339.7	3712.87	2670.93	6383.8
1994	4471.1	3813.80	2790.40	6604.2

Kısıtsız modelimiz:

$$U:\ C = b_0 + b_1W + b_2P$$

olsun ve $b_1=b_2$ durumunu test edelim.

1. Adım

   Hipotezlerimiz:

   $H_0: b_1=b_2$

   $H_1: b_1 \neq b_2$

2. Adım

   Z= W + P ise

   Kısıtsız model:

   $$C=b_0 + b_1Z$$

olur.

3. Adım

Kısıtsız model tahminimiz: $C= b_0 + b_1W + b_2P$

	Tahminci	StHata	t	P
Sabit	-200.26	99.42	-2.01	0.058
W	0.6636	0.1176	5.64	0.000
P	0.1073	0.2495	0.43	0.672
S = 34.08	R^2 = 0.998	Düz.R^2 = 0.997	F=4096.5	p=0.000

HKT_U=23223, Sd_U=20

Kısıtlı model tahminimiz: $C= b_0 + b_1Y$

	Tahminci	StHata	t	P
Sabit	-239.28	39.81	-6.01	0.000
Z	0.714061	0.007734	92.32	0.000
S = 33.41	R^2 = 0.998	Düz.R^2 = 0.997	F=8523.6	p=0.000

HKT_R=23437, Sd_R=21

4. Adım

$F=[(HKT_R-HKT_U)/(Sd_R-Sd_U)] / (HKT_U/Sd_U)$

$F=[(23437-23223)/(21-20)] / (23223/20)$

F=0.18

5. Adım

$\alpha = 0.05$, $f_1=k-m=3-2=1$, $f_2=23-3=20$

$F_{0.05,\,1,\,20} = 4.35$

6. Adım

Hesaplanan F değeri (0.18), tablo F değerinden (4.35) küçük, H_0 hipotezini reddetmeyiz.

Buna göre, marjinal tüketim eğilimi ve gelir dışı marjinal tüketim eğilimi birbirine eşittir.

Örnek 4-21

Aşağıdaki verileri kullanarak ölçeğe sabit getirili Cobb Douglas üretim fonksiyonu üzerinde çalışacağız.

Firma	**Q**	**K**	**L**
1	2350	1570	2334
2	2470	1850	2425
3	2110	1150	2230
4	2560	1940	2463
5	2650	2450	2565
6	2240	1340	2278
7	2430	1700	2380
8	2530	1860	2437
9	2550	1880	2446
10	2450	1790	2403
11	2290	1480	2301
12	2160	1240	2253
13	2400	1660	2367

14	2490	1850	2430
15	2590	2000	2470

Ölçeğe sabit getirinin varlığını, Wald testiyle belirleyelim:

Cobb Douglas üretim fonksiyonunu:

$$\ln Q = b_0 + b_1 \ln K + b_2 \ln L$$

olarak tanımlayalım. Burada, Q, çıktı; K, sermaye; L, işgücüdür. Ölçeğe sabit getiri söz konusuysa:

$$b_1 + b_2 = 1$$

olmalıdır.

Modeli tahmin edelim:

$$\ln Q = 0.50 + 0.188 \ln K + 0.758 \ln L$$

t (0.11) (1.36) (1.07)

$S = 0.01285$ $R^2 = 0.969$ Düz.$R^2 = 0.964$ $F=186.82$

$HKT_U = 0.001981$ $Sd_U=12$

Denklemin tahmincileri düşük t değerlerine sahiptir. Ancak F ve R^2 oldukça yüksektir. Bu, çoklu doğrusal bağlantının belirtisidir. Gerçekten de lnQ, lnK ve lnL arasında 1'e çok yakın korelasyon vardır. Biz denklemi bu haliyle kullanmayacağımız için, işlemlerimizi sürdürelim.

Ölçeğe sabit getiride, b1+b2=1 olması gerekir:

$$b_1+b_2=1$$

$$b_2 = (1 - b_1)$$

Bunu denklemde yerine koyalım:

$$\ln Q = b_0 + b_1 \ln K + (1 - b_1) \ln L$$

$$\ln Q = b_0 + b_1 \ln K + \ln L - b_1 \ln L$$

$$\ln Q - \ln L = b_0 + b_1 \ln K - b_1 \ln L$$

$$\ln Q - \ln L = b_0 + b_1(\ln K - \ln L)$$

$\ln Q^* = \ln Q - \ln L$ ve $\ln K^* = \ln K - \ln L$ ise

$$\ln Q^* = b_0 + b_1 \ln K^*$$

Denklemimizi tahmin edebilmek için:

(ln Q–ln L) ve (ln K–ln L) değişkenlerini hesaplamamız gerekiyor:

Firma	lnQ-lnL	lnK-lnL
1	0.006832	-0.39651
2	0.018387	-0.27065
3	-0.05531	-0.66224
4	0.038627	-0.23869
5	0.032601	-0.04587
6	-0.01682	-0.53063
7	0.020791	-0.33647
8	0.037452	-0.27019
9	0.041639	-0.26318
10	0.01937	-0.2945
11	-0.00479	-0.4413
12	-0.04215	-0.59715
13	0.013845	-0.35481
14	0.024392	-0.27271

15	0.04744	-0.21107

Şimdi:

$\ln Q^* = b_0 + b_1 \ln K^*$

denklemini tahmin edelim:

$\ln Q^* = 0.0726 + 0.175 \ln K^*$

t (9.44) (8.26)

$S = 0.01235$ $R^2 = 0.846$ Düz. $R^2 = = 0.834$

$HKT_R = 0.001983$ $Sd_R = 13$

Şimdi bu işlemin istatistiki olarak geçerliliğini belirleyelim.

Hipotezlerimiz:

$H_0: b_1 + b_2 = 1$

$H_1: b_1 + b_2 \neq 1$

$F = [(HKT_R - HKT_U)/(Sd_R - Sd_U)] / (HKT_U/Sd_U)$

$F = [(0.001983 - 0.001981)/(13-12)] / (0.001981/12)$

$F = 0.012$

$\alpha = 0.05$, $f_1 = 2-1 = 1$, $f_2 = 12$

$F_{0.05,1,12} = 4.75$

Hesaplanan F değeri (0.012), tablo F değerinden (4.75) küçük olduğundan, H_0 hipotezini reddetmeyiz.

Buna göre, üzerinde çalıştığımız Cobb Douglas üretim fonksiyonu, ölçeğe sabit getiri özelliği taşımaktadır.

Modelimizi, Q=f(K,L) formuna dönüştürebiliriz:

$\ln Q^* = 0.0726 + 0.175 \ln K^*$

$\ln Q - \ln L = 0.0726 + 0.175 (\ln K - \ln L)$

$\ln Q = 0.0726 + \ln L + 0.175 \ln K - 0.175 \ln L$

$\ln Q = 0.0726 + +0.175 \ln K + 0.825 \ln L$

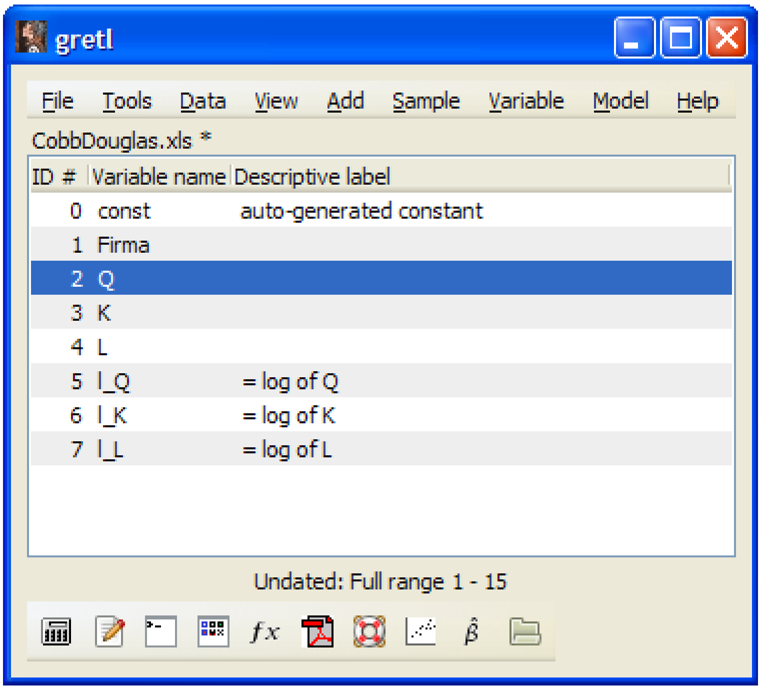
Son durum dikkate alındığında, üretimin kısmi sermaye esnekliği 0.175, kısmi işgücü esnekliği ise 0.825'tir. Kısmi esnekliklerin toplamı tam 1 olduğunudan, ölçeğe sabit getiri sağlanmaktadır.

Fonksiyonumuzu üssel olaral yeniden yazalım:

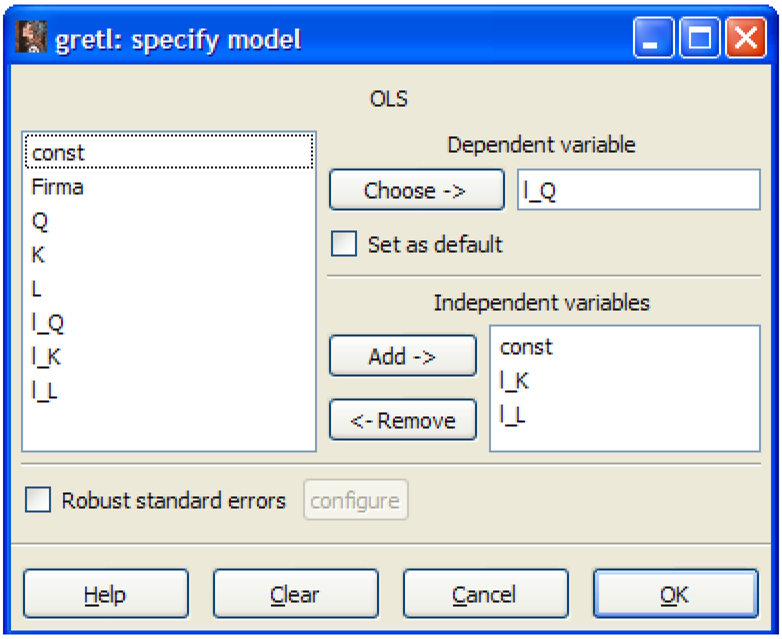
$Q = 1.075\, K^{0.175}\ L^{0.825}$

Denklemdeki b_0, etkinlik parametresidir. b_0, istatistki açıdan anlamlı olduğuna göre, teknoloji düzeyinin bir göstergesi olarak güvenle kullanılabilir.

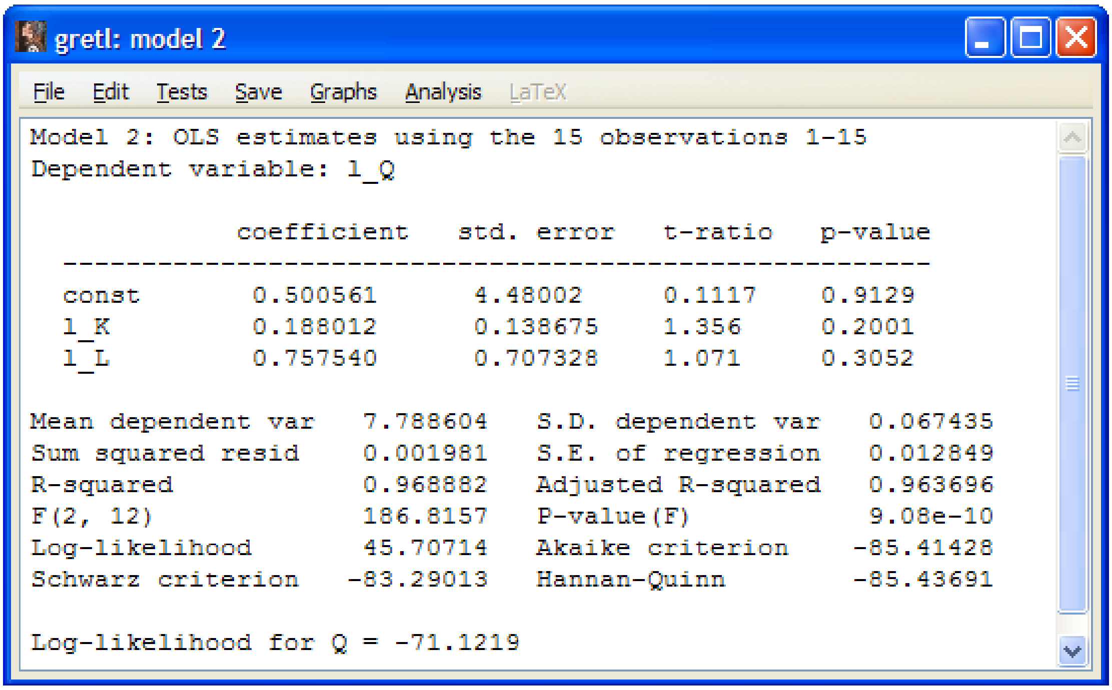
Bu problemin çözümünü Gretl yardımıyla yapabiliriz. Değişken listemizde logaritmik dönüşümler hazırlanmış durumdadır.



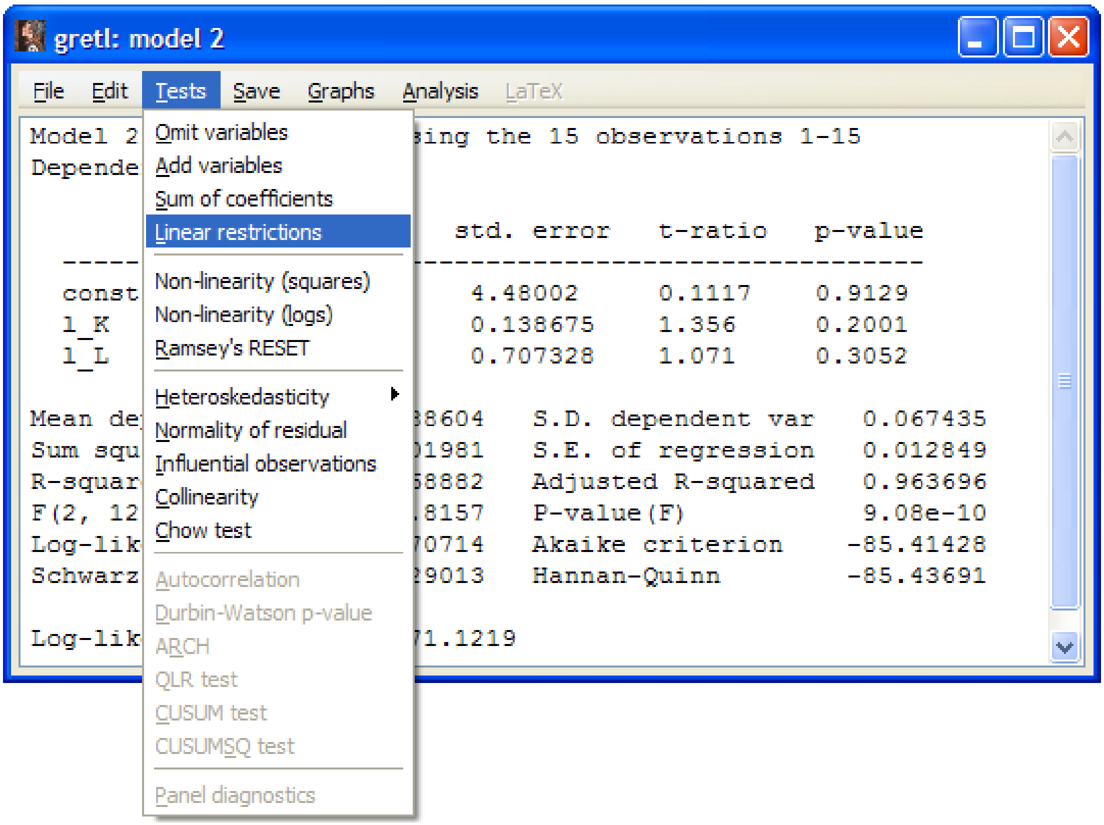
EKK tanımlamamız:



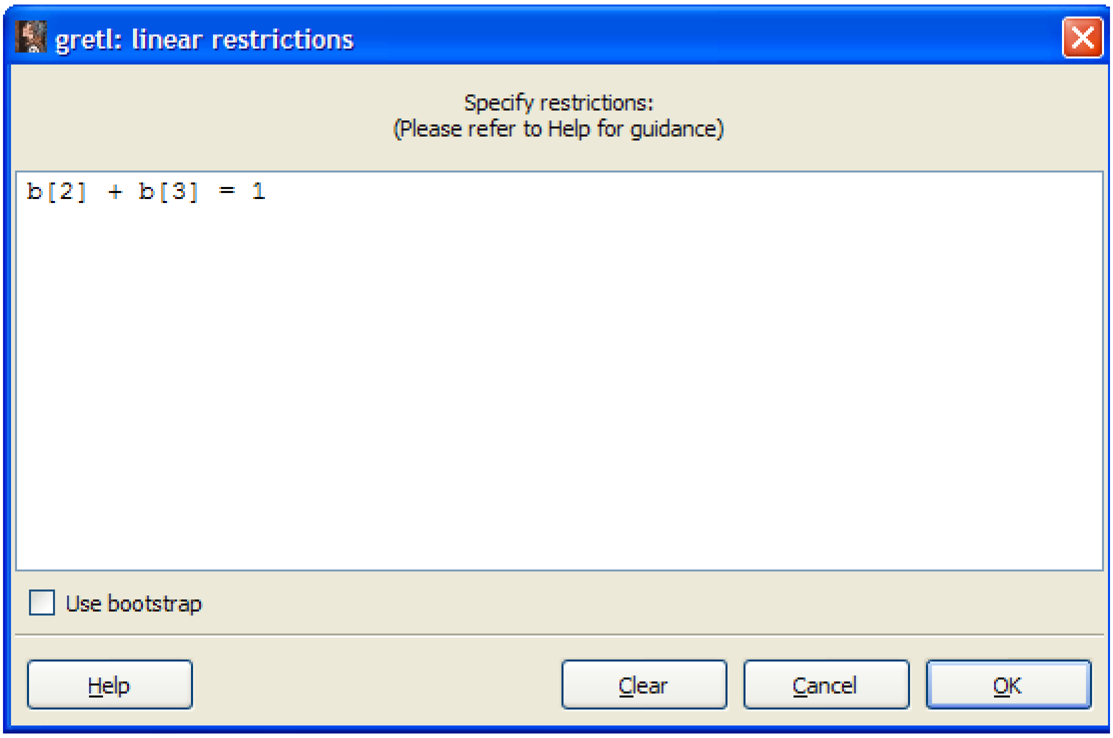
EKK tahmin sonuçlarımız:



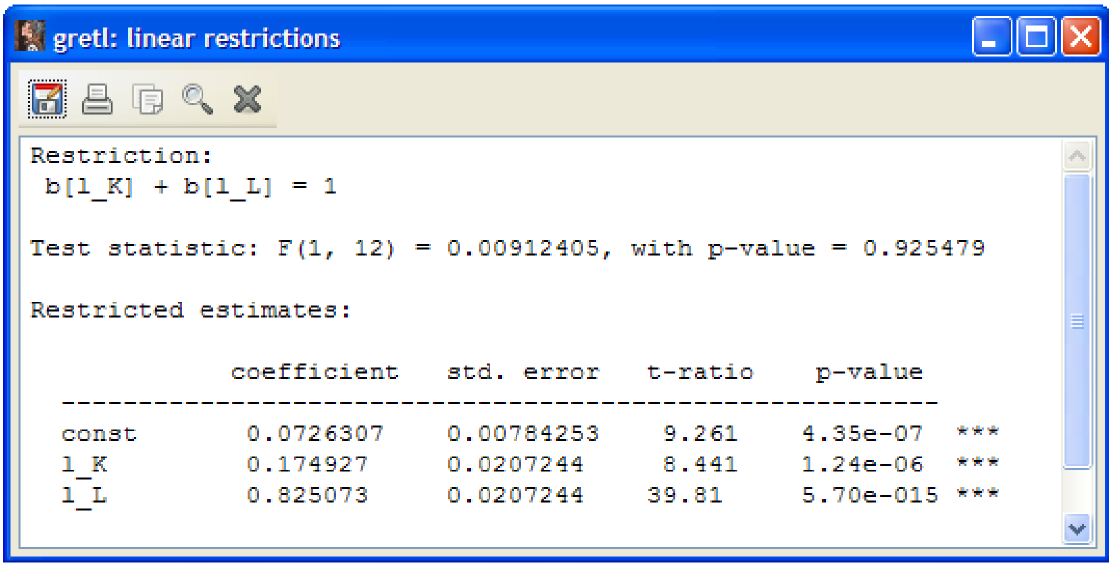
Katsayıların toplamının 1 olmasını test etmek için, *Tests/Linear restrictions* menü dizisi izlenir.



lnK ve lnL değişkenlerinin kasayıları, b[2] ve b[3] sırasıyla modelde yer alır. Bu nedenle kısıt tanımlamasını: b[2] + b[3] = 1 şeklinde ifade ederiz. Bu, aynı zamanda sıfır hipotezimizi oluşturur.



F testine göre, sıfır hipotezini kabul edilir. Bu, tanımlanan kısıta uygun bir model tahmini yapmanın mümkün olduğu anlamına gelmektedir. Tahmine ait sonuçlar da sunulmuştur.



Örnek 4-22

Bu örnekte, sabit ikame esnekliğini doğrudan elde edebileceğimiz CES fonksiyonunu tahmin edeceğiz. CES fonksiyonu:

$$\ln Y = \ln \theta - v/\rho \ln [\delta K^{-\rho} + (1-\delta) L^{-\rho}]$$

şeklinde ifade edilebilir.

Burada θ, etkinlik katsayısı (teknoloji düzeyi); ρ, ikame parametresidir.

ρ=0 için bu fonksiyona Taylor yaklaşımı uygulandığında:

$$\ln Y = \ln \theta + v\delta \ln K + v(1-\delta) \ln L + \rho v \delta (1-\delta)[-0.5 (\ln K - \ln L)^2]$$

denklemini elde ederiz. Basitleştirecek olursak:

$\ln Y = b_0 + b_1 \ln K + b_2 \ln L + b_3 [-0.5 \ln^2 (K/L)]$

elde ederiz. Denklem tahmin edildiğinde:

$b_0 =$	$\ln \theta$	$\theta = e^{bo}$
$b_1 =$	$v\delta$	$\delta = b_1/(b_1+b_2)$
$b_2 =$	$v(1-\delta)$	$v = b_1 + b_2$
$b_3 =$	$\rho v\delta (1-\delta)$	$\rho = b_3(b_1+b_2)/(b_1b_2)$

Denklemdeki değişkenler:

X1: lnK; X2:lnL; X3: $-0.5 \ln^2 (K/L)$

Aşağıdaki verileri kullanarak, CES fonksiyonunu elde edelim.

Değişkenlerimizi hazırlayalım: X1: lnK; X2:lnL; X3: $-0.5 \ln^2 (K/L)$

Firma	Q	K	L	lnY (lnQ)	X1 (lnK)	X2 (lnL)	X3 $[-0.5 \ln^2 (K/L)]$
1	2350	1570	2334	7.76217	7.35883	7.75534	-0.078609
2	2470	1850	2425	7.81197	7.52294	7.79359	-0.036625
3	2110	1150	2230	7.65444	7.04752	7.70976	-0.219281
4	2560	1940	2463	7.84776	7.57044	7.80914	-0.028487
5	2650	2450	2565	7.88231	7.80384	7.84971	-0.001052
6	2240	1340	2278	7.71423	7.20042	7.73105	-0.140783
7	2430	1700	2380	7.79565	7.43838	7.77486	-0.056607
8	2530	1860	2437	7.83597	7.52833	7.79852	-0.036502
9	2550	1880	2446	7.84385	7.53903	7.80221	-0.034632
10	2450	1790	2403	7.80384	7.48997	7.78447	-0.043366
11	2290	1480	2301	7.73631	7.29980	7.74110	-0.097374
12	2160	1240	2253	7.67786	7.12287	7.72002	-0.178295
13	2400	1660	2367	7.78322	7.41457	7.76938	-0.062944
14	2490	1850	2430	7.82004	7.52294	7.79565	-0.037184
15	2590	2000	2470	7.85941	7.60090	7.81197	-0.022275

CES fonksiyonunun EKK tahmin sonuçları:

$\ln Y = b_0 + b_1 \ln K + b_2 \ln L + b_3 [-0.5 \ln^2 (K/L)]$

	Katsayı	St.Hata	t	p
Sabit	-35938	6123	-5.87	0.000
lnK	-792.3	261.6	-3.03	0.011
lnL	5704	1025	5.57	0.000
$-0.5 \ln^2 (K/L)$	1579.1	294.7	5.36	0.000
S = 15.26	$R^2 = 0.993$	Düz.$R^2 = 0.991$	F=507.22	p=0.000

Tahmin sonuçlarını denklem formunda yazalım:

$\ln Y = b_0 + b_1 \ln K + b_2 \ln L + b_3 [-0.5 \ln^2 (K/L)]$

$\ln Y = -35938 - 792.3 \ln K + 5704 \ln L + 1579.1 [-0.5 \ln^2 (K/L)]$

$b_0=$ ln θ =-35938

$b_1=$ vδ = -792.3

$b_2=$ v(1-δ) = 5704

$b_3=$ ρvδ (1-δ) = 1579.1

$\theta = e^{bo}$ = 0

$\delta = b_1/(b_1+b_2)$ = -0.16131

$v = b_1 + b_2$ = 1286.6

$\rho = b_3(b_1+b_2)/(b_1b_2)$= -1.716

İkame esnekliği, -1.716'dır.

4.3.8. Doğrusal ve Logaritmik Modeller Arasından Seçim

4.3.8.1. MWD Testi

Aynı değişkenlerle kurulmuş bir ilişki, hem doğrusal hem de logaritmik formda tahmin edilebilir. Ancak bunlardan sadece birini kullanmamız gerekir. Hangi formu seçeceğimize karar verirken, R^2 en basit karşılaştırma kriteri olabilir; ancak denklemlerin sol taraflarının aynı olma zorunluluğu nedeniyle, logaritmik denklem, log-lin veya log-log ise, doğrusal denklemle karşılaştıramayız. Doğrusal ve logaritmik formdaki denklemler arasında seçim yaparken, McKinnon White Davidson (MWD) testinden yararlanılabiliriz.

MWD testi, aşağıdaki aşamalarla gerçekleştirilir.

1. **Adım: Hipotezleri hazırla:**

   H_0: Doğrusal form: $Y = a_0 + a_1X + e_0$

   H_1: Logaritmik form: $\ln Y = b_0 + a_1 \ln X + e_1$

2. **Adım: Doğrusal formu tahmin et; $\hat{Y}$'leri hesapla ve D'ye eşitle.**

   $D=\hat{Y}$

3. **Adım: Logaritmik formu tahmin et; ln $\hat{Y}$'leri hesapla ve F'ye eşitle.**

   $F=\ln\hat{Y}$

4. **Adım: Z_1'i hesapla**

   $Z_1 = \ln D - F$

5. **Adım:**

   $Y = a_0 + a_1X + Q_0Z_1$

   denklemini tahmin et. Denklemdeki Q_0, Z_1'in tahmincisidir. %1, %5 veya %10 düzeyinde Q_0'ı test edip, sıfırdan farklıysa, logaritmik formu seç.

6. **Adım: Z_2'yi hesapla**

   $Z_2 = \text{antilog}_e(F) - D$

7. Adım:

$$\ln Y = b_0 + b_1 \ln X + Q_1 Z_2$$

denklemini tahmin et. Denklemdeki Q_1, Z_2'nin tahmincisidir. %1, %5 veya %10 düzeyinde Q_1'i test edip, sıfırdan farklıysa, doğrusal formu seç.

Örnek 4-23

Aşağıdaki verileri kullanarak, MWD testini uygulayalım.

Talep Y	Fiyat X
150	2
140	7
110	8
80	10
50	11
40	13
35	18

Önce doğrusal ve logaritmik denklemleri elde edelim:

1. Adım

Hipotezlerimiz:

H_0: Doğrusal form: $Y = a_0 + a_1 X + e_0$

H_1: Logaritmik form: $\ln Y = b_0 + a_1 \ln X + e_1$

2. Adım

Doğrusal tahmin sonuçları:

Tahminci	Katsayı	StHata	t	p
b_0	171.61	19.32	8.88	0.000
b_1	-8.641	1.773	-4.87	0.005
S = 21.78		R^2 =0.826		

$\hat{Y}$'leri hesaplayip, D'ye eşitlediğimizde, D:

154.323	111.117	102.476	85.194	76.553	59.271	16.065

olur.

3. Adım

log-log tahmin sonuçları:

Tahminci	Katsayı	StHata	t	p
b_0	5.8019	0.4837	12	0.000
b_1	-0.7007	0.2176	-3.22	0.023
S = 0.3750		R^2 =0.675		

$\ln \hat{Y}$'leri hesaplayıp, F'ye eşitlediğimizde, F:

5.31619	4.43842	4.34486	4.18851	4.12173	4.00468	3.77666

olur.

4. Adım

$Z_1 = \ln D - F$

hesaplamasını yaptığımızda, Z_1:

-0.27714	0.272164	0.284769	0.256421	0.216253	0.07744	-1.00002

olur.

5. Adım

$Y = a_0 + a_1X + Q_0Z_1$

denklemini tahmin edelim:

	Katsayı	StHata	t	p
Sabit	178.1	22.46	7.93	0.001
X	-9.339	2.126	-4.39	0.012
Z_1	-15.6	22.51	-0.69	0.526
S = 23.01		R^2 =	0.845	

Görüldüğü gibi, Q_0 tahmincisi, istatistiki açıdan sıfırdan farklı değildir. Eğer sıfırdan farklı olsaydı, log-log formu tercih edecektik.

6. Adım

$Z_2 = \text{antilog}_e(F) - D$

hesaplamasını yaptığımızda, Z_2:

49.28366	-26.4759	-25.3948	-19.2695	-14.8872	-4.41673	27.60494

olur.

7. Adım

$\ln Y = b_0 + b_1 \ln X + Q_1Z_2$

denklemini tahmin edelim.

	Katsayı	StHata	t	p
Sabit	6.1697	0.3623	17.03	0
X	-0.8831	0.1657	-5.33	0.006
Z_2	-0.01023	0.00399	-2.56	0.062
S = 0.2579		R^2 =0.877		

Q_1 istatistik açıdan %10 düzeyinde anlamlıdır. Buna göre, doğrusal formu seçebiliriz.

4.3.8.2. P_E Testi

P_E testi de MWD testi gibi, aynı değişkenlerle kurulmuş bir ilişkinin doğrusal ve logaritmik formdaki tahminleri arasından seçim yaparken kullanabileceğimiz bir testtir.

Şimdi P_E testi sürecinin nasıl işlediğini görelim:

1. **Aşama**

   Hipotezleri hazırla:

   H_0: Doğrusal form: $Y = a_0 + a_1X + e_0$

H_1: Logaritmik form: $\ln Y = b_0 + a_1 \ln X + e_1$

Logaritmik denklemi tahmin et:

e_1'i hesaplayıp, doğrusal denkleme değişken olarak dahil et ve doğrusal formu yeniden tahmin et:

$Y = a_0 + a_1X + Q_0e_1$

Denklemdeki Q_0, e_1'in tahmincisidir.

Q_0'ın sıfırdan farklı olup olmadığını %1, %5 veya %10 düzeyinde test et.

2. **Aşama**

Hipotezleri hazırla:

H_0: Logaritmik form: $\ln Y = b_0 + b_1 \ln X + e_1$

H_1: Doğrusal form: $Y = a_0 + a_1X + e_0$

Doğrusal denklemi tahmin et:

e_0'ı hesapla

e_0'ı logaritmik denkleme değişken olarak dahil edip, logaritmik formu yeniden tahmin et:

$\ln Y = a_0 + a_1X + Q_1e_0$

Denklemdeki Q_1, e_0'ın tahmincisidir.

Q_1'in sıfırdan farklı olup olmadığını %1, %5 veya %10 düzeyinde test et.

3. **Aşama: Karar verme**

I. Aşamada Q_0=0 ve II. aşamada $Q_1 \neq 0$ ise doğrusal formu seç

I. Aşamada $Q_0 \neq 0$ ve II. aşamada Q_1=0 ise logaritmik formu seç

I. Aşamada $Q_0 \neq 0$ ve II. aşamada $Q_1 \neq 0$ ise hiçbirini seçme

I. Aşamada Q_0=0 ve II. aşamada Q_1=0 ise, test seçim için yeterli bilgi vermiyor.

Örnek 4-24

P_E testi için de, Örnek 4-23'deki verileri kullanacağız.

1. **Aşama**

Hipotezlerimiz:

H_0: Doğrusal form: $Y = a_0 + a_1X + e_0$

H_1: Logaritmik form: $\ln Y = b_0 + b_1 \ln X + e_1$

Logaritmik denklemin tahmin sonuçları:

	Katsayı	StHata	t	P
Sabit	5.8019	0.4837	12	0.000
lnX	-0.7007	0.2176	-3.22	0.023
S = 0.3750		R^2 =0.675		

e_1'i hesaplayalım:

-0.30555	0.503225	0.355624	0.19352	-0.2097	-0.3158	-0.22132

e_1'i doğrusal denkleme değişken olarak dahil edip, doğrusal formu yeniden tahmin edelim:

$Y = a_0 + a_1X + Q_0e_1$

	Katsayı	StHata	t	p
Sabit	164.86	17.2	9.59	0.001
X	-7.957	1.588	-5.01	0.007
e_1	38.17	23.27	1.64	0.176
S = 0.1883		R^2 =0.896		

Q_0, istatistiki açıdan sıfırdan farklı değildir.

2. Aşama

Hipotezlerimiz:

H_0: Logaritmik form: $\ln Y = b_0 + b_1 \ln X + e_1$

H_1: Doğrusal form: $Y = a_0 + a_1X + e_0$

Doğrusal denklemi tahmin edelim:

$Y = a_0 + a_1X + e_0$

	Katsayı	StHata	t	p
Sabit	171.61	19.32	8.88	0.000
X	-8.641	1.773	-4.87	0.005
S = 21.78		R^2 =0.826		

e_0'ı hesaplayalım:

-4.3229	28.8826	7.5237	-5.1941	-26.553	-19.2708	18.9347

e_0'ı logaritmik denkleme değişken olarak dahil edip, logaritmik formu yeniden tahmin edelim:

$\ln Y = b_0 + b_1 \ln X + Q_1e_0$

	Katsayı	StHata	t	p
Sabit	5.7907	0.4277	13.54	0.000
lnX	-0.6954	0.1924	-3.61	0.022
e_0	0.010541	0.00681	1.55	0.197
S = 0.3316		R^2 =0.796		

Q_1, istatistiki açıdan, sıfırdan farklı değildir.

3. Aşama: Karar verme

I. Aşamada Q_0=0 ve II. aşamada Q_1=0 olduğundan, test seçim için yeterli bilgi vermemiştir. MWD testi, doğrusal formun seçilmesini önermişti. Bu durumda, hangi testin sonuçlarının dikkate alınacağı, araştırıcıya bırakılır.

4.3.8.3. J Testi

Ekonometrik modellemede, aynı bağımlı değişken, farklı bağımsız değişkenlerle ilişkilendirilebilmektedir. Örneğin bir araştırıcı, belli bir malın talebini sadece o malın fiyatıyla veya sadece gelir düzeyiyle ilişkilendirmek isteyebilir. Bunun için iki farklı model kurup, hangisinin daha uygun olduğu sorusuna cevap arayabilir. J testi böyle bir seçim için kullanılabilir.

J testinin adım adım uygulanışı:

I. Aşama:

1. Adım

Modelleri tanımla:

Model C: $Y_C = a_0 + a_1X + e_0$

Model D: $Y_D = b_0 + b_1Z + e_1$

2. Adım

Model D'yi tahmin et ve $\hat{Y}_D$'leri hesapla.

3. Adım

$\hat{Y}_D$'yi, model C'ye ekle:

$$Y_C = a_0 + a_1X + Q_0\hat{Y}_D$$

ve bu denklemi tahmin et. $\hat{Y}_D$ burada model C'ye alınmayan değişken(ler)i temsil etmektedir.

4. Adım

Q_0'ın %1, %5 veya %10 düzeyinde sıfırdan farklılığını test et.

5. Adım

Eğer Q_0 sıfırdan farklı değilse, model C'ye alınmayan değişkenler istatistiki açıdan anlamsız olacağından, model C tercih edilir. Buna göre, model C, model D'yi de kapsamaktadır.

II. Aşama:

1. Adım

Model C'yi tahmin et ve $\hat{Y}_C$'leri hesapla.

2. Adım

$\hat{Y}_C$'yi, model D'ye ekle:

$$Y_D = b_0 + b_1Z + Q_1\hat{Y}_C$$

ve bu denklemi tahmin et. $Y\hat{}_C$ burada model D'ye alınmayan değişken(ler)i temsil etmektedir.

3. Adım

Q_1'in %1, %5 veya %10 düzeyinde sıfırdan farklılığını test et.

4. Adım

Eğer Q_1 sıfırdan farklı değilse, model D'ye alınmayan değişkenler istatistiki açıdan anlamsız olacağından, model D tercih edilir. Buna göre, model D, model C'yi de kapsamaktadır.

5. Adım

Q_1'in %1, %5 veya %10 düzeyinde sıfırdan farklılığını test et.

III. Aşama:

Çizelge 4-4: J Testi Karar Tablosu

$Q_1 = 0$	$Q_0 = 0$	
	Doğru	**Yanlış**
Doğru	C ve D modellerinin ikisi de uygun	Model D'yi tercih et
Yanlış	Model C'yi tercih et	C ve D modellerinin hiçbiri de uygun

Örnek 4-25

Aşağıdaki veriler, bir malın talebi, fiyatı ve gelir düeyini içermektedir.

Talep Y	Fiyat X	Gelir Z
150	2	350
140	7	350
110	8	450
80	10	450
50	11	750
40	13	800
35	18	1000

Şimdi bu verileri kullanarak, J testini uygulayalım.

I. Aşama:

Önce modelleri tanımlayalım. Talep ile fiyat arasındaki ilişkiyi model C; talep ile gelir arasındaki ilişkiyi ise model D temsil etsin.

Model C: $Y_C = a_0 + a_1X + e_0$

Model D: $Y_D = b_0 + b_1Z + e_1$

1. Adım

Model D'yi tahmin edelim: $Y_D = b_0 + b_1Z + e_1$

	Katsayı	StHata	t	p
Sabit	187.99	21.03	8.94	0.000
Z	-0.17131	0.03294	-5.2	0.003
S = 20.63		R^2 =0.844		

$\hat{Y}_D$ 'leri hesaplayalım.

128.033	128.033	110.902	110.902	59.508	50.943	16.68

2. Adım

$\hat{Y}_D$'yi, model C'ye ekleyip tahmin edelim:

$Y_C = a_0 + a_1X + Q_0\hat{Y}_D$

	Katsayı	StHata	t	p
Sabit	76.12	78.2	0.97	0.385
X	-4.038	4.032	-1	0.373
$\hat{Y}_D$	0.5797	0.4616	1.26	0.278

3. Adım

Görüldüğü gibi, Q_0, sıfırdan farklı değildir.

II. Aşama:

1.Adım

Model C'yi tahmin edelim: $Y_C = a_0 + a_1X + e_0$

	Katsayı	StHata	t	p
Sabit	171.61	19.32	8.88	0.000
X	-8.641	1.773	-4.87	0.005
S = 21.78		R^2 =	0.826	

$\hat{Y}_C$'leri hesaplayalım:

154.323	111.117	102.476	85.194	76.553	59.271	16.065

2.Adım

$\hat{Y}_C$'yi, model D'ye ekleyip, tahmin edelim:

$Y_D = b_0 + b_1Z + Q_1\hat{Y}_C$

Tahminci	Katsayı	StHata	t	p
Sabit	104.92	85.57	1.23	0.287
Z	-0.0993	0.07908	-1.26	0.278
$\hat{Y}_C$	0.4673	0.4666	1	0.373

3.Adım

Q_1, istatistiki açıdan sıfırdan farklı değildir.

III. Aşama:

Karar verirken, J testi karar tablosunu dikkate almamız gerekiyor. Tabloya göre, Q_1 ve Q_0 her ikisi birden sıfır olduğunda, iki model de kullanıma uygundur. Bir başka ifadeyle, araştırıcı her iki denklemi de kullanabilir; tercih kendisine aittir.

5. KUKLA DEĞİŞKENLER

Ekonometrik analizlerde modellere, nicel değişkenler yanında, nitel değişkenler de dahil edilmesi ihtiyacı, kukla değişkenleri gündeme getirmiştir. Sayısal olarak ölçülebilir değişkenler için matematiksel yöntemlerin uygulanmasında bir güçlük yoktur. Ancak sayısal ölçekte yeri olmayan değişkenlerin önce sayısallaştırılması gerekir. Bunun için, kukla değişkenlerden yararlanılır.

Kukla değişken, sadece iki değer alır. Zorunlu olmamakla birlikte, hesaplama ve yorumlama kolaylığı açısından bu iki değerin, 0 ve 1 'dir. Birbirinden farklı başka ikili değer çiftlerinin kullanımı da mümkün olmakla birlikte, yorumlama kolaylığı açısından, yaygın olarak, 0 ve 1 kullanılmaktadır. Bu nedenle 0-1 değişkeni veya ikili değişken adları da verilmektedir.

Cinsiyet, bölge, olumlu/olumsuz, evet/hayır, önce/sonra, var/yok gibi nitel değişkenleri, kukla değişkenler olarak tanımlayabiliriz. Genellikle olumlu durumları temsil eden değer 1, olumsuz durumları temsil eden değer ise 0'dır. Örneğin:

- Evet/hayır değişkeninde, eveti 1, hayırı 0 ile;
- politika uygulama yıllarını 1, uygulama dışı yılları 0 ile;
- üretim dönemini 1, üretim dışı dönemi 0 ile;
- başarılı olanları 1, başarısızları 0 ile;

sayılaştırılabiliriz.

Şimdi beş çiftçinin son 3 yılda kredi alıp almadığını temsil eden bir kukla değişken tanımlayalım. 4 ve 5 no'lu çiftçiler son 3 yılda kredi almıştır. Buna göre kredi alanları 1, almayanları yılları 0 ile ifade edelim:

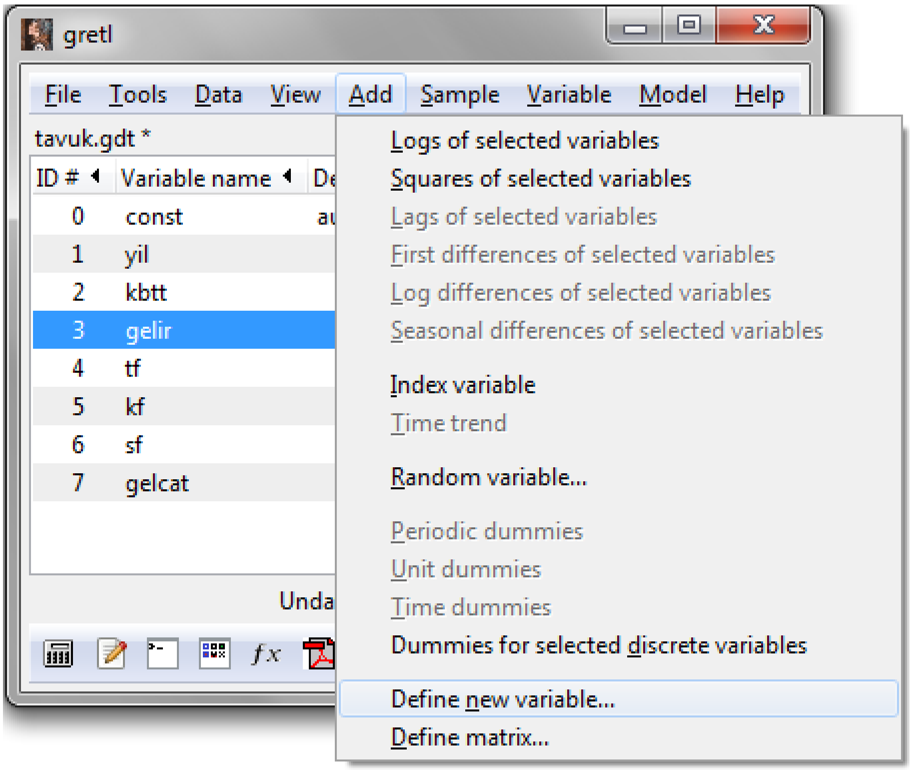
Çiftçi No	Durum	Kukla
1	Kredi Almadı	0
2	Kredi Almadı	0
3	Kredi Almadı	0
4	Kredi Aldı	1
5	Kredi Aldı	1

Tek bir kukla değişken, iki durumu içerebilir. Eğer ikiden fazla durum söz konusuysa; durum sayısının bir eksiği kadar kukla değişkene ihtiyaç duyulur. Örneğin A, B ve C bölgelerindeki üretimi temsil eden 2 kukla değişken olmalıdır. İlk kukla A bölgesini, ikinci kukla B bölgesini temsil eder. Üçüncü kuklaya ihtiyaç yoktur:

Kukla A	Kukla B	Açıklama
1	0	Veri, A'ya aittir
0	1	Veri, B'ye aittir
0	0	Veri, C'ye aittir

A bölgesine ait bir gözlem için kukla A'ya 1 dersek, kukla B doğal olarak 0 olur. B bölgesine ait bir gözlem için ise kukla B'ye 1 dersek, kukla A kuşkusuz 0 olur. C bölgesine ait bir gözlem için, ne A ne de B bölgesine ait olduğundan, hem kukla A hem de kukla B sıfır olur.

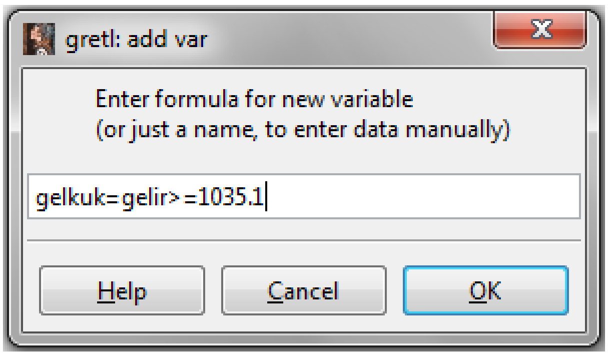
Gretl programı kukla değişken tanımlama açısından önemli kolaylıklar sağlamaktadır. Örneğin bir talep fonksiyonunda gelirin etkisini, düşük ve yüksek gelir grubu açısından analiz etmeyi planladığımızı varsayalım. Bunun için daha once kullandığımız Örnek 4-1'deki veri setinden yararlanacağız. Amacımız ortalama gelirin altına 0, üstüne ise 1 değerini atayacağımız bir kukla değişken tanımlamak olsun. Ortalama gelir, 1035.1'dir. Yeni kukla değişkenimizin adını gelkuk olarak belirleyelim. Buna göre gelir 1035.1'den küçük ise gelkuk değişken 0 değerini, 1035.1'e eşit veya bundan daha büyükse 1 değerini almalıdır. Gretl'daki işlem süreci, **Add/Define new variable…** ile başlar:



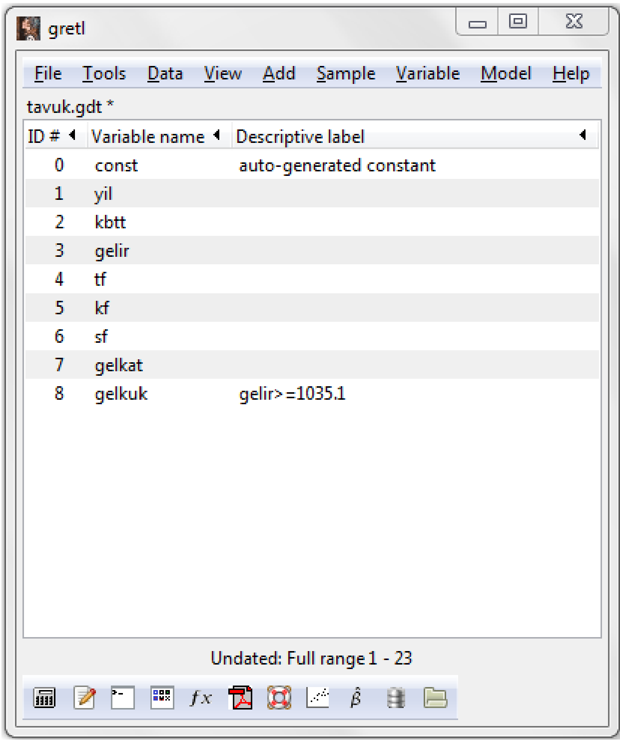
Yeni kukla değişkenimizi:

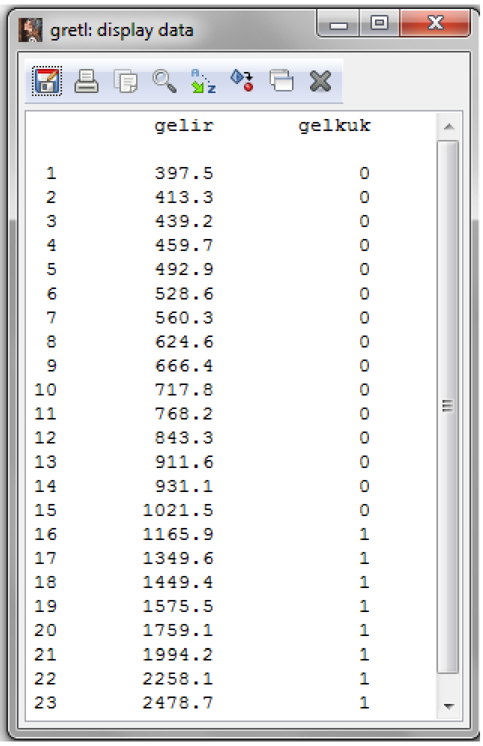
gelkuk = gelir>=1035.1

şeklinde tanımlamamız gerekir:



Bu tanımlamadaki ***gelir>=1035.1*** mantıksal bir ifadedir; o nedenle, ifade doğru ise 1, yanlış ise 1 sonucu ortaya çıkacaktır:





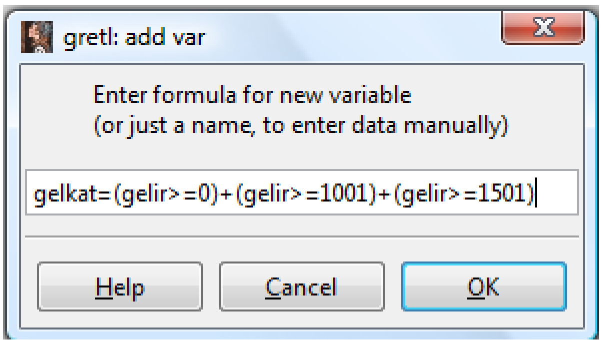
Eğer mantıksal ifadeyi:

gelkuk = gelir<1035.1

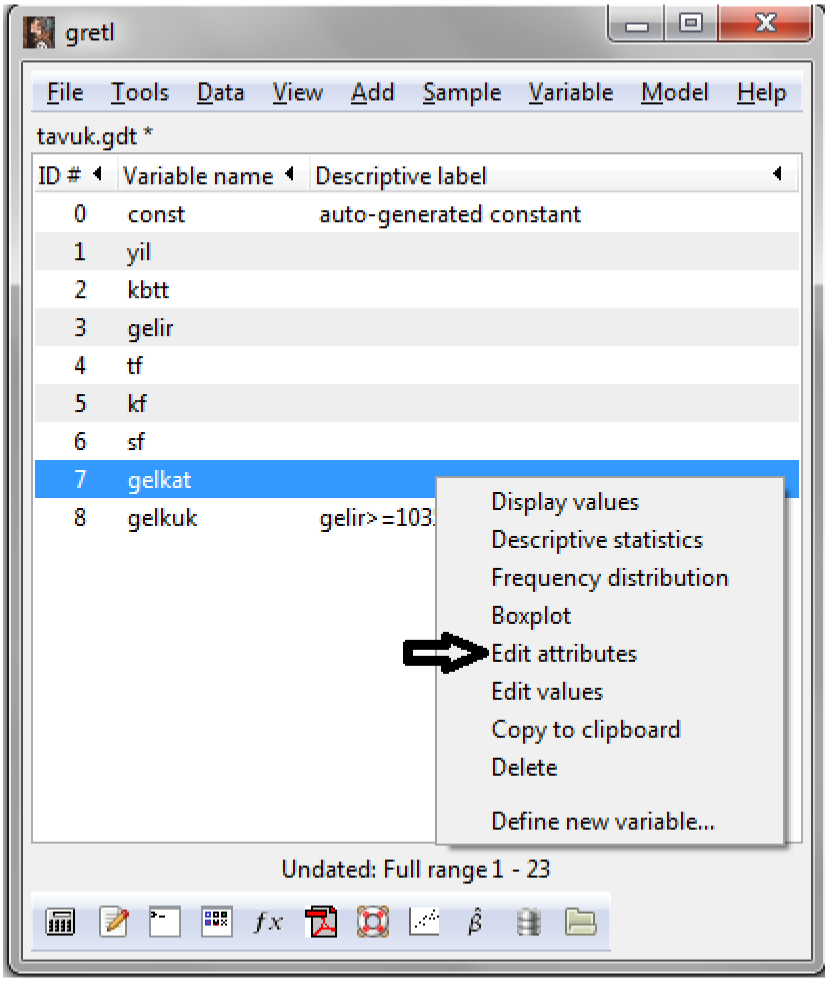
şeklinde tanımlasaydık, gelirin 1035.1'den küçük olduğu durumlarda gelkuk değişken 1 değerini, 1035.1'e eşit veya bundan daha büyük olduğu durumlarda 0 değerini alacaktı.

Eğer elimizdeki veri setinde kategorik değişken varsa, Gretl her kategori için bir kukla değişken tanımı yapabilir. Örneğin daha once gelir değişkenini; 0 ile 1000 arasında ise 1. kategori, 1001 ile 1500 arasında ise 2. kategori, 1500'den büyükse 3. kategori olarak tanımlamış ve bu değişkene ***gelkat*** adını vermiştik. Şimdi bu değişkenden yararlanarak, farklı şekillerde nasıl kukla değişkenler tanımlayabileceğimizi göreceğiz. Önce gelkat değişkenini Gretl'da nasıl tanımladığımız hatırlayalım:

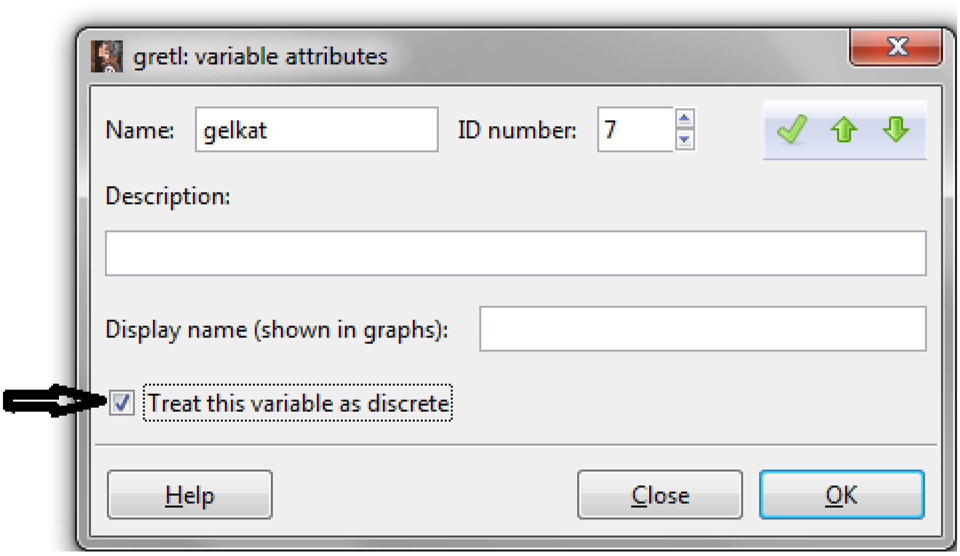
$$\text{gelkat} = \underbrace{(\text{gelir}>=0)}_{\text{Kategori 1}} + \underbrace{(\text{gelir} >= 1001)}_{\text{Kategori 2}} + \underbrace{(\text{gelir} >= 1501)}_{\text{Kategori 3}}$$



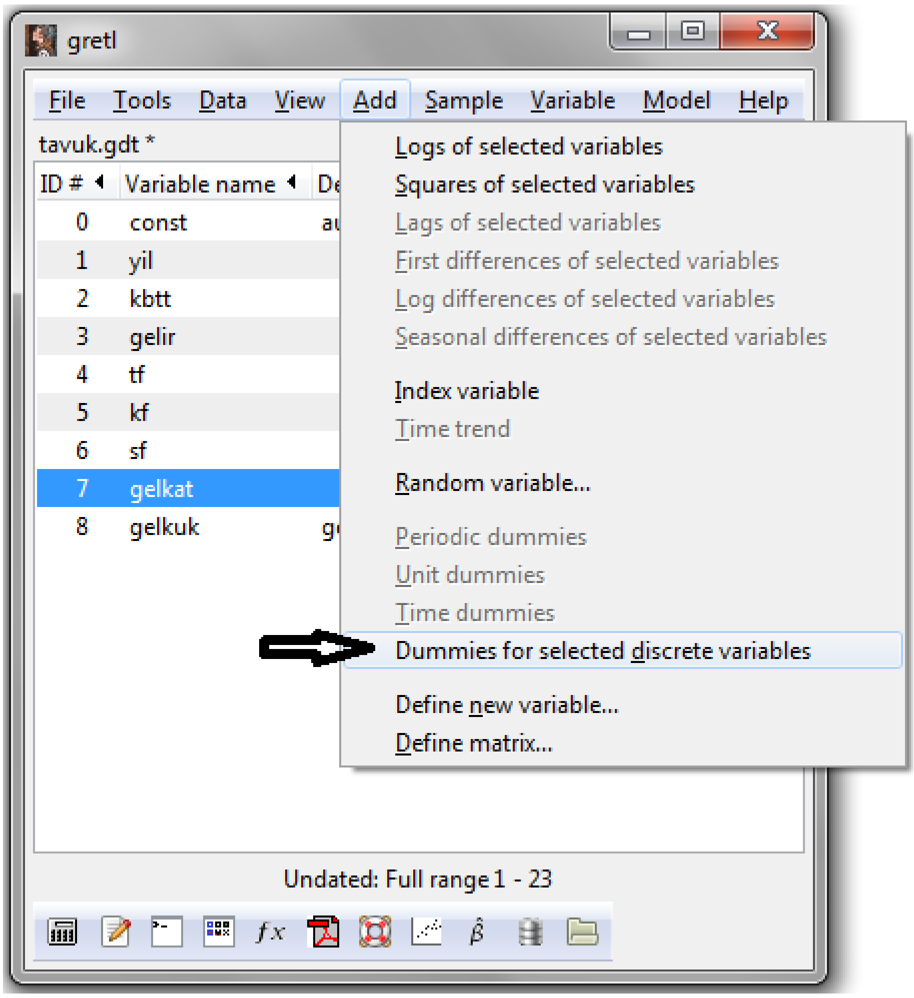
İzleyen adımda, Gretl'ın ***gelkat*** değişkenini kesikli değişken olarak tanımasını sağlamamız gerekiyor. Bunun için, gelkat değişkeni üzerindeyden sağ kliği tıkladığımızda açılan menüden ***Edit attributes*** seçeneğini tıklamalıyız:



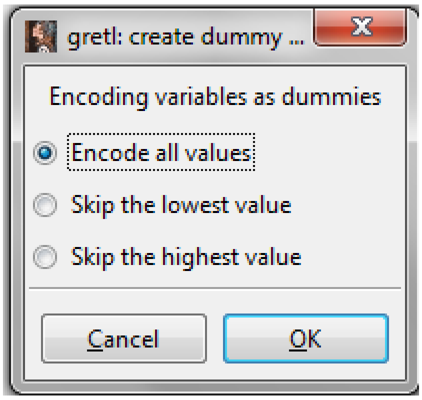
Aşağıda görülen ***Edit attributes*** ekranında ***Treat this variable as discrete*** seçeneğine onay işareti koyup, OK düğmesine basmamız gerekiyor.



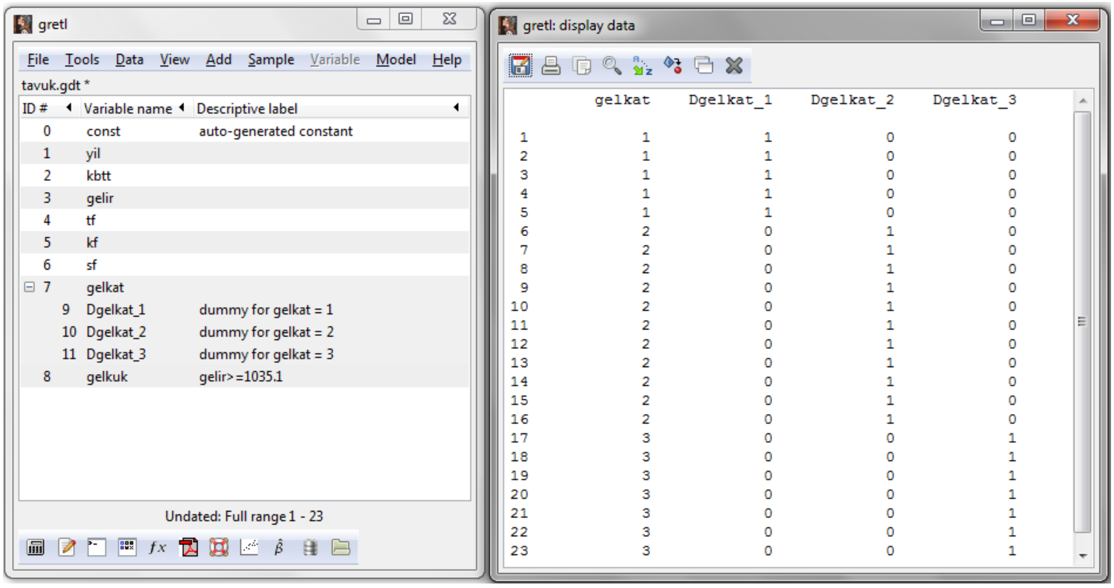
Artık ***gelkat*** değişkeninden kukla değişkenler elde etmeye hazırız. ***gelkat*** değişkeni işaretli iken ***Add/Dummies for selected variables*** menü dizisi izlenir.



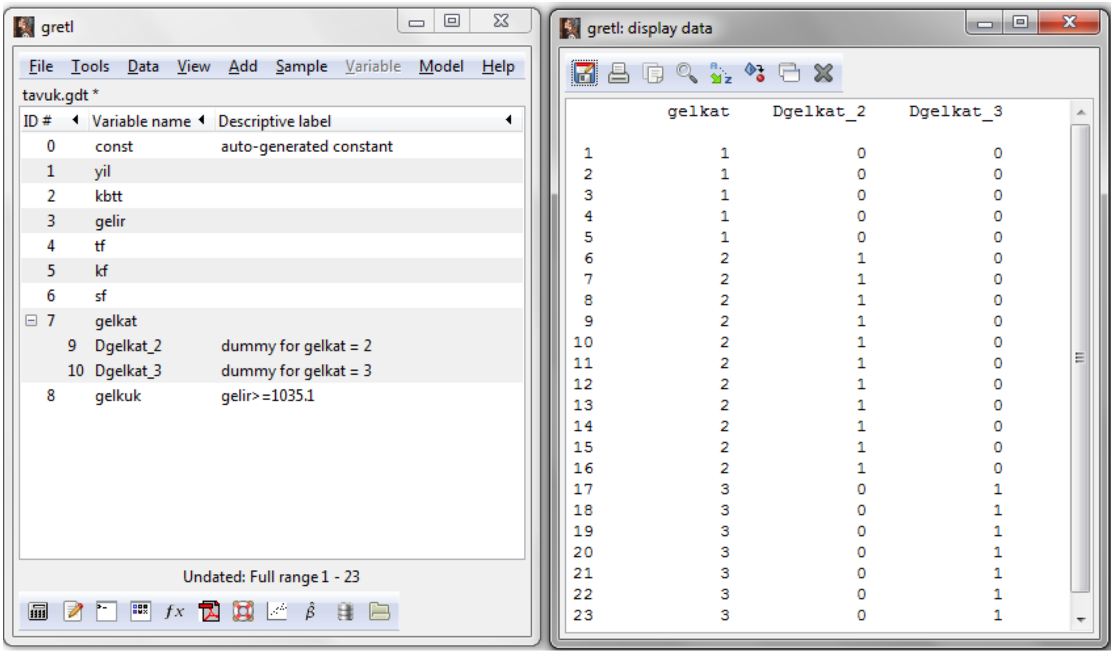
Bunun üzerine karşımıza:



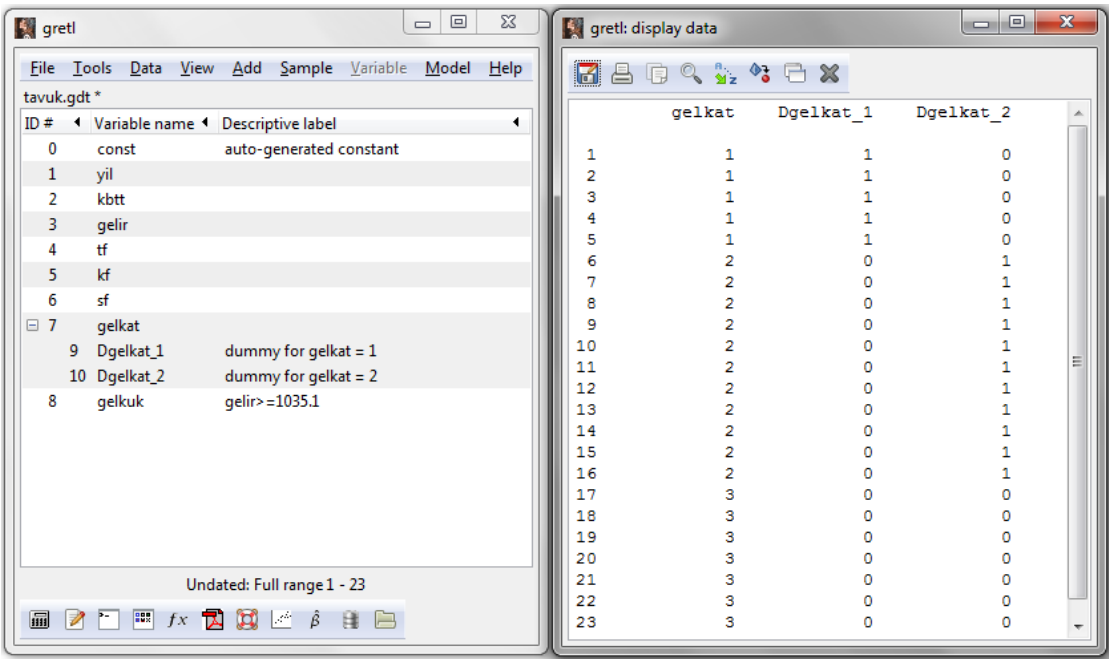
çıkacaktır. ***Encode all values*** seçeneğini işaretleyip OK düğmesine basarsak, 3 farklı gelir kategorisinin herbiri için birer kukla değişken hazırlanacak ve bu değişkenlere sırasıyla Dgelkat_1, Dgelkat_2 ve Dgelkat_3 adları verilecektir.



Skip the lowest value seçeneğini işaretleyip OK düğmesine basarsak, ilk kategori hariç diğer 2 gelir kategorisinin herbiri için birer kukla değişken hazırlanacak ve bu değişkenlere sırasıyla Dgelkat_2 ve Dgelkat_3 adları verilecektir.



Skip the highest value seçeneğini işaretleyip OK düğmesine basarsak, son kategori hariç diğer 2 gelir kategorisinin herbiri için birer kukla değişken hazırlanacak ve bu değişkenlere sırasıyla Dgelkat_1 ve Dgelkat_2 adları verilecektir..



5.1. Farklı Sabit Terim

Kukla değişkenleri ilk olarak, farklı sabit terimli modellerin elde edilmesinde kullanacağız.

Örnek 5-1

Bir işyerinde çalışanların maaşlarını (TL/ay), cinsiyetin bir fonksiyonu olarak tanımlayalım. Cinsiyet kukla değişkendir ve kadın 0, erkek 1 olarak sayısallaştırılmıştır. Doğrusal modelimiz:

$$Y = b_0 + b_1D$$

olsun. Modelimizi tahmin ettiğimizde eğer kukla değişkene ait tahminci (b_1) istatistiki açıdan sıfırdan farklı bulursak, kadın ve erkek için farklı modeller elde edebiliriz. Bir başka ifadeyle, kadın verkek için farklı sabit terimlere ulaşırız. Aksi halde gerek kadın gerekse erkek için tek bir sabit terim söz konusu olur. Bu modelde D yerine 0 koyduğumuzda, kadınlar için; 1 koyduğumuzda ise erkekler için maaş modeli elde ederiz:

Kadın maaş modeli: $Y_K = b_0 + b_1D$

$Y_K = b_0 + b_1(0)$

$Y_K = b_0$

Erkek maaş modeli: $Y_E = b_0 + b_1D$

$Y_E = b_0 + b_1(1)$

$Y_E = b_0 + b_1$

Yeni bir bakış açısıyla, erkek modeliyle kadın modeli farkı alırsak:

$$Y_E - Y_K = (b_0 + b_1) - (b_0)$$

$Y_E - Y_K = b_1$

b_1 tahmincisine ulaşırız. Eğer b_1 istatistiki olarak önemliyse, bu sonuçtan yararlanabiliriz. Aksi halde, fark, sıfır olacağından, böyle bir kullanımın anlamı olmaz.

Şimdi aşağıdaki verileri kullanarak, modellerimizi tahmin edelim.

Personel No	Maaş (Y)	Cinsiyet (D)
1	300	0
2	350	0
3	600	1
4	700	1
5	750	1
6	900	1
7	250	0
8	400	0
9	725	1
10	400	0

Tahmin sonuçları:

Bağımlı değişken: Maaş (Y)

	Tahminci	StHata	t	p
Sabit	340	40	8.5	0.000
D	395	56.57	6.98	0.000
S = 89.44	R^2 = 0.859			

Tahmin sonuçlarına göre kukla değişkenimizin tahmincisi istatistiki açıdan anlamlıdır. O halde, kadın ve erkek için farklı sabit terimler olacaktır.

Tahmin sonuçlarını denklem halinde yazalım:

$$Y = 340 + 395\ D$$

D'nin tahmincisi anlamlı olduğundan, model bize, erkeklerin maaşının kadınlardan 395 birim daha fazla olduğunu söylemektedir. Bir başka ifadeyle, maaşın cinsiyete göre türevi, erkek ve kadın maaşı arasındaki farkı verir. Bunu ayrıntılandırırsak:

Kadın modelimiz: $Y_K = 340 + 395\ (0)$

$Y_K = 340$

Erkek modelimiz: $Y_E = 340 + 395\ (1)$

$Y_E = 735$

Görüldüğü gibi, tek bir model tahmini olmakla birlikte, kadın ve erkek için ayrı modeller elde edilmiştir. Aslında bu, iki farklı sabit terimden başka bir şey değildir.

Eğer model kurmaksızın kadın ve erkeklerin maaşlarının aritmetik ortalamasını alsaydık:

Ortalama kadın maaşı: 340

Ortalama erkek maaşı: 735

olduğunu görecektik. Zaten model de bize aynı maaş düzeylerini vermişti.

Örnek 5-2

Maaş örneğimizi genişletip, personelin deneyimini (yıl) ekleyelim.

Personel No	Maaş Y	Cinsiyet D	Deneyim X
1	300	0	3

2	350	0	11
3	600	1	11
4	700	1	11
5	750	1	17
6	900	1	21
7	250	0	3
8	400	0	15
9	725	1	18
10	400	0	17

Yeni doğrusal modelimiz:

$Y = b_0 + b_1D + b_2X$

şeklindedir. Bu modeli, kadın ve erkekler için farklı modellere bölelim:

Kadın modeli için, D yerine 0 koyalım:

$Y_K = b_0 + b_1(0) + b_2X$

$Y_K = b_0 + b_2X$

Erkek modeli için, D yerine 1 koyalım:

$Y_E = b_0 + b_1(1) + b_2X$

$Y_E = b_0 + b_1 + b_2X$

$Y_E = (b_0 + b_1) + b_2X$

Bu kez, kukla değişkenin, farklı sabit terim amaçlı kullanımı daha açık bir biçimde görülmektedir. Kadın ve erkek maaş modellerinin sabit terimleri arasında, b_1 kadar fark vardır. Kuşkusuz bunun için, b_1'in istatistiki olarak sıfırdan farklı olması gerekir.

Tahmin sonuçları:

Bağımlı değişken: Maaş (Y)

	Tahminci	StHata	t	p
Sabit	211.28	41.28	5.12	0.000
D	318.82	39.55	8.06	0.000
X	13.135	3.411	3.85	0.006
S = 54.15	R^2 = 0.955	Düz.R^2 = 0.942	F=73.93	p=0.000

Model, istatistiki açıdan anlamlıdır. Tahmincilerimizin de tamamı, %1 düzeyinde sıfırdan farklıdır. O halde modelimizi güvenle yorumlayabiliriz.

Modelimizi denklem formunda yazalım:

$Y=211.28 + 318.82D + 13.135X$

Kadın ve erkek modellerimizi, tahmin sonuçlarını kullanarak yazabiliriz:

Kadın modeli (D=0):

$Y_K = 211.28 + 318.82(0) + 13.135X$

$Y_K = 211.28 + 13.135X$

Erkek modeli (D=1):

$Y_E = 211.28 + 318.82(1) + 13.135X$

$$Y_E = 530.10 + 13.135X$$

Görüldüğü gibi, kadın ve erkek modellerinin sabit terimleri faklı, deneyim terimi aynıdır. Buna göre, aynı deneyime sahip iki çalışan, cinsiyet farkı nedeniyle 318.82 TL/ay farklı maaş almaktadır. Bu fark, erkeklerin lehinedir.

Maaşın deneyime göre türevi, fazladan her bir deneyim yılının, maaşta meydana getireceği değişikliği verecektir. Bir yıl fazla deneyimin maaşa yaptığı etki, 13.135 TL/ay'dır. Bu değer yani eğim, sabit terimin farklı olmasının aksine, kadın ve erkek için aynıdır.

5.2. Farklı Eğim

Bazı araştırmalarda, farklı nitelikler için, farklı eğimler istenebilir. Böyle durumlar için de kukla değişkenlerden yararlanabiliriz.

Örnek 5-3

Cinsiyet ve deneyime göre maaş örneğimizi kullanarak, sabit terimi aynı, eğimi farklı modellerin elde edilmesini inceleyelim. Doğrusal modelimiz:

Denklem 5-1: $Y = b_0 + b_1X$

şeklindedir. Eğimin farklı olması, bu modeldeki X'e ait katsayının, kadın ve erkek için farklı olması anlamına gelmektedir. Buna göre b_1 tahmincisini, cinsiyete göre yeniden tanımlayalım:

Denklem 5-2: $b_1 = \beta_1 + \beta_2 D$

Eşitlikte, D yerine 0 koyarsak, kadın için eğimi:

$$b_1 = \beta_1 + \beta_2(0)$$

$$b_1 = \beta_1$$

D yerine 1 koyarsak, erkek için eğimi:

$$b_1 = \beta_1 + \beta_2(1)$$

$$b_1 = \beta_1 + \beta_2$$

olarak elde ederiz.

Şimdi Denklem 5-2'ü, Denklem 5-1'te yerine koyalım:

Denklem 5-3: $Y = b_0 + (\beta_1 + \beta_2 D)\, X$

Denklemi açarsak:

$$Y = b_0 + \beta_1 X + \beta_2\, DX$$

halini alır. Bu, aynı zamanda, tahmin edeceğimiz denklemdir. Modelimizin değişkenleri X ve DX'tir. X, deneyim değişkeniydi; DX ise, cinsiyet kuklasıyla, deneyim değişkeninin çarpımından elde edilir. Veri tablomuzun son durumu:

Personel No	Maaş Y	Cinsiyet D	Deneyim X	DX
1	300	0	3	0
2	350	0	11	0
3	600	1	11	11
4	700	1	11	11
5	750	1	17	17

6	900	1	21	21
7	250	0	3	0
8	400	0	15	0
9	725	1	18	18
10	400	0	17	0

Şimdi modelimizi tahmin edelim:

Bağımlı değişken: Maaş (Y)

	Tahminci	StHata	t	p
Sabit	276.83	41.36	6.69	0.000
X	7.251	3.728	1.95	0.093
DX	21.612	2.606	8.29	0.000
S = 52.78	R^2 = 0.957	Düz.R^2 = 0.945	F=77.99	p=0.000

Modelimizdeki tahmincilerin tamamı, %10 düzeyinde sıfırdan farklıdır. Denklem formunda yazarsak:

$$Y = b_0 + \beta_1 X + \beta_2 DX$$

$$Y = 276.83 + 7.251X + 21.612\ DX$$

Burada b_0=276.83, β_1=7.251, β_2=21.612'dir.

Şimdi kadın ve erkek modellerini hazırlayacağız. Önce Denklem 5-3'i hatırlayalım:

$$Y = b_0 + (\beta_1 + \beta_2 D)\ X$$

Tahmincileri yerlerine koyalım:

$$Y = 276.83 + (7.251 + 21.612\ D)\ X$$

Kadın modeli:

$$Y_K = 276.83 + [7.251 + 21.612(0)]\ X$$

$$Y_K = 276.83 + 7.251\ X$$

Erkek modeli:

$$Y_E = 276.83 + [7.251 + 21.612(1)]\ X$$

$$Y_E = 276.83 + 28.863\ X$$

Kadın ve erkek modellerini bir arada görelim:

$$Y_K = 276.83 + 7.251\ X$$

$$Y_E = 276.83 + 28.863\ X$$

Görüldüğü gibi, sabit terimler aynı, eğimler farklıdır. Buna göre, kadınların çalıştıkları fazladan her yıl, 7.251 TL/ay maaş farkına yol açarken; erkeklerin çalıştıkları fazladan her yıl, 28.863 TL/ay maaş farkına neden olmaktadır.

5.3. Farklı Sabit Terim ve Farklı Eğim

Yalnız sabit terim veya yalnız eğim farklılığının yanısıra, hem sabit terim, hem de eğim farklılığı olan modellere de ihtiyaç duyulabilir. Kukla değişkenleri bu amaçla kullanabiliriz.

Örnek 5-4

Maaş örneğimizi yeniden ele alıp, farklı sabit terim ve farklı eğimli modelleri elde edelim.

Denklem 5-4: $Y = b_0 + b_1X$

şeklindedir.

Sabit terim farklılığını:

Denklem 5-5: $b_0 = \alpha_1 + \alpha_2 D$

olarak ifade edebiliriz.

Kadın sabiti:

$b_0 = \alpha_1 + \alpha_2 D$

$b_0 = \alpha_1 + \alpha_2(0)$

$b_0 = \alpha_1$

Erkek sabiti:

$b_0 = \alpha_1 + \alpha_2 D$

$b_0 = \alpha_1 + \alpha_2(1)$

$b_0 = \alpha_1 + \alpha_2$

olur.

Eğimin farklı olması, bu modeldeki X'ye ait katsayının, kadın ve erkek için farklı olması demektir. Bu nedenle b_1 tahmincisini, cinsiyete göre yeniden tanımlamamız gerekir:

Denklem 5-6: $b_1 = \beta_1 + \beta_2 D$

Eşitlikte, D yerine 0 koyarsak, kadın için eğimi:

$b_1 = \beta_1 + \beta_2(0)$

$b_1 = \beta_1$

D yerine 1 koyarsak, erkek için eğimi:

$b_1 = \beta_1 + \beta_2(1)$

$b_1 = \beta_1 + \beta_2$

olarak elde ederiz.

Şimdi Denklem 5-5 ve Denklem 5-6'i, Denklem 5-4'da yerine koyalım:

Denklem 5-7: $Y = (\alpha_1 + \alpha_2 D) + (\beta_1 + \beta_2 D)\,X$

Denklemi açarsak:

$Y = \alpha_1 + \alpha_2 D + \beta_1 X + \beta_2\, DX$

olur. Bu, tahmin edeceğimiz denklemdir. Modelimizin değişkenleri D, X ve DX'tir. X deneyim, D cinsiyet değişkeniydi; DX ise, cinsiyet kuklasıyla, deneyim değişkeninin çarpımıdır.

Modelimizin tahmin sonuçları:

Bağımlı değişken: Maaş (Y)

	Tahminci	StHata	t	p
Sabit	246.99	40.81	6.05	0.001
D	159.07	94.22	1.69	0.142

X	9.491	3.571	2.66	0.038
DX	11.595	6.37	1.82	0.119
S = 46.94	R^2 = 0.971	Düz.R^2 = 0.956	F=66.68	p=0.000

Denklem, % 1 için önemlidir. Tahmincilerimizin tamamı %15 için önemlidir. Yaygın bir hata payı olmamakla birlikte, buradaki değerlendirmelerimiz açısından α=0.15 kabul edilmiştir.

Tahmin sonuçlarına göre denklemimiz:

$$Y = \alpha_1 + \alpha_2 D + \beta_1 X + \beta_2 DX$$

$$Y = 246.99 + 159.07D + 9.491X + 11.595DX$$

Burada α_1=246.99, α_2=159.07, β_1= 9.491, β_2 =11.595'tir. Bunları Denklem 5-7'da yerine koyalım:

$$Y = (\alpha_1 + \alpha_2 D) + (\beta_1 + \beta_2 D)X$$

$$Y = (246.99+159.07D) + (9.491+11.595D)X$$

Artık sabiti ve eğimi farklı kadın ve erkek denklemlerimizi elde edebiliriz.

Kadın denklemi (D=0):

$$Y_K = [246.99+159.07(0)] + [9.491+11.595(0)]X$$

$$Y_K = 246.99 + 9.491X$$

Erkek denklemi (D=1):

$$Y_E = [246.99+159.07(1)] + [9.491+11.595(1)]X$$

$$Y_E = 406.06 + 21.086X$$

Kadin ve erkek denklemlerini bir arada görelim:

$$Y_K = 246.99 + 9.491X$$

$$Y_E = 406.06 + 21.086X$$

Farkedileceği üzere, kadın ve erkek modellerinde gerek sabit terim, gerekse eğim farkldır. Buna göre:

Kadınlar ilk işe başladıklarında (X=0), 246.99 TL/ maaş almaktadır. Çalıştıkları her fazladan yıl için, kadınlara, 9.491 TL/ay maaş artışı yapılmaktadır.

Erkekler ilk işe başladıklarında (X=0), 406.06 TL/ maaş almaktadır. Çalıştıkları her fazladan yıl için, erkeklere, 21.086 TL/ay maaş artışı yapılmaktadır.

5.4. Kategorilerin Oranı

Hatırlanacağı üzere, doğrusal modeldeki kukla değişken, iki kategori arasındaki farkı mutlak olarak belirlemekteydi. Kukla değişkenin temsil ettiği kategoriler arasındaki oransal ilişki, log-lin denklemlerle belirlenebilir.

Modelimiz:

$$\ln Y = b_0 + b_1 D + b_2 X$$

olsun. Burada Y bağımlı değişken, X sürekli bağımsız değişken ve D kukla değişkeni temsil etmektedir. Kukla değişkenin 1 olduğu durum için modelimiz:

$$\ln Y_1 = b_0 + b_1 + b_2 X$$

Kukla değişkenin 0 olduğu durum için modelimiz:

$$\ln Y_0 = b_0 + b_2 X$$

İki model arasındaki fark:

$\ln Y_1 - \ln Y_0 = b_0 + b_1 + b_2X - (b_0 + b_2X)$

$\ln Y_1 - \ln Y_0 = b_1$

Hatırlanacağı üzere iki logaritmik terim arasındaki fark aşağıdaki gibi ifade ediliyordu:

$\ln(Y_1/Y_0) = b_1$

b_1'in antilogaritmasını aldığımızda:

$Y_1/Y_0 = \text{antilog}_e\,(b_1)$

olur.

Örnek 5-5

Cinsiyet ve deneyime göre maaş modelimizi, bu kez log-lin olarak olarak tahmin edelim. Önce verilerimizi hatırlayalım:

Personel No	Maaş Y	Cinsiyet D	Deneyim X
1	300	0	3
2	350	0	11
3	600	1	11
4	700	1	11
5	750	1	17
6	900	1	21
7	250	0	3
8	400	0	15
9	725	1	18
10	400	0	17

Modelimiz:

$\ln Y = b_0 + b_1D + b_2X$

Tahmin sonuçları:

Bağımlı değişken: Maaş (Y)

	Tahminci	StHata	t	p
Sabit	5.5317	0.05727	96.6	0.000
D	0.6115	0.05487	11.14	0.000
X	0.028726	0.004732	6.07	0.000
S = 0.07512	R^2 = 0.978	Düz.R^2 = 0.971	F=152.53	p=0.000

Modelimizi denklem formunda yazalım:

$\ln Y = 5.5317 + 0.6115D + 0.028726X$

Kadın modeli: $\ln Y_K = 5.5317 + 0.028726X$

Erkek modeli: $\ln Y_E = 6.1432 + 0.028726X$

İki model arasındaki fark:

$\ln Y_E - \ln Y_K = 0.6115$

$\ln (Y_E / Y_K) = 0.6115$

$Y_E / Y_K = \text{antilog}_e (0.6115)$

$Y_E / Y_K = 1.84$

Anlaşılacağı üzere, erkek ve kadın kategorileri arasındaki maaş oranı, 1.84'tür. Bir başka ifadeyle, erkek maaşı, kadın maaşının 1.84 katıdır.

Burada yapılan, basit olarak kukla değişkene ait katsayının antilogaritmasının alınmasıdır.

5.5. Oransal Kategori Etkisi

Eğer sorumuz "kategorik değişkendeki bir birim değişmenin, bağımlı değişkende meydana getireceği oransal değişme nedir?" ise, bunun cevabını, yine kukla değişken yardımıyla verebiliriz.

Modelimiz:

$\ln Y = b_0 + b_1D + b_2X$

olsun. D, kukla değişken; X ise sürekli değişkendir. Bu denklemden yararlanarak, kukla değişkendeki 1 birim değişmenin Y'de meydana getirdiği oransal değişimi (=OD) hesaplayalım:

OD = (Bağımlı değişkendeki oransal değişme) / (kategori farkı)

$OD = (\Delta Y/Y) / \Delta D$

şeklinde formüle edebiliriz. D, iki kategori içerir. ΔD, kukla değişkenin iki kategorisi arasındaki farktır. Kukla değişken üzerinde çalıştığımız için kuşkusuz bu fark, 1'dir. Kısmi türev alacağımızdan, Δ yerine ∂ koyarak işlemimize devam edebiliriz. Gerekli düzenlemeleri yaptıktan sonra:

$OD = (\Delta Y/Y) / \Delta D = (\partial y/\partial D).(1/Y)$

olur. $\partial Y/\partial D$, kısmi türevdir. Logaritmik fonksiyonlarda türev alma kurallarını uyguladığımızda:

$$OD = \frac{\partial y}{\partial D}.\frac{1}{y} = b_1 y \frac{1}{y} = b_1$$

olur.

Hatırlanacağı üzere iki logaritmik terim arasındaki fark aşağıdaki gibi ifade ediliyordu:

$\ln(Y1/Y0) = b1$

b1'in antilogaritmasını aldığımızda:

$Y1/Y0 = \text{antiloge}\ (b1)$

olur.

Örnek 5-5'deki modeli ve tahmin sonuçlarını tekrar ele alalım:

$\ln Y = 5.5317 + 0.6115D + 0.028726X$

Denklemde D'nin tahmincisi, 0.6115'tir ve bu, oransal değişim değerine eşittir. Yani, kukla değişkendeki bir birim değişmenin, maaşta yol açacağı değişim oranıdır. Bir başka ifadeyle, kadın ve erkek maaşları arasındaki oransal farktır. Burada referans olarak, kukla değişkenin sıfır değeri alınır. Hatırlanacağı gibi, sıfır, kadını temsil etmektedir. O halde, kadın maaşından erkek kategorisine geçildiğinde, yaklaşık %61.15'lik bir maaş farkı meydana gelecektir.

Örnek 5-6

Burada hedonik model uygulaması yapacağız. Ev fiyatı ile onu açıklayan değişkenler arasındaki ilişkiyi log-lin formunda tahmin edeceğiz.

Evfiyat:	Evin fiyatı (1000 TL)
Evalan:	Evin kullanım alanı (m^2)
bahcealan:	Bahçenin alanı (m^2)
banyosay:	Banyo sayısı
yatodsay:	Yatak odası sayısı

Evno	Banyo sayisi	Yatak Odasi sayisi	evfiyat	Ev alan	Bahce alani	Evno	Banyo sayisi	Yatak Odasi sayisi	Ev fiyat	Ev alan	Bahce alani
1	1	2	107	73.6	336.4	22	1	2	135	93.6	505.4
2	1	2	133	72	178	23	1	2	147	72.8	192.2
3	1	2	141	76.8	653.2	24	1	2	165	101.4	641.6
4	1	3	165	92.9	274.7	25	2	3	175	166.1	493.9
5	1	3	170	108	552	26	1	3	190	124.8	795.2
6	1	2	173	94.2	680.8	27	2	3	191	183.4	671
7	1	3	182	100	610	28	1	3	195	98.9	591.1
8	2	3	200	147.2	532.8	29	2	2	205	123.2	461.8
9	1	3	220	120	585	30	1	2	210	101.7	508.3
10	2	3	226	130.2	529.8	31	1	2	215	121.6	683.4
11	2	3	260	210.9	369.1	32	2	2	228	144.7	414.3
12	2	2	275	152.8	586	33	2	4	242	197.4	549.9
13	2	3	280	142.1	667.9	34	2	3	250	160	405
14	2	3	289	175.3	230.4	35	1	3	250	116.8	518.2
15	2	3	295	152.8	629.2	36	2	3	255	147.8	412.2
16	2	3	300	164.3	712.7	37	2	3	255	175.6	642
17	2	3	310	167.5	902.5	38	2	3	265	154.2	683.3
18	2	3	315	171.4	646.6	39	2	4	265	163.3	711.7
19	2	4	350	215	1482.5	40	2	2	275	150	740.6
20	2.5	4	365	220.6	814.7	41	2	3	285	173.4	858.3
21	2.5	4	503	326.9	1004.5	42	2	3	365	190	1958

Modelimizi log-lin olarak tahmin edelim:

Bağımlı değişken: l_evfiyat

	Katsayı	*St.Hata*	*t*	*p*	
sabit	4.48476	0.121427	36.9337	<0.00001	***
evalan	0.00293657	0.001128	2.6034	0.01321	**
bahcealan	0.000238036	0.000091	2.6053	0.01315	**
banyosay	0.173049	0.085717	2.0188	0.05079	*
yatodsay	0.0295241	0.054845	0.5383	0.59358	

R-kare	0.777128	Düzeltilmiş R-kare	0.753034
F(4, 37)	32.25367	P(F)	1.34e-11
Log-likelihood	20.77997	Akaike kriteri	-31.55994
Schwarz kriteri	-22.87159	Hannan-Quinn	-28.37532

Hedonik modelimizin tahmin sonuçlarını inceleyelim:

- Model istatistiki açıdan anlamlıdır. Güvenle kullanılabilir.
- Yatak odası sayısı dışındaki tüm değişkenler istatistiki olarak önemlidir.
- Evin kullanım alanı arttıkça evin fiyatı artmaktadır. Evin alanı 10 m^2 arttığında ev fiyatı %2.9 (0.0029*10=0.029) artmaktadır.
- Evin bahçe alanı arttıkça evin fiyatı artmaktadır. Evin bahçesi 10 m^2 arttığında ev fiyatı %0.238 (0.000238*10=0.00238) artmaktadır.
- Banyo sayısı 1 arttığında, ev fiyatı %17.3 artmaktadır.
- Yatak odası sayısı ev fiyatını etkilememektedir.

5.6. Mevsimselliğin Belirlenmesi

Zaman serisi modellerinde, mevsimsel değişkenliği belirlemek amacıyla, mevsim kuklalarından yararlanılır. Üç aylık bir zaman serisinde, mevsim değişkenini:

D_{1t}	=	1. üç ayda = 1
		Diğerleri= 0
D_{2t}	=	2. üç ayda = 1
		Diğerleri= 0
D_{3t}	=	3. üç ayda = 1
		Diğerleri= 0

olarak tanımlayalım.

Buna göre modelimiz:

$$Y_t = b_0 + b_1D_1 + b_2D_2 + b_3\,D_3 + b_4X$$

olsun. X, kukla olmayan bir değişkendir. 4 adet üç aylık dönemi, 3 kukla değişkenle temsil ettiğimize dikkat edelim. Denklemdeki;

- b_1, Y_t'nin 1. dönemdeki beklenen değerinin, denklemde dikkate alınmayan 4. üç aylık dönemdeki beklenen değerinden farkını;
- b_2, Y_t'nin 2. dönemdeki beklenen değerinin, denklemde dikkate alınmayan 4. üç aylık dönemdeki beklenen değerinden farkını;
- b_3, Y_t'nin 3. dönemdeki beklenen değerinin, denklemde dikkate alınmayan 4. üç aylık dönemdeki beklenen değerinden farkını;

göstermektedir.

Mevsim kuklalarının modele alınması, Y'deki ve mevsimsel düzeltmesi yapılmamış bağımsız değişkenlerdeki mevsim etkilerini giderir. Zaman serileri sık sık mevsimsellik olarak adlandırılan periyodizite gösterir. Örneğin 3'er aylık perakende satış verileri, 4. dönemde sıçrama gösterir. Mevsimsellik, mevsim kuklaları yardımıyla giderilebilir. Böylece, trendde olduğu gibi, tahminlemeden önce, serilerin mevsimsel düzeltmesi yapılmış olur. Y ve X_4 tahminleme sürecinden önce mevsim etkisinden arındırılmadığı sürece, bu işlemden yararlanılabilir. Tahmin öncesi mevsim etkisinin giderilmesi, belirlenemeyen yön ve miktarda verilerin bozulmasına yol açmaktadır. Halbuki mevsim kuklalarının sınırlamaları tesbit edilebilmektedir.

Verilerde istatistiki öneme sahip mevsim etkisinin bulunduğunu belirlemek için, tüm kukla değişkenlerini aynı anda test edilmesi gerekir. Bir başka ifadeyle, mevsim kuklaları tek tek değil, aynı anda hipotez testine tabi tutulmalıdır. Bu nedenle, t testi deği, F testi yapılır. Örneğimiz için sıfır hipotezimiz:

H_0: $b_1 = b_2 = b_3 = 0$ (Mevsimsellik yok)

H_1: H_0 doğru değil. (Mevsimsellik var)

olarak hazırlanır. Buna göre sıfır hipotezi, mevsimsellik bulunmadığını iddia eder.

O halde, kısıtlı fonksiyonumuz:

$$Y_t = b_0 + b_4X$$

olur. Mevsim kuklalarının hepsinin birden denkleme alınmasının gerekli olup, olmadığına karar vermek için Wald testini testini kullanarak, kısıtlı fonksiyon ile kısıtsız fonksiyonun karşılaştırması yapılabilir. Ancak istatistiki açıdan önemsiz olan mevsim kuklalarının modelden çıkarılmaması gerekir.

Örnek 5-7

Mevsimlere göre elektrik tüketiminin farklılık gösterip göstermediğini, 3'er aylık verilerle belirleyelim. Kişi başına elektrik tüketimini (KBELTUK), reel elektrik fiyatı (RFIYAT) ve mevsim kuklalarıyla (D_{ilk}=ilkbahar kuklası, D_{yaz}=Yaz kuklası, D_{son}=Sonbahar kuklası) ilişkilendirerek tahminleyeceğiz. Referans dönem olarak kış mevsimini alacağız:

$$KBELTUK = b_0 + b_1\, RFIYAT + b_2\, D_{ilk} + b_3\, D_{yaz} + b_4\, D_{son}$$

DÖNEM	KBELKTUK	RFIYAT	D_{ilk}	D_{yaz}	D_{son}	DÖNEM	KBELKTUK	RFIYAT	D_{ilk}	D_{yaz}	D_{son}
1978.1	0.567394	8.433842	0	0	0	1986.1	0.527911	10.21662	0	0	0
1978.2	0.477106	8.635546	1	0	0	1986.2	0.456645	9.980737	1	0	0
1978.3	0.501628	8.486284	0	1	0	1986.3	0.496015	10.12132	0	1	0
1978.4	0.556246	8.071069	0	0	1	1986.4	0.500847	9.96098	0	0	1
1979.1	0.626602	7.762458	0	0	0	1987.1	0.56636	9.836842	0	0	0
1979.2	0.493575	7.621531	1	0	0	1987.2	0.468485	9.607815	1	0	0
1979.3	0.512611	8.084164	0	1	0	1987.3	0.487869	9.708267	0	1	0
1979.4	0.554249	7.681072	0	0	1	1987.4	0.530209	9.618004	0	0	1
1980.1	0.575572	7.782004	0	0	0	1988.1	0.578119	9.058543	0	0	0
1980.2	0.489966	8.665279	1	0	0	1988.2	0.481734	8.828634	1	0	0
1980.3	0.495495	11.34229	0	1	0	1988.3	0.526707	8.953877	0	1	0
1980.4	0.51272	11.35606	0	0	1	1988.4	0.519669	8.649469	0	0	1
1981.1	0.540351	11.23329	0	0	0	1989.1	0.597334	7.979247	0	0	0
1981.2	0.461138	10.80261	1	0	0	1989.2	0.477401	7.631036	1	0	0
1981.3	0.497156	10.34498	0	1	0	1989.3	0.522422	7.710697	0	1	0
1981.4	0.505729	9.742386	0	0	1	1989.4	0.518885	7.505085	0	0	1
1982.1	0.545145	11.18603	0	0	0	1990.1	0.596365	7.379663	0	0	0
1982.2	0.452551	11.24103	1	0	0	1990.2	0.480142	7.226234	1	0	0
1982.3	0.473617	11.40056	0	1	0	1990.3	0.550786	7.49683	0	1	0
1982.4	0.493991	11.56722	0	0	1	1990.4	0.525166	7.29912	0	0	1
1983.1	0.513455	12.39384	0	0	0	1991.1	0.56901	7.575715	0	0	0
1983.2	0.459437	12.46561	1	0	0	1991.2	0.470671	7.470507	1	0	0
1983.3	0.485679	12.08131	0	1	0	1991.3	0.487536	7.440769	0	1	0
1983.4	0.493787	11.68338	0	0	1	1991.4	0.54133	7.386149	0	0	1
1984.1	0.498806	10.92543	0	0	0	1992.1	0.55232	7.519335	0	0	0
1984.2	0.448823	10.50135	1	0	0	1992.2	0.476694	7.326788	1	0	0
1984.3	0.504401	10.92892	0	1	0	1992.3	0.569147	7.363625	0	1	0
1984.4	0.521642	10.78022	0	0	1	1992.4	0.530199	7.284585	0	0	1
1985.1	0.558129	11.23412	0	0	0	1993.1	0.570125	7.234281	0	0	0
1985.2	0.450108	10.99859	1	0	0	1993.2	0.466722	7.135704	1	0	0
1985.3	0.501914	11.2016	0	1	0	1993.3	0.526554	7.277372	0	1	0
1985.4	0.5058	10.93668	0	0	1	1993.4	0.519754	7.25116	0	0	1

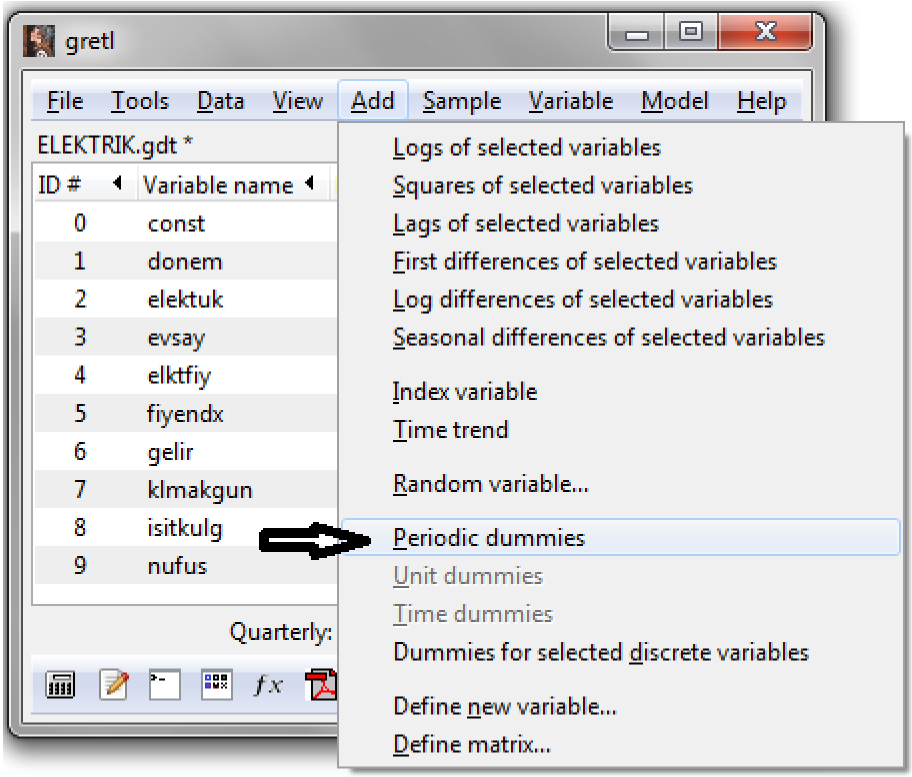
$$KBELTUK = b_0 + b_1\, RFIYAT + b_2\, D_{ilk} + b_3\, D_{yaz} + b_4\, D_{son}$$

	Katsayı	StHata	t	p
Sabit	0.64723	0.0129	50.16	0.000
RFIYAT	-0.00929	0.001315	-7.07	0.000
D_{ilk}	-0.09292	0.006184	-15.03	0.000
D_{yaz}	-0.05144	0.006185	-8.32	0.000
D_{son}	-0.04137	0.006183	-6.69	0.000

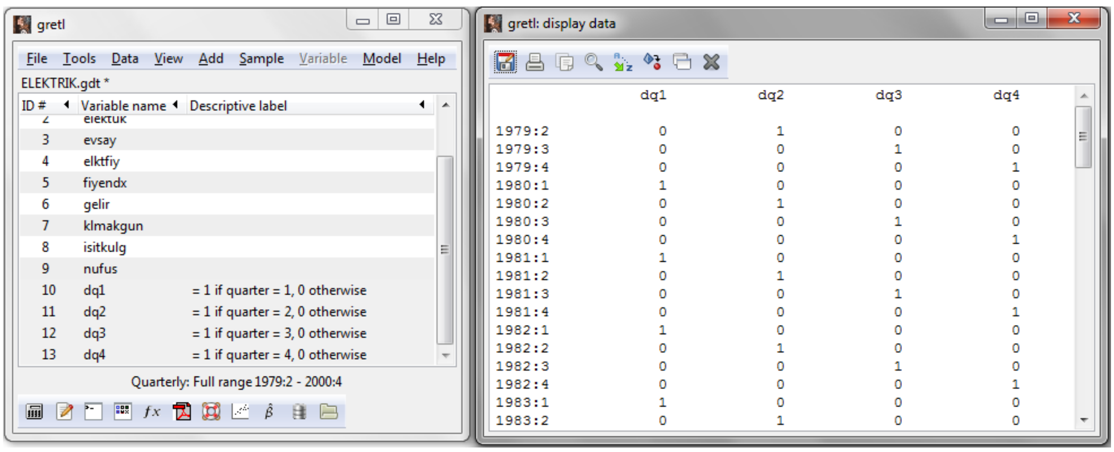
S = 0.0175 R^2 = 0.823 Düz.R^2 = 0.811 F=68.76 p=0.000

Modeldeki tüm katsayılar istatistiki açıdan anlamlıdır. Mevsim kuklalarının sıfırdan farklı olması, bu mevsimlerdeki elektrik tüketiminin referans dönem olan kış mevsiminden farklı olduğu anlamına gelmektedir. Mevsim kuklalarımızın tümünün negatif işaretli olması, ilkbahar, yaz ve sonbahar mevsimlerinde elektrik tüketiminin kış mevsimine göre daha düşük olduğu anlamına gelmektedir. Bir başka ifadeyle, kışın elektrik tüketimi diğer mevsimlere göre daha fazladır. Sonuçta, elektrik tüketimi mevsimsellik göstermektedir diyebiliriz.

Gretl mevsim kuklalarını otomatik olarak hazırlayabilir. ***Add/Periodic dummies*** seçeneğini tıkladığımızda:



Birinci 3 aylık dönemin kukla değişkeni dq1 adıyla, ikinci 3 aylık dönemin kukla değişkeni dq2 adıyla, üçüncü 3 aylık dönemin kukla değişkeni dq3 adıyla ve dördüncü 3 aylık dönemin kukla değişkeni dq4 adıyla hazırlanır.



5.6.1. *Uygulama*

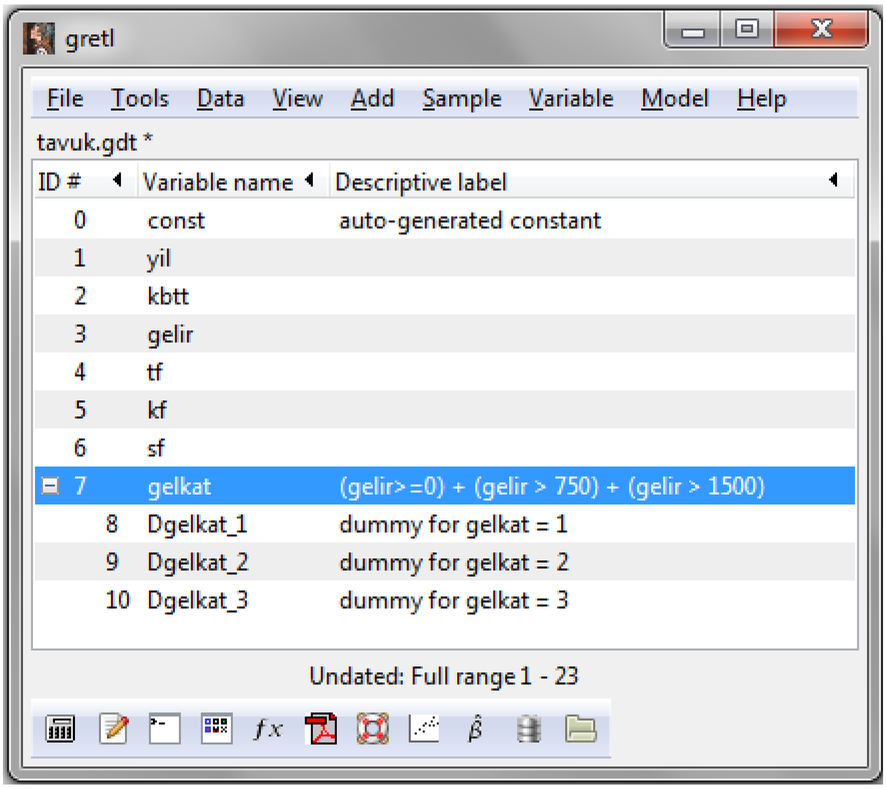
Tavuk eti talebiyle ilgili uygulamamızı farklı açılardan bakarak geliştirebiliriz. gelkat değişkenini gelir:

0 ile 750 arasında ise 1. kategori

751 ile 1500 arasında ise 2. kategori

1500'den büyükse 3. kategori

olarak tanımlayalım:



Önce kişi başına tavuk eti tüketiminin (kbtt), tavuk eti fiyatı (tf), kırmızı et fiyatı (kf) ve gelir ile doğrusal ilişkisini ele alalım:

$$kbtt = b_0 + b_1 tf + b_2 kf + b_3 gelir$$

fonksiyonunu tahmin edelim:

```
gretl: model 5
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 5: OLS, using observations 1-23
Dependent variable: kbtt

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       38.6472      3.64960      10.59     2.08e-09 ***
  tf          -0.541084    0.157970     -3.425    0.0028   ***
  kf           0.174055    0.0625307     2.784    0.0118   **
  gelir        0.0108762   0.00238125    4.567    0.0002   ***

Mean dependent var   39.66957   S.D. dependent var   7.372950
Sum squared resid    75.75855   S.E. of regression   1.996820
R-squared            0.936653   Adjusted R-squared   0.926651
F(3, 19)             93.64503   P-value(F)           1.45e-11
Log-likelihood      -46.34424   Akaike criterion     100.6885
Schwarz criterion    105.2305   Hannan-Quinn         101.8308
```

Bu modelin tahmin sonuçlarından, tavuk eti fiyatı (tf) bir birim arttığında tavuk eti talebinin 0.54 birim azalacağını; kırmızı et fiyatı (kf) bir birim arttığında tavuk eti talebinin 0.17 birim artacağını; gelir bir birim arttığında ise tavuk eti talebinin 0.01 birim artacağını anlıyoruz.

Modeldeki gelir değişkeninin yerine gelkat değişkenini koyarsak:

$$kbtt = b_0 + b_1 tf + b_2 kf + b_3 gelkat$$

```
gretl: model 11
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 11: OLS, using observations 1-23
Dependent variable: kbtt

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const        34.1902     3.15438      10.84     1.42e-09 ***
  tf          -0.569674    0.145769     -3.908    0.0009   ***
  kf           0.249215    0.0494255     5.042    7.23e-05 ***
  gelkat       5.77366     1.09087       5.293    4.15e-05 ***

Mean dependent var   39.66957   S.D. dependent var   7.372950
Sum squared resid    64.23420   S.E. of regression   1.838681
R-squared            0.946289   Adjusted R-squared   0.937809
F(3, 19)             111.5823   P-value(F)           3.04e-12
Log-likelihood      -44.44656   Akaike criterion     96.89313
Schwarz criterion    101.4351   Hannan-Quinn         98.03542
```

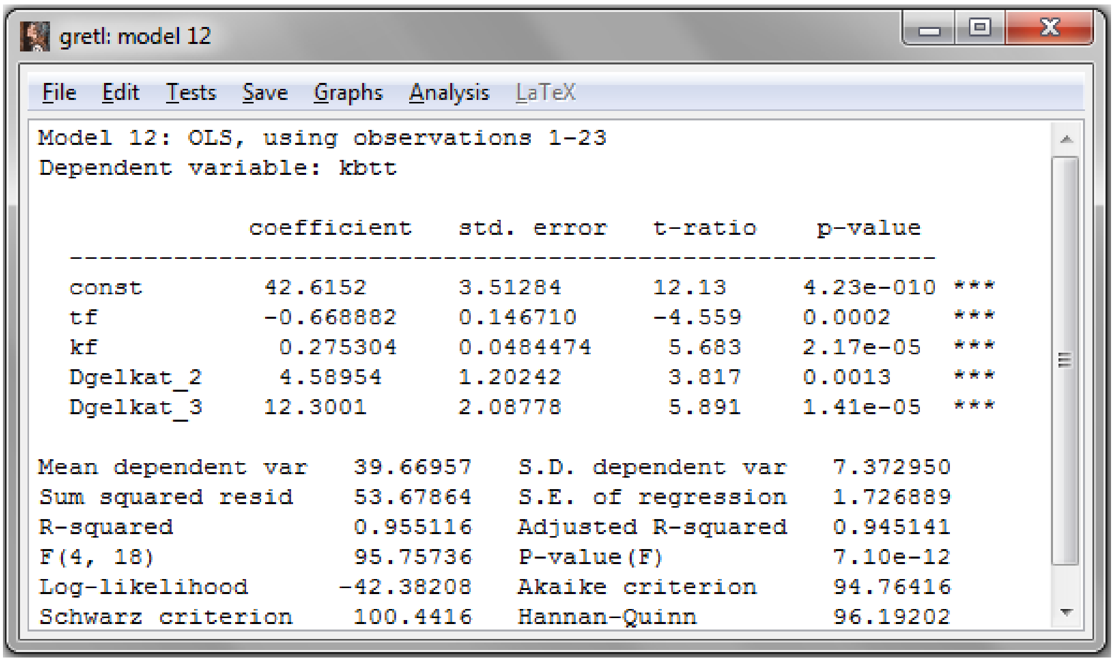
Bir öncekinden farklı olarak bu modelde gelkat değişkeninin etkisine bakacağız. Gelir kategorisinin her bir kademe yükselişinde tavuk eti tüketiminin 3.74 birim arttığını görüyoruz. Bir başka ifadeyle, 2. Gelir kategorisinde tüketiciler 1. Gelir kategorisine göre 5.77 birim daha fazla tavuk eti tüketmektedirler. Aynı şekilde 3. Gelir kategorisinde tüketiciler 2. Gelir kategorisine göre 5.77 birim daha fazla tavuk eti tüketmektedirler.

Şimdi de gelir kategorisi değişkenini kukla değişkenlerle ifade edip, farklı bir yorum açısı elde etmeye çalışalım. Bunun için ilk kategoriyi referans olarak alıp, 2. ve 3. Kategorilere kukla değişken tanımlaması yapacağız:

Bu durum için modelimiz

$$kbtt = b_0 + b_1 tf + b_2 kf + b_3 Dgelkat\_2 + b_4 Dgelkat\_3$$

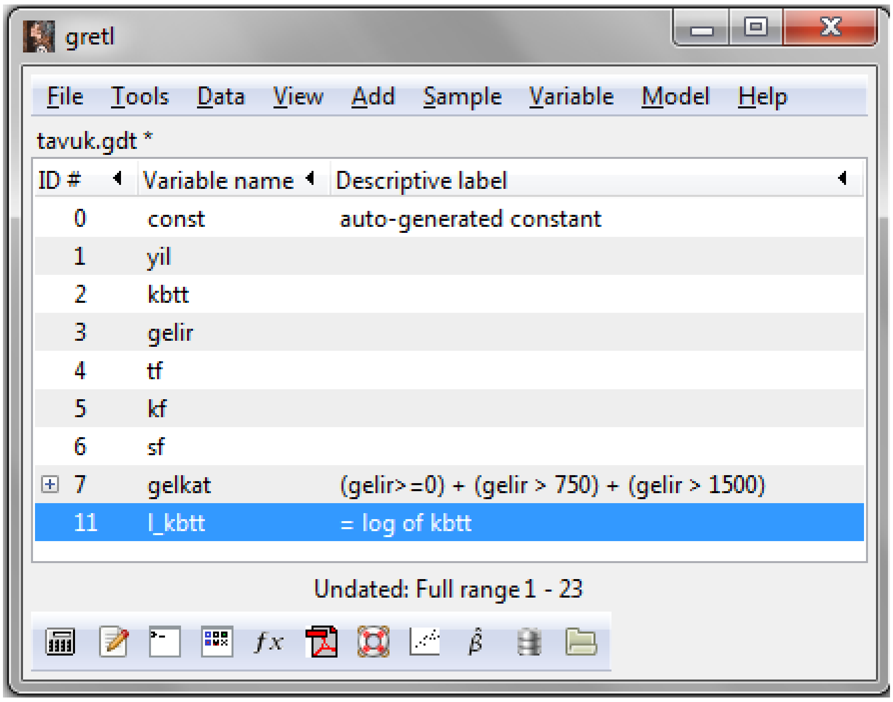
olur. Tahmin sonuçları:



```
gretl: model 12
File Edit Tests Save Graphs Analysis LaTeX
Model 12: OLS, using observations 1-23
Dependent variable: kbtt

               coefficient   std. error   t-ratio    p-value
  -----------------------------------------------------------
  const          42.6152     3.51284      12.13     4.23e-010 ***
  tf             -0.668882   0.146710     -4.559    0.0002    ***
  kf              0.275304   0.0484474     5.683    2.17e-05  ***
  Dgelkat_2       4.58954    1.20242       3.817    0.0013    ***
  Dgelkat_3      12.3001     2.08778       5.891    1.41e-05  ***

Mean dependent var   39.66957   S.D. dependent var   7.372950
Sum squared resid    53.67864   S.E. of regression   1.726889
R-squared            0.955116   Adjusted R-squared   0.945141
F(4, 18)             95.75736   P-value(F)           7.10e-12
Log-likelihood      -42.38208   Akaike criterion     94.76416
Schwarz criterion    100.4416   Hannan-Quinn         96.19202
```

Tahmin sonuçlarını sadece kukla değişkenler açısından değerlendirdiğimizde, 2. gelir kategorisinin referans kategori olan 1. Kategoriye göre 4.59 birim daha fazla tavuk eti tükettiğini, 3. gelir kategorisinin ise 1. Kategoriye göre 12.3 birim daha fazla tavuk eti tükettiğini söyleyebiliriz.

Şimdi aynı modelleri log-lin olarak tahmin edeceğiz. Değişken listemiz:



$$Ln(kbtt) = b_0 + b_1 tf + b_2 kf + b_3 gelkat$$

modelinin tahmin sonuçları:

```
gretl: model 13
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 13: OLS, using observations 1-23
Dependent variable: l_kbtt

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       3.61614      0.0839788    43.06     2.08e-020 ***
  tf         -0.0186410    0.00388079   -4.803    0.0001    ***
  kf          0.00756769   0.00131585    5.751    1.53e-05  ***
  gelkat      0.144910     0.0290421     4.990    8.14e-05  ***

Mean dependent var    3.663887   S.D. dependent var    0.187659
Sum squared resid     0.045528   S.E. of regression    0.048951
R-squared             0.941235   Adjusted R-squared    0.931957
F(3, 19)              101.4414   P-value(F)            7.12e-12
Log-likelihood        38.95102   Akaike criterion     -69.90203
Schwarz criterion    -65.36006   Hannan-Quinn         -68.75974

Log-likelihood for kbtt = -45.3184
```

gelkat değişkeni kesikli olduğundan, log-lin modellerde bu tür değişkenlerin tahmincileri yüzde olarak yorumlanabilmektedir. Buna göre gelir kategorileri arasında %14.5 tavuk eti tüketimi farkı bulunmaktadır. Bir başka ifadeyle, 2. gelir kategorisi 1. gelir kategorisinden %14.5 daha fazla tavuk eti tüketmektedir. Aynı şekilde 3. gelir kategorisi de 2. gelir kategorisinden %14.5 daha fazla tavuk eti tüketmektedir.

Log-lin modelimizi bu kez gelkat kuklalarıyla kuralım. Referans grup 1. gelir kategorisidir:

$$Ln(kbtt) = b_0 + b_1 tf + b_2 kf + b_3 Dgelkat\_2 + b_4 Dgelkat\_3$$

gretl: model 14

File Edit Tests Save Graphs Analysis LaTeX

```
Model 14: OLS, using observations 1-23
Dependent variable: l_kbtt

              coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const        3.79516      0.100323     37.83     1.31e-018 ***
  tf          -0.0199175    0.00418986   -4.754    0.0002    ***
  kf           0.00790337   0.00138360    5.712    2.04e-05  ***
  Dgelkat_2    0.129675     0.0343397     3.776    0.0014    ***
  Dgelkat_3    0.299506     0.0596245     5.023    8.83e-05  ***

Mean dependent var   3.663887   S.D. dependent var   0.187659
Sum squared resid    0.043781   S.E. of regression   0.049318
R-squared            0.943491   Adjusted R-squared   0.930933
F(4, 18)             75.13329   P-value(F)           5.58e-11
Log-likelihood       39.40111   Akaike criterion    -68.80221
Schwarz criterion   -63.12474   Hannan-Quinn        -67.37434

Log-likelihood for kbtt = -44.8683
```

Kukla değişkenler de kesikli olduğundan, log-lin tahmin sonuçlarında yer alan bu tür değişkenlerin katsayıları yüzde olarak yorumlanmaktadır. Tahmin sonuçları, 2. gelir kategorisinin, referans grup olan 1. gelir kategorisinden %12.96 daha fazla tavuk eti tükettiğini; 3. gelir kategorisinin ise, referans grup olan 1. gelir kategorisinden %29.95 daha fazla tavuk eti tükettiğini göstermektedir.

5.6.2. *Yapısal Değişimin Belirlenmesi (Chow Test)*

Bağımlı ve bağımsız değişkenler arasındaki ilişki yapısal değişime uğrayabilir. Yani ilişki belli bir dönemden diğerine değişim gösterebilir. Buna yapısal *istikrarın değişimi* veya *yapısal kesinti* adı da verilmektedir. Örneğin 24 Ocak 1980 kararlarının yapısal değişime yolaçtığı düşünülebilir. Eğer yapısal değişim söz konusuysa, 1980 öncesine göre, 1980 ve onu izleyen yılların ekonomik değişkenlerde farklılık meydana getirmesi gerekir. Bir zaman serisinde yapısal değişim olup olmadığını anlamak üzere Chow testi yapılabilir.

Chow test aşağıdaki süreçle gerçekleştirilir:

Adım 1

Yapısal değişimin gerçekleştiği yıl belirlenir ve veri seti bu yıl itibariyle ikiye bölünür. Yapısal değişim yılından önceki dönemler 1. grup (n1), yapısal değişim yılı ve izleyen yıllar 2. grubu oluşturur (n2=n-n1).

Adım 2

Aynı değişkenlerle 1. ve 2. grup için ayrı ayrı model tahminleri yapılır ve her iki modelin hata kareleri toplamları SSE_1 ve SSE_2 olarak adlandırılır.

Adım 3

$SSE_U = SSE_1 + SSE_2$ hesaplanır.

Adım 4

Aynı değişkenlerle tüm veri setini kullanarak model tahminlemesi yapılır ve bu modelin hata kareleri toplam SSE_R olarak adlandırılır.

Adım 5

$$F_c = \frac{(SSE_R - SSE_1 - SSE_2)/k}{(SSE_1 + SSE_2)/(n-2k)}$$ hesaplanır.

Adım 6

$F_{k,n-2k}$ tablo değeriyle F_c karşılaştırılıp:

H_o: Yapısal değişme yok

hipotezi test edilir.

Örnek 5-8

Tavuk eti talebini (KBTT) fiyat (TF) ve gelirle (GEL) ilişkilendirdiğimiz modelde yapısal değişikliğin başlangıç yılı olarak 1989'u alalım (Çizelge 4-1). Buna göre 1. grup veri seti 1977-1988, 2. grup ise 1989-1999 yıllarını kapsayacaktır.

Sıfır hipotezimiz: H_o: Yapısal değişme yok

Modelimiz: KBTT = $b_0 + b_1$ TF+ b_2 GEL

Grup 1 için tahmin edilen model:

	Katsayı	StHata	t	p
TF	-0.309728	0.206680	-1.498588	0.1682
GEL	0.028312	0.001910	14.82641	0.0000
Sabit	30.02743	8.312380	3.612375	0.0056
	R^2 = 0.962	Düz.R^2 = 0.954	SSE_1=7.697837	

Grup 2 için tahmin edilen model:

	Katsayı	StHata	t	p
TF	-0.133995	0.106029	-1.263762	0.2419
GEL	0.010611	0.001684	6.301230	0.0002
Sabit	37.04818	4.158456	8.909118	0.0000
	R^2 = 0.927	Düz.R^2 = 0.909	SSE_2=17.71	

Tüm veriler için tahmin edilen model:

	Katsayı	StHata	t	p
TF	-0.213592	0.121905	-1.752115	0.0951
GEL	0.014884	0.002193	6.785304	0.0000
Sabit	34.51561	3.855779	8.951655	0.0000
	R^2 = 0.911	Düz.R^2 = 0.902	SSE_2=106.65	

Buna göre:

SSE_1 =7.698 SSE_2 =17.71 SSE_R =106.65

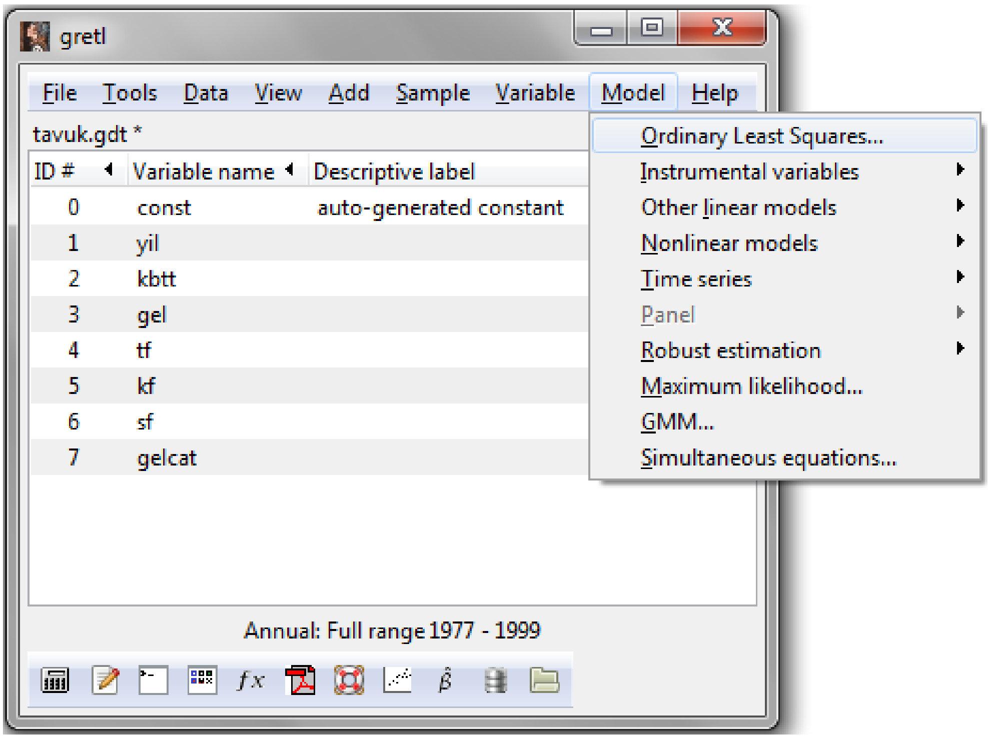
$$F_c = \frac{(106.65 - 7.698 - 17.71)/2}{(7.698 + 17.71)/(21-4)} = 27.18$$

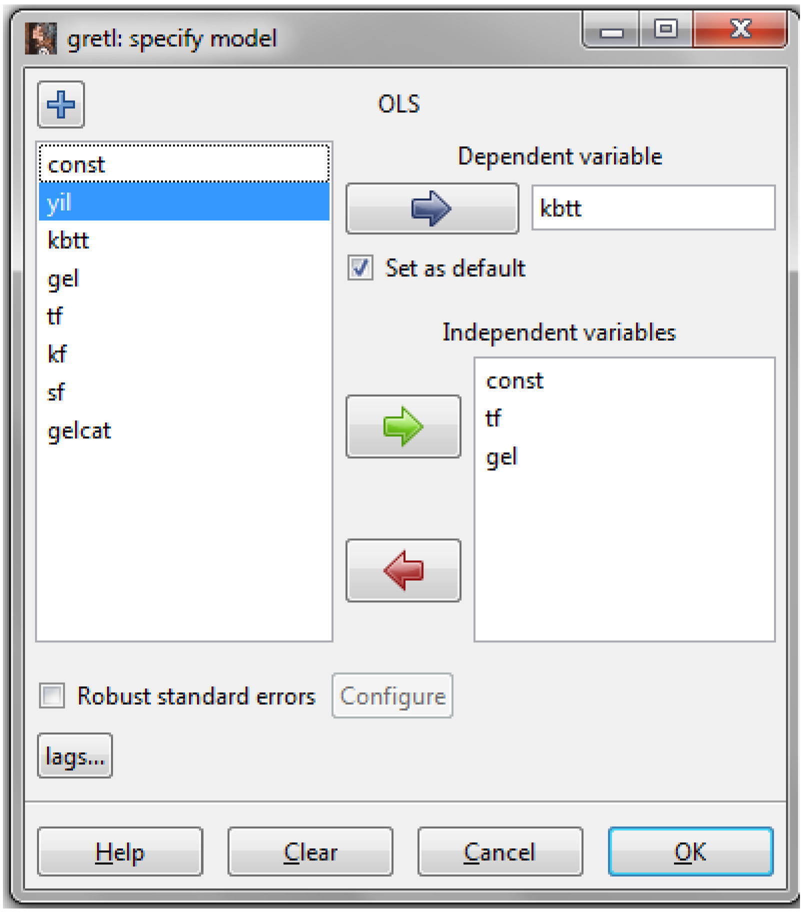
$F_{\alpha, k, n-2k} = F_{0.05, 2, 17} = 3.60$

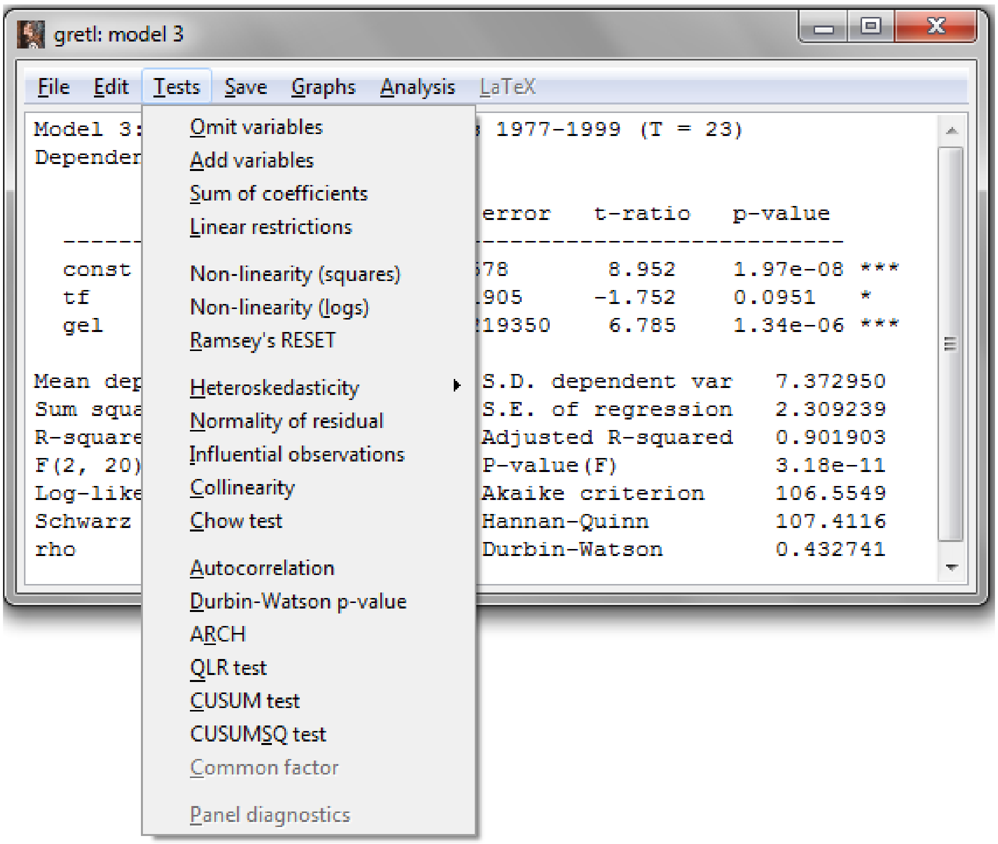
H_0: Yapısal değişme yok hipotezi reddedilir.

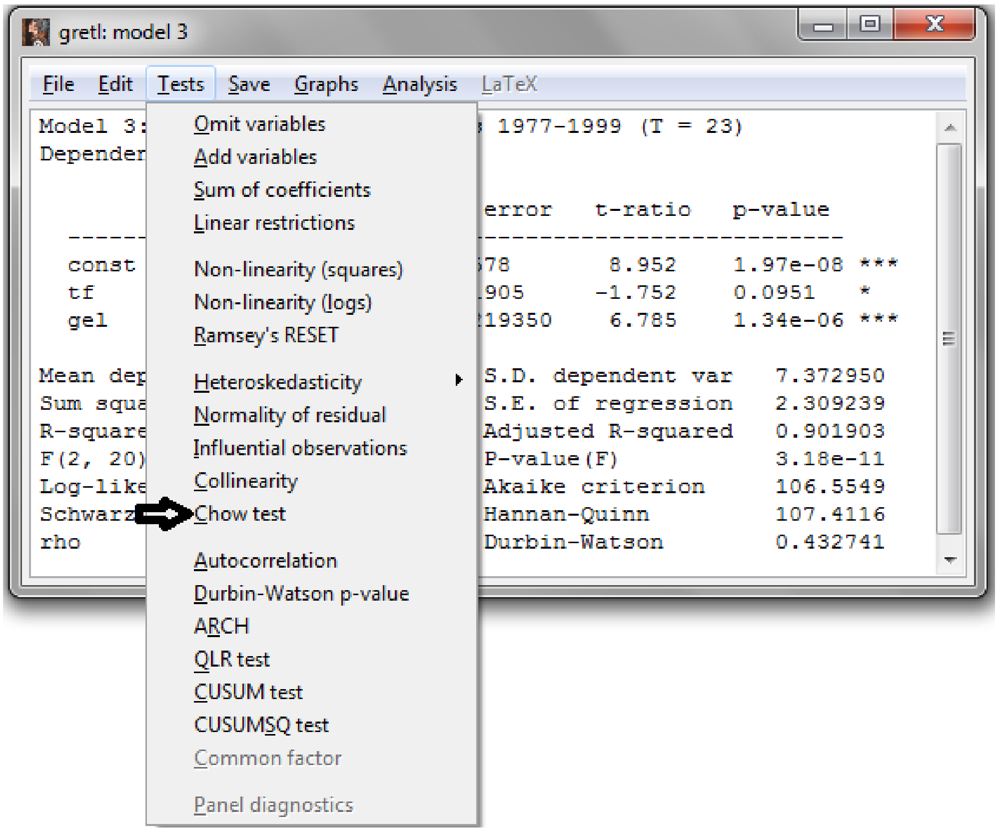
Görüldüğü gibi sıfır hipotezi reddedilmektedir. O halde 90 yılından itibaren bir yapısal değişiklik söz konusudur.

Chow testi Gretl yardımıyla yapabiliriz. Model tahminini yaptıktan sonra Tests menüsünü açtığımızda

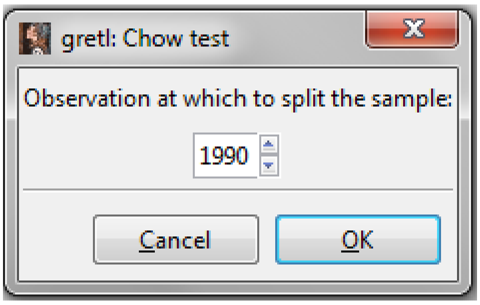



gretl: specify model
OLS
const
yil
kbtt
gel
tf
kf
sf
gelcat
Dependent variable
kbtt
Set as default
Independent variables
const
tf
gel
Robust standard errors
Configure
lags...
Help
Clear
Cancel
OK


gretl: model 3
File Edit Tests Save Graphs Analysis LaTeX
Omit variables
Add variables
Sum of coefficients
Linear restrictions
Non-linearity (squares)
Non-linearity (logs)
Ramsey's RESET
Heteroskedasticity
Normality of residual
Influential observations
Collinearity
Chow test
Autocorrelation
Durbin-Watson p-value
ARCH
QLR test
CUSUM test
CUSUMSQ test
Common factor
Panel diagnostics
Model 3:
1977-1999 (T = 23)
error t-ratio p-value
const 8.952 1.97e-08 ***
tf -1.752 0.0951 *
gel 6.785 1.34e-06 ***
S.D. dependent var 7.372950
S.E. of regression 2.309239
Adjusted R-squared 0.901903
P-value(F) 3.18e-11
Akaike criterion 106.5549
Hannan-Quinn 107.4116
Durbin-Watson 0.432741



Aşağıdaki ekranda kırılmanın hangi yılda olduğunu belirtiyoruz:



```
gretl: Chow test output

Augmented regression for Chow test
OLS, using observations 1977-1999 (T = 23)
Dependent variable: kbtt

              coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const       32.9692        10.9589        3.008     0.0079    ***
  tf          -0.363796       0.276175     -1.317     0.2052
  gel          0.0267205      0.00214594   12.45      5.71e-010 ***
  splitdum     1.28250       12.5230        0.1024    0.9196
  sd_tf        0.291055       0.309822      0.9394    0.3607
  sd_gel      -0.0166635      0.00275381   -6.051     1.30e-05  ***

Mean dependent var   39.66957   S.D. dependent var   7.372950
Sum squared resid    26.79319   S.E. of regression   1.255416
R-squared            0.977596   Adjusted R-squared   0.971007
F(5, 17)             148.3609   P-value(F)           2.11e-13
Log-likelihood      -34.39110   Akaike criterion     80.78220
Schwarz criterion    87.59517   Hannan-Quinn         82.49564
rho                  0.298917   Durbin-Watson        1.374828

Chow test for structural break at observation 1990
  F(3, 17) = 16.8898 with p-value 0.0000
```

5.6.3. Brown-Durbin-Evans CUSUM Yöntemi

Chow testinde, belli bir dönem ya da noktayı esas almak suretiyle, keskin bir değişim söz konusu olup olmadığı irdelenir. Kuşkusuz değişimin yavaş yavaş gerçekleştiği durumlar da olabilir. O nedenle değişimin yavaş olduğu ve bilinmeyen bir noktaya kadar devam ettiği durumları belirlemek üzere CUSUM yönteminden yararlanılır.

CUSUM yönteminin akışı aşağıda verilmiştir:

1. Hipotezleri hazırla:

   H_0: Parametreler istikrarlıdır
   H_1: Parametreler istikrarlı değildir

2. $Y_i = \beta X_i + \varepsilon_i$

   Modelini tahmin et ve hata terimlerini hesapla:

   $e_{t+1} = Y_{t+1} - X_{t+1} b^{(t)}$

3. Aşağıdaki hesaplamayı yap:

   $w_{t+1} = e_{t+1} / (1+X_{t+1}(X^{(t)\prime}X^{(t)})^{-1}X_{t+1}')^{½}$. hesaplamasını yap

4. W_t'nin t'ye göre (K+1'den N'e kadar) grafiğini çiz.
   Grafiğe: $[K \pm c(N-K)^{½}]$ and $[N, \pm 3c(N-K)^{½}]$
   güven aralığını ekle. Burada c çeşitli önem düzeyleri için aşağıdaki değerleri alır:

	Önem Düzeyi		
	1%	*5%*	*10%*
c	1.143	0.948	0.850

5. Eğer W_t, güven sınırlarının dışına çıkarsa, parametreler istikrarlıdır şeklindeki sıfır hipotezini reddet.

Örnek 5-9

Yukarıdaki örneği kullanarak (Çizelge 4-1):

Bağımlı değişken: kbtt

	Katsayı	StHata	t	p	
Sabit	34.5156	3.85578	8.9517	<0.00001	***
Gel	0.0148836	0.0021935	6.7853	<0.00001	***
Tf	-0.213592	0.121905	-1.7521	0.09507	*

R-kare	0.910821	Düzeltilmiş R-kare	0.901903
F(2, 20)	102.1340	P(F)	3.18e-11
Log-likelihood	-50.27744	Akaike kriteri	106.5549
Schwarz kriteri	109.9614	Hannan-Quinn	107.4116

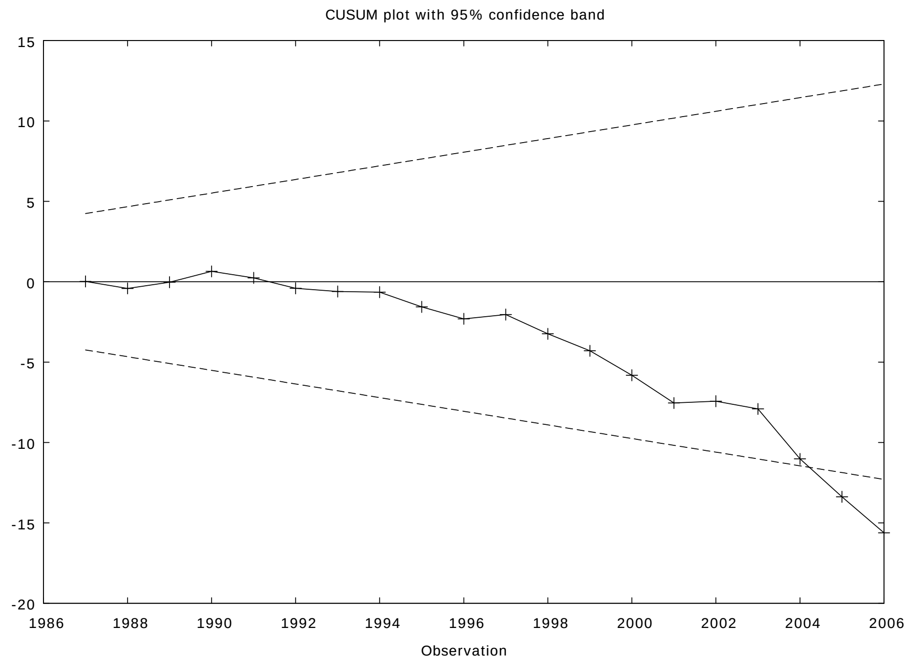
CUSUM parametre kararlılık testi

Ölçeklendirilmiş kalıntı ortalaması = -1.44399
sigmahat = 1.84889

Eklemeli ölçeklendirilmiş kalıntı toplamları (W_t değerleri)
('*' %95 güven aralığının dışına çıktığını gösterir)

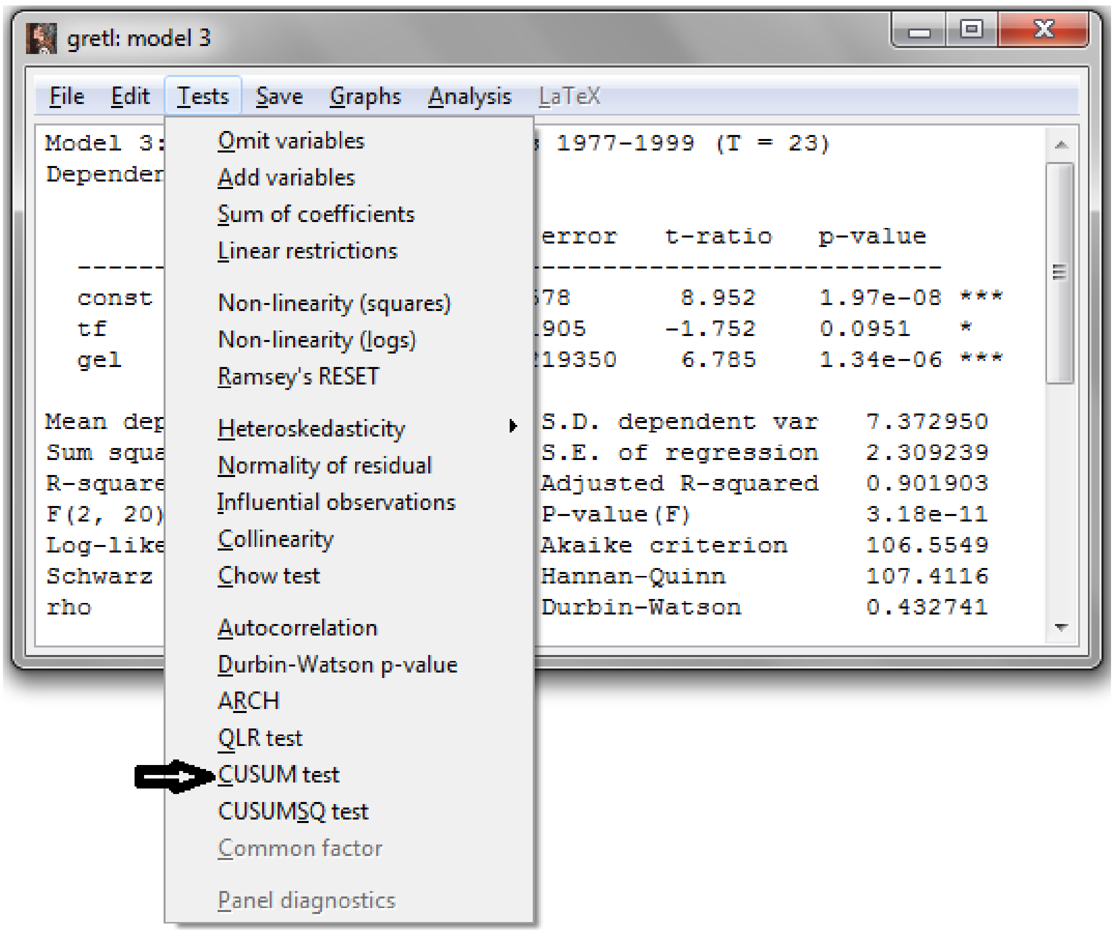
1987	0.020
1988	-0.417
1989	-0.032
1990	0.640
1991	0.239
1992	-0.409
1993	-0.606
1994	-0.648
1995	-1.565
1996	-2.308
1997	-2.046
1998	-3.243
1999	-4.299
2000	-5.818
2001	-7.547
2002	-7.433
2003	-7.914
2004	-11.021
2005	-13.376*
2006	-15.620*

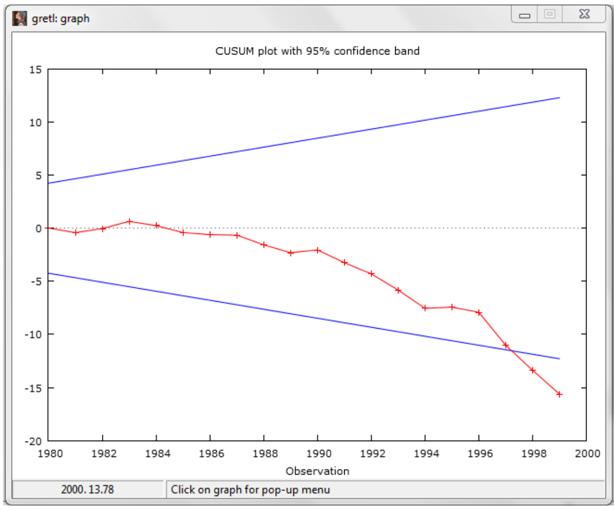
Harvey-Collier t(19) = -3.49277 with p 0.002435
CUSUM testi, sıfır hipotezinin reddiyle sonuçlanmaktadır.

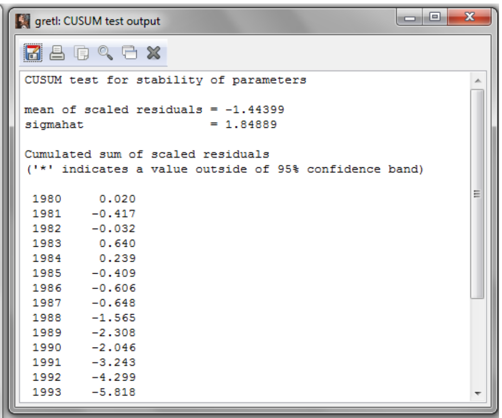


Grafiği incelediğimizde, güven sınırları dışına taşma olduğunu görebiliriz. Buna göre parametrelerde istikrarsızlık bulunmaktadır. Bir başka ifadeyle zaman içinde parametrelerde değişim meydana gelmiştir.

Bu testi Gretl'da model tahminini yaptıktan sonra aşağıdaki sırayla gerçekleştirebiliriz:







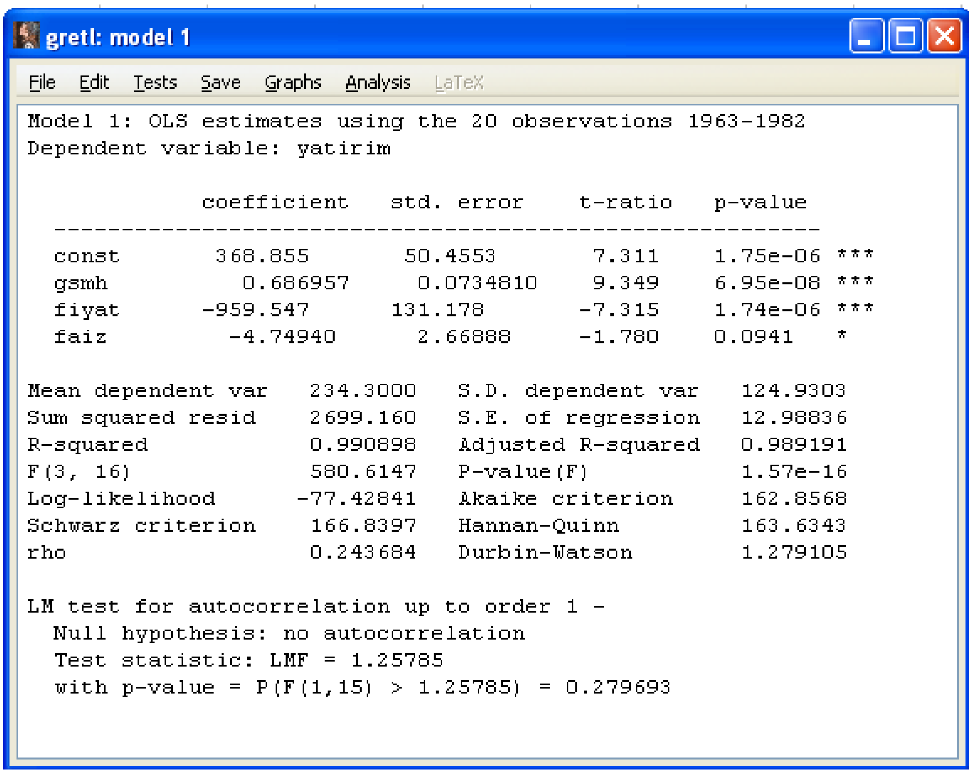
5.6.4. Quandt likelihood ratio (QLR) testi

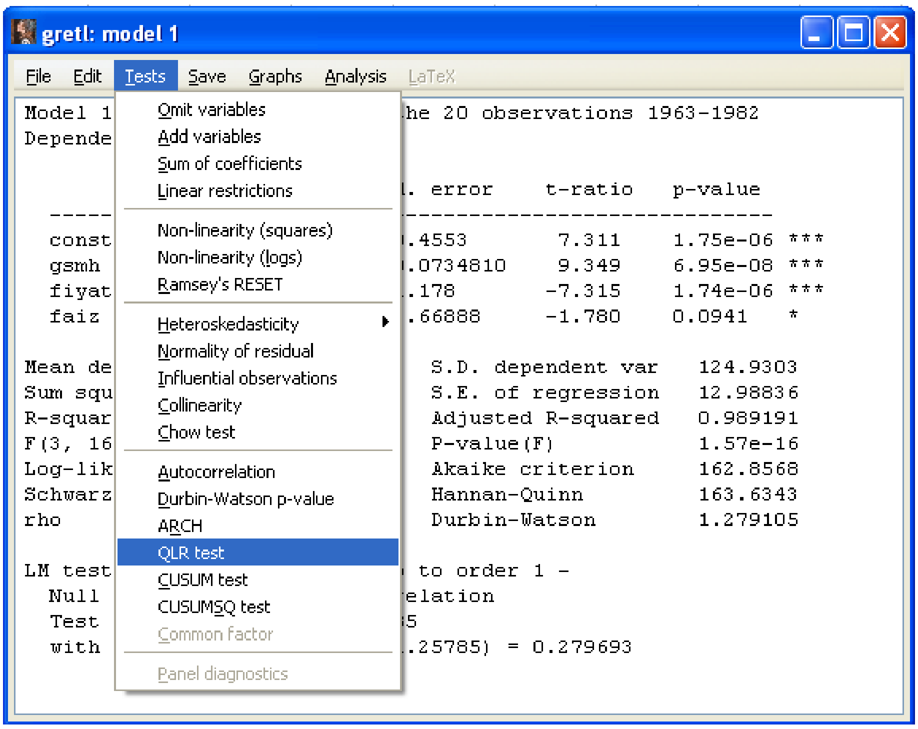
QLR testi, bir zaman serisi modelinde yapısal kırılmanın hangi dönemde gerçekleştiğini belirler. Chow testinde belli bir dönem için yapısal kırılmanın varlığı test edilmektedir. Eğer amaç belli bir yıla ilişkin analiz ve yorum yapmak ise Chow test yeterlidir. Ancak kırılmanın hangi yılda gerçekleştiği belirlenecekse QLR testine başvurulur. QLR testi ele alınan zaman aralığının başından ve sonundan %15'erlik bölümler çıkarılarak yapılır. Geriye kalan %70'lik bölümde her yıl için bir Chow testi yapılır ve bunlar arasından en yüksek F değerini veren dönem, kırılmanın olduğu dönem olarak belirlenir.

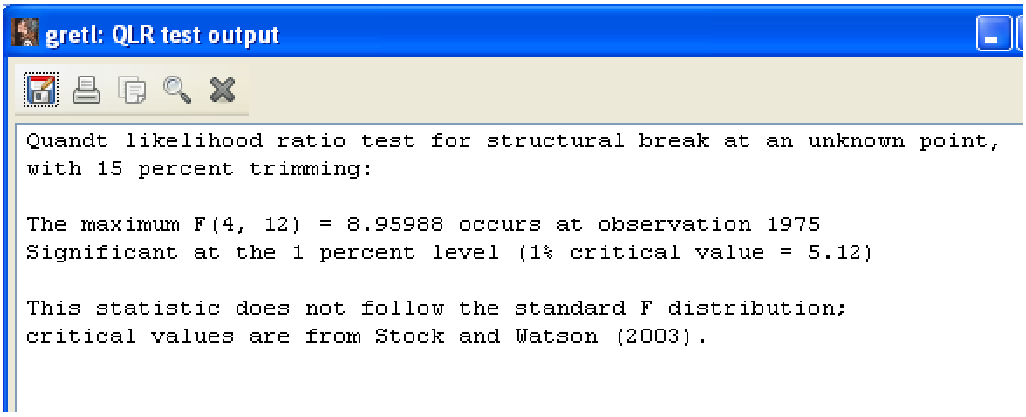
Örnek 5-10

Aşağıdaki yatırım problemine QLR testini uygulayalım:

Yıl	gsmh	yatirim	fiyat	faiz
1963	596.7	90.90	0.72	3.23
1964	637.7	97.40	0.73	3.55
1965	691.1	113.50	0.74	4.04
1966	756	125.70	0.77	4.5
1967	799.6	122.80	0.79	4.19
1968	873.4	133.30	0.83	5.16
1969	944	149.30	0.87	5.87
1970	992.7	144.20	0.91	5.95
1971	1077.6	166.40	0.96	4.88
1972	1185.9	195.00	1.00	4.5
1973	1326.4	229.80	1.06	6.44
1974	1434.2	228.70	1.15	7.83
1975	1549.2	206.10	1.26	6.25
1976	1718	257.90	1.32	5.5
1977	1918.3	324.10	1.40	5.46
1978	2163.9	386.60	1.50	7.46
1979	2417.8	423.00	1.63	10.28
1980	2631.7	401.90	1.78	11.77
1981	2954.1	474.90	1.95	13.42
1982	3073	414.50	2.07	11.02


gretl: model 1
File Edit Tests Save Graphs Analysis LaTeX
Model 1: OLS estimates using the 20 observations 1963-1982
Dependent variable: yatirim
coefficient std. error t-ratio p-value
const 368.855 50.4553 7.311 1.75e-06 ***
gsmh 0.686957 0.0734810 9.349 6.95e-08 ***
fiyat -959.547 131.178 -7.315 1.74e-06 ***
faiz -4.74940 2.66888 -1.780 0.0941 *
Mean dependent var 234.3000 S.D. dependent var 124.9303
Sum squared resid 2699.160 S.E. of regression 12.98836
R-squared 0.990898 Adjusted R-squared 0.989191
F(3, 16) 580.6147 P-value(F) 1.57e-16
Log-likelihood -77.42841 Akaike criterion 162.8568
Schwarz criterion 166.8397 Hannan-Quinn 163.6343
rho 0.243684 Durbin-Watson 1.279105
LM test for autocorrelation up to order 1 -
Null hypothesis: no autocorrelation
Test statistic: LMF = 1.25785
with p-value = P(F(1,15) > 1.25785) = 0.279693


gretl: model 1
File Edit Tests Save Graphs Analysis LaTeX
Omit variables
Add variables
Sum of coefficients
Linear restrictions
Non-linearity (squares)
Non-linearity (logs)
Ramsey's RESET
Heteroskedasticity
Normality of residual
Influential observations
Collinearity
Chow test
Autocorrelation
Durbin-Watson p-value
ARCH
QLR test
CUSUM test
CUSUMSQ test
Common factor
Panel diagnostics




```
Quandt likelihood ratio test for structural break at an unknown point,
with 15 percent trimming:

The maximum F(4, 12) = 8.95988 occurs at observation 1975
Significant at the 1 percent level (1% critical value = 5.12)

This statistic does not follow the standard F distribution;
critical values are from Stock and Watson (2003).
```

Görüldüğü gibi QLR testi 1975 yılında yapısal bir kırılma olduğunu göstermektedir.

5.6.5. *Vekil Değişkenler*

Ekonometrik modele dahil etmeyi planladığımız açıklayıcı değişkenlerden bazılarına ait verileri bulmamız mümkün olmayabilir. Bazen de birkaç değişkenin yerini tutabilecek tek bir değişkene ihtiyaç duyulabilir. Bu durumlarda o değişken ya da değişkenlerin yerini tutabilecek değişkenlerden yararlanabiliriz.

Vekil değişken kullanımının basit bir örneği, ailenin ortak görüşü alınmak istendiğinde, sadece erkek veya sadece kadının görüşünü değişken olarak modele dahil edilmesiyle ortaya çıkabilir.

Modelimiz:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + u$$

olsun.

Y değişkeninin X_2, ..., X_k değişkenleri tarafından açıklandığını varsayalım. Ancak X_2 değişkenine ait veri bulunmamaktadır. Bu durumda Y'nin X_3, ..., X_k ile regresyonu, yanlı tahmincilere ve geçersiz hipotez testlerine yol açacaktır. Bazan X_2 yerine onu temsil edecek bir vekil değişken kullanmak iyi bir çözüm olabilir. Vekil değişken, yerine geçeceği değişkenle doğrusal ilişkili olduğu kabul edilir. Burada Z'yi X_2'nin vekil değişkeni olarak kullanılacaktır.

$$X_2 = \lambda + \mu Z$$

Vekil değişken ilişkisi; teori, genel kabul veya deneyimle doğrulanmış olmalıdır. Bunu doğrudan belirlemek mümkün değildir zira X2 verisi yoktur. Eğer uygun bir vekil bulunabilirse, regresyon modeli aşağıdaki gibi yeniden yazılabilir:

$$\begin{aligned} Y &= \beta_1 + \beta_2(\lambda + \mu Z) + \beta_3 X_3 + \ldots + \beta_k X_k + u \\ &= (\beta_1 + \beta_2 \lambda) + \beta_2 \mu Z + \beta_3 X_3 + \ldots + \beta_k X_k + u \end{aligned}$$

Böylece tüm değişkenleri gözlemlenebilen bir model elde etmiş oluruz. Eğer vekil ilişkisi tek ve biz bu ilişkiyi tahmin edebiliyorsak, tahmin sonuçlarına güvenebiliriz. Doğru bir vekil değişken kullanımıyla:

- X_3, ..., X_k değişkenlerinin katsayıları, *Y'nin* X_2, ..., X_k ile regresyon ilişkisinden elde edileceklerle aynı olacaktır.
- X_3, ..., X_k değişkenlerinin katsayılarının standart hataları ve t değerleri, Y'nin X_2, ..., X_k ile regresyon ilişkisinden elde edileceklerle aynı olacaktır.

- R^2 Y'nin X_2, ..., X_k ile regresyon ilişkisinden elde edilecek olanla aynı olacaktır.
- Z'nin katsayısı b_2 μ'nün bir tahmini olacaktır. Eğer μ'yü bilmiyorsak, b_2'nin tahmincisini bilemeyiz.
- Bununla birlikte, X_2'nin katsayısını tahmin edemesek bile, Z'nin t değeri, olası X_2'nin t değeri ile aynı olacaktır.

b_1, revize modelimizde (b_1+b_2) şeklinde olduğundan tahmincisini elde edemeyiz. Ancak bizim için bu çok da önemli değildir. X_2 ile Z arasındaki ilişkinin tam olmaktan ziyade yaklaşık olması daha gerçekçidir. Eğer Z, X_2 için zayıf bir vekilse, ölçme hatası problemi gündeme gelir. Ayrıca diğer X değişkenleri X_2'nin vekili gibi davranmaya başlayacak ve gerekli değişkenlerin modele alınmamış olması hatasını ortaya çıkaracaktır.

Vekil değişken kullanımını, eğitim modeli örneği üzerinden inceleyeceğiz. Eğitim süresinin (ES); yetenek puanı (YP) ve aile eğitim altyapısı (INDEX) ilişkili olduğunu kabul edelim.

$$ES = \beta_1 + \beta_2 YP + \beta_3 INDEX + u$$

Yetenek puanının IQ testi ile veya benzeri bir yöntemle ölçülebileceğini düşünebiliriz. Ancak aile eğitim altyapısını farklı bir açıdan ele almamız gerekir. Aile eğitim altyapısını temsil etmek üzere bir vekil değişken bulmaya çalışalım. Hemen akla gelebilecek bir vekil değişken, annenin eğitim süresidir (AES). Kuşkusuz babanın eğitim süresi (BES) de vekil değişken olabilir; ya da her ikisinin birden dikkate alınabileceği bir indeks değişken geliştirebiliriz:

$$INDEX = \lambda + \mu_1 AES + \mu_2 BES$$

Böylece eğitim süresini; yetenek puanı (YP), anne eğitim düzeyi (AES) ve baba eğitim düzeyi (BES) ile ilişkilendirmiş oluruz:

$$\begin{aligned} ES &= \beta_1 + \beta_2 YP + \beta_3(\lambda + \mu_1 AES + \mu_2 BES) + u \\ &= (\beta_1 + \beta_3\lambda) + \beta_2 YP + \beta_3\mu_1 AES + \beta_3\mu_2 BES + u \end{aligned}$$

Bu denklemin tahmin sonuçları:

	Tahminci	StHata	t	p
Sabit	4.914654	0.506353	9.706	0.000
YP	0.129501	0.009954	13.009	0.000
AES	0.069403	0.042297	1.641	0.101
BES	0.110268	0.031195	3.535	0.000
	$R^2 = 0.37$	Düz.$R^2 = 0.3667$	F=110.83	

Şimdi eğitim süresi ile sadece yetenek puanı (YP) arasındaki regresyon analizi sonuçlarını görelim:

	Tahminci	StHata	t	p
Sabit	5.770845	0.466847	12.361	0.000
YP	0.154538	0.009156	16.879	0.000
	$R^2 = 0.334$	Düz.$R^2 = 0.3329$	F=284.89	

İki modelin regresyon sonuçlarını karşılaştırdığımızda, eğer aile eğitim altyapısını gözardı edersek, YP'nın yanlı olacağını ve daha yüksek tahminleneceğini görebiliriz. Bu bizim beklentimizdir. Eğitim süresi, anne ve babanın eğitimi ile ve YP ile pozitif ilişkilidir.

Şimdi başka vekil değişkenler üzerinde duralım. Örneğin ailenin kütüphane varlığını gösteren bir kukla değişken KUTUPHANE ve kardeş sayısı değişkenleri de vekil değişkenler olabilir. Kütüphane değişkeninin pozitif olması beklenir ancak istatistiki olarak anlamlı değildir. Kardeş sayısının (KARDESSAY) negatif işarete sahip olması beklenir. Model tahmini bu yönde olmakla birlikte, istatistiki olarak anlamlı değildir.

	Tahminci	StHata	t	p
Sabit	5.236995	0.566554	9.244	0.000
YP	0.127785	0.010054	12.71	0.000
AES	0.061998	0.042756	1.45	0.148
BES	0.104504	0.031493	3.318	0.001
KUTUPHANE	0.115127	0.196984	0.584	0.559
KARDESSAY	-0.05095	0.039956	-1.275	0.203
	R^2 = 0.37	Düz.R^2 = 0.3667	F=66.87	

6. REGRESYON HATALARI

6.1. Tahmincilerin Özellikleri

$\hat{b}_1$, $\hat{b}_2$ ve s^2 gibi EKK tahmincilerinin bazı ideal özelliklere sahip oldukları kabul edilir. Bu özellikler **Gauss-Markov Teoremi** diye bilinmektedirler.

Gauss-Markov Teoremi: EKK tahmincileri, sapmasız doğrusal tahminciler[1] sınıfında, minimum varyanslıdırlar, yani en iyidirler. Bu teorem, literatürde En İyi Doğrusal Sapmasız Tahminci (EDST) olarak bilinmektedir.

Bu teoremi $\hat{b}_2$'ni ele alarak açıklayalım: EKK $\hat{b}_2$ tahmincisinin β_2'nin EDST'si olabilmesi için şu şartları gerçekleştirmesi lazımdır:

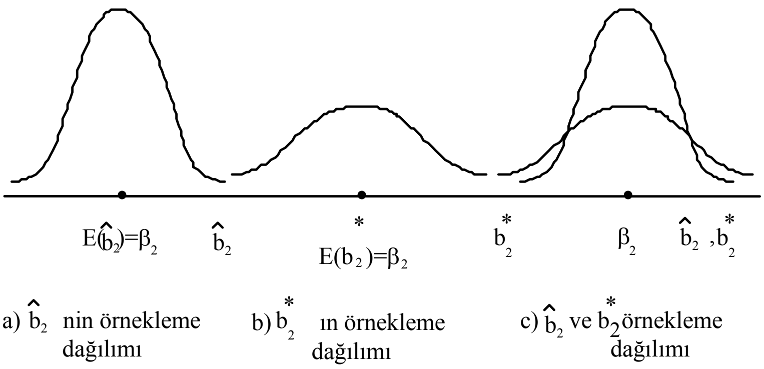
1. **Doğrusal** olmalıdır, regresyon modelindeki bir stokastik değişken olan Y'nin doğrusal fonksiyonu olmalıdır.

2. **Sapmasız** olmalıdır, yani ortalaması veya beklenen değeri E($\hat{b}_2$), gerçek β_2 değerine eşit olmalıdır: $E(\hat{b}_2)=\beta_2$

3. Doğrusal sapmasız tahminciler sınıfında **minimum varyanslı** olmalıdır; minimum varyanslı sapmasız bir tahminciye **etkin tahminci** denir.

Teorem ekonometride teorik ve uygulamalı açıdan önemlidir. Teoremin anlamını β_2 için Şekil 6-1'de görmek mümkündür.

Şekil 6-1.a.da EKK tahmincisi $\hat{b}_2$'nin örnekleme dağılımını görüyoruz, yani anakütleden çekilen tekrarlı örnekleme denemeleri sonucu $\hat{b}_2$'nin aldığı değerlerin dağılımı yer alıyor ve $\hat{b}_2$'nin ortalaması E($\hat{b}_2$) gerçek anakütle tahmincisi β_2'ye eşittir. $\hat{b}_2$'nin bu özelliğine β_2'nin **sapmasız tahmincisi** denir.



Şekil 6-1: EKKY tahmincisi $\hat{b}_2$ ile başka bir yöntem tahmincisi b_2^*'nin örnekleme dağılımları

Şekil 6-1.b'de EKKY dışında bir yöntemle (Ençok Benzerlik Yöntemi=EÇBY gibi) elde edilmiş bir β_2 tahmincisinin örnekleme dağılımı yer alıyor. b_2^*'ın da sapmasız olduğunu yani beklenen değerinin anakütle gerçek değerine eşit olduğunu farzedelim. Ayrıca $\hat{b}_2$ ve b_2^*'ın doğrusal tahminciler olduğuna yani bağımsız

---

[1] $\hat{b}_1$ ve $\hat{b}_2$'lerin değerlerine tahmin, bunları bulmak için kullanılan formüllere de tahminci dediğimizi hatırlarsak; $\hat{b}_1$ tahmini deyince $\hat{b}_1$'nin değerini, $\hat{b}_1$ tahmincisi denildiğinde bunu bulmak için kullandığımız formülü kastediyoruz.

değişken Y'nin doğrusal fonksiyonları olduğunu varsayalım. Bu durumda Şekil 6-1.c'den açıkça görüleceği gibi, $\hat{b}_2$ ve b_2^* nin ikisi de sapmasız dağılımlı olmakla beraber, (her ikisinin ortalaması da gerçek b_2 ye eşit), b_2^*'nin dağılması yani varyansı, $\hat{b}_2$ den daha fazladır. b_2^* daha geniş bir alana yayıldığından $\hat{b}_2$'ni b_2^* a tercih etmemiz gerekir, yani EDST tahminciyi ekonometrik yorumlarda dikkate alacağız. Aynı açıklamalar β_1 içinde geçerlidir.

Ekonometrik bir araştırmada gayemiz sadece $\hat{b}_1$ ve $\hat{b}_2$'nin değerlerini tahmin ise (buna nokta tahmin diyoruz), V1, V3 ve V4 varsayımları yeterlidir. Fakat ekonometride β_i'ler tahmin edildikten sonra bunlar için aralık tahminleri yapılır ve güvenilirlikleri test edilir. Aralık tahminleri ve hipotez testleri yapabilmemiz için anakütle hata terimi u_i'nin normal dağılımlı olduğu varsayımı (V2) da gereklidir. Zira bu varsayım sağlanmadığında dahi $\hat{b}_1$ ve $\hat{b}_2$ EKK tahmincileri EDST olmakla beraber, dağılımları u_i'nin dağılımı hakkında yapılan varsayımlara bağlı olduğundan, hipotez testlerinde u_i'nin dağılımının şekli önemlidir. İster büyük örnek ($n \geq 30$), isterse küçük örnekler ($n < 30$) için testlerde (Z ve t testi), örnek'in normal dağılım gösteren bir anakütleden alındığı varsayımına dayanılır. Ayrıca u_i'ler normal dağılıyorsa, $\hat{b}_1$ ve $\hat{b}_2$ de β_1 ve β_2 ortalamalı, $s^2(\hat{b}_1)$ ve $s^2(\hat{b}_2)$ varyanslı normal dağılır. Zira normal dağılan değişkenlerin herhangi bir doğrusal fonksiyonu da kaide gereği normal dağılır. Ayrıca

$(n-2), (s^2/\sigma^2)$

ifadesi u normal dağılıyorsa n-2 serbestlik derecesi ile χ^2 (ki-kare) dağılımı gösterir. Sonuç olarak u anakütle hata teriminin normal dağıldığı varsayımı sayesinde β_1, β_2 ve σ^2 hakkında, aralık tahminleri ve hipotez testleri yapmak mümkün olmaktadır. Diğer önemli bir sonuç da u'nun sıfır ortalamalı ve σ^2 varyanslı normal dağılması halinde, Y_i'de aşağıdaki ortalama ve varyanslı normal dağılır:

$E(Y_i) = \beta_1 + \beta_2 X_i$ 6-1

$Var(Y_i) = \sigma^2$ 6-2

Bu takdirde Y_i değişkenin dağılışını,

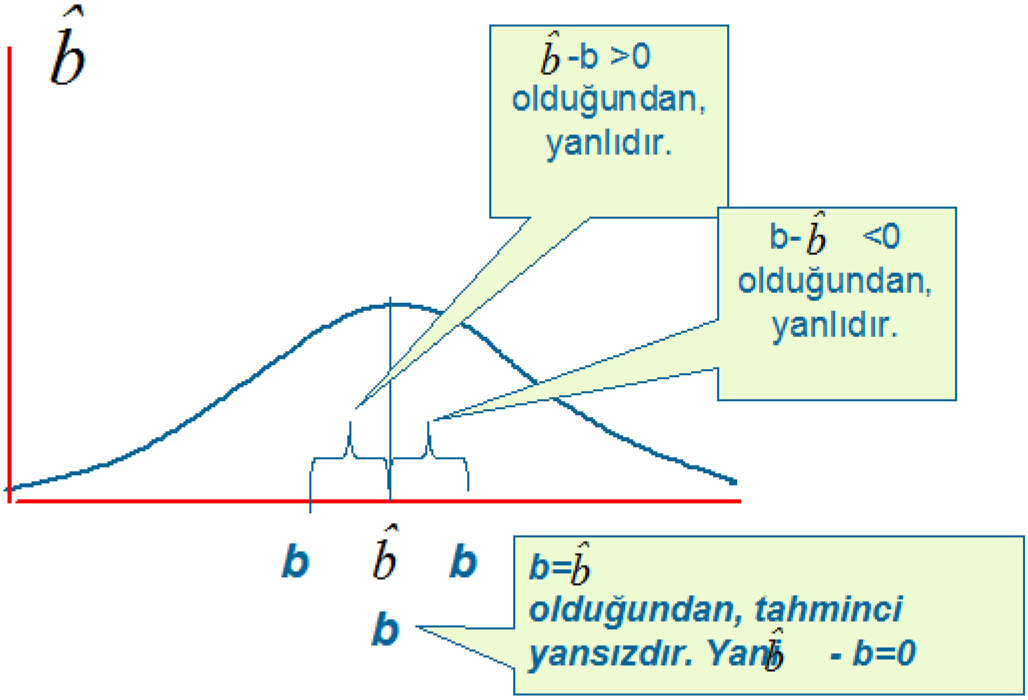
$Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$

şeklinde ifade edebiliriz.

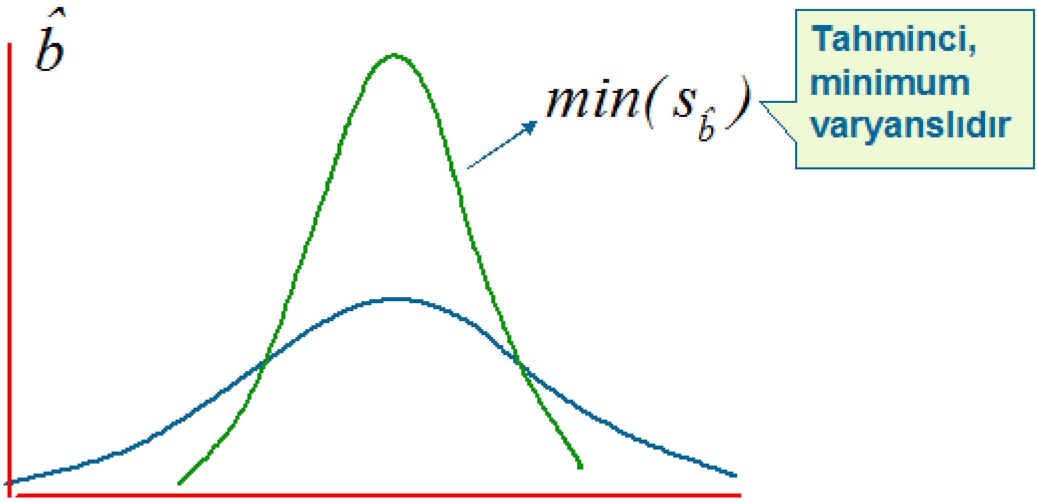
Buraya kadar EKK nokta tahmincilerinin sahip oldukları özellikler incelenmiştir. Nokta tahmincilerinin genel olarak küçük ve büyük örnekler için özellikleri konusunda EK. 2.6. ya müracaat edilebilir.

- Yansızlık (unbiasedness)
- Etkinlik (Efficiency)
- Tutarlılık (consistency)

6.1.1. Yansızlık



6.1.2. Etkinlik



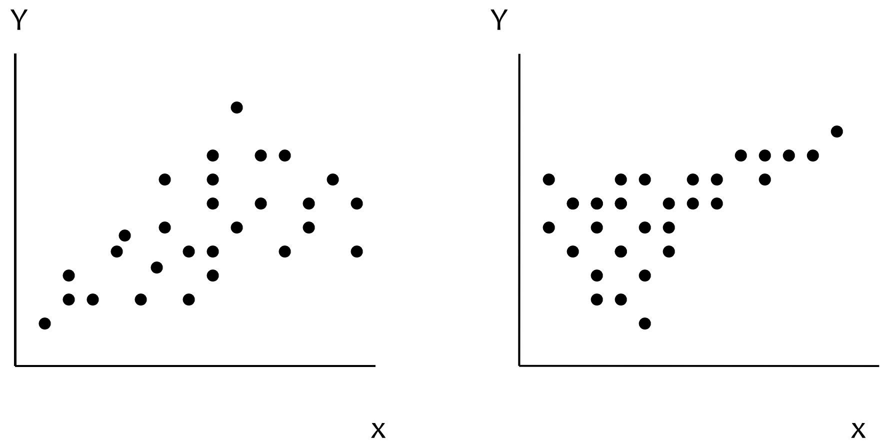
6.1.3. Tutarlılık

Gözlem sayısı arttıkça; $\hat{b}$, anakitle parametresi olan b'ye yaklaşır:

$$\lim_{n \to \infty} \hat{b} = b$$

6.2. Farklı Varyans

Farklı varyans problemine yatay kesit verilerinde rastlanır. Tüm seri boyunca varyansın aynı olmaması, farklı varyanslılığa yol açar. Bu problemin varolduğu örneklere ilişkin grafikler aşağıda verilmiştir:



Problemin doğası, farklı varyans probleminin nedeni olabilir. Örneğin tüketim harcamaları ve gelir ilişkilerinde, düşük ve yüksek gelir düzeylerinde tüketim harcamalarının gelir içindeki pay önemli derecede farklıdır. Deneme yanılma süreçlerinde de farklı varyans ortaya çıkabilir. Gerçekten bilgisayar klavyesi kullanmayı öğrenen bir kişinin başlangıçta yaptığı hata sayısı, ilerleyen zamanlarda giderek düşer. Bazı veri setlerinde uç (çok küçük veya çok büyük) değerlerin bulunması da yine farklı varyanslılığın nedenlerindendir.

6.2.1. Farklı Varyansın Sonuçları

Ekonometrik bir modelde farklı varyans problemi söz konusu ise:

- Modelin standart hatası daha yüksektir
- Tahmincilerin standart hataları daha yüksektir
- Tahmincilerin t değerleri daha düşüktür
- Modelin F değeri daha küçüktür
- R^2 daha düşüktür
- Tahminciler ve model için yapılacak hipotez testlerinin sonuçlarına güvenilmez

Bunlara bağlı olarak farklı varyans problemi olan modellerde:

- Tahminciler yansız ve tutarlıdır
- Tahminciler etkin değildir
- Tahmin veya kestirimler etkin değildir

Farklı varyans problemiyle karşılaşıldığında, hipotez testlerine güvenilmemesi gerekir. Ancak, test sonucu tahminci veya modelin anlamlı olduğunu gösteriyorsa, güvenle kullanılabilirler. Burada sorun, istatistiki açıdan anlamlı bulunmayanlarla ilgilidir. Yani, aslında istatistiki olarak anlamlı olup, farklı varyanstan dolayı anlamsız sonucu elde edilmiş tahminci veya modeller olabilir. Farklı varyans probleminin giderilmesi de, bu nedenle gereklidir.

6.2.2. Farklı Varyansı Belirleme Yöntemleri

6.2.2.1. Goldfeld Quandt Testi

Farklı varyans probleminin varlığını belirlemede kullanılan yöntemlerden biri, Goldfeld-Quandt (GQ) testidir. Bu test aşağıdaki süreçle gerçekleştirilir:

Veriler, farklı varyansa neden olduğu düşünülen değişkene göre büyükten küçükten sıralanır.

Veri seti ortadan ikiye bölünür ve gözlem sayısı tek ise ortada kalan gözlem (c) çıkartılır. Böylece n büyüklüğündeki bir veri setinden, ½(n-c)=m veriye sahip iki veri seti elde edilir.

İki veri seti için ayrı ayrı EKK tahminlemesi yapılır.

Eğer hatalar aynı varyanslı (homoskedastic) ise, bu iki veri setinin hata varyansları birbirine eşit olacaktır. Hipotezlerimiz:

H_o: Farklı varyans yok

H_l: Farklı varyans var

Sıfır hipotezine göre iki hata varyansının birbirine oranı 1 olacaktır. GQ testi, iki EKK denkleminden elde edilen hata kareleri toplamlarını birbirine oranlayıp, 1'e eşitliğini irdeler.

$GQ = SSE_2 / SSE_1$

Son olarak, F testi yapılır. F tablo değeri:

$F_{m, m, \alpha}$

şeklindedir.

$GQ > F_{m, m, \alpha}$ ise H_o reddedilir.

Örnek 6-1

Aşağıda, hanehalkı anketlerinden elde edilmiş tüketim harcamaları ve gelir verileri sunulmuştur. GQ testi ile farklı varyans lup olmadığını belirleyelim.

Gıda Harcaması	Gelir	Gıda Harcaması	Gelir
24.8	80	14.4	40
16.8	60	8	20
19.2	60	14	40
24	80	16.8	60
8	20	18	60
12.8	40	8	20
20	60	19.2	80
12	40	22.8	80
20	80	7.2	20
14	40	8.4	20

Çözüm:

Modelimizi, gıda harcaması gelirin bir fonksiyonudur şeklinde tanımlayalım. Gelir, farklı varyansa yolaçabilecek potansiyel değişkendir. Şimdi GQ testinin aşamalarını sırasıyla gerçekleştirelim.

Verileri gelire göre büyükten küçüğe sıralayalım.

Veri seti I		Veri seti II	
Gıda Harcaması	Gelir	Gıda Harcaması	Gelir
19.2	83	12	43
24.8	81	14.4	43
24	79	12.8	42
20	78	14	40
22.8	77	14	38
20	62	8	22
19.2	61	8	21
16.8	60	8	20
18	60	7.2	19
16.8	58	8.4	18

Veri sayısı çift olduğundan 10'arlı iki veri seti elde ederiz: Veri seti I ve veri seti II.

Veri seti I ve II için EKK tahminleri:

EKK I

	Tahminci	*StHata*	*t*	*p*
Sabit	6.095	4.596	1.33	0.221
Gelir	0.20121	0.06511	3.09	0.015
S = 2.031	$R^2 = 0.823$	SSE_1=32.995	F=9.55	p=0.015

EKK II

	Tahminci	*StHata*	*t*	*p*
Sabit	3.0135	0.9928	3.04	0.016
Gelir	0.25054	0.03061	8.18	0.000
S = 1.039	$R^2 = 0.893$	SSE_2=8.641	F=66.97	p=0.000

$GQ = SSE_2 / SSE_1$
$GQ = 32.995 / 8.641$
$GQ = 3.818$

H_0: Farklı varyans yok
H_1: Farklı varyans yok

$F_{m, m, \alpha} = F_{8, 8, 0.05} = 3.44$
$GQ = 3.818 > F_{8, 8, 0.05} = 3.44$

olduğundan H_0 reddedilir. Buna göre farklı varyans vardır.

6.2.2.2.White Testi

White testi farklı varyansın varlığını blirlemede kullanılan yaygın yöntemlerden biridir. Bu testi, iki değişkenli bir regresyonda aşağıdaki süreçle uygulayabiliriz:

White testi farklı varyansın varlığını blirlemede kullanılan yaygın yöntemlerden biridir. Bu testi, iki değişkenli bir regresyonda aşağıdaki süreçle uygulayabiliriz:

Orijinal denklem tahmin edilir.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e \qquad (1)$$

Tahmin edilen regresyon denkleminin (1) hata değerleri (e) elde edilir.

$$e = Y - (b_0 + b_1X_1 + b_2X_2 + b_3X_3) \qquad (2)$$

Hata değerlerinin karesi bağımlı değişken olarak alınıp, orijinal denklemdeki bağımsız değişkenlerin kendileri, kareleri ve her bir bağımsız değişkenin diğer bağımsız değişkenlerle çarpımı açıklayıcı değişkenleri listesi oluşturulur:

$$e^2 = b_0+b_1X_1+b_2X_2+b_3X_3+b_4X_1^2+b_5X_2^2+b_6X_3^2+ b_7X_1X_2+b_8X_1X_3+b_9X_2X_3+u \qquad (3)$$

(3) no'lu denklemi bütün olarak test etmek üzere ki kare testi yapılır. Test istatistiği, nR^2 dir. Burada n gözlem sayısı, R^2 belirleme katsayısıdır. Test istatistiği, (3) no'lu denklemdeki bağımsız değişken sayısı kadar serbestlik derecesine göre ki kare tablo değeriyle karşılaştırılır. Hipotezlerimiz:

H_0: Farklı varyans yok

H_1: Farklı varyans var

Eğer: $nR^2 > \chi^2$ ise H_0 reddedilir.

Örnek 6-2

Aşağıda, hanehalkı anketlerinden elde edilmiş tüketim harcamaları ve gelir verileri sunulmuştur. Şimdi White testi ile farklı varyans olup olmadığını belirleyelim.

Tüketim Harcamaları Y	Harcanabilir Gelir X1	Varlık X2
11	12	47
20	12	86
13	14	52
14	15	58
15	16	63
16	17	67
17	18	70
17	18	70
17	19	76
18	19	76
18	20	79
20	23	90
32	36	144
33	37	149
41	50	204

Orijinal denklem:

Bağımlı değişken: y

	Tahminci	StHata	t	p	
Sabit	2.93061	0.475437	6.1640	0.00005	***
x1	-0.0302937	0.0831927	-0.3641	0.72209	
x2	0.20129	0.0207403	9.7052	<0.00001	***

Mean dependent var	20.13333	S.D. dependent var	8.433493
Sum kare resid	7.392754	S.E. of regression	0.784897
R-kare	0.992576	Düzeltilmiş R-kare	0.991338
F(2, 12)	802.1426	P(F)	1.67e-13
Log-likelihood	-15.97745	Akaike kriteri	37.95491
Schwarz kriteri	40.07906	Hannan-Quinn	37.93228

Tahmin edilen regresyon denkleminin hata değerleri:

-1.03	0.12	0.03	-0.15	-0.13	0.09	0.52	0.52	-0.65	0.35	-0.22	-0.35	1.17	1.19	-1.48

Hata değerlerinin karesi bağımlı değişken olarak alınıp, orijinal denklemdeki bağımsız değişkenlerin kendileri, kareleri ve her bir bağımsız değişkenin diğer bağımsız değişkenlerle çarpımı açıklayıcı değişkenleri listesi oluşturulur:

$$e^2 \;=\; b_o + b_1X_1 + b_2X_2 + b_3X_1^2 + b_4X_2^2 + b_5X_1X_2 + u$$

	Katsayı	St.Hata	t	p	
sabit	4.49701	1.18755	3.787	0.0043	***
X1	-2.9083	0.770638	-3.774	0.0044	
X2	0.629779	0.169643	3.712	0.0048	***
sq_x1	0.02385	0.031606	0.7546	0.4698	
X2_X3	0.024581	0.013778	1.784	0.1081	
sq_x2	-0.0071	0.002212	-3.212	0.0106	

Hipotezlerimiz:

H_o: Farklı varyans yok

H_1: Farklı varyans var

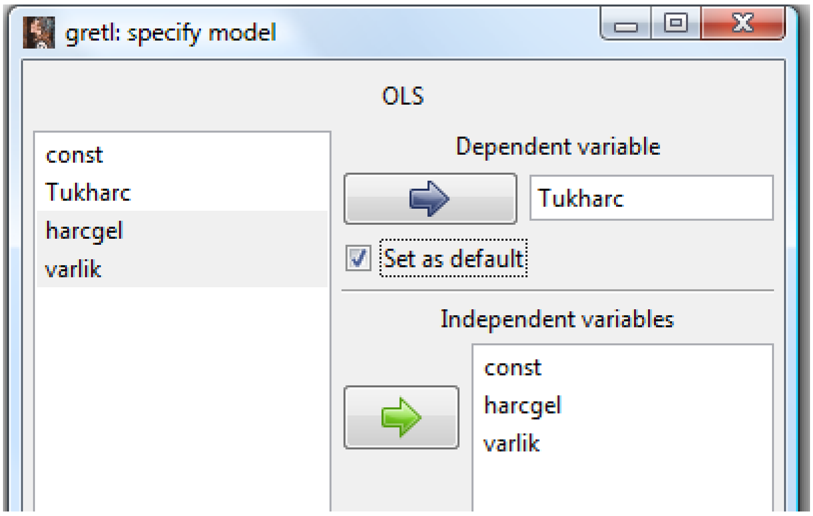
Test istatistiği:

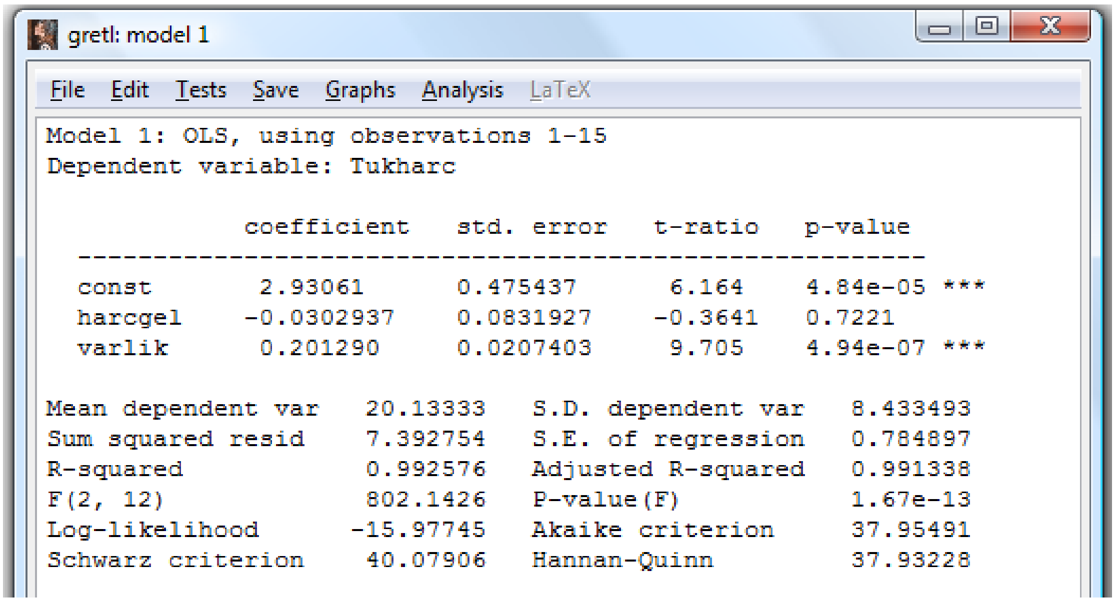
$LM = nR^2 = 15(0.923) = 13.845$

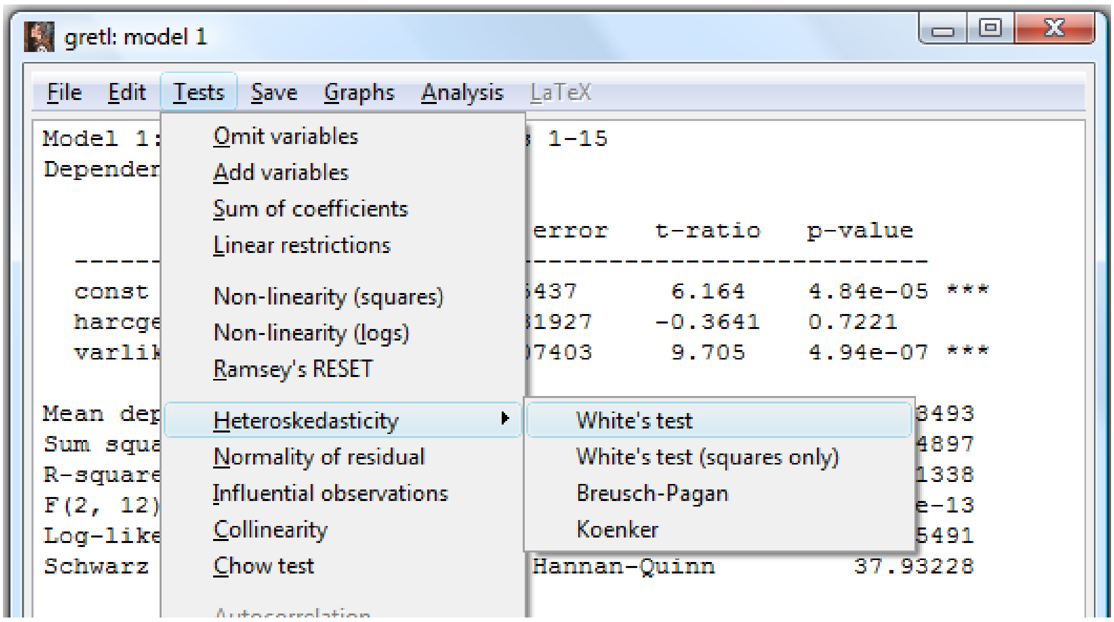
$X^2_{(15,\ 0.05)} = 25$

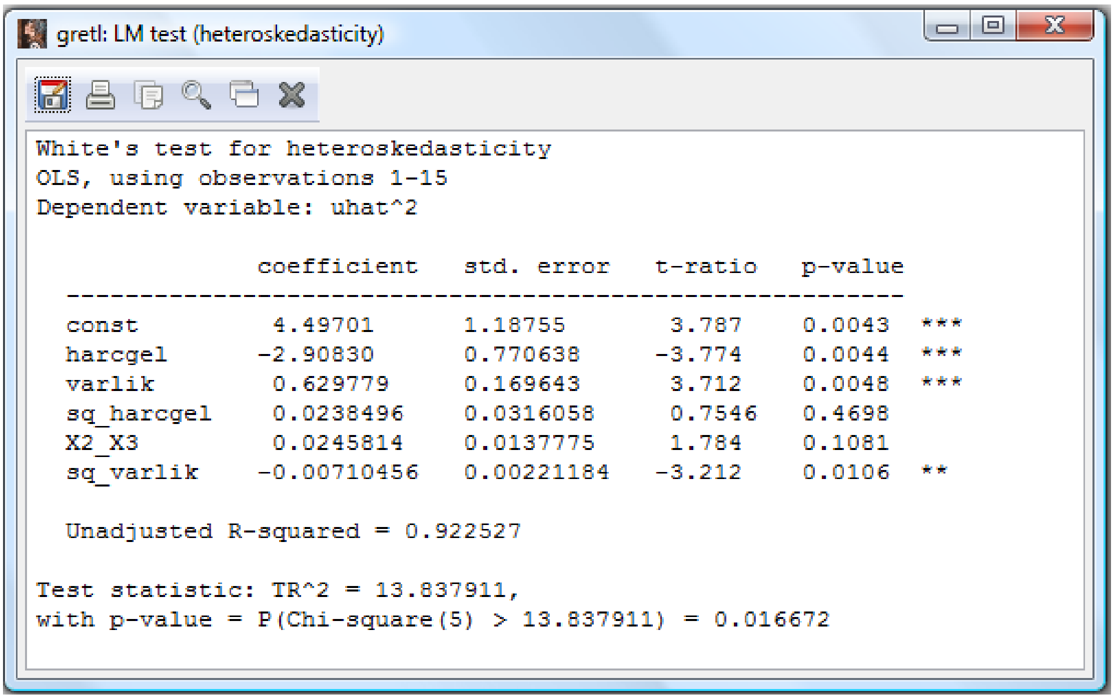
P=0.0166725

$nR^2 = 13.845 < \chi^2 = 25$ olduğundan H_o reddedilmez. Yani farklı varyans problemi yoktur.


gretl: specify model
OLS
const
Tukharc
harcgel
varlik
Dependent variable
Tukharc
Set as default
Independent variables
const
harcgel
varlik


gretl: model 1
File Edit Tests Save Graphs Analysis LaTeX
Model 1: OLS, using observations 1-15
Dependent variable: Tukharc
coefficient std. error t-ratio p-value
const 2.93061 0.475437 6.164 4.84e-05 ***
harcgel -0.0302937 0.0831927 -0.3641 0.7221
varlik 0.201290 0.0207403 9.705 4.94e-07 ***
Mean dependent var 20.13333 S.D. dependent var 8.433493
Sum squared resid 7.392754 S.E. of regression 0.784897
R-squared 0.992576 Adjusted R-squared 0.991338
F(2, 12) 802.1426 P-value(F) 1.67e-13
Log-likelihood -15.97745 Akaike criterion 37.95491
Schwarz criterion 40.07906 Hannan-Quinn 37.93228


gretl: model 1
File Edit Tests Save Graphs Analysis LaTeX
Omit variables
Add variables
Sum of coefficients
Linear restrictions
Non-linearity (squares)
Non-linearity (logs)
Ramsey's RESET
Heteroskedasticity
Normality of residual
Influential observations
Collinearity
Chow test
White's test
White's test (squares only)
Breusch-Pagan
Koenker



gretl: LM test (heteroskedasticity)

```
White's test for heteroskedasticity
OLS, using observations 1-15
Dependent variable: uhat^2

               coefficient    std. error    t-ratio    p-value
  -------------------------------------------------------------
  const          4.49701      1.18755        3.787     0.0043   ***
  harcgel       -2.90830      0.770638      -3.774     0.0044   ***
  varlik         0.629779     0.169643       3.712     0.0048   ***
  sq_harcgel     0.0238496    0.0316058      0.7546    0.4698
  X2_X3          0.0245814    0.0137775      1.784     0.1081
  sq_varlik     -0.00710456   0.00221184    -3.212     0.0106   **

  Unadjusted R-squared = 0.922527

Test statistic: TR^2 = 13.837911,
with p-value = P(Chi-square(5) > 13.837911) = 0.016672
```

6.2.3. Farklı Varyansı Giderme Yolları

6.2.3.1. Ağırlıklı EKK Yöntemi

Eğer farklı varyansın varlığı tesbit edildiyse, bunu gidermenin yollarından biri, ağırlıklı en küçük kareler yöntemini kullanmaktır. Böylece EKK'ye göre daha etkin tahminciler elde edilecektir.

Farklı varyansa yolaçtığı düşünülen bir değişkene göre AEKK

AEKK yönteminde, modelde farklı varyansa yol açtığı düşünülen değişkene göre ağırlıklandırma söz konusudur. Birden fazla açıklaycı değişken bulunan modellerde, hangi değişkenin farklı varyansa yol açtığını belirlemek, modelcinin görevidir. Genellikle gelir, kapasite, üretim hacmi gibi değişkenlerin farklı varyansın nedeni olduğu düşünülür.

Modelimiz:

$$Y = a + bX$$

olsun. Tek açıklayıcı değişkenimiz olduğuna göre, ağırlık değişkenimiz X'tir. O halde modelimizi X'e göre ağırlıklandıralım:

$$Y/X = a/X + bX/X$$

$$Y/X = a(1/X) + b$$

AEKK tahminin yapabilmek için kullanacağımız değişkenler, Y/X ve 1/X şeklindedir. Burada Y/X bağımlı değişken, 1/X ise bağımsız değişkendir. Bu modeli EKK ile tahmin ettikten sonra, denklemin her iki tarafını X ile çarparak bağlangıç denklemine döndüreceğiz.

Örnek 6-3

Bu örnekte ortalama maaş ile çalışan sayısı arasındaki ilişkiyi inceleyeceğiz.

Çalışan sayısı	Ortalama Maaş	Y/X	1/X	Çalışan sayısı	Ortalama Maaş	Y/X	1/X

X	Y			X	Y		
100	5.2	0.052	0.01	300	10.3	0.034333	0.003333
100	5.1	0.051	0.01	300	10.3	0.034333	0.003333
100	5.3	0.053	0.01	300	10.5	0.035	0.003333
100	5.2	0.052	0.01	400	10.3	0.02575	0.0025
100	5.4	0.054	0.01	400	13.7	0.03425	0.0025
100	5.5	0.055	0.01	400	13.1	0.03275	0.0025
200	7.1	0.0355	0.005	400	13.2	0.033	0.0025
200	7.2	0.036	0.005	400	13.2	0.033	0.0025
200	7.4	0.037	0.005	400	13.9	0.03475	0.0025
200	7.3	0.0365	0.005	500	16.1	0.0322	0.002
200	7.1	0.0355	0.005	500	18.3	0.0366	0.002
200	7.1	0.0355	0.005	500	22.1	0.0442	0.002
300	9.5	0.031667	0.003333	500	19.1	0.0382	0.002
300	9.8	0.032667	0.003333	500	27.6	0.0552	0.002
300	9.9	0.033	0.003333	500	13.1	0.0262	0.002

Önce Y = a+ bX için EKK çözümünü alalım:

Bağımlı değişken: Y

	Tahminci	*StHata*	*t*	*p*	
Sabit	0.793333	1.0791	0.7352	0.46834	
X	0.0339	0.0032536	10.4192	<0.00001	***

R-kare	0.794962	Düzeltilmiş R-kare	0.787639
F(1, 28)	108.5601	P(F)	3.86e-11
Log-likelihood	-69.26377	Akaike kriteri	142.5275
Schwarz kriteri	145.3299	Hannan-Quinn	143.4240

Farklı varyans probleminin var olup olmadığını White testi ile belirleyelim:

Test istatistiği: LM = 7.10139
p = P(Chi-Square(2) > 7.10139) = 0.0287046

Görüldüğü gibi modelimiz farklı varyans göstermektedir. O halde AEKK'yı uygulayabiliriz:

Bağımlı değişken: Y/X

	Tahminci	*StHata*	*T*	*p*	
Sabit	0.0283415	0.00197537	14.3474	<0.00001	***
1/X	2.26172	0.365108	6.1947	<0.00001	***

R-kare	0.578147	Düzeltilmiş R-kare	0.563081
F(1, 28)	38.37385	P(F)	1.09e-06
Log-likelihood	112.9531	Akaike kriteri	-221.9062

Schwarz kriteri	-219.1038	Hannan-Quinn	-221.0097

Farklı varyans probleminin var olup olmadığını yine White testi ile belirleyelim:

Test istatistiği: LM = 3.49802
p = P(Chi-Square(2) > 3.49802) = 0.173946

Bu kez farklı varyans söz konusu değildir. O halde denklemimiz:

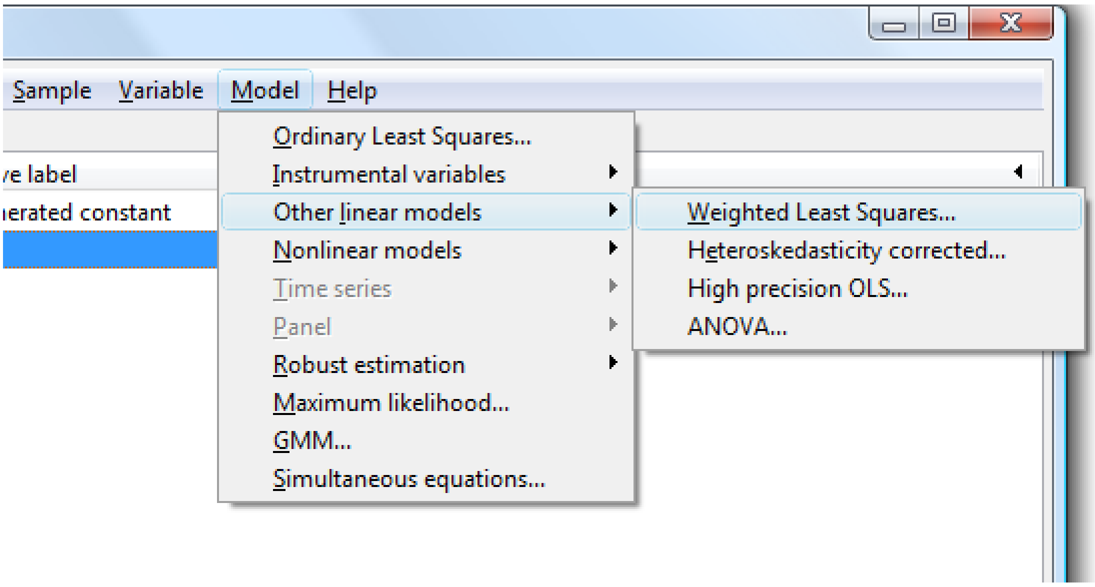
$$Y/X = 0.0283415 + 2.26172\,(1/X)$$

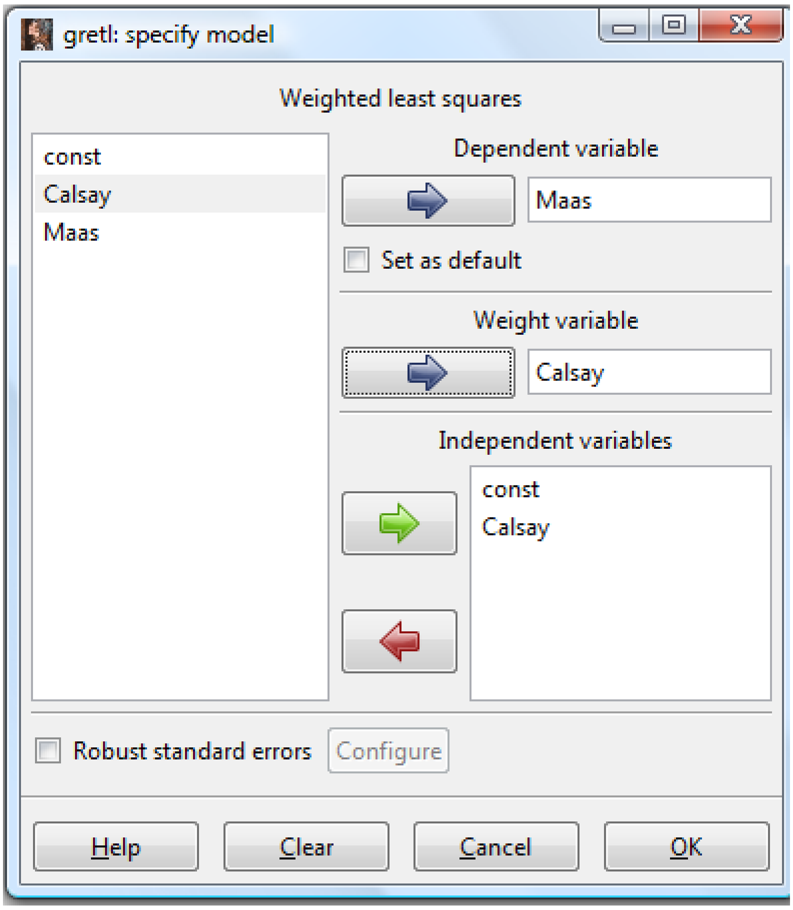
Şimdi denklemimizin her iki tarafını X ile çarpalım:

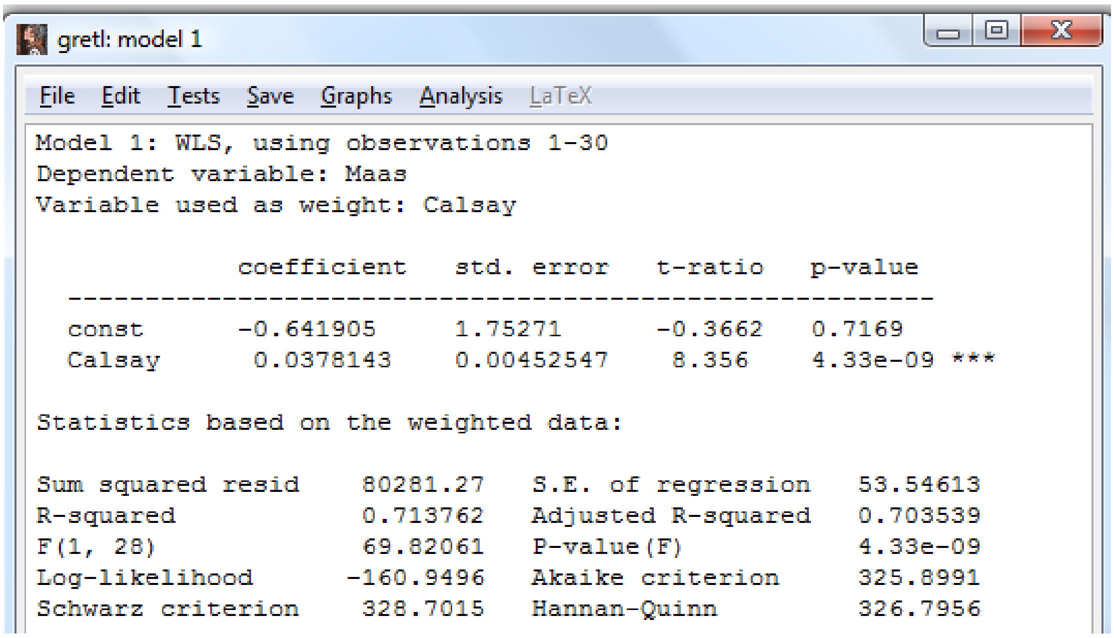
$$(Y/X)*X = 0.0283415*X + 2.26172\,(1/X)*X$$

$$Y = 0.0283415*X + 2.26172$$

$$Y = 2.26172 + 0.0283415\,X$$







Örnek 6-4

Modelimiz :

Envanter = f(Satis, borc orani)

Amacımız, **satış**tan kaynaklanan farklı varyans olup olmadığını belirlemektir.

Firma	Envanter I	Satis S	Borc orani R	Firma	Envanter I	Satis S	Borc orani R
1	10	85	17	19	15	122	11
2	10	101	17	20	15	123	11
3	10	103	17	21	15	125	11
4	11	105	16	22	16	128	10
5	11	106	16	23	16	128	10
6	11	106	16	24	16	131	10
7	12	108	15	25	17	133	10
8	12	109	15	26	17	134	9
9	12	111	14	27	17	135	9
10	12	111	14	28	17	136	9
11	12	112	14	29	18	139	8
12	13	113	14	30	18	143	8
13	13	114	13	31	19	147	8
14	13	114	13	32	19	151	8
15	14	116	12	33	19	157	8
16	14	117	12	34	20	163	7
17	14	118	12	35	20	171	7
18	15	120	11				

Modelimiz:

$$Envanter = b_0 + b_1 Satis + b_2 BorcOrani$$

EKK çözümü:

Model 4: EKK tahminleri 35 gözlem 1-35
Bağımlı değişken: Envanter

	Tahminci	*StHata*	*t*	*p*	
Sabit	15.5598	1.64227	9.4746	<0.00001	***
Satis	0.0544438	0.00855192	6.3663	<0.00001	***
BorcOran	-0.649529	0.0515543	-12.5989	<0.00001	***

Mean dependent var	14.65714	S.D. dependent var	3.038410
Sum kare resid	3.506859	S.E. of regression	0.331043
R-kare	0.988828	Düzeltilmiş R-kare	0.988129
F(2, 32)	1416.099	P(F)	5.89e-32
Log-likelihood	-9.401873	Akaike kriteri	24.80375
Schwarz kriteri	29.46979	Hannan-Quinn	26.41446

Farklı varyans probleminin var olup olmadığını yine White testi ile belirleyelim:

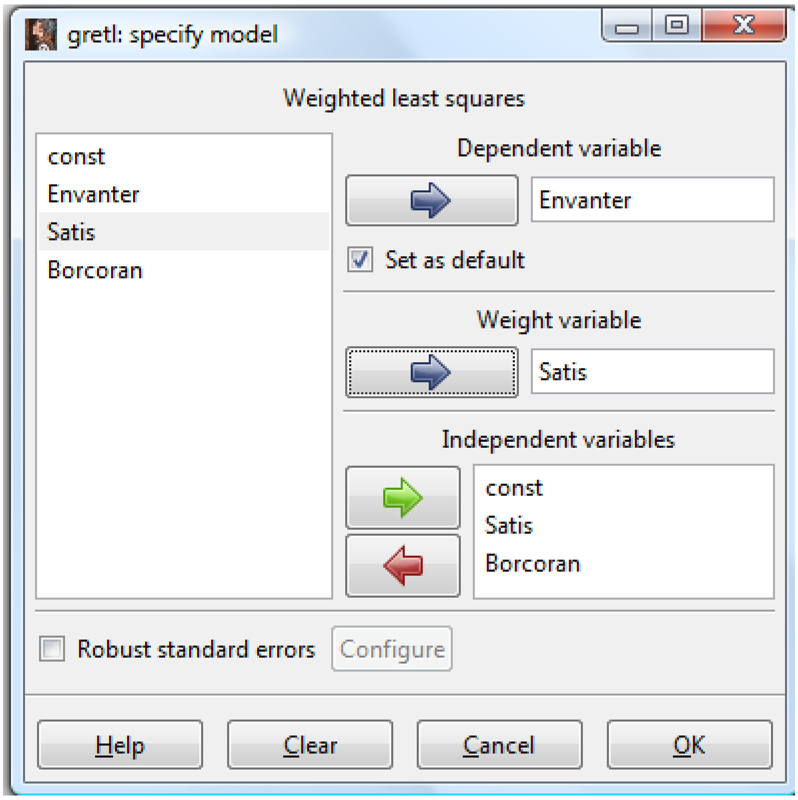
Test istatistiği: LM = 19.8558
p = P(Chi-Square(5) > 19.8558) = 0.00133006

Görüldüğü gibi farklı varyans söz konusudur. O halde AEKK tahminimizi Gretl ile yapalım:

AEKK çözümü (Ağırlık değişkeni=Satis)

Bağımlı dğişken: Envanter

	Tahminci	*StHata*	*t*	*p*	
Sabit	15.5379	1.53977	10.0911	<0.00001	***
Satis	0.0548642	0.00790073	6.9442	<0.00001	***
BorcOran	-0.652296	0.0495972	-13.1519	<0.00001	***



Bağımlı değişken: Envanter
Ağırlık değişkeni: Satis

	Tahminci	*StHata*	*t*	*p*	
Sabit	15.5379	1.53977	10.0911	<0.00001	***
Satis	0.0548642	0.00790073	6.9442	<0.00001	***
BorcOran	-0.652296	0.0495972	-13.1519	<0.00001	***

Örnek 6-5

Aşağıdaki verileri kullanarak, harcanabilir gelir değişkenine göre AEKK uygulaması yapalım.

Hane	Tüketim C	Harcanabilir Gelir Y	Varlık W
1	47	120	47
2	53	140	52
3	56	150	58
4	59	160	63
5	62	170	67
6	65	180	70
7	65	180	70
8	68	190	76
9	71	200	79
10	76	220	76
11	87	260	90
12	98	300	86
13	117	360	144
14	140	450	149
15	180	500	204

Modelimiz:

Tüketim = f(harcanabilir gelir, varlık)

Tuketim = $b_0 + b_1$ HarcGelir + b_2Varlik

EKK çözümü:

Bağımlı değişken: Tuketim

	Tahminci	*StHata*	*T*	*p*	
Sabit	6.56965	2.27248	2.8910	0.01355	**
HarcGelir	0.191083	0.03496	5.4657	0.00014	***
Varlik	0.346641	0.0929128	3.7308	0.00287	***

Mean dependent var	82.93333	S.D. dependent var	36.90038
Sum kare resid	165.0673	S.E. of regression	3.708855
R-kare	0.991341	Düzeltilmiş R-kare	0.989898
F(2, 12)	686.9151	P(F)	4.22e-13
Log-likelihood	-39.27135	Akaike kriteri	84.54270
Schwarz kriteri	86.66685	Hannan-Quinn	84.52007

Farklı varyans probleminin var olup olmadığını yine White testi ile belirleyelim:

Test istatistiği: LM = 11.641
p = P(Chi-Square(5) > 11.641) = 0.0400518

Görüldüğü gibi farklı varyans söz konusudur. O halde AEKK tahminimizi Gretl ile yapalım:

Burada gelir değişkeni farklı varyansa yolaçmaktadır. Bu nedenle Gretl Aekk çözümünde ağırlık değişkeni olarak gelir alınmıştır:

Bağımlı değişken: Tuketim

	Tahminci	*StHata*	*T*	*p*	
Sabit	5.57068	2.99051	1.8628	0.08714	*
HarcGelir	0.173967	0.0369835	4.7039	0.00051	***
Varlik	0.402115	0.0947004	4.2462	0.00113	***

R-kare	0.990274	Düzeltilmiş R-kare	0.988654
F(2, 12)	610.9328	P(F)	8.46e-13
Log-likelihood	-83.82256	Akaike kriteri	173.6451
Schwarz kriteri	175.7693	Hannan-Quinn	173.6225

Standart Sapmayla Farklı Varyansın Düzeltilmesi

Farklı varyansın varlığı belirlendikten sonra uygulanabilecek düzeltme yöntemlerinden bir diğeri de standart sapmayı kullanarak ağırlıklandırmadır. Bu yöntemin işleyişini aşağıdaki gibidir:

- Önce farklı varyansa yolaçan değişkene göre verilerimizi sıralayıp, tam ortadan ikiye bölelim ve iki farklı veri seti elde edelim. Bunların adları; 1. veri seti ve 2. veri seti olsun.
- 1. veri setine ait EKK çözümlerini alalım. Tüm değişkenlerin sadece ilk yarısındaki verilerini, denklemin standart sapmasına bölelim.
- 2. veri setine ait EKK çözümlerini alalım. Tüm değişkenlerin sadece ikinci yarısındaki verilerini, denklemin standart sapmasına bölelim.
- Değişkenlerin yeni halleriyle EKK çözümlerini alalım.

Örnek 6-6

Veri setimiz, çalışan sayısına göre farklı varyans göstermektedir. Bu nedenle sıralamayı çalışan sayısına göre yapıyoruz. Ardından, ilk 18 gözleme 1. veri seti, 2. 18 gözleme ise 2. veri seti diyelim.

1. veri seti		2. veri seti	
Çalışan sayısı X	Ortalama maaş Y	Çalışan sayısı X	Ortalama maaş Y
100	8.4	300	10.3
100	8.4	300	10.3
100	8.6	300	10.5
100	8.7	400	10.3
100	8.9	400	10.6
100	9.0	400	10.9
200	8.9	400	11.3
200	9.1	400	11.5
200	9.3	400	11.7
200	9.3	500	11.6
200	9.4	500	11.8
200	9.6	500	10.8
300	9.5	500	11
300	9.8	500	16
300	9.9	500	25

Önce veri setini bütün olarak alıp, farklı varyans durumunu belirleyelim:

Model 9: EKK tahminleri 30 gözlem 1-30
Bağımlı değişken: maas

	Tahminci	*StHata*	*T*	*p*	
Sabit	6.725	1.06617	6.3076	<0.00001	***
calsay	0.0131833	0.00321462	4.1011	0.00032	***

Mean dependent var	10.68000	S.D. dependent var	3.095536
Sum kare resid	173.6078	S.E. of regression	2.490036
R-kare	0.375260	Düzeltilmiş R-kare	0.352948
F(1, 28)	16.81862	P(F)	0.000320
Log-likelihood	-68.90218	Akaike kriteri	141.8044
Schwarz kriteri	144.6068	Hannan-Quinn	142.7009

Farklı varyans probleminin var olup olmadığını yine White testi ile belirleyelim:

Test istatistiği: LM = 4.89517
p = P(Chi-Square(2) > 4.89517) = 0.0865022

Görüldüğü gibi farklı varyans söz konusudur.

1. veri setinin EKK çözümü:

EKK 1-15 gözlemleri
Bağımlı değişken: maas

	Tahminci	*StHata*	*T*	*p*	
Sabit	8.14286	0.157026	51.8567	<0.00001	***
calsay	0.00542857	0.000805528	6.7391	0.00001	***

Mean dependent var	9.120000	S.D. dependent var	0.476895
Sum kare resid	0.708571	S.E. of regression	0.233464
R-kare	0.777459	Düzeltilmiş R-kare	0.760340
F(1, 13)	45.41613	P(F)	0.000014
Log-likelihood	1.610082	Akaike kriteri	0.779837
Schwarz kriteri	2.195937	Hannan-Quinn	0.764752

Modelin standart sapması, 0.233464'tür. Şimdi 1. veri setindeki tüm gözlemleri 0.233464'e böleceğiz:

1. veri seti	
X	Y
428.3316	35.97985
428.3316	35.97985
428.3316	36.83651
428.3316	37.26485
428.3316	38.12151
428.3316	38.54984
856.6631	38.12151
856.6631	38.97817
856.6631	39.83484
856.6631	39.83484
856.6631	40.26317
856.6631	41.11983
1284.995	40.6915
1284.995	41.97649

2. veri setinin EKK çözümü:

EKK 16-30 gözlemleri
Bağımlı değişken: maas

	Tahminci	*StHata*	*T*	*p*	
Sabit	3.05	5.18779	0.5879	0.56666	
calsay	0.021881	0.0121604	1.7994	0.09520	*

R-kare	0.199395	Düzeltilmiş R-kare	0.137810
F(1, 13)	3.237722	P(F)	0.095203
Log-likelihood	-39.10651	Akaike kriteri	82.21301
Schwarz kriteri	83.62911	Hannan-Quinn	82.19793

Modelin standart sapması, 3.5244'tür. Şimdi 2. veri setindeki tüm gözlemleri 3.5244'e böleceğiz:

2. veri seti	
x	y
2.922483	85.12087
2.922483	85.12087
2.979231	85.12087
2.922483	113.4945
3.007604	113.4945
3.092725	113.4945
3.20622	113.4945
3.262967	113.4945
3.319714	113.4945
3.29134	141.8681

3.348088	141.8681
3.064351	141.8681
3.121099	141.8681
4.53978	141.8681
7.093406	141.8681

Şimdi 1. ve 2. veri setini birleştirelim. Artık x ve y değişkenlerimiz aşağıdaki gibi olmuştur.

x	y
428.3316	35.97985
428.3316	35.97985
428.3316	36.83651
428.3316	37.26485
428.3316	38.12151
428.3316	38.54984
856.6631	38.12151
856.6631	38.97817
856.6631	39.83484
856.6631	39.83484
856.6631	40.26317
856.6631	41.11983
1284.995	40.6915
1284.995	41.97649
1284.995	42.40482
2.922483	85.12087
2.922483	85.12087
2.979231	85.12087
2.922483	113.4945
3.007604	113.4945
3.092725	113.4945
3.20622	113.4945
3.262967	113.4945
3.319714	113.4945
3.29134	141.8681
3.348088	141.8681
3.064351	141.8681
3.121099	141.8681

Şimdi bu verilerle EKK çözümü alalım:

Bağımlı değişken: maas

	Tahminci	*StHata*	*T*	*p*	
Sabit	108.527	6.55579	16.5543	<0.00001	***
calsay	-0.0759491	0.0111036	-6.8400	<0.00001	***

R-kare	0.625599	Düzeltilmiş R-kare	0.612227
F(1, 28)	46.78611	P(F)	1.97e-07
Log-likelihood	-140.5258	Akaike kriteri	285.0516
Schwarz kriteri	287.8539	Hannan-Quinn	285.9481

Farklı varyans probleminin var olup olmadığını yine White testi ile belirleyelim:

Test istatistiği: LM = 0.631341
p = P(Chi-Square(2) > 0.631341) = 0.7293

Görüldüğü gibi farklı varyans söz konusu değildir.

6.2.3.2.Farklı Varyansı Düzeltilmiş Tahminciler (FVDT)

Farklı varyans bulunan modellerde Gretl:

a) Modelimizin EKK tahminin yapar.
b) Hata varyansını tahminini elde etmek üzere bir yardımcı regresyon tahmin eder.
c) Tahmin edilen varyansın tersini kullanarak, ağırlıklı en küçük kareler yöntemini uygular

Yardımcı regresyonda, EKK ile elde edilen hata teriminin karelerinin logaritması alındıktan sonra bu değişken bağımlı değişken olarak alınır ve daha önceki bağımsız değişkenler ile bunların kareleri sağ taraf değişkenleri olarak kabul edilir. Bu denklemden elde edilen Y tahmin değerleri ağırlık değişkeni olarak kabul edilip, orijinal model ağırlıklı EKK ile yeniden tahmin edilir.

Şimdi Örnek 6-4 verilerini bu kez de HSK için kullanalım. FVDT sonuçları:

Bağımlı değişken: Envanter

	Tahminci	*StHata*	*t*	*p*	
Sabit	12.1436	1.69664	7.1574	<0.00001	***
Satis	0.0745232	0.00948692	7.8554	<0.00001	***
BorcOran	-0.569459	0.0456567	-12.4726	<0.00001	***

R-kare	0.994415	Düzeltilmiş R-kare	0.994066
F(2, 32)	2848.983	P(F)	8.95e-37
Log-likelihood	-73.69227	Akaike kriteri	153.3845
Schwarz kriteri	158.0506	Hannan-Quinn	154.9953

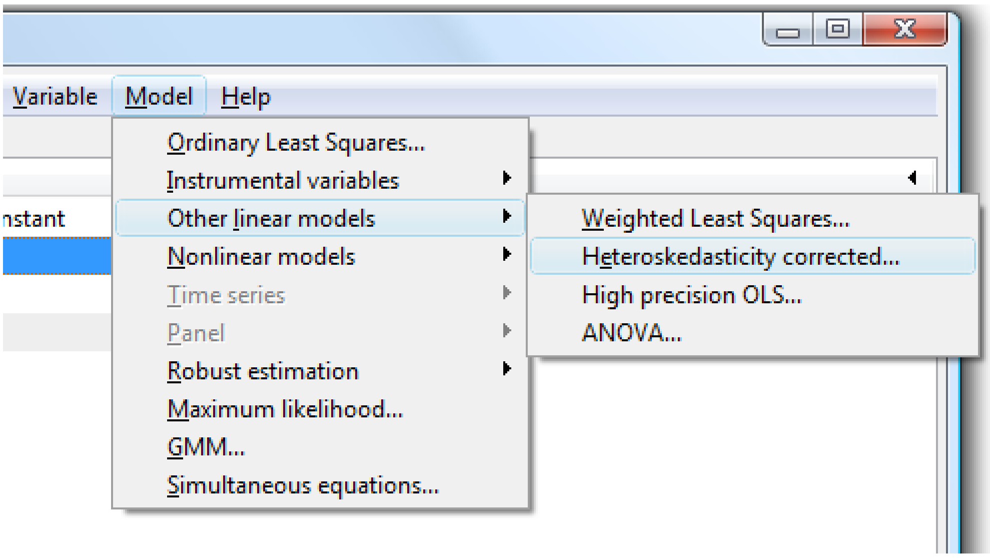
FVDT sonuçlarını incelediğimizde, gerek tahmincilerin, gerekse standart hataların EKK'nınkilerden farklı olduğunu görebiliriz.

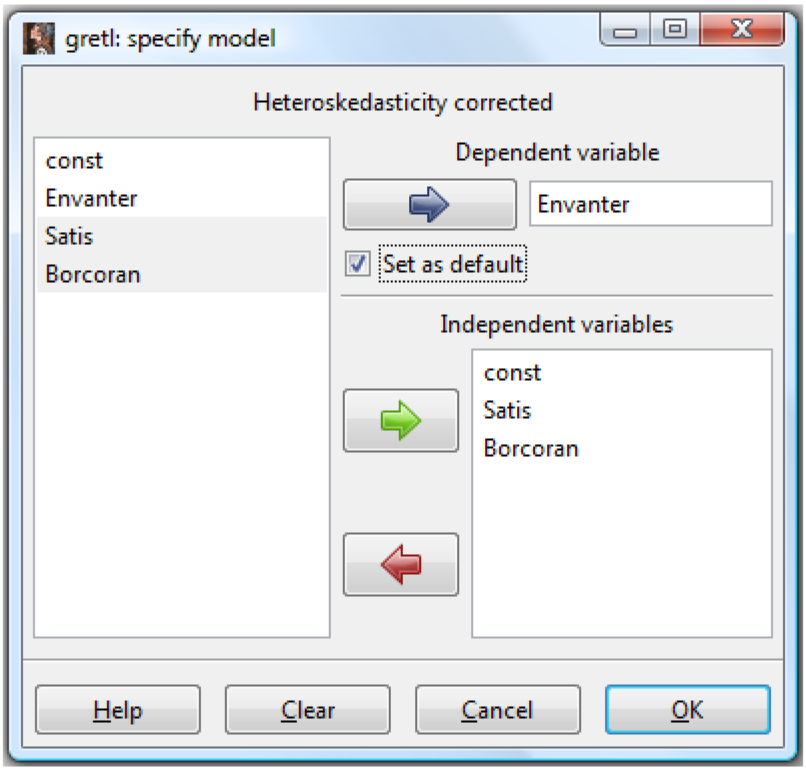
Örnek 6-6'daki modeli FVDT ile çözelim:

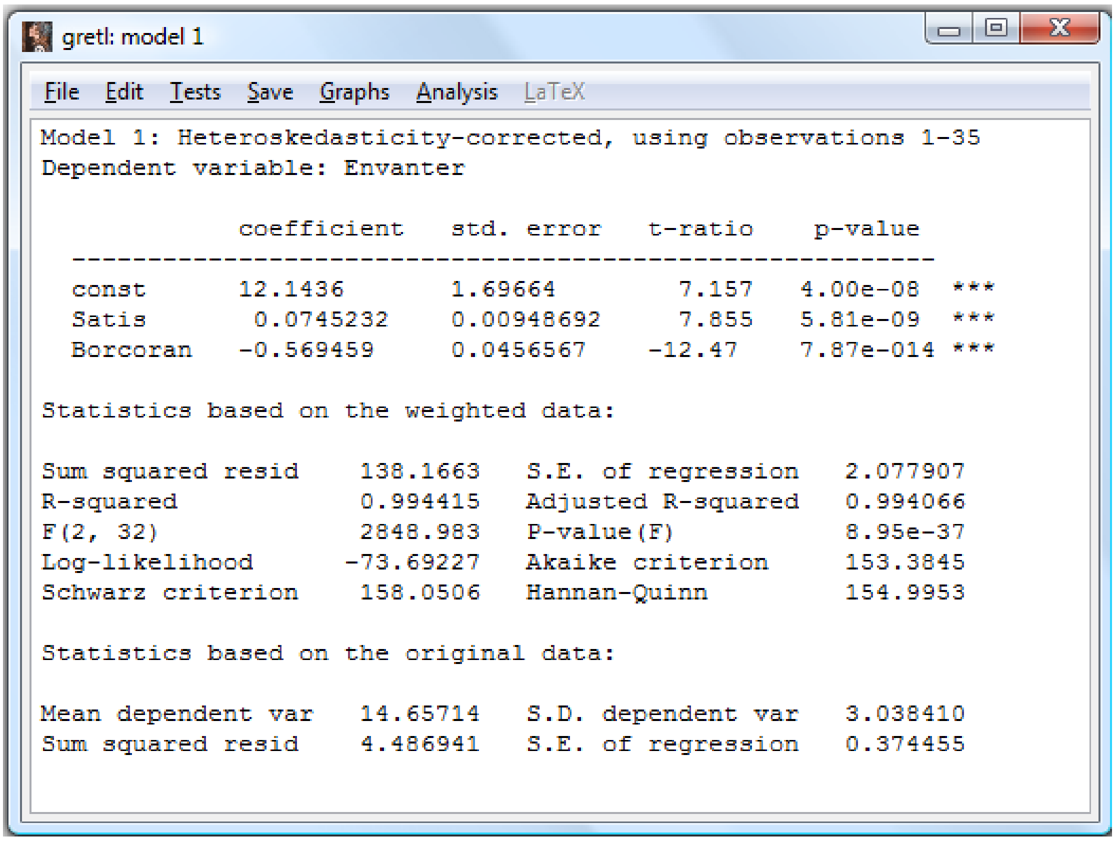
Bağımlı değişken: maas

	Tahminci	*StHata*	*T*	*p*	
Sabit	7.73309	0.244106	31.6793	<0.00001	***
calsay	0.0079247	0.00100921	7.8524	<0.00001	***

R-kare	0.687708	Düzeltilmiş R-kare	0.676555
F(1, 28)	61.65970	P(F)	1.49e-08
Log-likelihood	-45.37366	Akaike kriteri	94.74732
Schwarz kriteri	97.54971	Hannan-Quinn	95.64383


Variable
Model
Help
Ordinary Least Squares...
Instrumental variables
Other linear models
Nonlinear models
Time series
Panel
Robust estimation
Maximum likelihood...
GMM...
Simultaneous equations...
Weighted Least Squares...
Heteroskedasticity corrected...
High precision OLS...
ANOVA...
nstant


gretl: specify model
Heteroskedasticity corrected
const
Envanter
Satis
Borcoran
Dependent variable
Envanter
Set as default
Independent variables
const
Satis
Borcoran
Help
Clear
Cancel
OK




```
Model 1: Heteroskedasticity-corrected, using observations 1-35
Dependent variable: Envanter

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const      12.1436       1.69664        7.157    4.00e-08  ***
  Satis       0.0745232    0.00948692     7.855    5.81e-09  ***
  Borcoran   -0.569459     0.0456567    -12.47     7.87e-014 ***

Statistics based on the weighted data:

Sum squared resid     138.1663   S.E. of regression   2.077907
R-squared             0.994415   Adjusted R-squared   0.994066
F(2, 32)              2848.983   P-value(F)           8.95e-37
Log-likelihood       -73.69227   Akaike criterion     153.3845
Schwarz criterion     158.0506   Hannan-Quinn         154.9953

Statistics based on the original data:

Mean dependent var    14.65714   S.D. dependent var   3.038410
Sum squared resid     4.486941   S.E. of regression   0.374455
```

6.3. OTOKORELASYON

Doğrusal regresyon, birbirini izleyen hata terimleri arasında ilişki bulunmadığını varsayar. Bu varsayımın ihlali, otokorelasyon problemine yolaçar. Otokorelasyona, özellikle zaman serisi modellerinde rastlanır.

Doğru tanımlanmış bir regresyon modeli:

$$Y_t = b_0 + b_1X_t + e_t$$

ise, ard arda iki hata değeri arasındaki basit korelasyon katsayısını sıfırdan farklı olması:

$$E(r_{e(i),e(i+1)}) \neq 0$$

otokorelasyonun varlığına işaret eder.

Birinci dereceden otokorelasyon, en sık rastlanan otokorelasyondur. Anlamı, t dönemine ait hata terimi ile, bir önceki döneme (t-1) ait hata terimi arasındaki korelasyon katsayısının (ρ) istatistiki açıdan önemli olmasıdır:

$$e_t = \rho e_{t-1} + u_t$$

Burada:

e_t = t döneminin hata terimi

e_t = t-1 döneminin hata terimi

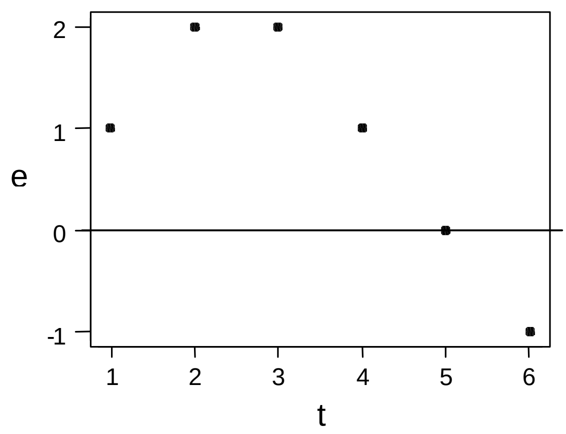
ρ = Fonksiyonel ilişkiyi temsil eden tahminci

u_t = Denklemin hata terimi

denklemdeki ρ (ro), birinci dereceden otokorelasyon katsayısı olarak adlandırılır ve bir gözlemin hata teriminin, izleyen gözlemin hata terimini doğrudan etkileme düzeyinin gösterir. ρ, -1 ile 1 arasındadır ve ρ'nun büyüklüğü, denklemdeki otokorelasyonun şiddetini belirler. ρ, sıfır ise, otokorelasyon yok demektir. ρ'nun mutlak değeri 1'e yaklaştıkça, otokorelasyonun bulunduğu anlaşılır. ρ'nun işareti, pozitif veya negatif otokorelasyona işaret eder.

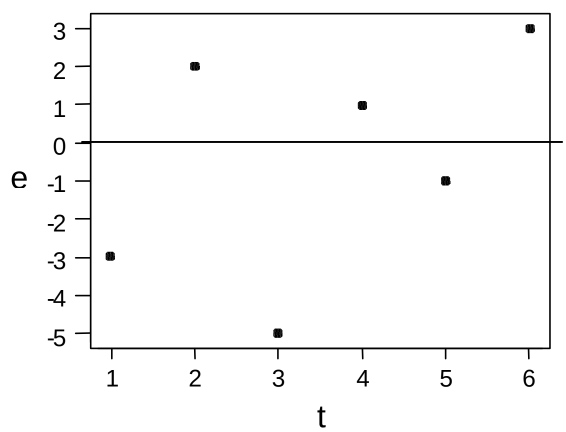
Pozitif Otokorelasyon:

Bir dönemin hata terimi (e_t) tesadüfi olarak büyük bir değer almışsa, takip eden gözlemler bu büyük değerin etkisini aynı yönde taşıyabilir. Bu, pozitif otokorelasyondur. Örneğin ekonomiye yapılacak dışsal bir şok etki, gelecek dönemlerde de kendisini hissettirecektir. Pozitif otokorelasyonda, birbirini izleyen hata terimlerinin işaretleri aynıdır.



Negatif Otokorelasyon:

Ard arda hata terimleri, sürekli işaret değiştiriyorsa, negatif otokorelasyından söz edilir.



6.3.1. Otokorelasyonun Nedenleri

- Önemli bir değişkenin modele alınmamış olması

- Yanlış fonksiyonel form seçimi
- GSMH, fiyat indeksleri, üretim, istihdam ve işsizlik gibi dönemsellik gösteren değişkenlerin modellerde bulunması
- Örümcek ağı etkisi

6.3.2. Birinci Dereceden Otokorelasyonun Yol Açtığı Sonuçlar

- Denklemin varyansı düşük
- R^2 yüksek
- Tahminciler yansız, tutarlı
- Tahminciler etkin değil
- Tahmin veya kestirimler etkin değil
- Yüksek t, güvensiz t-testi
- Yüksek F değeri, güvensiz F-testi

6.3.3. Birinci Dereceden Otokorelasyonu Belirleme Yöntemleri

6.3.3.1. Durbin Watson d Testi

Durbin Watson d testi, birinci dereceden otokorelasyonun belirlenmesinde kullanılan bir testtir. Bu testi uygulayabilmek için:

- Regresyon modelinde sabit terim olmalıdır
- Otokorelasyon, birinci derecedendir: $e_t = \rho e_{t-1} + u_t$

Burada:

e_t = t döneminin hata terimi

e_t = t-1 döneminin hata terimi

ρ = Fonksiyonel ilişkiyi temsil eden tahminci

u_t = Denklemin hata terimi

Regresyon modelinde, bağımlı değişkenin gecikmeli değerleri bulunmamalıdır.

n gözleme sahip bir modelin d istatistiği:

$$d = \sum_2^n (e_i - e_{i-1})^2 / \sum_1^n e_i^2$$

formülüyle hesaplanır. Burada, e, hata terimi değerlerini temsil etmektedir. Hesaplanan d istatistiği:

2 ise, otokorelasyon yok

0 ise, pozitif otokorelasyon var

4 ise, negatif otokorelasyon var

sonucuna varılır.

Durbin Watson testini gerçekleştirebilmek için, aşağıdaki süreç izlenir:

Adım 1. $Y_t = b_0 + b_1X_t + e_t$
denklemini tahmin et

Adım 2. Hipotezleri hazırla:
H_0: Otokorelasyon yok
H_1: Otokorelasyon var

Adım 3. Diagramı hazırla:

H_0'ı reddet (Pozitif Otokorelasyon)	Kararsız	H_0'ı kabul et	Kararsız	H_0'ı reddet (Negatif Otokorelasyon)

0 d_L d_U 2 $4\text{-}d_U$ $4\text{-}d_L$ 4

Adım 4. Hata değerlerini hesapla.

Adım 5 d istatistiğini hesapla:

$d = \sum_2^n (e_i - e_{i-1})^2 / \sum_1^n e_i^2$

Adım 6.

n sayısı, bağımsız değişken sayısı (k') ve belli bir α düzeyine göre, Durbin Watson tablosunda d_L ve d_U değerlerini bulup, diyagramda yerinde koy.

Adım 7.

Otokorelasyonun varlığını tesbit et.

Örnek 6-7

Daha önce ele aldığımız balık eti tüketimi probleminde, otokorelasyon olup olmadığını Durbin Watson testiyle belirleyelim.

Çizelge 6-1: 1973-98 Yılları Kişi Başına Balık Eti Tüketimi, Reel Balık Fiyatı, Reel K.Et Fiyatı, Kişi Başına Gelir Değerleri

Yıl	Kişi Başına K.Et Tüketimi (Y)	Reel K.Et Fiyatı (X1)	Kişi Başına Gelir (X2)
1973	85.1	20.4	6.036
1974	87.8	20.2	6.113
1975	88.9	21.3	6.271
1976	94.5	19.9	6.378
1977	99.9	18	6.727
1978	99.5	19.9	7.027
1979	104.2	22.2	7.28
1980	106.5	22.3	7.513
1981	109.7	23.4	7.728
1982	110.8	26.2	7.891
1983	113.7	27.1	8.134
1984	113	29	8.322
1985	116	33.5	8.562
1986	108.7	42.8	9.042
1987	115.4	35.6	8.867
1988	118.9	32.2	8.944
1989	127.4	33.7	9.175
1990	123.5	34.4	9.381
1991	117.9	48.5	9.735
1992	105.4	66.1	9.829
1993	103.2	62.4	9.722

1994	104.2	58.6	9.769
1995	103.7	56.7	9.725
1996	105.7	55.5	9.93
1997	105.5	57.3	10.419
1998	106.5	53.7	10.625
1999	107.3	52.6	10.905
2000	103.3	61.1	10.97

Adım 1. $Y_t = b_0 + b_1X1_t + b_2X2_t + e_t$ denklemini tahmin edelim:

	Katsayı	StHata	t	p
Sabit	37.54	10.04	3.74	0.001
X1	-0.8826	0.1647	-5.36	0.000
X2	11.891	1.762	6.75	0.000
S = 6.081	R^2 = 0.658	Düz. R^2=0.631	F=24.05	p=0.000

Adım 2. Hipotezlerimiz:

H_0: Otokorelasyon yok

H_1: Otokorelasyon var

Adım 3. Hata değerlerini hesaplayalım.

Adım 4. d istatistiğini hesaplayalım:

e_t	e_{t-1}	e_t^2	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
-6.2055	*	38.508	*	*
-4.5977	-6.2055	21.138	1.6078	2.585021
-4.4056	-4.5977	19.409	0.1921	0.036902
-1.3136	-4.4056	1.726	3.092	9.560464
-1.7406	-1.3136	3.030	-0.427	0.182329
-4.031	-1.7406	16.249	-2.2904	5.245932
-0.3094	-4.0310	0.096	3.7216	13.85031
-0.6918	-0.3094	0.479	-0.3824	0.14623
0.9225	-0.6918	0.851	1.6143	2.605964
2.5556	0.9225	6.531	1.6331	2.667016
3.3604	2.5556	11.292	0.8048	0.647703
2.1019	3.3604	4.418	-1.2585	1.583822
6.2198	2.1019	38.686	4.1179	16.9571
1.4204	6.2198	2.018	-4.7994	23.03424
3.8465	1.4204	14.796	2.4261	5.885961
3.43	3.8465	11.765	-0.4165	0.173472
10.5071	3.4300	110.398	7.0771	50.08534
4.7753	10.5071	22.804	-5.7318	32.85353
7.4108	4.7753	54.920	2.6355	6.94586
9.3272	7.4108	86.997	1.9164	3.672589

5.1339	9.3272	26.357	-4.1933	17.58376
2.221	5.1339	4.933	-2.9129	8.484986
0.5672	2.2210	0.322	-1.6538	2.735054
-0.9296	0.5672	0.864	-1.4968	2.24041
-5.3556	-0.9296	28.683	-4.426	19.58948
-9.9827	-5.3556	99.653	-4.6271	21.41005
-13.4831	-9.9827	181.793	-3.5004	12.2528
-10.7537	-13.4831	115.642	2.7294	7.449624
0.000		924.358	-4.5482	270.466

$\sum_2^n (e_i - e_{i-1})^2 = 270.466$

$\sum_1^n e_i^2 = 924.358$

$d = \sum_2^n (e_i - e_{i-1})^2 / \sum_1^n e_i^2$

$d = 270.466 / 924.358$

$d = 0.29$

Adım 5.

$n=28$, $k'= 2$, $\alpha=0.05$

$d_L=1.255$, $d_U=1.560$

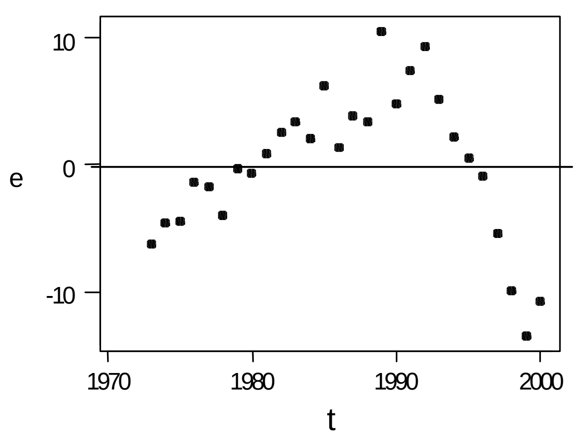
H_0'ı reddet (Pozitif Otokorelasyon)	Kararsız	H_0'ı kabul et	Kararsız	H_0'ı reddet (Negatif Otokorelasyon)

0 **0.29** 1.255 1.560 2 2.44 2.745 4

Adım 6. Otokorelasyonun varlığını tesbit et.

Hesapladığımız d değerini, diyagramda yerleştirdiğimizde, H_0 reddedilir. Otokorelasyon tipi ise, pozitiftir.

Şimdi, hata değerlerinin grafiğini alalım:



Görüldüğü gibi, ardıl hata değerlerinin işaretleri aynıdır. Grafik, pozitif otokorelasyonu doğrulamaktadır.

6.3.3.2. LM (Breusch-Godfrey) Testi

LM testi, sadece birinci dereceden değil, aynı zamanda daha yüksek derecden otokorelasyonun belirlenmesinde kullanılabilir. Bu testi, aşağıdaki adımları izleyerek gerçekleştirebiliriz:

Adım 1

$Y_t = b_0 + b_1X1_t + b_2X2_t + b_3X3_t + \ldots + b_pXp_t + e_t$

modelini EKK ile tahmin edip, hata terimini (e_t) hesapla.

Adım 2

Hata terimini, $X1_t$, $X2_t$ … b_pXp_t ve e_{t-1}'in fonksiyonu olarak tanımlayıp, tahmin et.

$e_t = b_0 + b_1X1_t + b_2X2_t + b_3X3_t + \ldots + b_pXp_t + b_{p+1}e_{t-1}$

Adım 3

$LM = (n-1)\,R^2$

hesaplamasını yap.

Adım 4

Serbestlik derecesi 1 ve belli bir α düzeyi için χ^2 tablo değerini bul ($\chi^2_{\alpha,1}$).

Adım 5

Eğer $LM > \chi^2_{\alpha,1}$ ise, "H_0: Otokorelasyon yok" hipotezini reddet.

Örnek 6-8

Örnek 6-7'deki veri setimiz için LM testini uygulayalım.

Adım 1

$Y_t = b_0 + b_1X1_t + b_2X2_t + e_t$ denklemini tahmin edelim:

	Katsayı	StHata	t	p
Sabit	37.54	10.04	3.74	0.001
X1	-0.8826	0.1647	-5.36	0.000
X2	11.891	1.762	6.75	0.000
S = 6.081	R^2 = 0.658	Düz. R^2=0.631	F=24.05	p=0.000

Adım 2

Hata terimini, $X1_t$, $X2_t$ … b_pXp_t ve e_{t-1}'in fonksiyonu olarak tanımlayıp, tahmin edelim:

$e_t = b_0 + b_1X1_t + b_2X2_t + b_3e_{t-1}$

	Katsayı	StHata	t	p
Sabit	4.353	5.678	0.77	0.451
Fiyat (X1)	-0.03542	0.08823	-0.4	0.692
Gelir (X2)	-0.3595	0.9741	-0.37	0.715
e_{t-1}	0.9089	0.1128	8.05	0.000

S = 3.160	R^2 = 0.740	Düz. R^2=0.707	F=21.86	p=0.000

Adım 3

$LM = (n\text{-}1)\ R^2$

$LM = (28\text{-}1)\ 0.740 = 19.089$

Adım 4

α=0.05, sd=1, χ^2 =3.841.

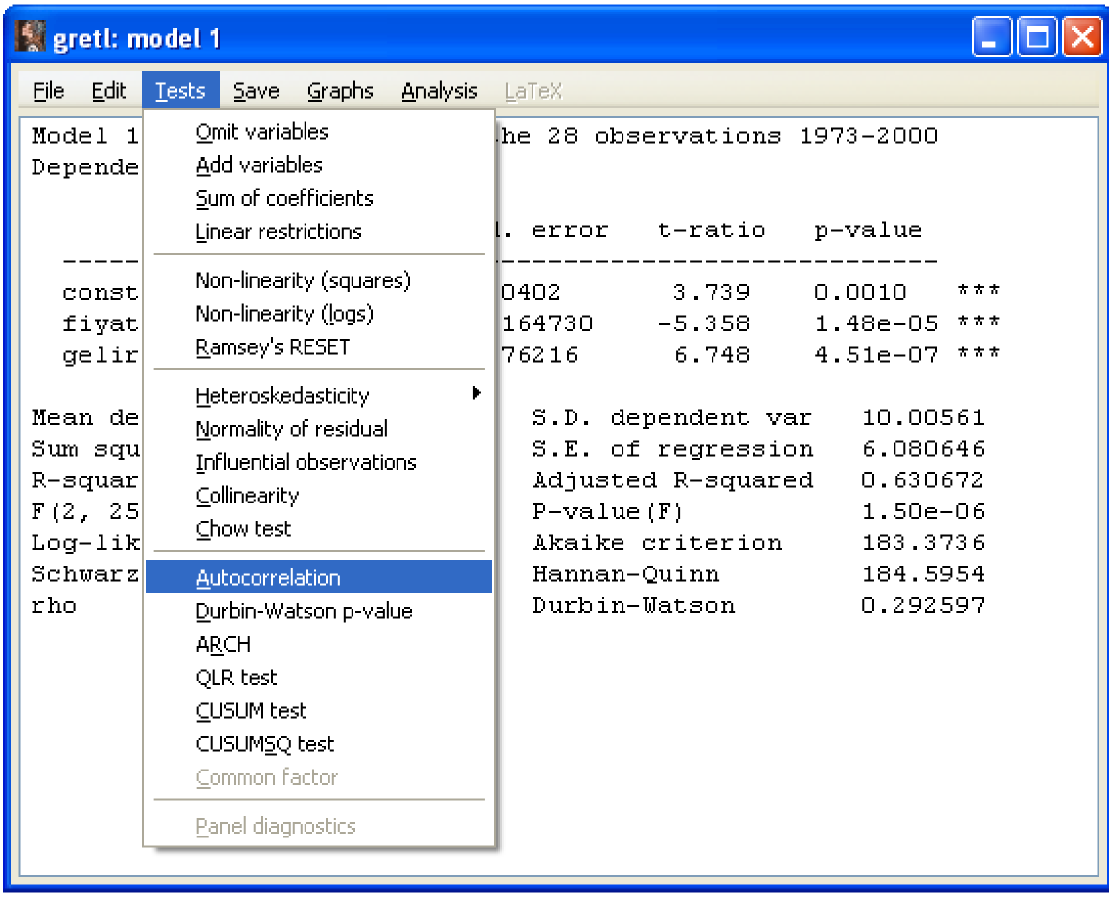
Adım 5

19.089 > 3.841 olduğundan "H_0: Otokorelasyon yok" hipotezini reddederiz.

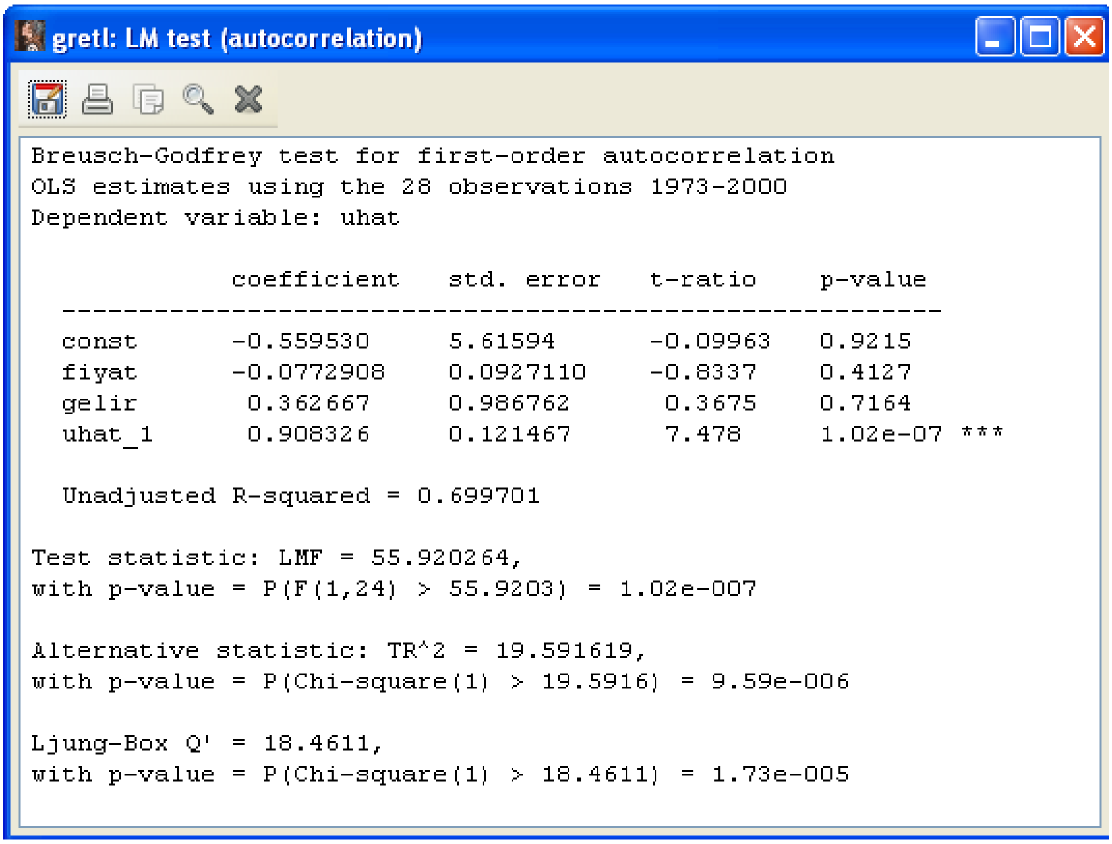
gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS estimates using the 28 observations 1973-2000
Dependent variable: kettuk

             coefficient    std. error    t-ratio    p-value
  ---------------------------------------------------------
  const       37.5360       10.0402        3.739     0.0010    ***
  fiyat       -0.882623      0.164730     -5.358     1.48e-05  ***
  gelir       11.8911        1.76216       6.748     4.51e-07  ***

Mean dependent var    106.6500   S.D. dependent var    10.00561
Sum squared resid     924.3564   S.E. of regression    6.080646
R-squared             0.658030   Adjusted R-squared    0.630672
F(2, 25)              24.05287   P-value(F)            1.50e-06
Log-likelihood       -88.68678   Akaike criterion      183.3736
Schwarz criterion     187.3702   Hannan-Quinn          184.5954
rho                   0.880470   Durbin-Watson         0.292597
```


gretl: model 1
File Edit Tests Save Graphs Analysis LaTeX
Model 1
Depende
he 28 observations 1973-2000
. error t-ratio p-value
const 0402 3.739 0.0010 ***
fiyat 164730 -5.358 1.48e-05 ***
gelir 76216 6.748 4.51e-07 ***
Mean de S.D. dependent var 10.00561
Sum squ S.E. of regression 6.080646
R-squar Adjusted R-squared 0.630672
F(2, 25 P-value(F) 1.50e-06
Log-lik Akaike criterion 183.3736
Schwarz Hannan-Quinn 184.5954
rho Durbin-Watson 0.292597
Omit variables
Add variables
Sum of coefficients
Linear restrictions
Non-linearity (squares)
Non-linearity (logs)
Ramsey's RESET
Heteroskedasticity
Normality of residual
Influential observations
Collinearity
Chow test
Autocorrelation
Durbin-Watson p-value
ARCH
QLR test
CUSUM test
CUSUMSQ test
Common factor
Panel diagnostics


gretl: autocorrelation
Lag order for test: 1
Help
Cancel
OK



gretl: LM test (autocorrelation)

```
Breusch-Godfrey test for first-order autocorrelation
OLS estimates using the 28 observations 1973-2000
Dependent variable: uhat

              coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       -0.559530     5.61594      -0.09963   0.9215
  fiyat       -0.0772908    0.0927110    -0.8337    0.4127
  gelir        0.362667     0.986762      0.3675    0.7164
  uhat_1       0.908326     0.121467      7.478     1.02e-07 ***

  Unadjusted R-squared = 0.699701

Test statistic: LMF = 55.920264,
with p-value = P(F(1,24) > 55.9203) = 1.02e-007

Alternative statistic: TR^2 = 19.591619,
with p-value = P(Chi-square(1) > 19.5916) = 9.59e-006

Ljung-Box Q' = 18.4611,
with p-value = P(Chi-square(1) > 18.4611) = 1.73e-005
```

6.3.4. Otoregresif Modellerde Otokorelasyonu Belirleme Yöntemleri

Durbin-h Testi

Durbin h otokorelasyon testi, büyük örneklerde ve modelde bir gecikmeli bağımlı değişken bulunduğu durumda kullanılır. Bu testin aşamaları:

Adım 1:

Modeli EKK ile tahmin et

Adım 2:

Birinci dereceden Durbin Watson d değerini hesapla:

$d = \sum_2^n (e_i - e_{i-1})^2 / \sum_1^n e_i^2$

Adım 3:

Durbin h istatistiğini hesapla:

$h = d[n'/(1 - n's^2_b)]^{0.5}$

Formülde:

$n'=(n-1)$

s^2_b= 1 Gecikmeli bağımlı değişkene ait tahmincinin varyansı

Adım 4:

Çift taraflı z testi ile h değeri test edilir. α=0.01 için 2.58, α=0.05 için Z=1.96, α=0.10 için 1.645 kritik değeri kullanılır ve eğer h<-Z veya h>Z ise "H_0: Otokorelasyon yok" hipotezi reddedilir.

Örnek 6-7'deki veri setimiz için Durbin h testini uygulayalım:

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: OLS estimates using the 27 observations 1974-2000
Dependent variable: kettuk

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        12.1280      6.64435      1.825    0.0810    *
  fiyat        -0.438968    0.0985122   -4.456    0.0002    ***
  gelir         4.29577     1.26207      3.404    0.0024    ***
  kettuk_1      0.700055    0.0843552    8.299    2.27e-08  ***

Mean dependent var    107.4481   S.D. dependent var    9.243350
Sum squared resid     219.6028   S.E. of regression    3.089975
R-squared             0.901143   Adjusted R-squared    0.888249
F(3, 23)              69.88673   P-value(F)            1.04e-11
Log-likelihood       -66.60712   Akaike criterion      141.2142
Schwarz criterion     146.3976   Hannan-Quinn          142.7555
rho                  -0.055531   Durbin's h           -0.313650
```

Gretl, otoregresif modellerde Durbin Watson yerine Durbin h değerini verir. Yukarıda Durbin h değeri -0.31365 olarak bulunmuştur. Bu değerin mutlak değeri 1.96'dan küçük olduğu için, otokorelasyon yok şeklinde tanımlanan sıfır hipotezi reddedilmez. Yani otokorelasyon yoktur. Bunu LM testi ile doğrulayalım:

gretl: LM test (autocorrelation)

```
Breusch-Godfrey test for first-order autocorrelation
OLS estimates using the 27 observations 1974-2000
Dependent variable: uhat

             coefficient    std. error   t-ratio    p-value
  ---------------------------------------------------------
  const      -0.518503      7.06685     -0.07337    0.9422
  fiyat      -0.00173838    0.100789    -0.01725    0.9864
  gelir       0.0132931     1.28943      0.01031    0.9919
  kettuk_1    0.00448580    0.0878083    0.05109    0.9597
  uhat_1     -0.0608261     0.232545    -0.2616     0.7961

  Unadjusted R-squared = 0.003100

Test statistic: LMF = 0.068417,
with p-value = P(F(1,22) > 0.0684172) = 0.796

Alternative statistic: TR^2 = 0.083706,
with p-value = P(Chi-square(1) > 0.0837063) = 0.772

Ljung-Box Q' = 0.0782343,
with p-value = P(Chi-square(1) > 0.0782343) = 0.78
```

LMF, TR^2 ve Ljung-Box Q' testlerine ait p değerleri göz önüne alındığında, LM testinin de sıfır hipotezinin reddedilmemesi sonucunu verdiği anlaşılır. O halde üzerinde çalıştığımız otoregresif modelde otokorelasyon yoktur.

6.3.5. Otokorelasyonun giderilmesi

Otokorelasyonu gidermek üzere:

- Cochrane-Orcutt (CORC) Yöntemi
- Hildreth –Lu
- Praise-Winsten

yöntemlerinden yararlanılabilir.

Cochrane-Orcutt (CORC) Yöntemi

Cochrane-Orcutt tahminlemesi, hata terimindeki otokorelasyonu giderme amacına dönük olarak kullanılır. Bu yöntem iki aşamalı iteratif bir tekniktir. Genelleştirilmiş EKK uygulamasıdır.

Otokorelasyon olduğunu daha önce tespit ettiğimiz Örnek 6-7'deki veri setimize bu kez de CORC yöntemini uygulayacağız. Önce EKK çözüm sonuçlarını hatırlayalım:

	Katsayı	StHata	t	p
Sabit	37.54	10.04	3.74	0.001
Fiyat (X1)	-0.8826	0.1647	-5.36	0.000
Gelir (X2)	11.891	1.762	6.75	0.000
S = 0.053	R^2 = 0.709	Düz. R^2=0.686	F=90.38	p=0.000

DW=0.299

n=28, k'= 2, α=0.05

d_L=1.255, d_U=1.560

H_0'ı reddet (Pozitif Otokorelasyon)	Kararsız	H_0'ı kabul et	Kararsız	H_0'ı reddet (Negatif Otokorelasyon)

0 0.299 1.255 1.56 2 2.44 2.745 4

CORC sonuçları ise:

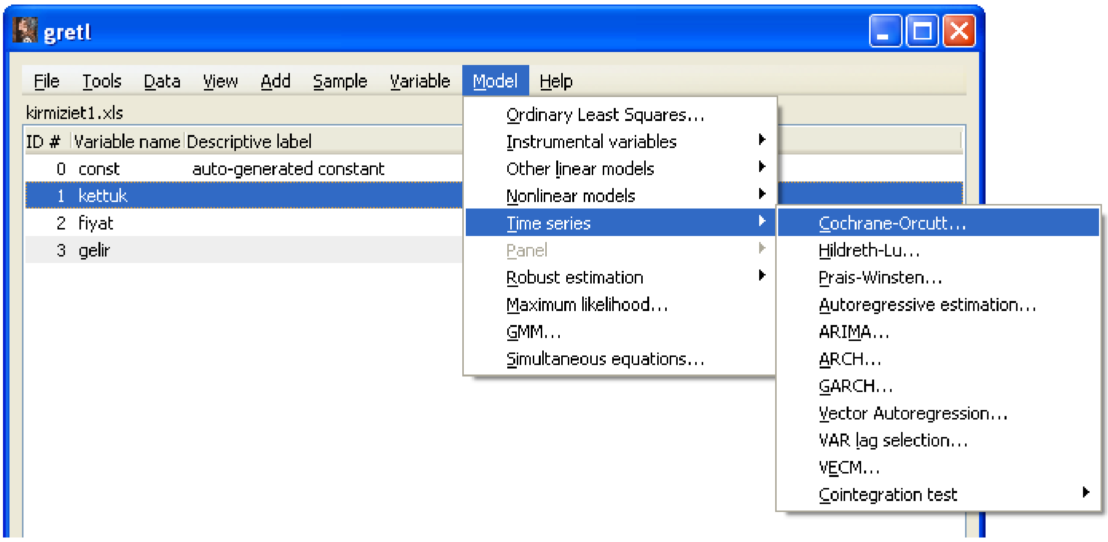
	Katsayı	StHata	t	p
Sabit	114.442	27.5995	4.147	0.000
Fiyat (X1)	-0.62327	0.114688	-5.434	0.000
Gelir (X2)	2.86762	2.96223	0.968	0.343
S = 2.76	R^2 = 0.923	Düz. R^2=0.917	F=17.2	P=0.000

Durbin-Watson istatistiği = 2.44

H_0'ı reddet (Pozitif Otokorelasyon)	Kararsız	H_0'ı kabul et	Kararsız	H_0'ı reddet (Negatif Otokorelasyon)

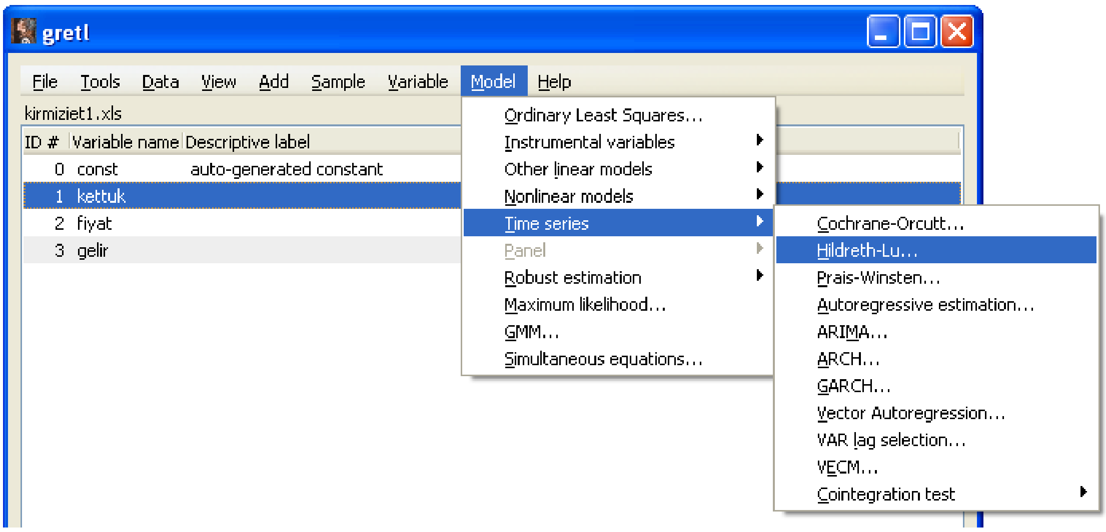
0 1.255 1.56 2 **2.44** 2.745 4

Görüldüğü gibi hesaplanan DW, "otokorelasyon yok" sınırında yeralmaktadır. CORC ve EKK tahmincileri gerek katsayı gerekse istatistiki anlamlılıkları açısından farklıdır. X2 EKK'da sıfırdan farklı iken, CORC bu değişkene ait tahmincinin önemsiz olduğunu ifade etmektedir. Gerçekten de hatırlanacağı üzere otokorelasyon problemi etkin olmayan tahminciler nedeniyle, hipotez testlerini geçersiz kılıyordu. CORC, etkin tahmincileri sağladığından, artık hipotez testlerine güvenebiliriz. CORC tahminlemesi Gretl yardımıyla aşağıdaki gibi gerçekleştirilir.



Hildreth-Lu

Gretl'da aşağıdaki süreç izlenir:



Hildreth-Lu tahmini (1974-2000)
Bağımlı değişken: kettuk

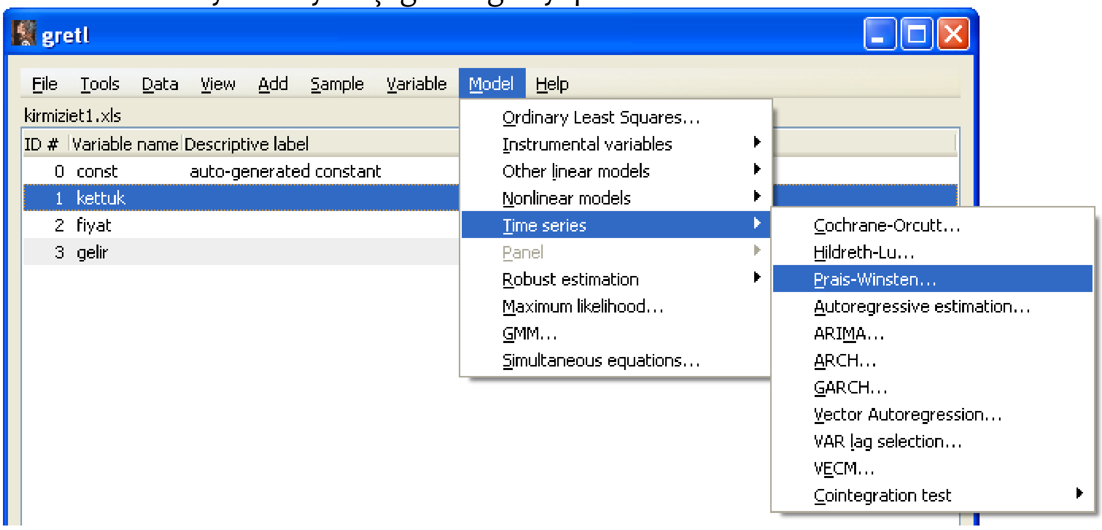
	Katsayı	*St.Hata*	*t*	*p*	
Sabit	114.34	27.5576	4.1491	0.00036	***
Fiyat (X1)	-0.623275	0.114698	-5.4340	0.00001	***
Gelir (X2)	2.87571	2.95922	0.9718	0.34085	

Statistics based on the rho-differenced data:

Mean dependent var	107.4481	S.D. dependent var	9.243350
Sum squared resid	183.1495	S.E. of regression	2.762468
R-squared	0.923190	Düzeltilmiş R-squared	0.916789
F(2, 24)	17.24515	P(F)	0.000023
rho	-0.232347	Durbin-Watson	2.447678

Praise-Winsten

Bu analizi Gretl yardımıyla aşağıdaki gibi yapmak mümkündür:



Prais-Winsten tahmini (1973-2000)
Bağımlı değişken: kettuk

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	52.8864	16.9795	3.1147	0.00458	***
Fiyat (X1)	-0.686824	0.122435	-5.6097	<0.00001	***
Gelir (X2)	8.79386	2.14452	4.1006	0.00038	***

Statistics based on the rho-differenced data:

Mean dependent var	106.6500	S.D. dependent var	10.00561
Sum squared resid	233.1015	S.E. of regression	3.053532
R-squared	0.915239	Düzeltilmiş R-squared	0.908458
F(2, 25)	44.49150	P(F)	5.80e-09
rho	-0.147229	Durbin-Watson	2.256279

Praise Winston tahmin sonuçlarına göre Gelir istatistiki açıdan anlamlıdır. CORC ve Hildreth-Lu'ya göre daha iyidir.

6.4. ÇOKLU BAĞLANTI

6.4.1. Tam ve Zayıf Çoklu Bağlantı

Eğer bağımsız değişkenler arasında tam bağlantı varsa doğrusal regresyonun varsayımlarından biri ihlal edilmiş olur. Bu varsayıma göre bağımsız değişkenlerden biri, diğer bağımsız değişkenlerden biri veya daha fazlasıyla tam bağlantı içinde olamaz şeklindedir. Tam bağlantı, iki değişken arasındaki korelasyon katsayısının -1 veya 1 olması durumudur. Tam bağlantıya nadiren rastlanır. Zayıf çoklu bağlantıda ise bağımsız değişkenlerden en az ikisi arasında yüksek bir korelasyon söz konusudur. Zayıf bağlantı, regresyonun ilgili varsayımı ihlal edilmesine yol açmaz. Çoklu bağlantı sadece iki değişken arasında değil, bazı durumlarda ikiden daha fazla değişken arasındaki bağlantıdan kaynaklanabilir.

6.4.2. Çoklu Bağlantının Sonuçları

Üzerinde çalıştığımız ekonometrik modelde çoklu bağlantı problemi varsa:

Tahminciler yansızdır: Zayıf bir çoklu bağlantı probleminde, tahminciler hala yansızdır; yani gerçek anakitle tahmincilerine yakındır.

Tahmincilerin standart hataları yüksektir: Bu, çoklu bağlantının ana sonuçlarıdır. İki ya da daha çok açıklayıcı değişken önemli ölçüde bağlantılı olduğundan, bu değişkenlerin etkilerini ayrı ayrı belirleme güçlüğü doğar.

Tahmincilerin t değerleri düşer: Tahmincilerin standart hatalarının yükselmesi, doğal olarak t değerlerinin düşmesine neden olur.

Tahminciler model tanımlamasındaki değişmelere karşı çok duyarlıdır: Çoklu bağlantı bulunan durumlarda, modelden açıklayıcı değişken çıkarılması veya modele yeni açıklayıcı değişken eklenmesi ya da yeni gözlemlerin dahil edilmesi tahmincileri önemli ölçüde değiştirecektir.

R^2 ve F yüksektir; buna karşın t testlerinde hipotezler reddedilir: R^2 ve F yüksek ise, tahmincilerin t testlerinde hipotezlerin reddedilmesi beklenir. Çoklu bağlantıda, F testi model istatistiki açıdan anlamlı sonucunu verirken, t testleri bunu doğrulamaz.

6.4.3. Çok Bağlantıyı Belirleme Yöntemleri

Çoklu bağlantıyı belirlemek amacıyla iki yöntem üzerinde duracağız. Bunlar, basit korelasyon katsayıları ve VIF yöntemidir.

Korelasyon Katsayıları

Açıklayıcı değişkenler arasındaki korelasyon katsayıları hesaplanıp, aralarında yüksek korelasyon katsayısına sahip olanlar belirlenir. Bu değişkenlerin çoklu bağlantıya yolaçtığı kabul edilir. Burada korelasyon katsayısı kaçın üzerindeyse çoklu bağlantı olduğu düşünülmesi gerekir sorusu gündeme gelecektir. Bazı araştırmacılar r'nin 0.80'nin üzerinde olması durumunda çoklu bağlantı olabileceğini varsayarlar. Ancak bu değeri α=0.01 düzeyinde t testiyle desteklemek daha uygundur. Korelasyon katsayısının 0.80'den büyük olması, mutlaka çoklu bağlantı olduğuna dair bir işaret olmayabilir.

R^2 Değişiminin İncelenmesi

Çoklu bağlantıyı belirlemede diğer bir yol, modele yeni bağımsız değişkenler ilave edildiğinde, R^2'de nasıl bir değişim olduğunu incelemektir. İlave değişken modelin R^2'sinde bir değişmeye yol açmıyorsa, çoklu bağlantı probleminden kuşkulanılması gerekir.

VIF Yöntemi

Kimi araştırmacılar basit olarak korelasyon katsayılarının incelemenin yetersiz olduğunu düşünürler ve daha formal bir yöntem olan VIF yöntemini önerirler. VIF, açıklayıcı değişkenlerin herbirinin diğer açıklayıcı değişkenler tarafından ne kadar açıklanabileceğini irdeler. Her açıklayıcı değişken için bir VIF değeri hesaplanır. VIF, çoklu bağlantının tahminciye ait varyansı ne kadar artırdığını belirler. VIF'in yüksek olması, tahminciye ait varyansın tni azaltacak şekilde yükseldiğini gösterir.

VIF Süreci

Denklemimiz:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5$$

olsun. Şimdi X_1 için VIF hesaplayalım.

EKK modelini X_1 olmaksızın tahmin et

$$Y = b_0 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5$$

X_1 tahmincisi için VIF hesaplamasını yap:

$$VIF(b_i) = \frac{1}{1 - R_i^2}$$

Buradaki R^2, düzeltilmemiş determinasyon katsayısıdır.

Eğer VIF değeri 10'dan büyükse, çoklu bağlantının nedeni X_1 dir.

Örnek 6-9

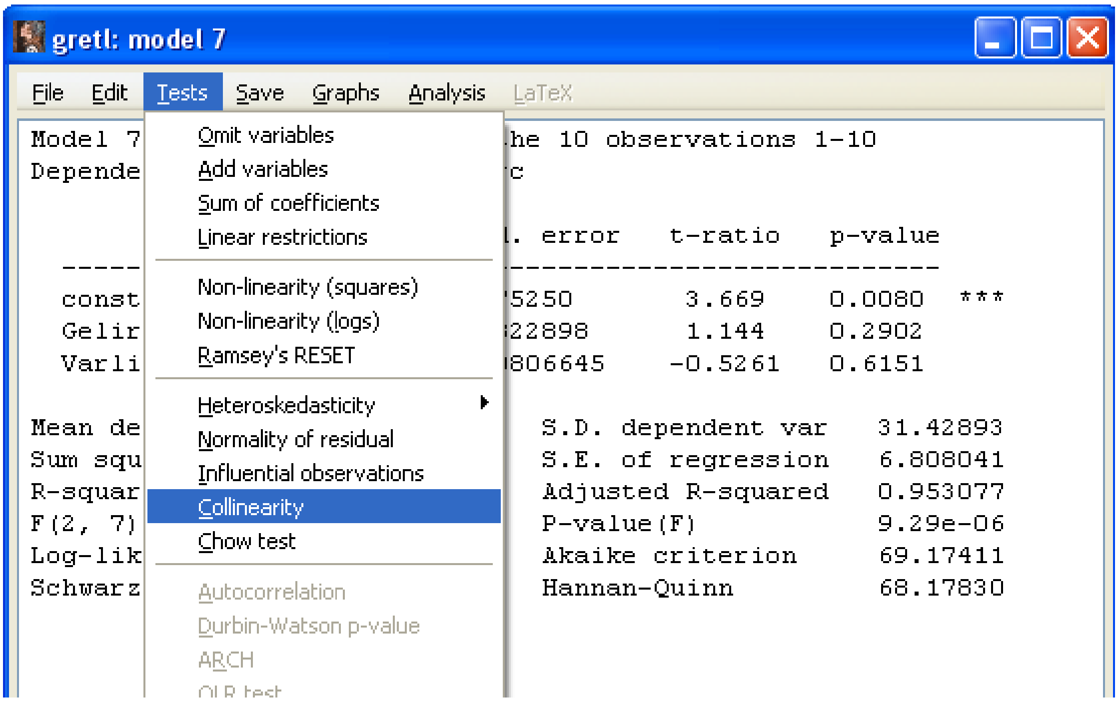
No	Tüketim Harcamaları	Gelir	Varlik
1	70	80	810
2	65	100	1009
3	90	120	1273
4	95	140	1425
5	110	160	1633
6	115	180	1876
7	120	200	2052
8	140	220	2201
9	155	240	2435
10	150	260	2686

Bağımlı değişken: Tuketim Harcamaları

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	24.7747	6.7525	3.6690	0.00798	***
Gelir	0.941537	0.822898	1.1442	0.29016	
Varlik	-0.0424345	0.0806645	-0.5261	0.61509	

R-squared	0.963504	Düzeltilmiş R-squared	0.953077
F(2, 7)	92.40196	P(F)	9.29e-06

F testi, modelin istatistiki açıdan anlamlı olduğunu göstermektedir. Aynı şekilde R^2 çok yüksek olduğu halde, modelde hiçbir değişken istatistiki açıdan anlamlı değildir. Bu durumda çoklu bağlantı probleminden şüphelenmemiz gerekir. VIF çoklu bağlantı testi yapalım:



Variance Inflation Factors

Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

Gelir 482.128
Varlik 482.128

VIF(j) = 1/(1 - R(j)^2), where R(j) is the multiple correlation Katsayı
between variable j and the other inBağımlı değişkens

Properties of matrix X'X:

1-norm = 37022026
Determinant = 2.35068e+009
Reciprocal condition number = 2.727131e-008

VIF değerinin 10'dan büyük olması durumunda, çoklu bağlantı problemi olduğuna dikkat çekilmektedir. Buna göre yukarıdaki analiz sonucundan, Gelir ve Varlık değişkenleri arasında çoklu bağlantı olduğu anlaşılmaktadır. Bu durumda Varlık değişkenini çıkarıp, modelimizi yeniden tahminleyelim:

Bağımlı değişken: Tuketim Harcamaları

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	24.4545	6.41382	3.8128	0.00514	***
Gelir	0.509091	0.0357428	14.2432	<0.00001	***

R-squared	0.962062	Düzeltilmiş R-squared	0.957319
F(1, 8)	202.8679	P(F)	5.75e-07

Örnek 6-10

Bu örnekte kırmızı et talebine ilişkin veriler üzerinde çalışacağız:

Kırmızı et talebi=f(KEtFiyat, TEtFiyat, ErkekGelir, KadinGelir)

modelini tahmin edelim:

Tahmin sonucunu incelediğimizde, standart hataların sıfır ve t değerlerininin tanımsız olduğunu görürüz. Bunun nedeni çoklu doğrusal bağlantıdır. Verilere dikkatle bakıldığında her gözlemde kadingelir değişkeninin erkekgelir değişkeninden 750 birim daha az olduğu görülür. Bu, tam çoklu bağlantı problemine yol açmıştır. VIF testi de bunu doğrulamaktadır:

```
Variance Inflation Factors

Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

       KEtFiyat    1.011
       TEtFiyat    1.023
     ErkekGelir    1.012

VIF(j) = 1/(1 - R(j)^2), where R(j) is the multiple
correlation coefficient
between variable j and the other independent variables

Properties of matrix X'X:

 1-norm = 6.3084348e+008
 Determinant = 4.3569263e+013
 Reciprocal condition number = 4.8531769e-010
```

6.4.4. Çoklu Bağlantıyı Giderme Yolları

Çoklu bağlantı probleminin giderilmesi için bazı çözüm yolları bulunmaktadır. Bunlar:

- Çoklu bağlantıya yol açtığı düşünülen veya yukarıda belirtilen yollardan biriyle tespit edilen bağımsız değişken(ler) modelden çıkarılabilir. Model dışı bırakılan değişkenlerin yanlış tanımlama hatasına yol açmamasına dikkat edilmelidir.
- Çoklu bağlantıya yol açan değişkenlerin toplanması veya birbirinin oranı olarak ifade edilmesiyle yeni bir değişken modele dahil edilebilir.
- Zaman serilerinde fark alma dönüşümü uygulanabilir.
- Gözlem sayısının artırılması çoklu bağlantı problemini giderebilir. Ancak her zaman böyle bir şansa sahip olunamayabilir. Kesit verilerinde gözlem sayısını artırmak daha kolay olmakla birlikte, zaman serilerinde oldukça güçtür.
- Çoklu bağlantı tespit edilen modellerin tahminlenmesinde, Ridge regresyondan veya temel bileşenler regresyon analizinden yararlanılabilir.

7. KUKLA VE KISITLI BAĞIMLI DEĞİŞKENE SAHİP MODELLER

7.1. Doğrusal Olasılık Modeli (LPM)

Doğrusal olasılık modelinde bağımlı değişken kukla, bağımsız değişkenler ise gerek kukla gerekse ölçülebilir değişkenlerdir. Örneğin ev sahibi olma durumunu, gelir ve eğitim düzeyinin bir fonksiyonu olarak ele alalım. Veriler aşağıdaki gibi ise:

EVSAHIBI (Ev sahibi=1)	GELIR TL	EGITIM	$\hat{Y}$
0	500	5	-0.032
0	450	8	0.088
1	3200	15	0.853
1	2200	22	0.982
1	2300	22	0.999
0	650	8	0.122
0	700	8	0.130
1	2750	15	0.777
1	2500	17	0.820
0	500	11	0.225
0	750	11	0.267
0	1000	11	0.309
1	1500	11	0.394
0	900	15	0.463
1	1200	15	0.514
1	1980	15	0.646
1	1850	17	0.710
1	3500	22	1.203
1	5000	17	1.243
1	4000	22	1.287

Bağımlı değişken: evsahibi

	Katsayı	*St.Hata*	*t*	*p*	
sabit	-0.329852	0.213846	-1.542	0.141368	
GELIR	0.000169418	00000.7869	2.153	0.045987	**
EGITIM	0.0427029	0.0198255	2.154	0.045886	**

Bağımlı değişkenin ortalaması = 0.6
Bağımlı değişkenin st sapması = 0.502625
Sum of kare residuals = 1.57263
Hataların st hatası= 0.304151
R-kare = 0.672368
Düzeltilmiş R-kare = 0.633823
F-istatistiği (2, 17) = 17.4437 (p = 7.6e-005)

Model tahmin sonuçlarına göre, gerek gelir gerekse eğitim ev sahibi olma olasılığı üzerinde etkilidir. Gelir düzeyi 100 arttığında ev sahibi olma olasılığı %1.69 artacaktır (0.000169418*100=0.0169). Eğitim düzeyi 1 yıl fazla olanın ev sahibi olma olasılığı ise %4.27 olacaktır. Bu arada evsahibi değişkeninin tahmin sonuçlarını inceleyecek olursak, 8 yıl eğitimli 700 maaşı olan birinin ev sahibi olma olasılığının %13, 17 yıl eğitimli 2.5 milyar maaşlı birinin ev sahibi olma olasılığının ise %82 olduğunu görürüz.

LPM'in yorumlanması kolay olmakla birlikte 3 önemli kusuru vardır:

- Bağımlı değişkenin tahmin değerleri 0'dan küçük veya 1'den büyük olabilir. Halbuki 0-1 arasında olmalıdır.
- Hata terimi farklı varyans gösterebilir.
- Hata terimi normal dağılış göstermeyebilir.

Şimdi yukarıdaki örneğin belirtilen kusurları içerip içermediğini irdeleyelim:

Yukarıdaki tabloda, tahmin değerlerinin eksi olabildiği gibi, 1'den büyük olabildiği de görülmektedir.

Faklı varyans (White testi)

Test istatistiği: TR^2 = 5.128552,

with p = P(Chi-square(4) > 5.128552) = 0.274360

Farklı varyans yoktur.

Normal dağılış

Ki kare (2)=1.557, p=0.46

Normal dağılış göstermektedir.

7.2. Logit Model

DOM'ne alternatif olarak logit model geliştirilmiştir. Logit model, DOM'nin olumsuz özellikleriyle karşılaşılmaz. Logit'in fonksiyonel formu:

$$\ln\left[\frac{P}{1-P}\right] = b_o + b_1 X + u$$

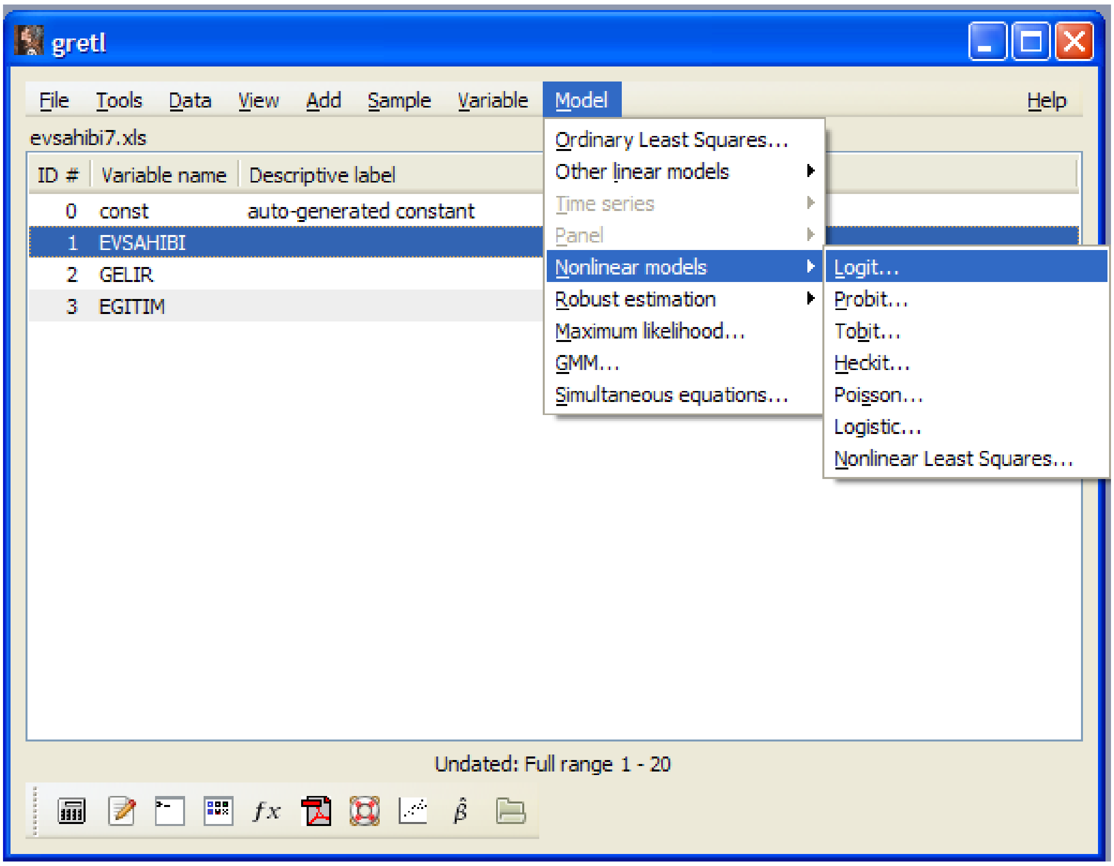
ya da

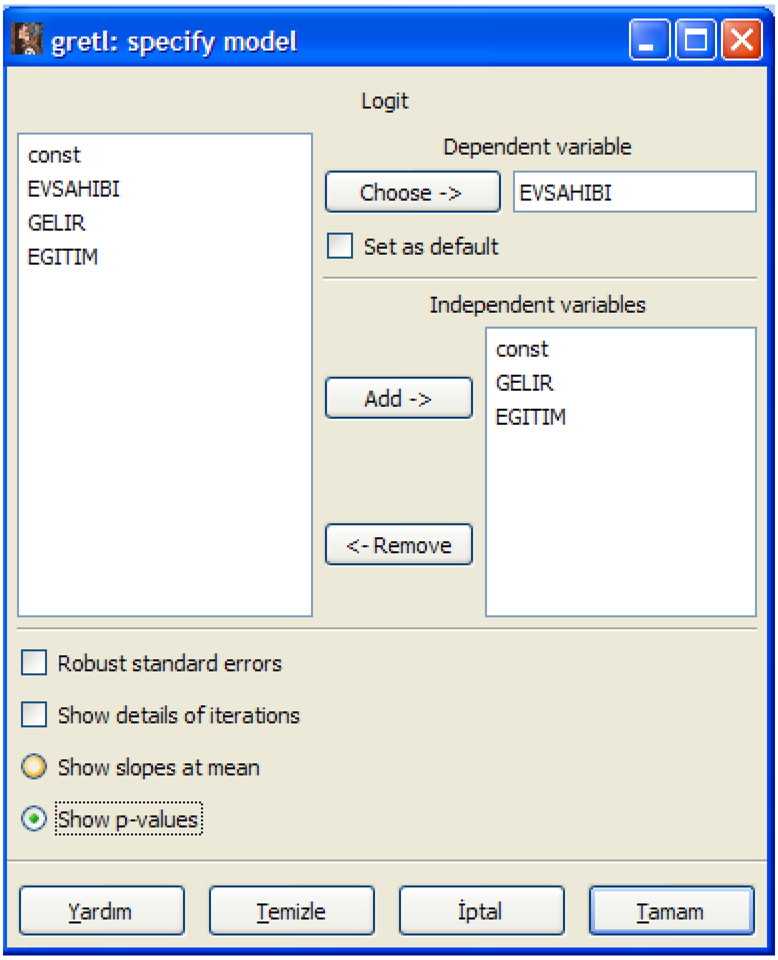
$$P = \frac{1}{1+e^{-(b_o+b_1 X+u)}}$$

EVSAHIBI (Ev sahibi=1)	GELIR TL	EGITIM	$\hat{Y}$
0	450	5	-0.032
0	750	5	0.088
0	900	5	0.853
0	1200	8	0.982
1	1500	11	0.999
1	2300	15	0.122
1	2500	15	0.130
1	1000	8	0.777
0	675	8	0.820

1	3000	11	0.225
1	3200	5	0.267
0	1750	11	0.309
0	900	5	0.394
0	500	8	0.463
1	4000	8	0.514
0	670	8	0.646
1	1100	15	0.710
1	1400	15	1.203
0	600	5	1.243
1	3000	15	1.287

Yukarıdaki örnek için bu kez Logit modelini deneyelim. Tahmin sonuçları aşağıdaki gibidir:



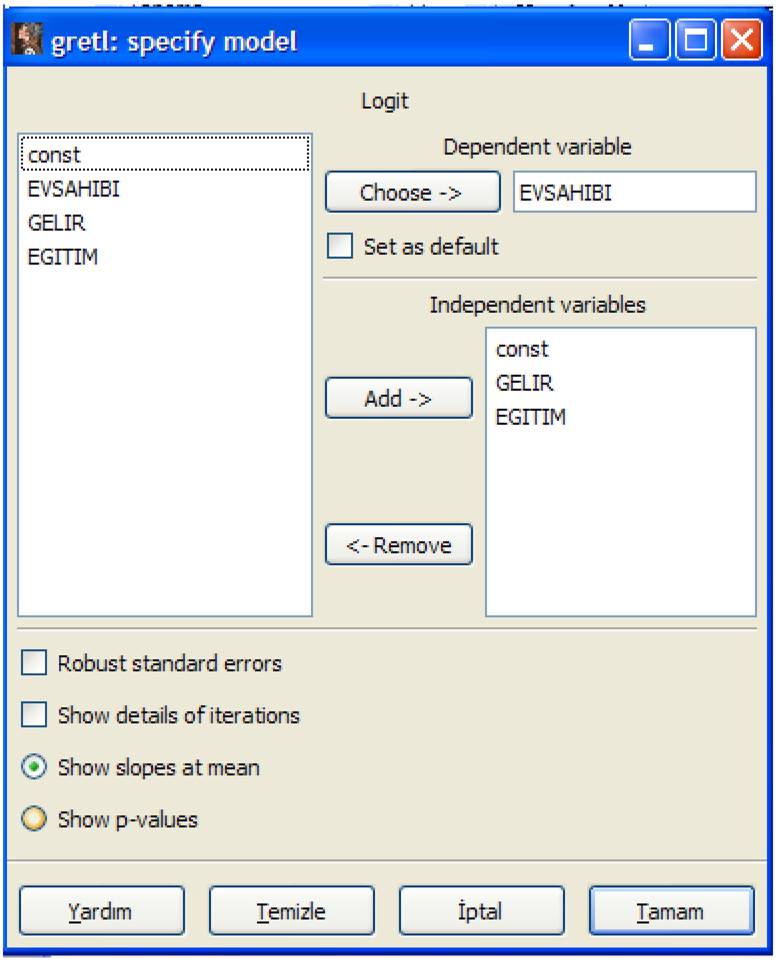


Model 2: Logit estimates using the 20 gözlem 1-20
Bağımlı değişken: EVSAHIBI

	Katsayı	*St.Hata*	*t*	*p*	
sabit	-7.46685	3.15865	-2.3639	0.01808	**
GELIR	0.00220737	0.00118127	1.8686	0.06167	*
EGITIM	0.46416	0.276789	1.6769	0.09355	*

Ortalama EVSAHIBI = 0.500
Doğru tahmin edilmiş gözlem sayısı = 18 (90.0%)
f(beta'x) at Ortalama Bağımsız vars = 0.244
McFadden's pseudo-R^2 = 0.649692
Log-likelihood = -4.8563
Likelihood ratio test: Chi-square(2) = 18.0133 (p 0.000123)

Likelihood ratio testi, modelin istatistiki olarak geçerli olduğunu göstermektedir. GELIR ve EGITIM %10 için anlamlıdır ve beklendiği gibi pozitif yönlü bir ilişkilerdir. Logit modelinin tahmincilerini yorumlamak için öncelikle marjinal etki (eğim=slope) veya esneklik hesaplaması yapmamız gerekir. GRETL, marjinal etkiyi hesaplamak için "show slopes at mean" seçeneğinin işaretlenmesini bekler.



	Katsayı	*St.Hata*	t	Slope*
sabit	-7.4669	3.15865	-2.3639	
GELIR	0.00221	0.00118	1.8686	0.00054
EGITIM	0.46416	0.27679	1.6769	0.11321

*Evaluated at the mean
Ortalama EVSAHIBI = 0.500
Doğru tahmin edilmiş gözlem sayısı =18 (90.0%) (**Tahminde %90 başarı sağlandı)**
f(beta'x) at Ortalama Bağımsız vars = 0.244
McFadden's pseudo-R^2 = 0.649692
Log-likelihood = -4.8563
Likelihood ratio test: Chi-square(2) = 18.0133 (p 0.000123)
Akaike bilgi kriteri (AIC) = 15.7126
Schwarz Bayesian kriteri (BIC) = 18.6998

Hannan-Quinn kriteri (HQC) = 16.2957

Eğim en sağdaki kolonda verilmektedir. Buna göre GELIR düzeyindeki 100 luk artışın, ev sahibi olma olasılığını %5.4 artırdığı görülmektedir. Bir yıl fazladan eğitim, ev sahibi olma olasılığını %11 artırmaktadır.

		Tahmin	
		0	1
Gerçek	0	9	1
	1	1	9

Bu tabloya göre:

- Gerçekte 0 olan 9 gözlem, yine 0 olarak tahmin edilmiştir.
- Gerçekte 1 olan 9 gözlem, yine 1 olarak tahmin edilmiştir.
- Gerçekte 1 olan 1 gözlem, 0 olarak tahmin edilmiştir.
- Gerçekte 0 olan 1 gözlem, 1 olarak tahmin edilmiştir.

Örnek

no	kutusut	gelir	cocuk0_7	kadincalisiyor
1	0	3400	1	1
2	0	1300	0	1
3	1	2200	0	1
4	1	2300	1	0
5	0	1200	1	0
6	0	700	0	0
7	0	1200	0	1
8	1	4300	1	1
9	0	1300	0	0
10	1	2100	1	0
11	0	1000	1	1
12	0	4300	0	0
13	1	600	1	1
14	1	4300	0	0
15	1	3100	1	1
16	0	900	0	0
17	1	3800	0	1
18	0	1300	0	0
19	1	3200	1	1
20	0	700	1	0
21	1	2300	1	1
22	0	500	0	1
23	0	700	0	0
24	1	4200	1	0
25	0	700	1	0
26	0	1200	0	0
27	1	3700	1	1

Model 12: Logit, using gözlem 1-27
Dependent variable: kutusut

	Coefficient	Std. Error	z-stat	p-value		Eğim
const	-4.58442	1.72254	-2.6614	0.00778	***	
gelir	0.00119478	0.000466245	2.5626	0.01039	**	0.000289
cocuko_7	2.04326	1.20763	1.6920	0.09065	*	0.456089
kadincalisiyor	1.37341	1.13937	1.2054	0.22804		0.32219

Mean dependent var	0.444444	S.D. dependent var	0.241921
McFadden R-squared	0.443173	Adjusted R-squared	0.227515
Log-likelihood	-10.32802	Akaike criterion	28.65603
Schwarz criterion	33.83938	Hannan-Quinn	30.19731

Doğru tahmin edilmiş gözlem sayısı = 21 (77.8%)
f(beta'x) at Ortalama Bağımsız vars = 0.242
Likelihood ratio test: Chi-square(3) = 16.4399 [0.0009]

Doğru tahmin edilmiş gözlem sayısı = 21 (77.8%)

f(beta'x) at Ortalama Bağımsız vars = 0.242

Likelihood ratio test: Chi-square(3) = 16.4399 [0.0009]

		Tahmin	
Gerçek		0	1
	0	12	3
	1	3	9

7.3. Probit Model

Probit model, parametreleri doğrusal olmayan (nonlinear) kesikli seçim modelidir. Bu modelin amacı; bağımlı değişken olan P_i seçim olasılığını, bağımsız değişkenlerle, P_i 0-1 arasında olacak şekilde ilişkilendirmektir. Probit modelde her gözlem için bir I_i fayda indeksi geliştirilir (Denklem 1):

$$I_i = \beta_1 + \beta_2 x_{i2} + .. + \beta_k x_{ik} \qquad \text{(Denklem 7-1)}$$

I_i ne kadar büyükse, i bireyinin $y_i = 1$ seçiminden elde edeceği faydanın o kadar büyük olacağı anlamına gelmektedir.

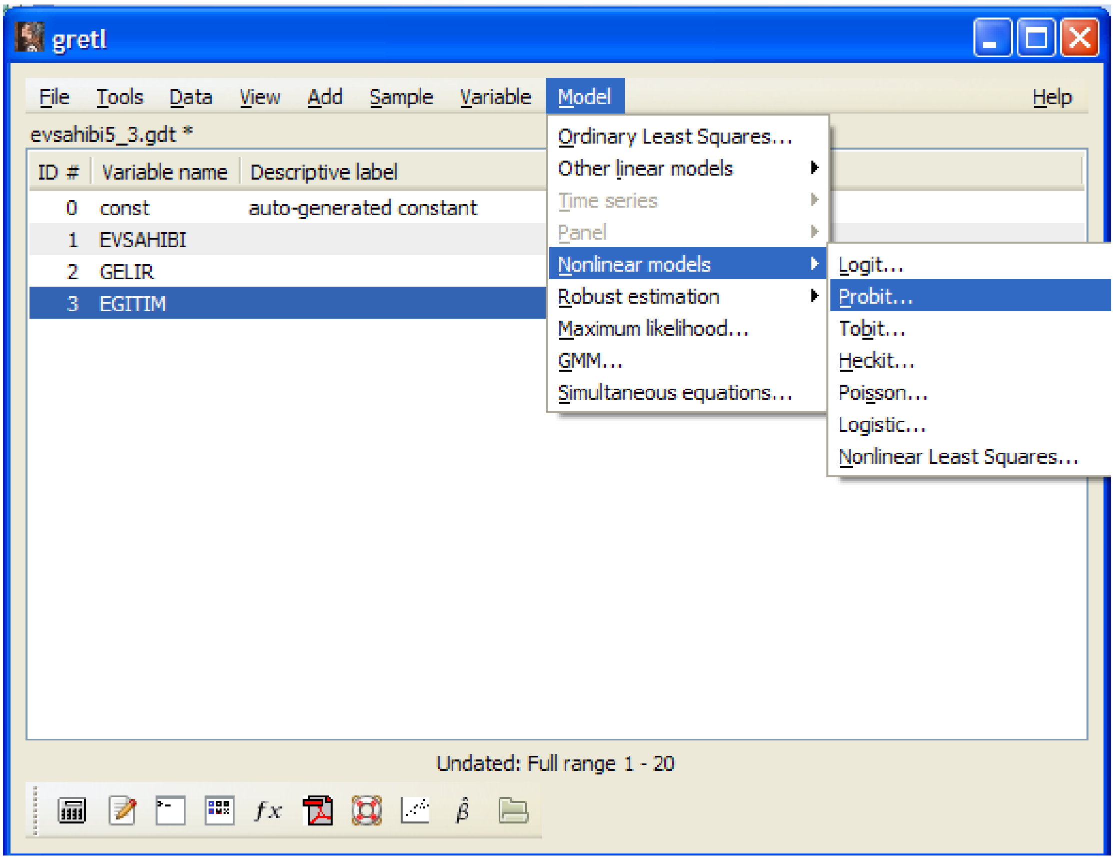
Probit modelin genel gösterilişi denklem 2'de sunulmuştur.

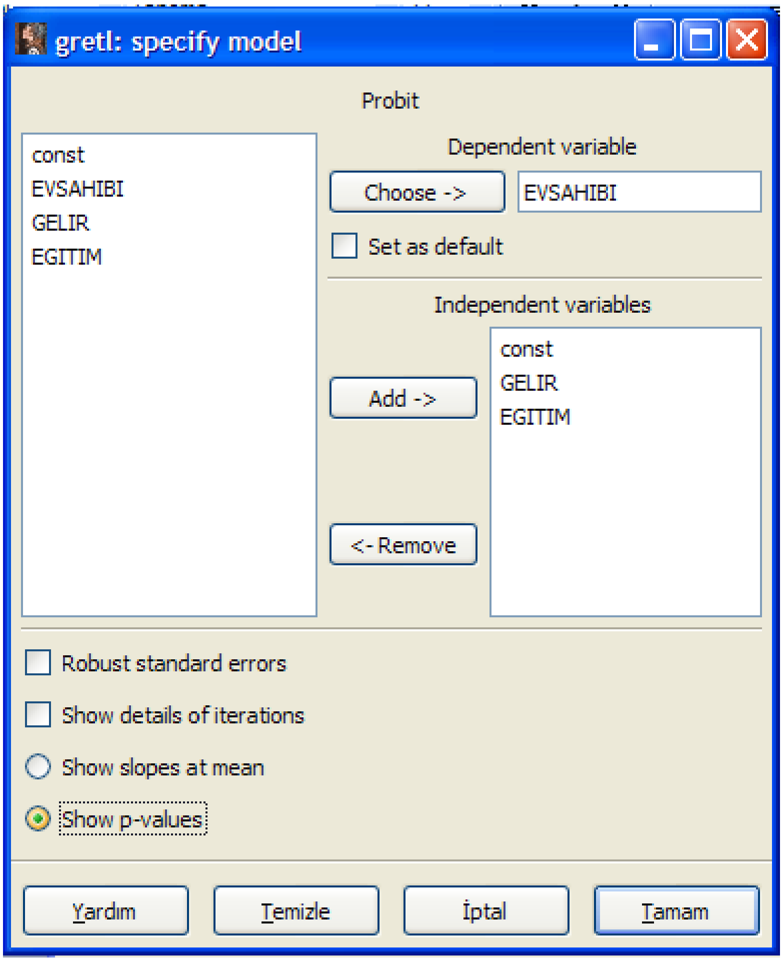
$$P_i = F(I_i) = F(\beta_1 + \beta_2 x_{i2} + .. + \beta_k x_{ik}) = F(x_i' \beta) \qquad \text{(Denklem 7-2)}$$

Burada $F(I_i)$; I_i olarak değerlendirilen standart normal (0,1) tesadüf değişkenine ait eklemeli olasılık fonksiyonudur.

Probit modelde tahminciler ML (maximum likelihood) yöntemiyle elde edilmektedir.

Şimdi yukarıdaki örneği Probit model olarak ele alalım. GRETL çözümü:



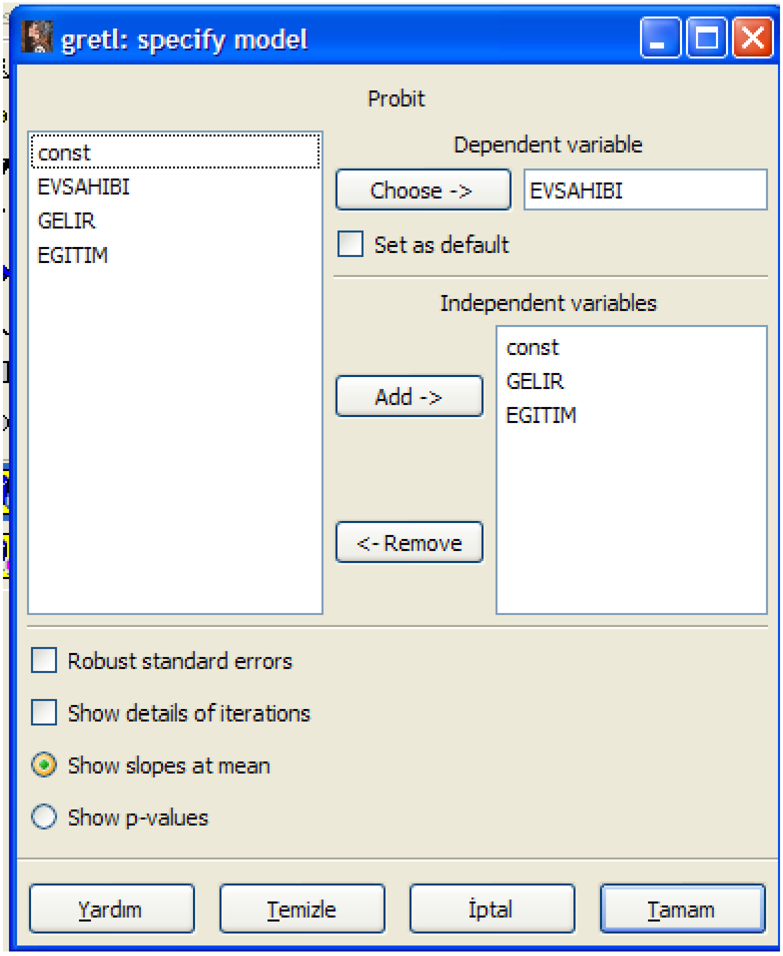


Model 4: Probit estimates using the 20 gözlem 1-20
Bağımlı değişken: EVSAHIBI

	Katsayı	*St.Hata*	*t*	*p*	
sabit	-4.38741	1.6997	-2.5813	0.00984	***
GELIR	0.00129754	0.000666206	1.9477	0.05146	*
EGITIM	0.272338	0.153569	1.7734	0.07616	*

Ortalama EVSAHIBI = 0.500
Doğru tahmin edilmiş gözlem sayısı = 18 (90.0%)
f(beta'x) at Ortalama Bağımsız vars = 0.392
McFadden's pseudo-R^2 = 0.659146
Log-likelihood = -4.72524
Likelihood ratio test: Chi-square(2) = 18.2754 (p 0.000108)
Akaike bilgi kriteri (AIC) = 15.4505
Schwarz Bayesian kriteri (BIC) = 18.4377

Hannan-Quinn kriteri (HQC) = 16.0336



Model 5: Probit estimates using the 20 gözlem 1-20
Bağımlı değişken: EVSAHIBI

	Katsayı	*St.Hata*	*t*	*Slope* *
sabit	-4.38741	1.6997	-2.5813	
GELIR	0.00129754	0.000666206	1.9477	0.000509127
EGITIM	0.272338	0.153569	1.7734	0.10686

*Evaluated at the mean
Ortalama EVSAHIBI = 0.500
Doğru tahmin edilmiş gözlem sayısı = 18 (90.0%)
f(beta'x) at Ortalama Bağımsız vars = 0.392
McFadden's pseudo-R^2 = 0.659146
Log-likelihood = -4.72524
Likelihood ratio test: Chi-square(2) = 18.2754 (p 0.000108)
Akaike bilgi kriteri (AIC) = 15.4505
Schwarz Bayesian kriteri (BIC) = 18.4377

Hannan-Quinn kriteri (HQC) = 16.0336

Likelihood ratio test sonucu, modelin istatistiki olarak geçerli olduğunu göstermektedir. GELIR ve EGITIM %10 için anlamlıdır ve beklendiği üzere pozitif yönlü bir ilişki söz konusudur. Probit modelinin tahmincilerini yorumlamak için öncelikle türev (eğim=slope) veya esneklik hesaplaması yapmamız gerekir. GRETL çözümünde eğim en sağdaki kolonda verilmektedir. Buna göre GELIR düzeyindeki 100 luk artışın, ev sahibi olma olasılığını %5; EGITIM'deki 1 yıllık artış ise ev sahibi olma olasılığını %10.6 artırmaktadır.

Örnek 7-1

Yeni tohumluğu kabul etme eğilimini etkileye faktörler, logit ile analiz edilecektir.

YENITOH	GELIR	ARAZI	EGITIM	DANISMA
0	381	25	5	0
0	290	20	5	0
0	438	34	5	1
0	329	25	5	0
0	469	10	5	1
0	424	12	5	1
1	400	75	5	0
1	482	9	5	1
1	298	28	8	1
1	351	80	8	0
1	309	23	8	0
1	358	56	8	0
0	291	4	8	1
0	434	17	8	1
1	304	35	8	1
1	390	15	8	1
1	449	47	8	1
1	430	50	8	1
0	383	14	11	0
0	416	9	11	0
1	305	75	11	0
1	304	4	11	0
1	348	52	15	0
1	408	18	15	0

Dependent : YENITOH

	Katsayı	*St.Hata*	*z-stat*	*p*	Eğim
Sabit	-3.99539	4.50981	-0.8859	0.37565	
GELIR	-0.00817434	0.0107525	-0.7602	0.44712	-0.00163958
ARAZI	0.0962269	0.0451036	2.1335	0.03289	0.0193008
EGITIM	0.533362	0.280193	1.9036	0.05697	0.10698
DANISMA	1.63061	1.36665	1.1931	0.23281	0.327062

McFadden R-kare 0.428300 Düzeltilmiş R-kare 0.121563

Log-likelihood	-9.319082	Akaike kriteri	28.63816
Schwarz kriteri	34.52843	Hannan-Quinn	30.20085

Doğru tahmin edilmiş gözlem sayısı = 20 (83.3%)
Likelihood ratio test: Chi-square(4) = 13.9631 [0.0074]

		Tahmin	
		0	**1**
Gerçek	**0**	9	1
	1	3	11

Arazi ve eğitim düzeyi yükseldikçe, yeni tohumluğu kabul etme eğilimi artmaktadır.

İşletme arazisinin 1 birim artması, yeni tohumluğu kabul etme eğilimini % 1.9 yükseltmektedir. Eğitimin 1 yıl artması yeni tohumluğu kabul etme eğilimini % 10.6 yükseltmektedir.

Örnek 7-2

Organik ürün tercih etmeyi etileyen tüketici özelliklerini logit model ile belirleyelim. Veriler:

Organik	Gelir	Egitim	Yas	SagBes	Organik	Gelir	Egitim	Yas	SagBes
1	1595	5	43	1	0	1111	1	47	1
1	1701	4	47	1	0	963	2	31	1
1	1320	3	47	1	0	876	2	49	1
1	1478	5	67	1	0	1096	2	24	0
1	1365	5	52	1	0	1018	1	43	0
0	1577	3	61	1	0	1500	2	50	0
0	1751	3	65	1	0	1124	2	37	0
1	1776	5	45	1	0	768	1	22	0
1	1964	4	56	0	0	963	1	41	0
1	1865	4	44	0	0	934	2	37	0
1	1317	5	56	0	0	1038	1	23	0
0	1419	4	65	1	0	1148	2	41	0
1	1853	3	40	1	0	600	2	23	1
1	1321	5	36	1	1	1017	1	28	1
1	1889	3	44	1	1	883	2	30	1
1	1569	4	61	1	1	1119	1	31	1
1	1636	5	42	1					

Model 4: Logit estimates using the 33 gözlem 1-33
Dependent : Organik

	Katsayı	*St.Hata*	*z-stat*	*p*	Eğim
sabit	-10.0318	6.12171	-1.6387	0.10127	
Gelir	0.012522	0.00799348	1.5665	0.11723	0.0022777

Egitim	3.51967	1.93197	1.8218	0.06849	0.640212
Yas	-0.436182	0.251164	-1.7366	0.08245	-0.0793394
SagBes	5.3513	3.21501	1.6645	0.09602	0.973376

McFadden R-kare	0.663326	Düzeltilmiş R-kare	0.444591
Log-likelihood	-7.695927	Akaike kriteri	25.39185
Schwarz kriteri	32.87439	Hannan-Quinn	27.90950

Doğru tahmin edilmiş gözlem sayısı = 28 (84.8%)
Likelihood ratio test: Chi-square(4) = 30.3256 [0.0000]

	Tahmin **0**	**1**
Gerçek **0**	14	2
1	3	14

Logit model:

- Eğitim arttıkça organik ürün tüketme eğilimi artmaktadır.
- İleri yaşlarda organik tüketme eğilimi azalmaktadır. Gençler daha çok organik tüketme eğilimindedir.
- Sağlıklı beslenme haberlerini dikkatle izleyenler daha fazla organin ürün tüketme eğilimindedir.
- Gelir istatistiki olarak anlamlı değildir. Bu, her gelir düzeyinde organik ürün tüketme eğiliminin olabileceği şeklinde yorumlanabilir.

7.4. Tobit Model

Bağımlı değişkenin 0 değerini aldığı durumlarda, doğrusal regresyonun "tahmincilerin yansızlığı (unbiasedness)" koşulu sağlanamamaktadır. Bu nedenle bağımlı değikenin 0 değerini de alabildiği durumlara uygun bir model olan ***tobit*** modelden yararlanılmıştır.

Tobit model, probit modelin bir uzantısıdır. Kısıtlı bağımlı değişken modelleri arasında yer almaktadır (Gujarati, 1995). Bu modelde bağımlı değişken olan Y'nin, pozitif ve negatif veya 0 değerleri arasında bir asimetri bulunmaktadır (Ramanathan, 1998). Tobit modelin genel formülasyonu, probit modelde olduğu gibi bir indeks fonksiyona dayalı olarak verilmektedir (denklem 3) (Greene, 1998).

$$Y_i^* = ß' x_i + u_i, \qquad \text{(Denklem 7-3)}$$

Eğer $Y_i^* <= 0$ ise $Y_i = 0$

Eğer $Y_i^* < 0$ ise $Y_i^* = 0$

Probit modelde olduğu gibi Tobit modelde de tahminciler ML (maximum likelihood) yöntemiyle hesaplanmaktadır.

8. EŞANLI DENKLEM SİSTEMLERİ

Ekonometrik modeller genellikle tek bağımlı değişken üzerine kurulmaktadır. Ancak çoğu ekonomik olayda birden fazla bağımlı değişken söz konusu olabilmektedir. Bu durumda birden fazla denklemin oluşturduğu bir sistemle çalışma zorunluluğu doğmaktadır. İşte bir denklem sisteminde, değişkenlerden bazıları denklemlerde hem bağımlı değişken hem de bağımsız değişken olarak yer alıyorsa, eşanlı denklem sisteminden söz edilir. Talep ve arz denklemlerinin tahmin edilmesi, bunun tipik bir örneğidir. Gerçekten de arz-talep denklem sisteminde, fiyat ve miktar aynı anda belirlenmektedir. Pek çok mikroekonomik ve makroekonomik model de eşanlı denklem sistemi özelliği taşıyabilmektedir. Bazı eşanlı denklem sistemi örnekleri aşağıda verilmiştir:

1.

Çalışma süresi = $\alpha_0 + \alpha_1$ Maaş+ α_2 Eğitim+ α_3Yaş + α_4 KucukCocukVarmı + u
Maaş = $\beta_0 + \beta_1$ Çalışma süresi + β_2 Eğitim + β_3 Deneyim + β_4 Deneyim^2 + v

Konut satın alma = $\alpha_0 + \alpha_1$ Tasarruf + α_2 Gelir + α_3 Eğitim + α_4 Yaş + u
Tasarruf = $\beta_0 + \beta_1$ Konut satın alma + β_2 Gelir + β_3 Eğitim + β_4 Yaş + v

Kişi başına cinayet = $\alpha_0 + \alpha_1$ Kişi başına polis sayısı + α_2 Kişi başına gelir+ u
Kişi başına polis sayısı = $\beta_0 + \beta_1$ Kişi başına cinayet + v

Eşanlı denklem modellerinde değişkenler; içsel (endojen) ve dışsal (egzojen) olmak üzere ikiye ayrılır:

İçsel Değişkenler: Eşanlı denklem sistemindeki bağımlı değişkenlere, içsel (endojen) değişkenler adı verilir. Bu değişkenler sistem tarafından belirlenir. Endojen değişkenler, "ortaklaşa belirlenen" değişkenler olarak da adlandırılır.

Dışsal Değişkenler: İçsel olmayan tüm değişkenler, dışsal değişken başlığı altında toplanır. Bunlar, sistem dışında önceden belirlenen değişkenlerdir. İçsel değişkenlerin geçmiş (gecikmeli) değerleri de önceden belirlenen değişkenler grubuna girer. Dışsal değişkenler önceden belirlenen değişkenler oldukları için, modeldeki hata terimlerinden bağımsızdırlar.

Eşanlı bir modelde her içsel değişken için bir denklem tanımlanır. Ekonomik kurumların davranışlarıyla belirlendiklerinden, bu denklemlere **davranış denklemleri** (veya **yapısal denklemler**) adı verilir. Yapısal denklemleri EKK ile tahmin etmek, yanlı ve tutarsız parametre tahminlerine yol açar. Tutarlı parametre tahminlerini elde edebilmek için, ilk olarak indirgenmiş form denklemleri belirlenir. İndirgenmiş form denklemleri, içsel değişkenlerin her birinin, dışsal değişkenlerin bir fonksiyonu olarak ifade edilmesiyle ortaya çıkar.

Şimdi aşağıda sunulan buğday arz ve talep modelini ele alalım.

$q_d = \alpha_0 + \alpha_1 p + \alpha_2 y + u$ (Denklem 8-1)
$q_s = \beta_0 + \beta_1 p + \beta_2 r + v$ (Denklem 8-2)
$q_d = q_s$ (Denklem 8-3)

q_d: Talep edilen buğday miktarı
q_s: Arz edilen buğday miktarı

p: Buğday fiyatı
y: Gelir
r: Yağış miktarı
u, v: Hata terimleri

Denklem 1 ve Denklem 2 talep ve arza ait yapısal denklemlerdir. Temel ekonomi teorisine göre; denge fiyatı ve denge miktarının, arz ve talebin eşit olduğu durumda ortaya çıkmaktadır. Denklem 3, bu denge durumunu ifade etmek üzere, 1 ve 2 no'lu denklemleri (arz ve talep denklemlerini) birbirine eşitlemektedir.

1, 2 ve 3 nolu denklemler, eşanlı denklem modelinin **yapısal denklemleri**, α ve β'lar **yapısal parametreler**dir. Miktar ve fiyat aynı anda belirlendiğinden, **içsel değişkenlerdir.** Bir başka ifadeyle, fiyat miktarı etkilerken, miktar da fiyatı etkilemektedir. Gelir ve yağış, model tarafından belirlenmeyeceği için **dışsal değişkenler**dir.

Model üç denklemden oluşmakla birlikte;

$$q_d = q_s = q$$

eşitliğini sağlayarak iki denkleme indirgeyebiliriz. Böylece eşanlı denklem modelimiz, iki endojen değişken (miktar ve fiyat) ve üç egzojen değişkenden (sabit terim, gelir ve yağış) oluşan iki denklemli bir model haline gelmiştir. Sistemdeki denklem sayısı G (=endojen değişken sayısı) ile, dışsal değişken sayısı ise K ile gösterilir.

Örnek 8-1

Şimdi, basit Keynes toplam talep modelini ele alacağız:

Denklem 8-4 $C_t = \alpha_0 + \alpha_1 DY_t + \alpha_2 DY_{t-1} + u$
Denklem 8-5 $I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + v$
Denklem 8-6 $DY_t = Y_t - T_t$
Denklem 8-7 $Y_t = C_t + I_t + G_t$

C: Tüketim harcaması
I: Yatırım
Y: GSMH
G: Hükümet harcamaları
T: Toplam vergiler
DY: Harcanabilir gelir

Denklem 8-6, GSMH (Y_t) eksi vergiler (T_t) olarak gösterilmiştir. Bu nedenle, bu denklem **özdeşlik** (veya **tanım denklemi**) olarak adlandırılır. 4 ve 5 no'lu denklemler **davranış denklemleri**, Denklem 8-7 no'lu denklem ise **denge koşulu**dur. Modelimiz, dört yapısal denklemden oluşmaktadır. Bu nedenle 4 endojen değişken (Y, C, I ve DY) söz konusudur. DY_{t-1}, bir önceki dönemin harcanabilir geliridir. t zamanında bu gecikmeli değişkenin değeri bilindiğinden, **önceden bilinen** (predetermined) değişken ismiyle anılır. Buna göre eşanlı denklem modelleri; davranışını açıklamaya çalıştığımız endojen değişkenler, değerleri sistem dışından verilen egzojen değişkenler ve gecikmeli bağımlı değişkenlerin oluşturduğu önceden bilinen değişkenlerden meydana gelir. Karışıklığı önlemek için tüm egzojen değişkenler, önceden bilinen değişkenler sınıfında ele alınabilir. Sonuç olarak modelimizde G=4 ve K=4'tür (sabit terim dahil). Şu anda örneğimizde olmayan diğer bir denklem tipi, **teknik denklem**dir. Örneğin modelimize; toplam arz ile sermaye ve işgücü gibi girdilerin ilişkilendirildiği bir teknik denklem ekleyebilirdik.

Genel olarak, eşanlı bir denklem sisteminde:

- Davranış denklemleri
- Teknik denklemler
- Denge koşulu
- Özdeşlik

denklemleriyle karşılaşabiliriz.

Eşanlı sistemdeki yapısal parametrelerin indirgenmiş formdan yararlanılarak tahmin edilmesinin mümkün olup olmadığı, tanımlama başlığı altında ele alınır. Eşanlı sistem indirgenmiş form yardımıyla tahmin edilebiliyorsa, denklem tanımlıdır, aksi halde denklem tanımsızdır denir. Bir başka ifadeyle, Eğer her yapısal parametre için birer katsayı tahmin ediliyorsa denklem tam tanımlı, birden fazla katsayı tahmin ediliyorsa aşırı tanımlı denir. Buna göre eşanlı denklem sistemi:

Tam tanımlı: Dışsal değişken sayısı = İçsel değişken sayısı – 1

Aşırı tanımlı: Dışsal değişken sayısı > İçsel değişken sayısı – 1

Eksik tanımlı: Dışsal değişken sayısı < İçsel değişken sayısı – 1

şeklinde tanımlanır.

Eşanlı denklem sisteminin tahminlenmesinde indirgenmiş form denklemlerinden yararlanlılır. Denklemlerimiz:

Denklem 8-8 $q_d = \alpha_0 + \alpha_1 p + \alpha_2 y + u$

Denklem 8-9: $q_s = \beta_0 + \beta_1 p + \beta_2 r + v$

olsun. Burada, q_d, Talep edilen buğday miktarı; q_s, arz edilen buğday miktarı; p, buğday fiyatı; y, gelir; r, yağış miktarı.

Denklem 8-8 ve **Denklem 8-9** no'lu denklemleri p için çözersek:

Denklem 8-10: $$p = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\alpha_2}{\alpha_1 - \beta_1} y + \frac{\beta_2}{\alpha_1 - \beta_1} r + \frac{v - u}{\alpha_1 - \beta_1}$$

Bir başka ifadeyle:

Denklem 8-11: $p = \lambda_0 + \lambda_1 y + \lambda_2 r + e_1$

elde edilir.

Bunu denklem **Denklem 8-8**'de yerine koyarsak, denge miktarını (q) elde ederiz:

$q = (\alpha_0 + \alpha_1\lambda_0) + (\alpha_1 + \alpha_2)y + \alpha_1\lambda_2 r + e_2$

veya

Denklem 8-12 $q = \mu_0 + \mu_1 y + \mu_2 r + e_2$

olur. Görüldüğü gibi, her içsel değişken için bir indirgenmiş form denklemi söz konusudur. İndirgenmiş form denklemleri, içsel değişkenlerin her birinin, sadece dışsal değişkenlerin bir fonksiyonu olarak ifade edilmesidir. Buna göre örneğimiz için indirgenmiş form denklemlerini bir kez daha yazalım:

$\mathbf{p} = \lambda_0 + \lambda_1 \mathbf{y} + \lambda_2 \mathbf{r} + e_1$

$$\mathbf{q} = \mu_0 + \mu_1\mathbf{y} + \mu_2\mathbf{r} + e_2$$

8.1. Hausman Eşanlılık Testi

Hausman eşanlılık testi, eşanlılığı belirlemek amacıyla yapılır. Bunun için tahminciler doğrudan karşılaştırmaya tabi tutulur. Hausman testini gerçekleştirmek üzere, satış ve reklama dönük iki yapısal model üzerinde duracağız:

Satış: Satılan Ürün Sayısı = **a** + **b** Reklam Sayısı + **c** Ürün Fiyatı

Reklam: Reklam Sayısı = **d** + **e** Satılan Ürün Sayısı + **f** Reel reklam fiyatı

$$S_t = a + bM_t + cP_{s,t} + u_t$$

$$M_t = d + eS_t + fP_{m,t} + w_t$$

S: Satılan ürün sayısı

M: Reklam sayısı

P_s : Reel satış fiyatı

P_m : Reel reklam fiyatı

u, w : Hata terimleri

Satılan ürün sayısı ve reklam sayısı, içsel değişkenler; satış fiyatı ve reklam fiyatı dışsal değişkenlerdir. Buna göre indirgenmiş form denklemleri:

Satılan ürün sayısı: $S_t = \beta_0 + \beta_1 P_{s,t} + \beta_2 P_{m,t} + \omega_t$

Reklam sayısı: $M_t = \alpha_0 + \alpha_1 P_{s,t} + \alpha_2 P_{m,t} + \nu_t$

şeklinde olacaktır. Hatırlanacağı gibi yapısal denklemler dışsal değişkenlere göre tanımlandığında indirgenmiş form denklemlerine ulaşılmaktadır.

Hausman testi oldukça kolaydır.

1. Önce M_t'nin indirgenmiş form tahminini yapıp, $\hat{M}_t$ değerlerini elde edelim.
2. Satış denklemine $\hat{M}_t$ değişkenini ekleyelim;

   $$S_t = a + bM_t + cP_{s,t} + h\hat{M}_t + u_t$$

   ve bu denklemi EKK ile tahmin edelim.
3. Hausman testi için H_0 hipotezini oluşturalım:

   H_0: Eşanlılık yok
4. $\hat{M}_t$, satış denklemine önemli bir katkıda bulunuyorsa h parametresi istatistiki olarak önemli bulunacaktır. Bu, Hausman testinde sıfır hipotezinin reddedilmesi anlamına gelecek ve eşanlı tahmin yapılması gerektiği sonucuna varılacaktır. Bir başka ifadeyle, eşanlılık varsa, 2AEKK yönteminin kullanılması gerekecektir.

8.2. İki Aşamalı En Küçük Kareler Yöntemi (2AEKK)

2AEKK, eşanlı denklem sistemindeki tam veya aşırı tanımlı denklemler için tutarlı yapısal parametre değerlerini tahmin etmede kullanılan bir yöntemdir. Tam tanımlı denklemlerde DEKK ve 2AEKK yöntemleri aynı sonuçları verir.

2AEKK, EKK'in iki aşamada uygulanması sürecidir. İlk aşamada, içsel değişkenlerin herbiri, sistemdeki dışsal değişkenlerle tahminlenir. Bu denklemler, indirgenmiş formdaki denklemlerdir. Dışsal değişkenlerin gözlenen değerleri indirgenmiş form denklemlerde yerlerine konarak içsel değişkenlerin tahmin edilen değerleri elde edilir. İkinci aşamada, tahmin edilecek içsel değişkenin (bağımlı değişken) gözlenen değerleri ile diğer içsel değişkenin tahmin edilen değerleri ve dışsal değişkenlerin gözlenen değerleri kullanılarak modelin yapısal denklemleri tahmin edilir. Artık içsel değişkenler hata terimleriyle korelasyon halinde değildir ve yapısal parametre tahminleri tutarlıdır.

2AEKK'in DEKK'den üstünlüğü, 2AEKK'in eşanlı denklem sistemindeki tam tanımlı denklemler kadar aşırı tanımlı denklemler için tutarlı parametre tahminlerinin elde edilmesi amacıyla kullanılabilmesidir.

Örnek 8-2

Aşağıda verilen arz-talep modeli için:

- Arz ve/veya talep fonksiyonlarının tam, eksik ya da aşırı tanımlanma durumlarını belirleyelim
- İndirgenmiş form denklemlerini bulalım
- Miktar (Q) için yapısal parametre formüllerini elde edelim

Talep: $Q=a_0 + a_1P + a_2Y + e$

Arz: $Q=b_0 + b_1P + a_2T + u$

Bağımlı değişken sayısı = 2 (Q ve P)

Dışsal değişken sayısı = 1 Y

Dışsal değişken sayısı, bağımlı değişken sayısının bir eksiğine eşit olduğundan, arz fonksiyonu tam tanımlıdır. Talep fonksiyonu da aynı şekilde tam tanımlıdır. Şimdi indirgenmiş form denklemlerini elde edeceğiz. Bunun için endojen değişkenlerin herbirini (Q ve P), dışsal değişkenlerin fonksiyonu olarak tanımlarız:

$$Q= \pi_0 + \pi_1 Y + \pi_2 T + e$$

$$P = \pi_3 + \pi_4 Y + \pi_5 T + e$$

Yıl	t	Miktar (Q)	Fiyat (P)	Gelir Y	Yıl	t	Miktar (Q)	Fiyat (P)	Gelir Y
1969	1	76	103	2386	**1984**	16	99	103	3152
1970	2	78	118	2408	**1985**	17	95	106	3274
1971	3	81	119	2434	**1986**	18	100	100	3371
1972	4	81	107	2491	**1987**	19	103	100	3464
1973	5	79	108	2476	**1988**	20	104	97	3515
1974	6	82	103	2517	**1989**	21	100	100	3619
1975	7	82	104	2643	**1990**	22	112	108	3714
1976	8	80	100	2650	**1991**	23	113	114	3837
1977	9	89	99	2636	**1992**	24	119	175	4062
1978	10	89	98	2696	**1993**	25	110	224	3973
1979	11	93	99	2697	**1994**	26	121	201	4025

1980	12	91	101	2725	**1995**	27	121	197	4144
1981	13	92	103	2796	**1996**	28	130	192	4285
1982	14	96	107	2849	**1997**	29	131	204	4449
1983	15	93	106	3009	**1998**	30	144	223	4509

1. Önce P'yi dışsal değişkenler olan Y ve T ile tahminleyelim.

$P = \pi_0 + \pi_1 Y + \pi_2 T + e$
$P = -191 + 130\,Y - 6.51\,T$

Bağımlı değişken: P

	Katsayı	*St.Hata*	*t*	*p*	
sabit	-191.11	68.9641	-2.7712	0.00999	***
Y	0.129929	0.0334745	3.8814	0.00061	***
time	-6.50667	2.6182	-2.4852	0.01944	**

$R^2 = 0.67465$
Düzeltilmiş $R^2 = 0.65055$
F-istatistiği (2, 27) = 27.994 ($p < 0.00001$)

2. Şimdi de Q'yu dışsal değişkenler olan Y ve T ile tahminleyelim.

$Q = \pi_3 + \pi_4 Y + \pi_5 T + e$
$Q = 28.7 + 19.9\,Y + 0.418\,T$

Bağımlı değişken: Q

	Katsayı	*St.Hata*	*t*	*p*	
sabit	28.713	12.2793	2.3383	0.02703	**
Y	0.0199196	0.0059602 6	3.3421	0.00245	***
time	0.417796	0.466181	0.8962	0.37806	

$R^2 = 0.94000$
Düzeltilmiş $R^2 = 0.93556$
F-istatistiği (2, 27) = 211.503 ($p < 0.00001$)

3. $\hat{P}$ ve $\hat{Q}$ elde edelim:

T	$\hat{P}$	$\hat{Q}$	**T**	$\hat{P}$	$\hat{Q}$
1	112.394	76.659	16	114.319	98.184
2	108.746	77.515	17	123.664	101.032
3	105.617	78.451	18	129.76	103.382

4	106.516	80.004	19	135.337	105.653
5	98.061	80.123	20	135.457	107.086
6	96.881	81.357	21	142.463	109.576
7	106.745	84.285	22	148.299	111.886
8	101.148	84.842	23	157.774	114.754
9	92.823	84.981	24	180.501	119.653
10	94.112	86.594	25	162.431	118.298
11	87.735	87.032	26	162.681	119.752
12	84.866	88.007	27	171.635	122.540
13	87.585	89.839	28	183.449	125.767
14	87.964	91.313	29	198.25	129.451
15	102.246	94.918	30	199.54	131.064

4. Q'yu diğer içsel değişken ve orijinal denklemdeki Y dışsal değişkeni ile tahmin edelim.

$$Q = \alpha_0 + \alpha_1 \hat{P} + \alpha_2 Y + e$$

$$Q = 16.4 - 0.0642\ \hat{P} + 28.3\ Y$$

	Katsayı	*St.Hata*	*t*	*p*	
sabit	16.442	4.547	3.62	0.001	**
ptah	-0.06421	0.07165	-0.9	0.378	***
Y	28.262	3.685	7.67	0	

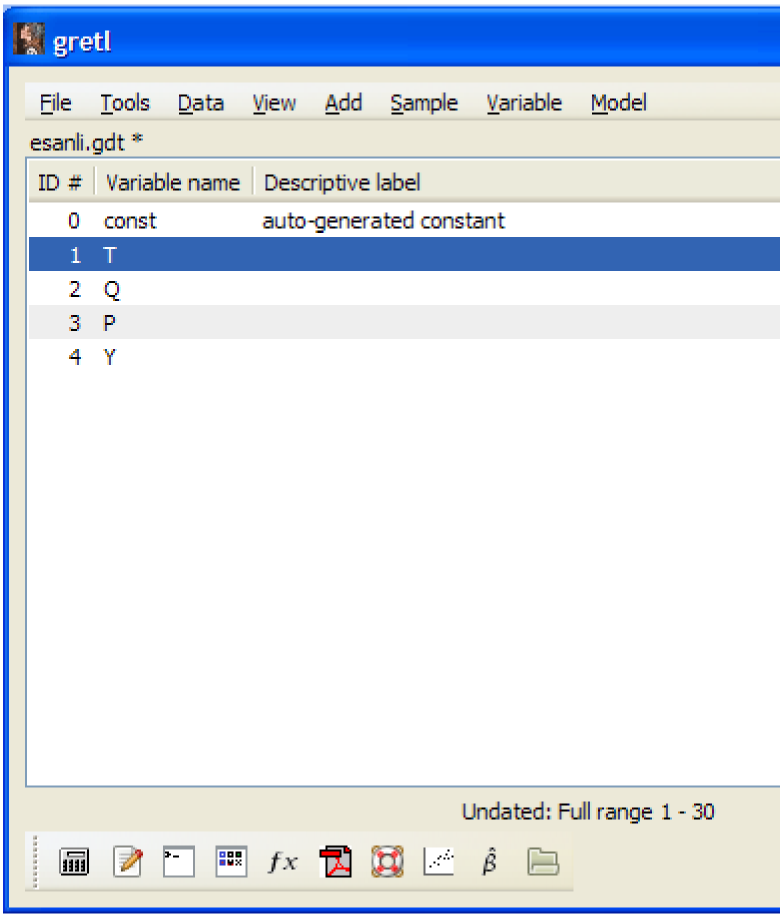
S = 4.538 R-Sq = 94.0% R-Sq(adj) = 93.6%

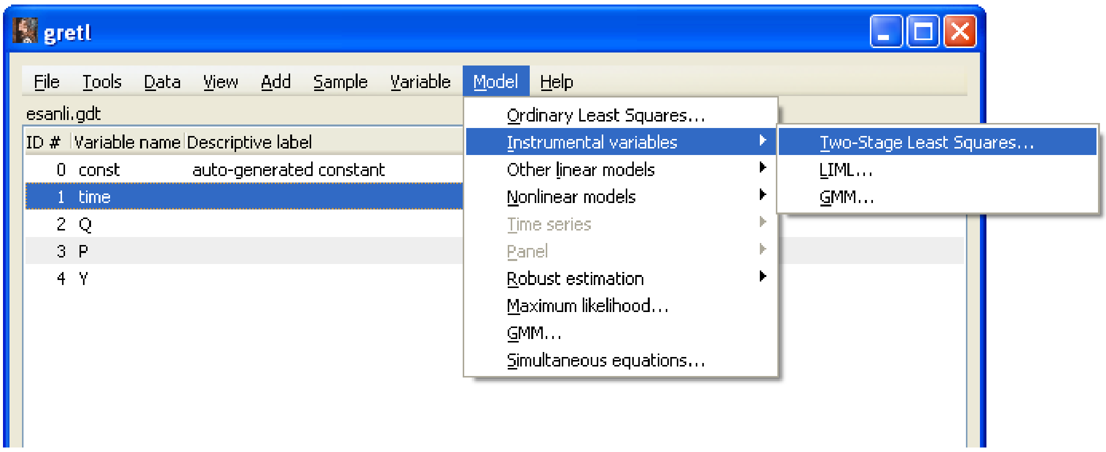
Fiyat ($\hat{P}$) değişkeninin tahmincisi istatistiki açıdan önemsizdir. Bunun anlamı fiyat-miktar ilişkilerinde talep fonksiyonu yerine arz fonksiyonunu tercih etmemiz gerekebileceğidir. Gerçekten de arz için çözdüğümüzde, 2AEKK sonuçları beklenen sonucu vermektedir:

	Katsayı	*St.Hata*	*t*	*p*	
sabit	58.0123	4.08858	14.189	0.0000	***
Time	1.41534	0.19876	7.121	0.0000	***
P	0.153311	0.04941	3.103	0.004458	***

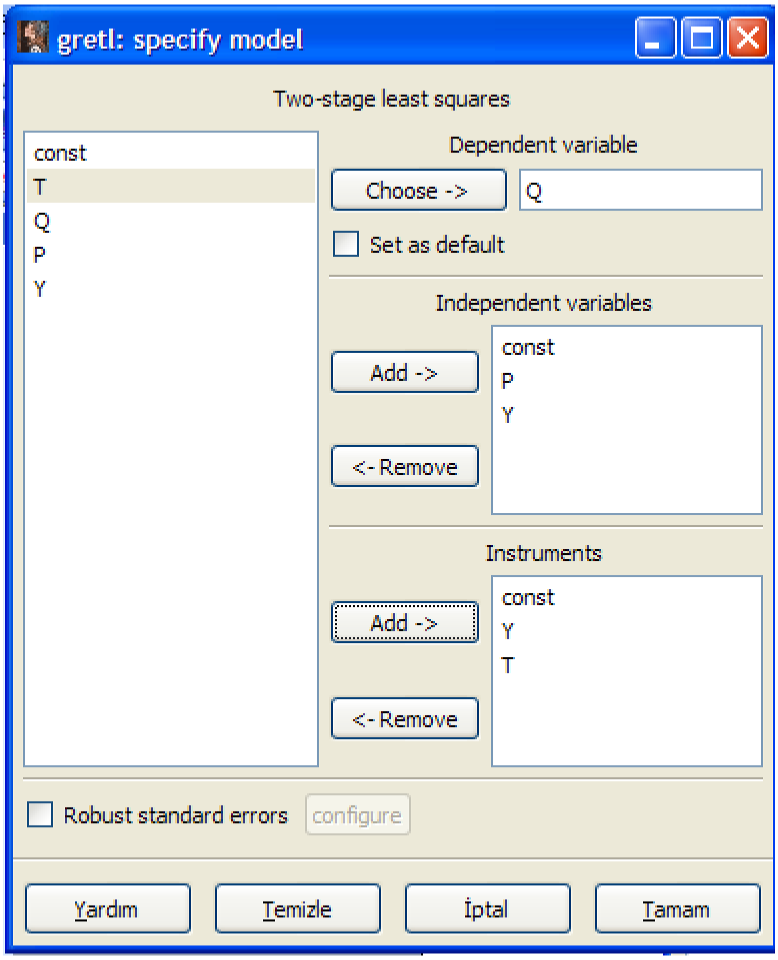
R-kare = 0.931137
Düzeltilmiş R-kare = 0.926036
F-istatistiği (2, 27) = 182.54 (p < 0.00001)

Şimdi bu örneği GRETL ile çözeceğiz. GRETL'da 2AEKK tahminini yaparken, indirgenmiş form denklemlerini tahmin etmemiz gerekmez. Doğrudan 2AEKK tahminlemesi yapılır. Bunun için içsel değişkenlerden birine ait denklemi dikkate almamız yeterlidir. İçsel değişkenlerin her biri için 2AEKK tahminlemesi yapılabilir.





GRETL'de 2AEKK tahminlemesi yaparken, Bağımsız değişkenler (InBağımlı değişkens) bölümüne ilgili denklemdeki sağ taraf değişkenleri; instruments bölümüne ise dışsal değişkenler eklenir. Burada önce sistemdeki talep denklemini dikkate alacağız. 2AEKK tahminlemesinden çözüm alabilmek için, instruments bölümündeki değişkenlerin sayısının, bağımsız değişkenler bölümündeki değişkenlerin sayısından az olmamalıdır.



Bağımlı değişken: Q
Alet değişkenleri: sabit Y T

	Katsayı	*St.Hata*	*t*	*p*	
sabit	16.4418	5.32852	3.0856	0.00203	***
P	-0.0642104	0.0839672	-0.7647	0.44445	
Y	0.0282624	0.00431855	6.5444	<0.00001	***

F-istatistiği (2, 27) = 150.798 (p < 0.00001)

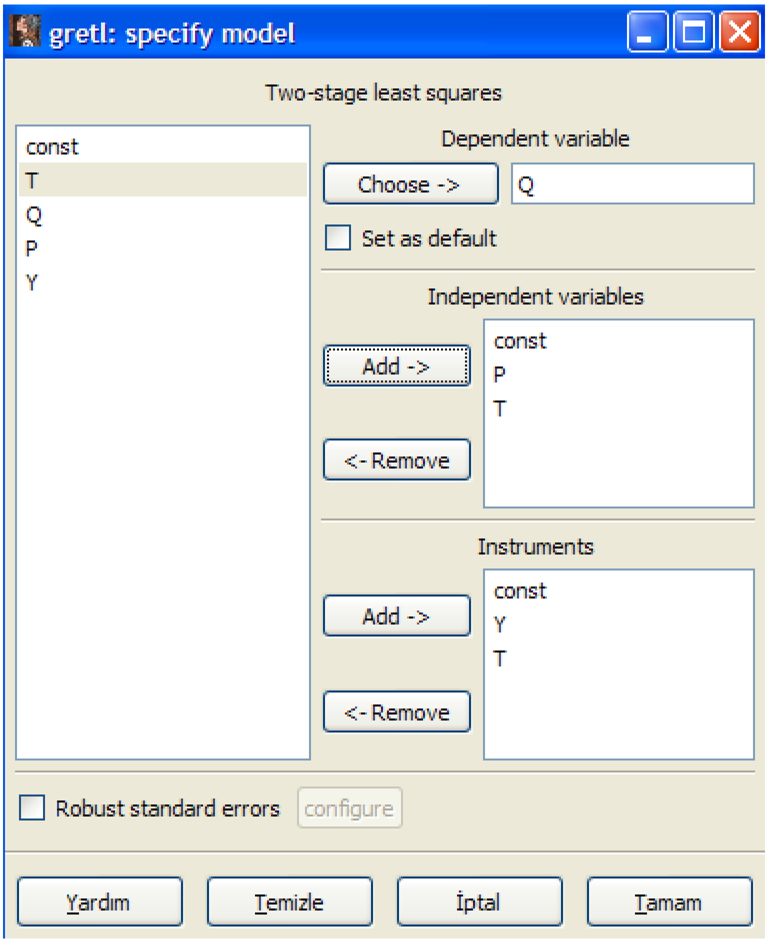
Hausman test –

Null hypothesis: EKK estimates are consistent
Asimtotik test istatistiği: Chi-square(1) = 3.07221
with p = 0.0796413
First-stage F-istatistiği (1, 27) = 6.17605

Talep yönüyle tahmin ettiğimiz 2AEKK sonuçlarına göre, P istatistiki açıdan önemsizdir. Oysa ki fiyatı önemli bulmamız gerekmektedir. Hausman testi, alfa

%510 için, sıfır hipotezini reddetmektedir. O halde, eşanlılık vardır. Bunun anlamı, EKK ile tek denklemli tahminleme yapılmaması gerektiğidir.

Şimdi bir de arz cephesinden bakalım:



Model 4: TSLS estimates using the 30 gözlem 1-30
Bağımlı değişken: Q
Alet değişkenleri: sabit Y T

	Katsayı	St.Hata	t	p	
sabit	58.0123	4.08858	14.1889	<0.00001	***
P	0.153311	0.0494096	3.1029	0.00192	***
T	1.41534	0.19876	7.1209	<0.00001	***

F-istatistiği (2, 27) = 182.54 (p < 0.00001)

Hausman test -
Null hypothesis: EKK estimates are consistent
Asimtotik test istatistiği: Chi-square(1) = 3.32314
with p = 0.0683113
First-stage F-istatistiği (1, 27) = 15.0655

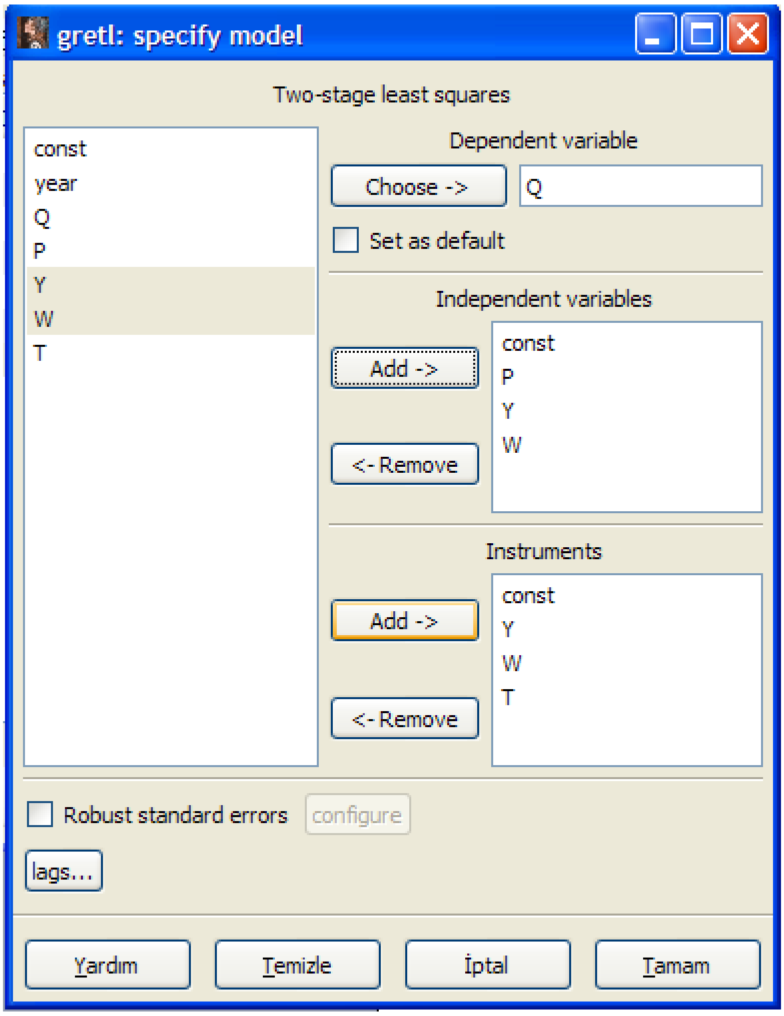
Görüldüğü gibi bu kez fiyat istatistik açıdan anlamlıdır. Arz denkleminden yararlanmak yerinde olacaktır. Hausman testi yine eşanlılığın olmadığını göstermektedir.

Örnek 8-3

Talep: $Q=a_0 + a_1P + a_2Y + a_3W + e$
Arz: $Q=b_0 + b_1P + b_2T + u$
Miktar (Q) için 2AEKK tahminini elde edelim.

Yil	T	Q (cikti)	P (fiyat)	Y (gelir)	W (top.varlik)
1970	1	81	119	2434.0	269.1
1971	2	81	107	2491.0	284.6
1972	3	79	108	2476.0	295.3
1973	4	82	103	2517.0	314.8
1974	5	82	104	2643.0	325.4
1975	6	80	100	2650.0	338.0
1976	7	89	99	2636.0	354.4
1977	8	89	98	2696.0	373.3
1978	9	93	99	2697.0	386.8
1979	10	91	101	2725.0	410.7
1980	11	92	103	2796.0	442.1
1981	12	96	107	2849.0	479.3
1982	13	93	106	3009.0	515.5
1983	14	99	103	3152.0	559.6
1984	15	95	106	3274.0	587.3
1985	16	100	100	3371.0	638.3
1986	17	103	100	3464.0	696.8
1987	18	104	97	3515.0	722.7
1988	19	100	100	3619.0	769.8
1989	20	112	108	3714.0	854.9
1990	21	113	114	3837.0	966.8
1991	22	119	175	4062.0	1086.1
1992	23	110	224	3973.0	1174.2
1993	24	121	201	4025.0	1295.6
1994	25	121	197	4144.0	1428.4
1995	26	130	192	4285.0	1598.7
1996	27	131	204	4449.0	1775.3
1997	28	144	223	4509.0	1957.7

Önce Talep açısından 2AEKK tahminini yapacağız:



Model 3: TSLS estimates using the 28 gözlem 1970-1997
Bağımlı değişken: Q
Alet değişkenleri: sabit Y W T

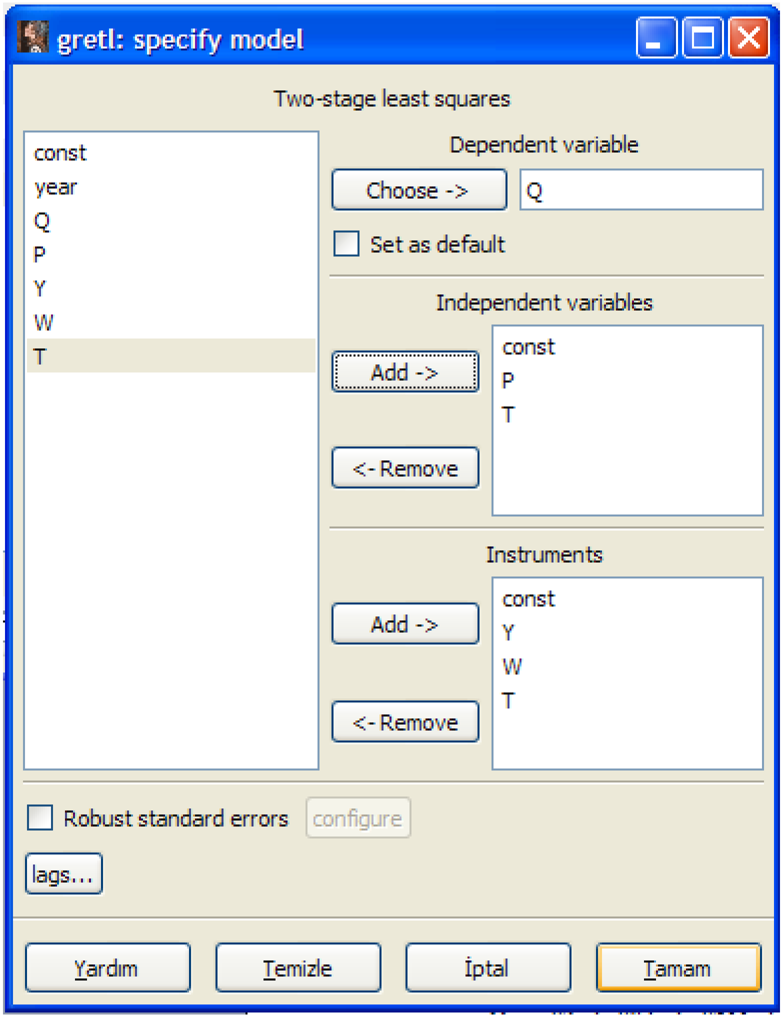
	Katsayı	St.Hata	t	p	
sabit	112.032	52.3197	2.1413	0.03225	**
P	-0.360657	0.294687	-1.2239	0.22100	
Y	-0.00537623	0.0145926	-0.3684	0.71256	
W	0.0710684	0.0420895	1.6885	0.09131	*

F-istatistiği (3, 24) = 60.0807 (p < 0.00001)
Durbin-Watson istatistiği = 0.941309

Hausman test -
Null hypothesis: EKK estimates are consistent
Asimtotik test istatistiği: Chi-square(1) = 5.34083
with p = 0.0208316
First-stage F-istatistiği (1, 24) = 1.36252

Hausman testine göre, eşanlılık var. EKK kullanılmamalıdır.

Şimdi de arz yönüyle 2AEKK tahminlemesini yapalım:



Bağımlı değişken: Q
Alet değişkenleri: sabit Y W T

	Katsayı	St.Hata	t	p	
sabit	59.5831	3.67236	16.2248	<0.00001	***
P	0.171239	0.0426064	4.0191	0.00006	***
T	1.34373	0.209449	6.4155	<0.00001	***

Hausman test -
Null hypothesis: EKK estimates are consistent
Asimtotik test istatistiği: Chi-square(1) = 34.8322
with p = 3.59374e-009

Eşanlılık var. EKK tahminlemesi uygun değil.

Sargan over-identification test -
Null hypothesis: all instruments are valid
Test istatistiği: LM = 0.174672
with p = P(Chi-Square(1) > 0.174672) = 0.675993
First-stage F-istatistiği (2, 24) = 20.7567

Sargan testi, dışsal değişken sayısının tahmine zarar vermediğini söylemektedir.

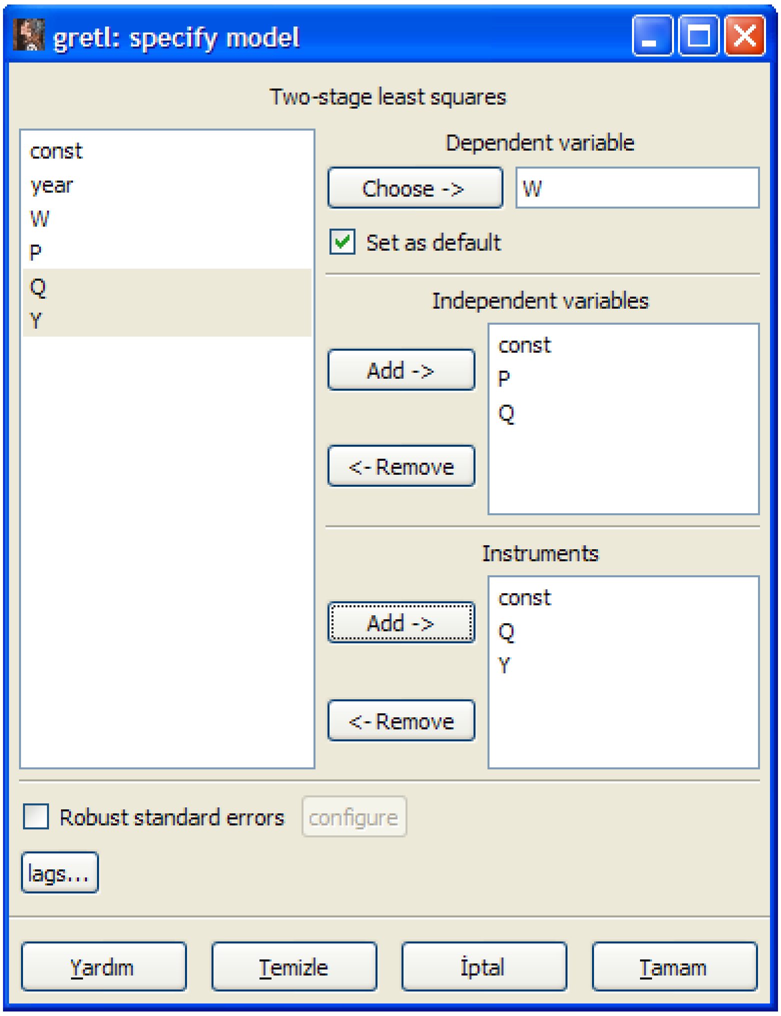
Örnek 8-4

Aşağıda iki denklemden oluşan basit bir ücret-fiyat modeli sunulmuştur:

Ücret: $W=a_0 + a_1P + a_2Q + e$

Fiyat: $P=b_0 + b_1W + b_2Y + u$

Yil	w (ucret)	p (fiyat)	q (cikti/saat)	y (gsmh)
1980	2.09	88.7	80.9	506
1981	2.14	89.6	83.0	523.3
1982	2.22	90.6	86.6	563.8
1983	2.28	91.7	89.6	594.7
1984	2.36	92.9	92.8	635.7
1985	2.46	94.5	95.9	688.1
1986	2.56	97.2	98.4	753
1987	2.68	100	100.0	796.3
1988	2.85	104.2	103.2	868.5
1989	3.04	109.8	102.9	935.5
1990	3.23	116.3	103.0	982.4
1991	3.45	121.3	106.2	1063.4
1992	3.7	125.3	110.1	1171.1
1993	3.94	133.1	112.0	1306.6
1994	4.24	147.7	108.5	1412.9
1995	4.53	161.2	110.5	1528.8
1996	4.86	170.5	114.4	1702.2
1997	5.25	181.5	116.2	1899.5
1998	5.69	195.4	116.8	2127.6
1999	6.16	217.4	115.5	2368.5

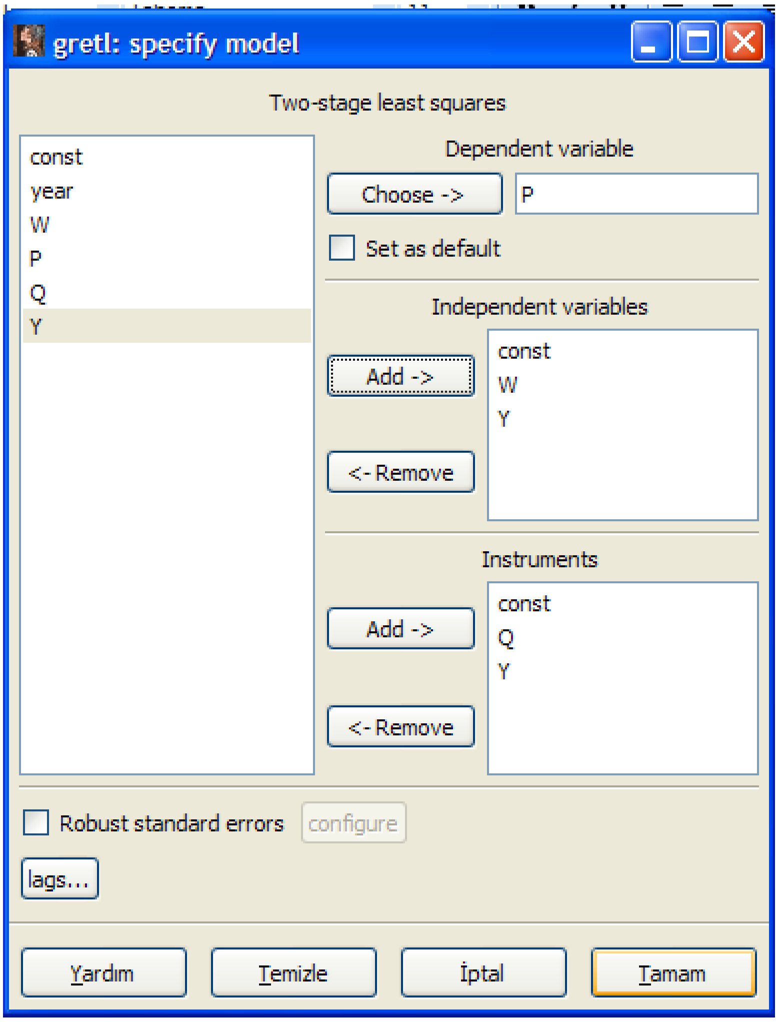


TSLS estimates using the 20 gözlem 1980-1999
Bağımlı değişken: W
Alet değişkenleri: sabit Q Y

	Katsayı	*St.Hata*	*t*	*p*	
sabit	-1.957	0.147866	-13.2349	<0.00001	***
P	0.0270591	0.000566015	47.8063	<0.00001	***
Q	0.0197607	0.00198713	9.9443	<0.00001	***

F-istatistiği (2, 17) = 5754.11 (p < 0.00001)

Hausman test -
Null hypothesis: EKK estimates are consistent
Asimtotik test istatistiği: Chi-square(1) = 1.82966
with p = 0.176168
First-stage F-istatistiği (1, 17) = 1331.24



Model 0: TSLS estimates using the 20 gözlem 1980-1999
Bağımlı değişken: P
Alet değişkenleri: sabit Q Y

	Katsayı	*St.Hata*	*t*	*p*	
sabit	152.958	83.9749	1.8215	0.06854	*
W	-110.573	87.848	-1.2587	0.20815	
Y	0.320136	0.198419	1.6134	0.10665	

F-istatistiği (2, 17) = 148.579 (p < 0.00001)

Hausman test -
Null hypothesis: EKK estimates are consistent
Asimtotik test istatistiği: Chi-square(1) = 83.0754
with p = 7.89833e-020
First-stage F-istatistiği (1, 17) = 2.46357

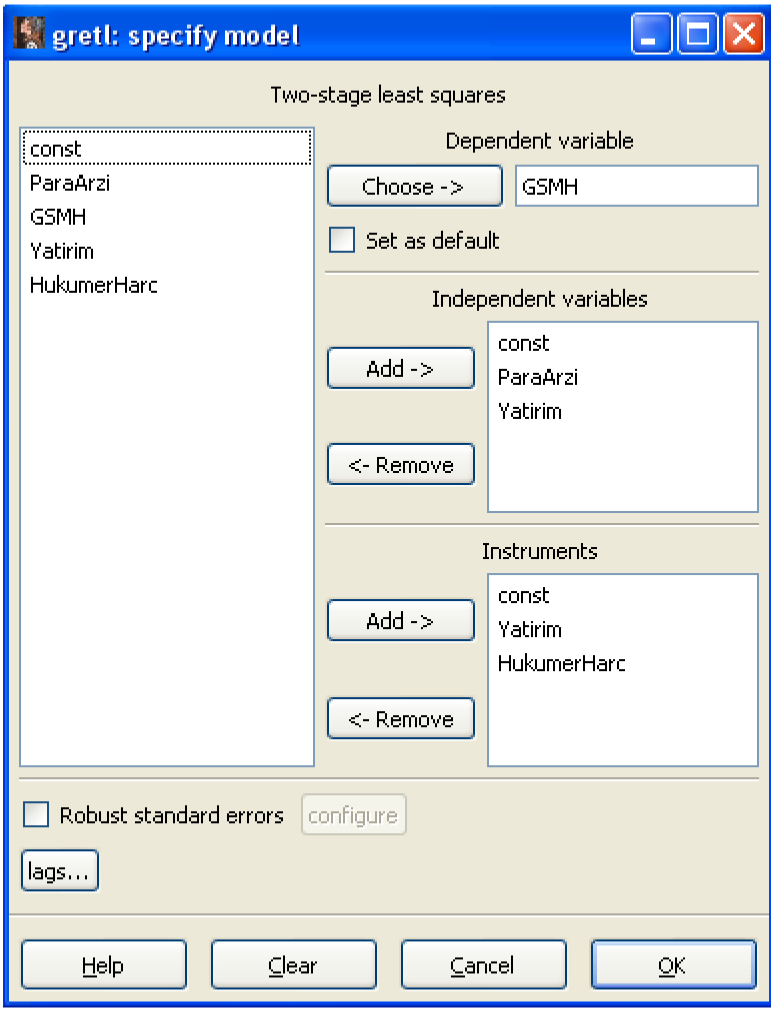
Örnek 8-5

$M = a_0 + a_1 Y + a_2 G + e$
$Y = b_0 + b_1 M + a_2 I + u$

M =para arzi
Y = GSMH
I= yatirim
G =hukumet harcamalari

Y (GSMH) için en küçük kareler tahminini elde edelim.

yil	M (para arzi)	Y (GSMH)	I (yatirim)	G (hukumet harcamalari)
1982	150.9	563.8	85.20	118.00
1983	156.5	594.7	90.20	123.70
1984	163.7	635.7	96.60	129.80
1985	171.4	688.1	112.00	138.40
1986	175.8	753	124.50	158.70
1987	187.4	796.3	120.80	180.20
1988	202.5	868.5	131.50	198.70
1989	209	935.5	146.20	207.90
1990	219.7	982.4	140.80	218.90
1991	233.9	1063.4	160.00	233.70
1992	255.3	1171.1	188.30	253.10
1993	270.5	1306.6	220.00	269.50
1994	283.2	1412.9	214.60	302.70
1995	295.4	1528.8	190.90	338.40
1996	313.8	1702.2	243.00	361.30
1997	338.7	1899.5	303.30	396.20
1998	361.5	2127.6	351.50	435.60
1999	382.1	2368.5	386.20	476.10



Model 2: TSLS estimates using the 18 gözlem 1982-1999
Bağımlı değişken: GSMH
Instrumented: ParaArzi
Instruments: Sabit Yatirim HukumerHarc

	Katsayı	*St.Hata*	*z-stat*	*p*	
Sabit	-468.17	77.0038	-6.0798	<0.00001	***
ParaArzi	5.67437	0.750485	7.5609	<0.00001	***
Yatirim	1.51902	0.615737	2.4670	0.01363	**

Mean dependent var	1188.811	S.D. dependent var	547.2380
Sum squared resid	36922.36	S.E. of regression	49.61341
R-squared	0.992771	Düzeltilmiş R-squared	0.991808
F(2, 15)	1032.137	P(F)	8.64e-17
Log-likelihood	-158.0764	Akaike criterion	322.1528
Schwarz criterion	324.8239	Hannan-Quinn	322.5211
Rho	0.713676	Durbin-Watson	0.592157

Hausman test -
Null hypothesis: EKK estimates are consistent

Asimtotik test istatistiği: Chi-square(1) = 44.2625
with p = 2.8717e-011
Weak instrument test -
First-stage F-istatistiği (1, 15) = 92.8872

Hausman testi, eşanlı tahminlemenin gerekli olduğunu göstermektedir.

9. ZAMAN SERILERI YÖNTEMLERI

9.1. AR Modelleri

Bir değişkenin geleceğe dönük kestirimi, başka değşkenlere baş vurmaksızın sadece değişkenin kendisinden yararlanılarak yapılabilir. Bu kestirimler belli bir teorik modele dayanmaz. Değişkenin geçmişte gösterdiği hareket, gelecekte göstereceği hareketi kestirmek için kullanılabilir. Yüksek frekanslı veriler (aylık, günlük gibi) uygun tahmin yöntemini değiştirebilecek karmaşık zaman serisi özellikleri gösterebilir.

Otoregresyon (AR), önemli zaman serisi yaklaşımlarından biridir. AR modeli, basit olarak bir zaman serisinin halihazır değerinin, bu serinin bir veya daha önceki değerlerine karşı doğrusal regresyonudur. Bir başka ifadeyle y_t gibi bir zaman serisi değişkeninin belli bir oranının gelecek döneme, y_{t+1}'in belli bir oranının daha sonraki döneme aktarıldığı bir süreçtir. Her dönemin bir sonraki döneme aktarılması nedeniyle, otoregresyon uzun solukludur. AR modelleri aşağıdaki gibi gösterilir:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + \ldots + \phi_1 Y_{t-p} + \varepsilon_t$$

Burada Y_t zaman serisi, ε_t hata terimi (beyaz gürültü=white noise), Φ_1 , ..., Φ_p modelin tahmincileri, δ ise aşağıda gösterildiği şekilde zaman serisinin ortalamasıdır:

$$\delta = \sum_{i=1}^{p} (1 - \phi_i)$$

p, AR modelinin derecesidir. AR modelleri, bu amaçla geliştirilmiş yöntemlerin yanısıra, EKK ile de tahmin edilebilir.

Hareketli ortalama (MA) sürecinde, hata terimi gelecek dönemlere taşınır. MA modelleri:

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

şeklinde gösterilir. Burada $\boldsymbol{Y_t}$ zaman serisi, μ zaman serisinin ortalaması, θ_1, ..., θ_q modelin tahmincileridir. q, MA modelinin derecesidir. MA modeli, zaman serisinin halihazır değerinin, bu serinin bir veya daha önceki değerlerinin tesadüfi şoklarına karşı doğrusal regresyonudur. Tesadüfi şokların, tipik olarak normal dağılış gösterdiği varsayılır. Bu modelin altında yatan düşünce, tesadüfi şokların zaman serisine ait gelecekteki değerleri yönlendirdiğidir.

MA tahminlemesi, AR tahminlemesinden daha zordur. Zira hata terimleri uydurulan modele bağlıdır. Bu, MA modellerinin EKK yerine tekrarlamalı doğrusal olmayan yöntemlerle tahmin edilebileceği anlamına gelmektedir. MA modellerinin yorumu biraz daha karmaşıktır.

Bazan tek başına AR veya MA, bazan da her ikisi birden aynı modelde bulunabilir.

9.2. Durağanlık

Çoğu zaman serisi yöntemi, verilerin durağan olmasını gerektirir. Durağanlık en basit anlamda, gözlemlerin sabit bir ortalama etrafında dalgalanmasıdır. Durağanlık; ortalama, varyans ve otokorelasyon yapısının zaman içinde değişme göstermediği anlamına gelir. Bir başka ifadeyle durağan bir zaman serisi; trendsiz, ortalaması ve varyansı sabit, düz görüntülü bir seridir. Durağan olmayan bir zaman serisinin sabit bir ortalaması olmayıp, ele alınan dönem boyunca bazan azalır bazan da artar. EKK ile tahminleme, zaman serisi verisinin durağan ve eş varyanslı olmasını gerektirir. Durağan olmayan bir seri:

$$Y_t = \theta Y_{t-1} + \varepsilon_t$$

modelinde $\theta = 1$ durumunu (birim kök) gösterir. Eğer zaman serisi durağan değilse, seriyi durağanlaştırmak için aşağıdaki tekniklere başvurulur:

Verilerin farkını alabiliriz: $\mathbf{Z_t}$ zaman serisi 1. dereceden farkı için yeni bir seri tanımlanır:

$$\Delta Z_t = Z_t - Z_{t-1}$$

ya da 2. dereceden farkı için

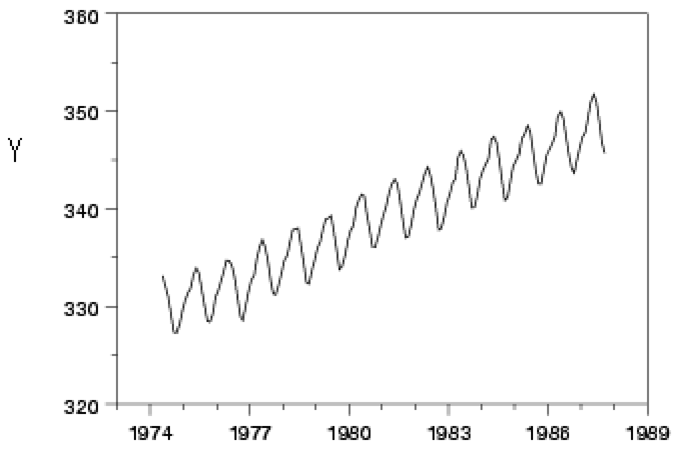
$$\Delta_2 Z_t = Z_t - Z_{t-2}$$

Farkı alınmış veriler, başlangıçtakinden daha az sayıda olacaktır. Birden daha çok fark alınabilir ancak genellikle bir fark yeterlidir.

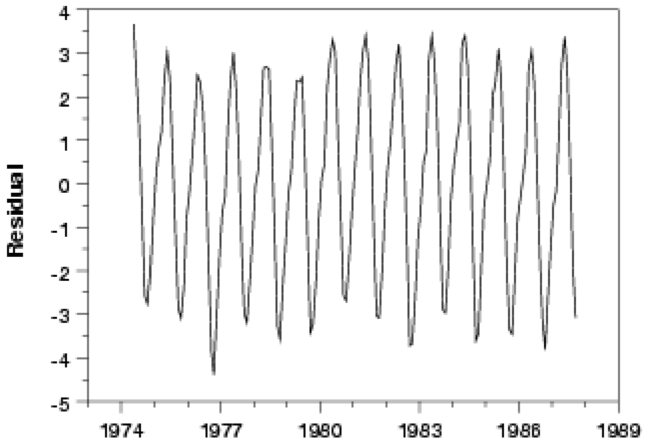
Eğer zaman serisi değişkeni trend gösteriyorsa, verilere basit bir doğru veya 2. dereceden bir eğri eğri uydurulabilir. Bu modelden elde edilen hataları kullanarak modelleme yapılır. Amaç uzun dönem trendini uzaklaştırmak olduğundan, basit bir doğru yeterli olabilir.

Varyans sabit değilse, serinin karekökü veya logaritması alınır. Böylece varyans kararlı bir hale gelir. Negatif veri problemi varsa, tüm verileri pozitif yapacak makul bir sabit sayı verilere eklenir. Daha sonra eklenen değer, modele göre tahmin edilen değerlerden çıkarılır.

Mevsimsellik durağanlığı bozmakla birlikte, bu genellikle zaman serisi modeline kolayca dahil edilebilir.



Grafikten, verilerin artan bir trend gösterdiği anlaşılmaktadır. Basit doğrusal bir tahmin, bu yukarı doğru trendi yok edebilecektir. Ayrıca grafik dönemsel bir davranışı da sergilemektedir.



Yukarıdaki grafik, orijinal verilerle tahmin edilmiş doğrusal modelin hatalarına aittir. Hataların gösterdiği desen sistematik bir dalgalanma gösterse de, doğrusal trend giderildikten sonra verilerin ortalaması ve varyansı sabitlenmiş görünmektedir.

Birim Kök Testi

Bu test, herhangi bir zaman serisinin durağanlık gösterip göstermediğini belirler. Modelimiz:

$$Y_t = \rho\, Y_{t-1} + u_t$$

olsun. Burada u_t, ortalaması sıfır, varyansı sabit, otokorelasyon göstermeyen hata terimidir. Bu özelliğe sahip hata terimine "beyaz gürültü" (white noise) adı verilir. Eğer Y_{t-1}'nin tahmincisi olan ρ 1'e eşitse birim kök vardır denir. Bunun anlamı, Y_t zaman serisinin durağanlık göstermediğidir. Yukarıdaki denklemi fark denklemi olarak yazarsak yeni durum:

$$\Delta Y_t = (\rho-1)\, Y_{t-1} + u_t$$

$$\Delta Y_t = \delta Y_{t-1} + u_t$$

$$\delta = (\rho-1)$$

Yeni duruma göre zaman serisinde durağanlık söz konusu değilse δ'nın sıfır olması gerekir.

Dickey Fuller veya Philips Peron birim kök testlerinde aşağıdaki regresyonlar uygulanır:

1) $\Delta Y_t = \delta Y_{t-1} + u_t$

2) $\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$

3) $\Delta Y_t = \beta_0 + \beta_1 t + \delta Y_{t-1} + u_t$

Denklemlerdeki t, trend değişkenidir.

Dickey Fuller veya Philips Peron birim kök testlerinde hipotezlerimiz:

H_0: Seri durağan değil [Birim kök var (δ = 0 veya ρ = 1)]

H_1: Seri durağan (Birim kök yok)

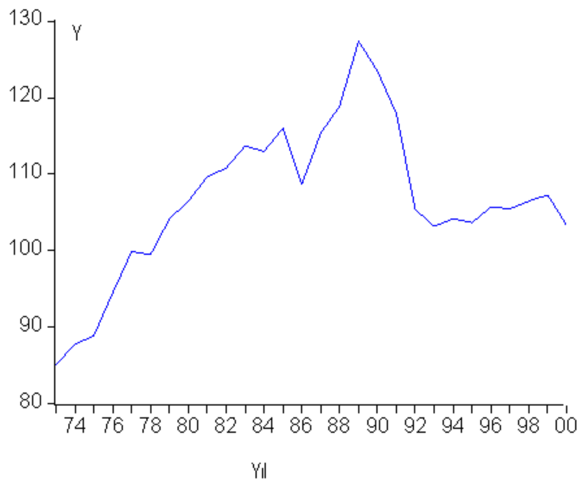
Hesaplanan DF τ (tau) değerinin mutlak değeri, tablo değerinin mutlak değerinden büyük olması durumunda H_0 hipotezi reddedilir. Yani seri durağandır denir.

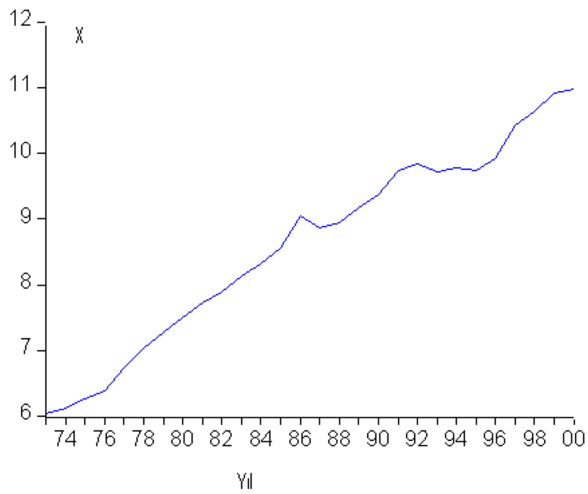
Örnek 9-1

Aşağıda verilen Y ve X gibi iki zaman serisi için durağanlık testi yapalım. Y, Y ülkesinin tüketim harcamaları, X, X ülkesinin borsa kapanış değerleridir.

Yıl	Y	X	ΔY	ΔX	Yıl	Y	X	ΔY	ΔX
1973	85.1	6.04	-	-	**1987**	115.4	8.87	-	-
1974	87.8	6.11	2.7	0.07	**1988**	118.9	8.94	3.5	0.07
1975	88.9	6.27	1.1	0.16	**1989**	127.4	9.18	8.5	0.24
1976	94.5	6.38	5.6	0.11	**1990**	123.5	9.38	-3.9	0.2
1977	99.9	6.73	5.4	0.35	**1991**	117.9	9.74	-5.6	0.36
1978	99.5	7.03	-0.4	0.3	**1992**	105.4	9.83	-12.5	0.09
1979	104.2	7.28	4.7	0.25	**1993**	103.2	9.72	-2.2	-0.11
1980	106.5	7.51	2.3	0.23	**1994**	104.2	9.77	1	0.05
1981	109.7	7.73	3.2	0.22	**1995**	103.7	9.73	-0.5	-0.04
1982	110.8	7.89	1.1	0.16	**1996**	105.7	9.93	2	0.2
1983	113.7	8.13	2.9	0.24	**1997**	105.5	10.42	-0.2	0.49
1984	113	8.32	-0.7	0.19	**1998**	106.5	10.63	1	0.21
1985	116	8.56	3	0.24	**1999**	107.3	10.91	0.8	0.28
1986	108.7	9.04	-7.3	0.48	**2000**	103.3	10.97	-4	0.06

Önce Y ve X'in grafiklerini çizeceğiz:



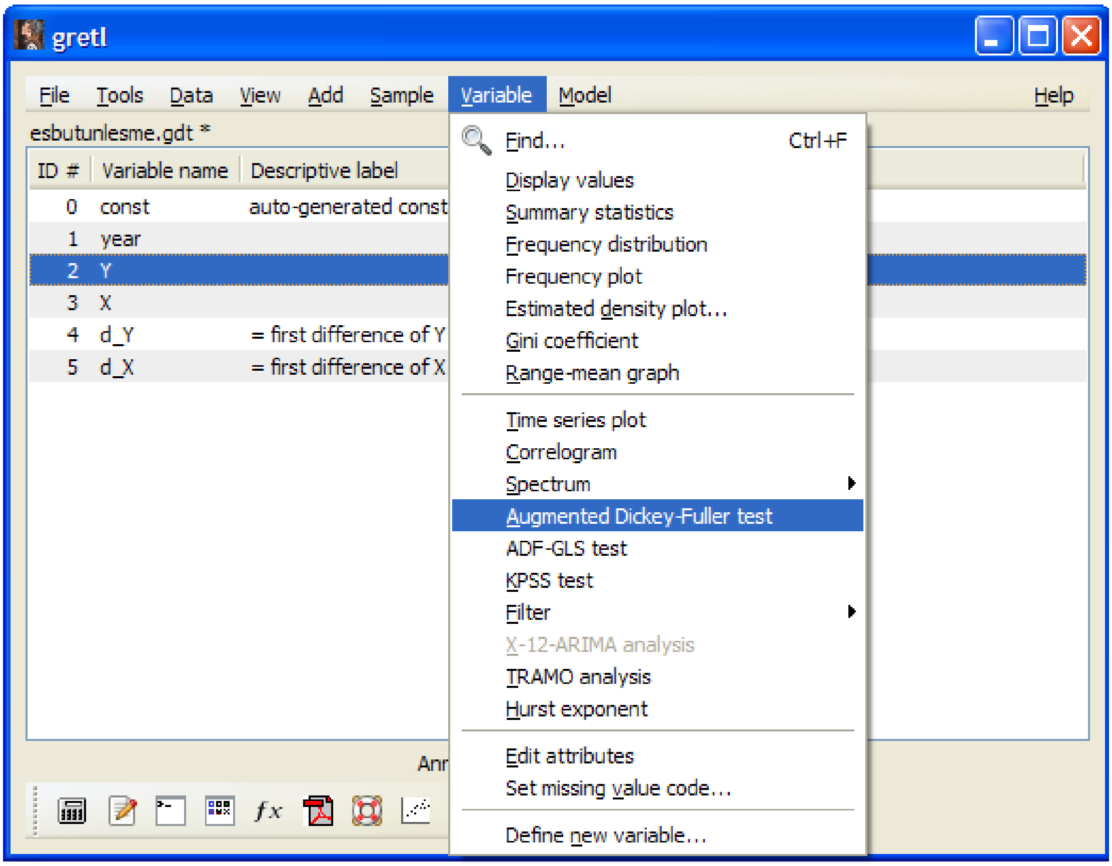


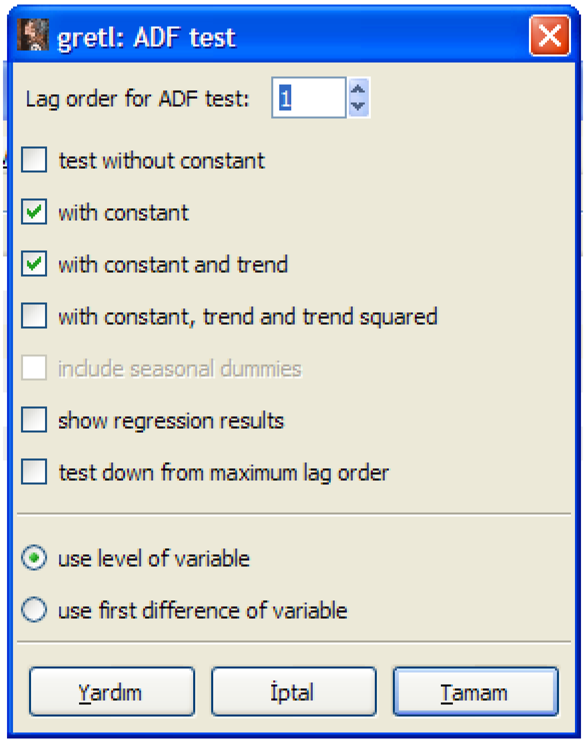
Önce Y ile X arasındaki ilişkiyi, Y'nin X'e göre regresyonuyla inceleyelim:

$$Y=77.3293+3.40599X \qquad R^2=0.26$$
$$t: \ (7.958) \quad (3.062)$$

$R^2=0.26$ olmakla birlikte, X'e ait tahminci istatistiki olarak anlamlıdır. Buna göre X, Y'yi açıklamaktadır. Ancak bu ilişkiyi bir de Y ve X'in durağanlıklarını test ettikten sonra inceleyeceğiz.

Y'nin durağanlığını belirlemek için, a) sabit b) Sabit ve trendli ADF modeline göre ADF testini uygulayalım.





Augmented Dickey-Fuller tests, order 1, for Y
sample size 26
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.087
Tahmin değeri (a - 1): -0.206303
test istatistiği: tau_c(1) = -2.30174
asimtotik p 0.1714

with Sabitant and trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.070
Tahmin değeri (a - 1): -0.171511
test istatistiği: tau_ct(1) = -1.71828
asimtotik p 0.7434

Görüldüğü gibi gerek sabitli, gerekse sabit ve trendli ADF istatistiği için p değeri %1, %5 veya %10'dan büyük olduğundan sıfır hipotezi reddedilmez. Yani Y serisi durağan değildir.

Aynı şekilde X'in durağanlığını belirlemek için, a) sabit b) Sabit ve trendli ADF modeline göre ADF testini uygulayalım.

Augmented Dickey-Fuller tests, order 1, for X
sample size 26
birim-kök sıfır hipotezi: a = 1

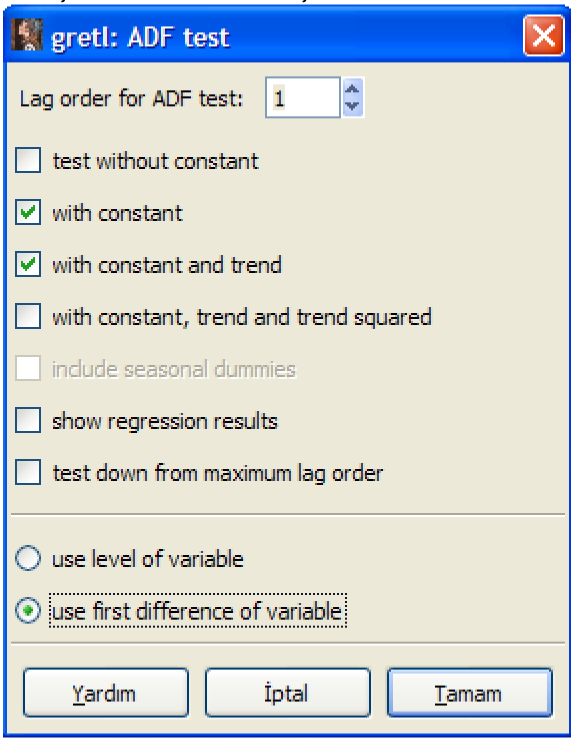
Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.012
Tahmin değeri (a - 1): -0.0228789
test istatistiği: tau_c(1) = -1.01976
asimtotik p 0.7485

with Sabitant and trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.057
Tahmin değeri (a - 1): -0.311646
test istatistiği: tau_ct(1) = -2.2346
asimtotik p 0.4696

Görüldüğü gibi ADF istatistiği için p değeri %1, %5 veya %10'dan büyük olduğundan sıfır hipotezi reddedilmez. Yani X serisi de durağan değildir.

Y ve X serilerini durağanlaştırmak için her iki değişkenin de birinci farklarını alalım ve durağanlıklarını test edelim:

ΔY için ADF testi sonuçları:



Augmented Dickey-Fuller tests, order 1, for d_Y
sample size 25
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: 0.099
Tahmin değeri (a - 1): -0.266283
test istatistiği: tau_c(1) = -2.70891
asimtotik p 0.07244

with Sabitant and trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: 0.081
Tahmin değeri (a - 1): -0.240621
test istatistiği: tau_ct(1) = -2.21921
asimtotik p 0.4782

ΔX için ADF testi sonuçları:

Augmented Dickey-Fuller tests, order 1, for d_X
sample size 25
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.024
Tahmin değeri (a - 1): -1.02268
test istatistiği: tau_c(1) = -3.568
asimtotik p 0.006434

with Sabitant and trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.043
Tahmin değeri (a - 1): -1.0526
test istatistiği: tau_ct(1) = -3.61442
asimtotik p 0.02855

Görüleceği üzere Y ve X'in birinci farkları durağandır. Artık Y ve X arasndaki ilişkiyi, durağanlığını belirlediğimiz ΔY ve ΔX üzerinden ele alalım:

$$\Delta Y = \underset{(0.958)}{1.31755} \underset{(-0.608)}{-3.52408}\,\Delta Y \qquad R^2 = 0.01$$

t: (0.958) (-0.608)

R^2 sıfıra çok yakın ve ΔY'nin tahmincisi istatistiki açıdan önemsizdir. Bir başka ifadeyle X, Y'yi açıklamamaktadır. Bu durumda durağan olmayan Y ve X serileri arasında yukarıda bulduğumuz ilişki sahtedir diyebiliriz. Gerçekten de Y ülkesinin tüketim harcamaları ile X ülkesinin borsa kapanış değerleri arasında herhangi bir ilişki olmaması gerekir. Bu, durağan olmayan serilerle yapılacak regresyon modellemelerinin, sahte regresyona yolaçabileceğini göstermektedir.

9.3. Eşbütünleşme ve Hata Düzeltme Modeli

Makroekonomik değişkenlerin büyük çoğunluğu durağan değildir. Bu değişkenlerin birinci derece farkları genellikle durağandır. Ekonomik değişkenler arasında uzun dönem ilişkiler olabilmektedir. Örneğin; tüketim-gelir, fiyatlar-ücretler, yurtiçi fiyatlar-yurtdışı fiyatlar gibi. Birim kök testiyle uzun dönem ilişkiye sahip değişkenler belirlenebilmektedir.

Bazı durumlarda Y ve X gibi iki seri durağan olmadığı halde uzun dönemde birlikte hareket edebilirler. Eğer:

$$Y_t = b_0 + b_1 X_t + e_t$$

modelinin tahmininden elde edilen e_t hata teriminin durağan olduğu belirlenirse, Y ve X'in eşbütünleşme gösterdiğinden söz edilir. Eğer eşbütünleşme varsa, bağımlı değişkeni açıklamak için uzun dönem süreci kullanılmalıdır. Y, uzun dönem dengesinin altında (ya da üstünde) kalıyorsa, takip eden dönemde Y'nin azalması (ya da artması) beklenir.

Eğer Y ve X eşbütünleşme gösteriyorsa, sadece değişkenlerin farkını alarak regresyon modellemesi yapmak yeterli olmayacaktır. Değişkenler arasındaki uzun dönem ilişkisinin de dikkate alınması gerekecektir. Bu amaçla hata düzeltme modelinden (error correction model=ECM) yararlanılır. ECM'de açıklayıcı değişken olarak uzun dönem dengesinden sapmalar da yer alır.

Uzun dönem ilişkisini:

$$Y_t = b_0 + b_1X_t + e_t$$

şeklinde tahmin edeceğimizi varsayalım. Bu durumda uzun dönem ilişkisinden sapmalar:

$$e_t = Y_t - b_0 - b_1X_t$$

olur. Bu sapmalar aşağıdaki ECM denkleminde yer alır:

$$\Delta Y_t = c_0 + c_1\Delta X_t + c_2e_{t-1} + u_t$$

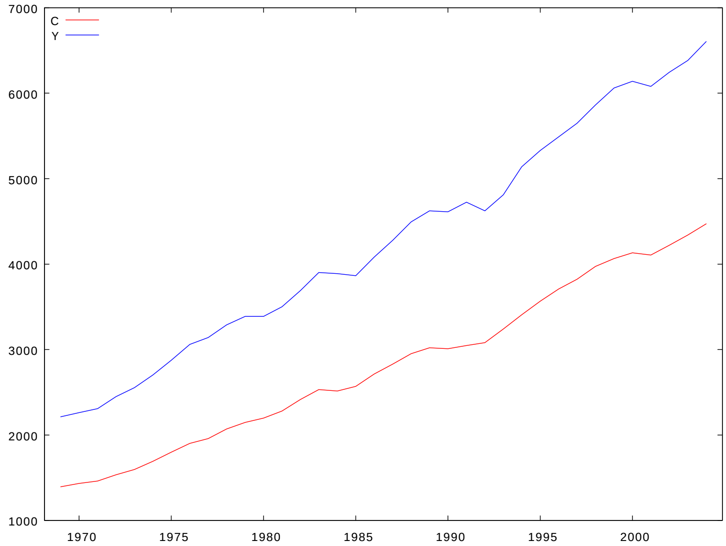
modeldeki tüm değişkenler durağan olduğundan, ECM modeli EKK ile tahmin edilebilir.

Örnek 9-2

Bir yatırımcı tüketim ve gelir arasındaki ilişkiyi öğrenmek istiyor. 1978-2002 arasındaki veriler aşağıdaki gibidir.

YIL	C	Y	YIL	C	Y
1969	1393.6	2212.3	**1987**	2829.8	4279.3
1970	1432.6	2261.7	**1988**	2951.6	4493.7
1971	1461.5	2309.08	**1989**	3020.2	4624
1972	1533.8	2449.1	**1990**	3009.7	4611.9
1973	1596.6	2554	**1991**	3046.4	4724.9
1974	1692.3	2702.9	**1992**	3081.5	4623.6
1975	1799.1	2874.8	**1993**	3240.6	4810
1976	1902	3060.2	**1994**	3407.6	5138.2
1977	1958.6	3140.2	**1995**	3565.5	5329.5
1978	2070.2	3288.6	**1996**	3708.7	5489.9
1979	2147.5	3388	**1997**	3822.3	5648.4
1980	2197.8	3388.2	**1998**	3972.7	5862.9
1981	2279.5	3500.1	**1999**	4064.6	6060.4
1982	2415.9	3690.3	**2000**	4132.2	6138.7
1983	2532.6	3902.3	**2001**	4105.8	6079
1984	2514.7	3888.2	**2002**	4219.8	6244.4
1985	2570	3865.1	**2003**	4339.7	6383.8
1986	2714.3	4081.1	**2004**	4471.1	6604.2

Önce değişkenlermizin grafiğini çizelim. Tüketim harcamaları ve gelir değişkenlerinin durağan olmadığı görülmektedir. Ancak bu değişkenlerin birlikte hareket ettikleri de açıkça görülmektedir.



Ekonomik teori, uzun dönemde tüketim harcamaları ile gelir arasındaki ilişkinin aşağıdaki gibi olması gerektiğini söyler:

$$C_t = b_0 + b_1 Y_t + e_t$$

Bu modeli tahmin ettiğimizde aşağıdaki sonuçlara ulaşırız:

Model 1: EKK tahminleri 36 gözlem 1969-2004
Bağımlı değişken: C

	Katsayı	*St.Hata*	*t*	*p*	
sabit	-221.425	19.2949	-11.4759	<0.00001	***
Y	0.71029	0.00431991	164.4223	<0.00001	***

$R^2 = 0.99874$
Düzeltilmiş $R^2 = 0.99871$

Görüldüğü gibi modelimizin tahmincileri anlamlıdır ve açıklayıcılığı çok yüksektir. Ancak "bu bir sahte regresyon mudur yoksa değişkenler arasında gerçekten bir uzun dönem ilişkisi mi vardır?" sorusunu sormamız gerekiyor. Bunun için önce Tuketim ve Gelir değişkenlerinin durağanlıklarını test edelim:

C için ADF test sonuçları

Augmented Dickey-Fuller tests, order 2, for C
sample size 33
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = bo + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.061
Tahmin değeri (a - 1): 0.00622487
test istatistiği: tau_c(1) = 0.619542
asimtotik p 0.9903

with Sabitant and trend
model: (1 - L)y = bo + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.014

Tahmin değeri (a - 1): -0.243856
test istatistiği: tau_ct(1) = -2.13187
asimtotik p 0.5273

Y için ADF test sonuçları

Augmented Dickey-Fuller tests, order 2, for Y
sample size 33
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.069
Tahmin değeri (a - 1): 0.00513383
test istatistiği: tau_c(1) = 0.377717
asimtotik p 0.9821

with Sabitant and trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.005
Tahmin değeri (a - 1): -0.375617
test istatistiği: tau_ct(1) = -2.30952
asimtotik p 0.4282

Görüldüğü gibi her iki değişken de durağan değildir. Şimdi

$$C_t = b_0 + b_1Y_t + e_t$$

Modelinden elde edilen e_t'leri hesaplayalım:

YIL	e_t	e_{t-1}	YIL	e_t	e_{t-1}
1969	43.65031		**1987**	11.68072	36.96021
1970	47.56198	43.65031	**1988**	-18.8055	11.68072
1971	42.80844	47.56198	**1989**	-42.7563	-18.8055
1972	15.65362	42.80844	**1990**	-44.6618	-42.7563
1973	3.944191	15.65362	**1991**	-88.2245	-44.6618
1974	-6.118	3.944191	**1992**	18.82784	-88.2245
1975	-21.4169	-6.118	**1993**	45.52977	18.82784
1976	-50.2046	-21.4169	**1994**	-20.5874	45.52977
1977	-50.4279	-50.2046	**1995**	1.434073	-20.5874
1978	-44.2349	-50.4279	**1996**	30.70354	1.434073
1979	-37.5377	-44.2349	**1997**	31.72257	30.70354
1980	12.62021	-37.5377	**1998**	29.76534	31.72257
1981	14.83875	12.62021	**1999**	-18.6169	29.76534
1982	16.14157	14.83875	**2000**	-6.63266	-18.6169
1983	-17.7399	16.14157	**2001**	9.371657	-6.63266
1984	-25.6248	-17.7399	**2002**	5.889678	9.371657
1985	46.08287	-25.6248	**2003**	26.77524	5.889678
1986	36.96021	46.08287	**2004**	1.627307	26.77524

Şimdi e_t'nin durağanlığını test edelim:

e_t için ADF durağanlık testi

Augmented Dickey-Fuller tests, order 2, for e_t
sample size 33

birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: 0.004
Tahmin değeri (a - 1): -0.652918
test istatistiği: tau_c(1) = -3.17436
asimtotik p 0.02155

with Sabitant and trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.017
Tahmin değeri (a - 1): -0.675437
test istatistiği: tau_ct(1) = -3.28392
asimtotik p 0.06887

ADF testi %10 için e_t'nin durağan olduğunu göstermektedir. O halde tüketim ve gelir arasında eş bütünleşme olduğundan söz edilebilir. Bir başka ifadeyle tüketim ve gelir arasında uzun dönem ilişkisi vardır. Buna göre uzun dönem marjinal tüketim eğilimi 0.71'dir.

Şimdi uzun dönemde düzeltme gerekip gerekmediğini, ECM ile belirleyelim:

$$\Delta C_t = a_0 + a_1 \Delta Y_t + a_2 e_{t-1} + u_t$$

olmalıdır. Bu modeli tahmin ettiğimizde:

Model 3: EKK tahminleri 35 gözlem 1970-2004
Bağımlı değişken: d_C

	Katsayı	*St.Hata*	*t*	*p*	
sabit	23.2987	8.1554	2.8568	0.00746	***
d_Y	0.51498	0.0557108	9.2438	<0.00001	***
e_1	-0.18697	0.152632	-1.2250	0.22953	

R^2 = 0.77476
Düzeltilmiş R^2 = 0.76068
F-istatistiği (2, 32) = 55.0343 (p < 0.00001)
Durbin-Watson istatistiği = 1.68792
First-order autocorrelation coeff. = 0.155973
Log-likelihood = -160.485

elde edilir. Modeldeki ΔY_t'ye ait tahminci (0.51), kısa dönemde dönemde gelirin tüketim üzerindeki etkisini göstermektedir. Yani kısa dönem marjinal tüketim eğilimi 0.51'dir. e_1'nin istatistiki açıdan önemli olmaması, tüketim ile gelir arasındaki dengesizliğin hemen o yılda düzeltildiğini, uzun dönemdeki dengenin korunduğunu göstermektedir.

9.4. ARCH

Tek denklemli regresyon modellemesi, çok yönlü ve geniş kullanımlı istatistik tekniklerden biridir. Genellikle bu tür modellemede, bir şans değişkeninin koşullu ortalaması modellenir. Ancak, bir değişkenin koşullu varyansı ya da istikrarsızlığını modellemek de mümkündür. İstikrarsızlığı tahminleme ve öngörüde bulunma isteğinin çeşitli nedenleri olabilir.

1. Herhangi bir mal ya da varlığı elde tutmanın riskini analiz etmek gerekebilir.
2. Öngörü güven aralıkları zamana bağlı olarak değişkenlik gösterebilir. Bu nedenle hata varyansını modelleyerek daha doğru aralıkların elde edilmesi mümkün olabilir.
3. Farklı varyans problemi uygun şekilde giderilebilirse, daha etkin tahminciler elde edilebilir.

ARCH modelleri, özellikle koşullu varyansı modellemek ve öngörmek için tasarlanır. ARCH modellemesinde, bağımlı değişkenin varyansı, bağımlı değişkenin geçmiş değerleri ve bağımsız (dışsal) değişkenlerin bir fonksiyonu olarak tanımlanır.

Engle 1982'de ARCH'ı bilim dünyasına sunmuştur. 1986'da Bollerslev genelleştirilmiş ARCH'ı tanıtmıştır. Bu modeller özellikle finansal zaman serilerinde olmak üzere, ekonometride yaygın şekilde kullanılmıştır.

9.4.1. ARCH'da model tanımlama

ARCH modelini geliştirirken iki farklı tanımlama yapılır. Birincisi, koşullu ortalama, ikincisi ise koşullu varyans içindir (genelleştirilmiş ARCH=GARCH). GARCH, koşullu varyansın kendi gecikmeli değeriyle ilişkilendirilmesini sağlar:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Bu GARCH (1,1) modelidir.

GARCH(1,1) Modelinin Tahminlenmesi

GARCH(1,1) tanımlamasında:

Denklem 9-1: $y_t = \gamma x_t + u_t$

Denklem 9-2: $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

Denklem 9-1'de verilen ortalama denklemi, dışsal değişkenlerin bir fonksiyonu olarak tanımlanmıştır. Denklem 9-2'de ise, koşullu varyans, hata teriminin gecikmeli değerinin ve varyansın gecikmeli değerinin bir fonksiyonu verilmiştir. Burada ω ortalamayı, u^2_{t-1} ARCH terimini, σ^2_{t-1} GARCH terimini temsil etmektedir. ARCH terimi, ortalama denkleminden sapma karelerinin gecikmesi olarak hesaplanır ve önceki dönemden kalan istikrarsızlığı temsil eder. Bir GARCH (p,q) modeli aşağıdaki gibi tanımlanır:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2$$

GARCH(p,q), p'inci dereceden GARCH ve q'ncu dereceden ARCH terimlerinin varlığına işaret eder. Buna göre GARCH (1,1), 1. dereceden GARCH ve 1. dereceden ARCH terimleri bulunduğunu gösterir. Basit bir ARCH modeli, Denklem 9-2'de gecikmeli öngörü varyansının bulunmadığı bir GARCH modelidir. Bir ekonomik kurum, uzun dönem ağırlıklı ortalamasını (sabit terim), geçmiş dönemin öngörülmüş varyansını (GARCH terimi) ve geçmiş dönemde gözlenmiş istikrarsızlık bilgisini (ARCH terimi) oluşturarak bu dönemin varyansını tahmin ettiği durumda, bu tanımlama sık sık, finansal açıdan yorumlanır. Eğer bir varlığın getirisi beklenmedik bir şekilde aşağı ya da yukarı yönlü büyük hareketlenme gösterirse, kurum bir sonraki dönemin varyans tahminini artıracaktır. Bu model, gelirdeki önemli dalgalanmaları, muhtemel yeni dalgalanmaların takip ettiği finansal gelir verilerinde karşılaşılan istikrarsızlık kümelenmeleriyle de tutarlıdır.

GARCH (1,1) modelinde üç parametre tahminlenir:

α_0 sabit.

α_1 geçen dönemin etkisi

β_1 geçen dönemin varyansının etkisi

$\alpha_1 + \beta_1$ dalgalanmanın şiddetini verir

$\alpha_1 + \beta_1 < 1$ dalgalanma azalma yönünde

$\alpha_1 + \beta_1 \cong 1$ çok az dalgalanma var

$\alpha_1 + \beta_1 > 1$ artan yönlü bir dalgalanma var

9.4.2. ARCH-M Modeli

Denklem 9-1'deki X'ler, dışsal veya önceden belirlenmiş değişkenlerdir. Eğer ortalama denklemine koşullu varyansı eklersek, ARCH-M modelini elde ederiz (ARCH-in-Mean=ARCH-M) (Engle, Lilien, Robins, 1987). Bazı ARCH-M modellerinde varyans yerine standart sapma da kullanılabilmektedir. ARCH-M modeli, varlığın sağladığı beklenen gelirin, beklenen riskle ilişkilendirildiği finansal uygulamalarda sıklıkla kullanılır. Beklenen riskin tahmin edilen katsayısı, gelir riskinin ölçülen değerini verir.

Örnek 9-3

log Günlük U.S. Doları/Japon Yeni birinci derece farkına bir GARCH(1,1) modeli uydurulduğunda, aşağıdaki çıktı elde edilmiştir:

Bağımlı değişken: R
Method: ML - ARCH
Date: 07/30/97 Time: 14:27
Sample(Düzeltilmiş): 2 5758
Included observations: 5757 after adjusting endpoints
Convergence achieved after 129 iterations

Ortalama denklemi				
	Katsayı	St.Hata	z-Statistic	Prob.
Sabit	-0.00517	0.006517	-0.793649	0.4274
Varyans denklemi				
	Katsayı	St.Hata	z-Statistic	Prob.
Sabit	0.016001	0.000304	52.64622	0.0000
ARCH(1)	0.207969	0.003492	59.56064	0.0000
GARCH(1)	0.784126	0.003822	205.1789	0.0000

R-kare	-0.000620	Mean dependent var	-0.020218
Düzeltilmiş R-kare	-0.001142	S.D. dependent var	0.604167
S.E. of regression	0.604512	Akaike info kriteri	1.596767
Sum kare resid	2102.343	Schwarz kriteri	1.601394
Log likelihood	-4592.295	Durbin-Watson stat	1.914825

ARCH çıktısı iki kısma ayrılır. Birinci kısım standart ortalama denklemini; ikinci kısım ise varyans denklemini verir. Denklemler katsayılar, standart hatalar, z istatistiği ve p değerlerini sunmaktadır. ARCH ve GARCH parametreleri, Denklem 9-2'deki parametrelere denk düşmektedir.

Bu örnekte ARCH ve GARCH tahmincilerinin toplamı pozitif ve 1'e çok yakındır. Bunun anlamı, dalgalanma şoklarının süreklilik arz ettiğidir.

9.4.3. ARCH Testi

Otokorelasyonun bir diğer şekli, zaman serilerinde geleceğe dönük kestirim yaparken ortaya çıkabilmektedir. Bazı durumlarda kestirim hatalarının varyansının sabit olmadığı, dönemden döneme değişebildiği görülmüştür. Bu, bir tür farklı varyanstır. Zaman serilerinde farklı varyansın modellenmesini Engel (1982) gerçekleştirmiş ve ARCH (**A**uto**R**egressive **C**onditional **H**eteroskedasticity) adını vermiştir. Yöntem aşağıdaki denklemden yola çıkmaktadır:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_p u_{t-p}^2$$

Bu denklem, p'inci dereceden ARCH modeli olarak bilinir. Otoregresif terimi, t zamanındaki hata varyansının, önceki kareli hata terimlerine bağlı olduğundan kullanılmıştır. Ayrıca, t zamanındaki varyans önceki dönemlerdeki koşullara bağlıdır ve bu nedenle koşullu farklı varyans söz konusudur.

ARCH testinde sıfır hipotezi:

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_p = 0$ (Koşullu farklı varyans yoktur)

şeklindedir.

Testin adımları:

1. Denklemi EKK ile tahmin et
2. Hata terimini hesaplayıp, karesini aldıktan sonra 1, 2,.., p gecikmelilerini elde et.
3. Hata teriminin karesini, 1, 2,.., p gecikmelere göre modelle
4. $(n-p)R^2$ hesaplamasını yap ve p serbestlik derecesine göre Ki kare değeriyle karşılaştır.

10. ZAMAN SERİLERİ

10.1. Trend Değişkenli Modeller

X ve Y gibi yıllık gözlemleri olan değişkenlerimiz olsun. Trend değişkeni içeren bir regresyon denklemi:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 t + e_t$$

Şeklinde olabilir. β_2 katsayısı, diğer değişkenler sabitken Y'deki yıllık değişimi ölçer.

Denklemimizi:

$$\ln(Y_t) = \alpha_0 + \alpha_1 \ln(X_t) + \alpha_2 t + u_t$$

olarak ta ifade edebiliriz. Bu durumda, trend değişkenine ait katsayı($\alpha 2$), büyüme oranı olarak yorumlanır. Örneğimizde, X'in belli bir değeri için:

$\alpha_2 > 0$ ise $100(\alpha_2)$ Y'deki yüzde büyüme oranıdır.

$\alpha_2 < 0$ ise $100(\alpha_2)$ Y'deki yüzde küçülme oranıdır

Hesaplanan bu değerler anlık büyüme oranlarıdır. Y'deki yıllık bileşik büyüme oranı, diğer değişkenler sabitken:

$$g = 100\,(\text{antilog}_e\,(\alpha_2) - 1)$$

formülüyle hesaplanır.

Örnek 10-1

Bu örnekte kırmızı et talebini ele alacağız. Bağımsız değişkenlerimiz: kırmızı et fiyatı, tavuk eti fiyatı ve gelir olarak belirlenmiştir.

ETTUK : Kişi başına kırmızı et tüketimi

ETFIY : Reel Taze ve dondurulmuş et fiyatı – tüketici fiyat indeksi (1986=100)

TAVFIY : Reel Taze ve dondurulmuş tavuk eti fiyatı – tüketici fiyat indeksi (1986=100)

GELIR : Reel Kişi başına harcanabilir gelir

PFood : Tüketici fiyat indeksi (1986=100)

PDFL : GSMH deflatörü (1986=100)

Yıl	ETTUK	ETFIY	TAVFIY	GELIR	PFOOD	PDFL
1975	79.2	42	51.3	4883	44	49
1976	84.2	39.3	51.3	5453	45.2	53
1977	80.1	41.8	51.6	5941	48.9	56.8
1978	75	61.1	59.9	6634	56.5	60.7
1979	63.7	80.3	66	7408	63.9	65.8
1980	63.1	87.2	71	8281	70.8	71.4
1981	64.8	89.5	84.5	9545	78.9	78.8
1982	64.2	88.9	86.8	10430	84.6	86.3
1983	63.9	89.6	89.9	10843	87.7	90.7
1984	60.9	95.5	95.8	11686	92.6	93.9
1985	61.7	97.9	91.8	12387	95.2	97
1986	61.2	100	100	12902	100	100
1987	58.2	109.1	106.1	13613	104.4	103.9

1988	58.2	110.8	107.6	14658	107.2	107.7
1989	56.6	113.2	120	15783	111.1	112.5
1990	54.5	117.6	126.1	16263	115.7	116.6
1991	53.4	118.3	123.5	16570	121.2	119.9
1992	51.7	116.8	123	16753	120.8	121.4
1993	49.6	123.1	125.6	16874	122.8	123
1994	50.6	124.1	119.3	17003	123.3	123.9

Yukarıdaki komutlarda öncelikle fiyatlar ve gelir reelleştirilmiştir. Bunun için bu değişkenler implicit fiyat index değişkenine bölünmüştür. Ekonomik teoriye göre kırmızı et fiyatına ait katsayı negatif olmalıdır. Gelir katsayısı ise normal mal için pozitif, adi mal için negatif olmalıdır. Tavuk fiyatı katsayısının pozitif olması ikame, negatif olması tamamlayıcı, sıfır olması ise ilişkisiz olduğu anlamına gelir. Regresyon analizi sonuçları:

Bağımlı değişken: beef

	Katsayı	StHata	T	p
Sabit	115.78	15.73	7.36	0.000
Rpbeef	-0.32517	0.04794	-6.78	0.000
Rpchkn	0.0302	0.1004	0.3	0.767
Rincome	-0.00059	0.001185	-0.5	0.627
Trend	-1.4681	0.2778	-5.28	0.000
S = 1.667	R^2 = 0.978	Düz. R^2=0.972	F=164.95	p=0.000

1975/94 döneminde trend değişkeni negatif ve anlamlıdır. Yani, Herşey sabitken k.et tüketimi her yıl ortalama 1.468 birim düşmektedir. K.et fiyatı negatif ve anlamlıdır. Yani fiyat artışı k.et tüketimini düşürmektedir. Tavuk eti fiyatı ve gelir değişkenlerinin katsayıları istatistiki açıdan anlamlı değildir. Buna göre tavuk eti fiyatı ile k.et tüketimi arasında her hangi bir ilişki yoktur. Aynı zamanda, k.et tüketimi gelirdeki değişimden etkilenmemektedir. Bu örnekte gelir ve trend değişkeni arasındaki yüksek korelasyon, katsayıları anlamsız bulmamıza neden olmuş olabilir.

Log-log denklemi:

Bağımlı değişken: lbeef

	Katsayı	StHata	t	p
Sabit	6.771	1.017	6.66	0.000
lrpbeef	-0.46122	0.0686	-6.72	0.000
lrpchkn	-0.0425	0.1509	-0.28	0.782
lrincome	-0.00567	0.06118	-0.09	0.927
trend	-0.02456	0.004184	-5.87	0.000
S = 0.0237	R^2 = 0.981	Düz. R^2=0.976	F=164.95	p=0.000

log-log denklemi tahmin ettiğimizde, trend değişkenine ait katsayıyı -0.02456, standart hatası ise buluruz. Denklemin standart hatası ise 0.004184'tür. Buna göre, diğer tüm değişkenler sabit tutulduğunda, kişi başına ortalama yıllık azalış:

$$g=[antilog_e(-0.02456)-1]*100= -2.426 \%$$

olarak hesaplanır.

10.2. Zaman Serilerinde Trendin Yok edilmesi

Ekonomik zaman serileri genellikle trend gösterir. Eğer iki zaman serisi ortak bir trende sahipse, bu değişkenler arasında ilişki olduğu düşünülür. Gözlenmeyen değişkenler, sözü edilen değişkenlerin ikisinin birden trend göstermesine yol açacaktır. Gözlenemeyen değişkenler olsa bile, bunları trend yardımıyla kontrol etmek mümkündür. Trendi yok etmek için kullanılabilecek bazı modeller:

Doğrusal trend:

$$y_t = a_0 + a_1 t + e_t,\ t = 1, 2, \ldots$$

Üstel trend:

$$\log(y_t) = a_0 + a_1 t + e_t,\ t = 1, 2, \ldots$$

İkinci dereceden trend:

$$y_t = a_0 + a_1 t + a_2 t^2 + e_t,\ t = 1, 2, \ldots$$

Tahmin edilen denkleme trend teriminin eklenmesi, trendsiz seri kullanmakla aynı anlama gelir. Bir serideki trendi yok etmek, modeldeki her bir değişkeni t ile ilişkilendirmek anlamına gelir. Trend yok edildiğinde, hata terimi de trendsiz bir seri haline gelir. Trendi yok etmek, R2'nin daha doğru hesaplamasını da sağlar. Trend çok iyi açıklandığından, zaman serisi regresyonları çok yüksek R2 gösterir. Trend yok edildiğinde, xt'nin yt'yi ne kadar iyi açıkladığı daha iyi ortaya konulabilir.

10.3. Dağıtılmış Gecikme Modelleri

Bağımsız değişkenlerden birinin etkisi, içinde bulunduğu dönemin dışında gelecekteki dönemlere de yayılıyorsa, bu değişkene zaman gecikmeli değişken adı verilir. Örneğin reklamın satış üzerine etkisini ele alalım. İnsanlar reklamı, sadece reklamın yapıldığı dönemde değil, takip eden dönemlerde de hatırlarlar. Yani, reklam kampanyası dönemi dışında da satış üzerinde etkilidir. Bu nedenle satış modellerinde, halihazır ve gecikmeli reklam değişkenlerine yer verilmesi gerekir. Tarımsal ürünlerin arzında da, gecikmeli değişkenlere yer verme zorunluluğu vardır. Tarımsal ürünlerin arzı, geçmiş yılın fiyatlarının bir fonksiyonudur. Bu yılın fiyatları ne olursa olsun, üretim sürecinin uzunluğu nedeniyle arzı değiştiremeyecek olan üreticiler, ancak gelecek yıl tepkilerini gösterebileceklerdir.

X_t gibi bir zaman serisi değişkeninin 1. gecikmesi, X_{t-1} ile; 2. gecikmesi X_{t-2} olarak gösterilir. Gecikmeli bir değişmenin hazırlanışını, bir örnekle açıklayalım:

T	X_t	X_{t-1}	X_{t-2}
1990	$X_{1990} = 275$	$X_{1990-1} = X_{1989} =$ *	*
1991	$X_{1991} = 260$	$X_{1991-1} = X_{1990} = 275$	*
1992	$X_{1992} = 345$	$X_{1992-1} = X_{1991} = 260$	275
1993	$X_{1993} = 300$	$X_{1993-1} = X_{1992} = 345$	260
1994	$X_{1994} = 220$	$X_{1994-1} = X_{1993} = 300$	345
1995	$X_{1995} = 410$	$X_{1995-1} = X_{1994} = 220$	300

1996	$X_{1996} = 350$	$X_{1996-1} = X_{1995} = 410$	220
1997	$X_{1997} = 250$	$X_{1997-1} = X_{1996} = 350$	410

Görüldüğü gibi, her gecikme, bir gözlemin kayıp veri olmasına yolaçmaktadır. Bir gecikmeli değişkende,1; iki gecikmeli değişkende 2; üç gecikmeli değişken 3 veri, örnek dışı bırakılır.

Dağıtılmış gecikme değişkenleri içeren modellere, zaman gecikmeli modeller denmektedir. Örneğin reklam X, satış miktarını Y, mevcut dönemde ve bir dönem sonra etkiliyorsa, modelimiz:

$$Y_t = b_0 + b_1X_t + b_2X_{t-1}$$

olur. b_1 tahmincisi, Y'nin X'e göre eğimidir ve "reklamın yapıldığı dönemde 1 birim artması, satış miktarını b_1 kadar değiştirecektir" şeklinde yorumlanır. b_2'nin yorumlanmasında ise dikkatli olunmalıdır. b_2, mevcut dönemde yapılan reklamın 1 birim artırılması, takip eden dönemde satışları b_2 kadar değiştireceğini ifade eder.

Dağıtılmış gecikme modellerini genel olarak:

Denklem 10-1: $\mathbf{Y_t = b_0 + b_1X_t + b_2X_{t-1} + b_3X_{t-2} + b_4X_{t-3} +...+ b_pX_{t-p}}$

şeklinde gösterebiliriz. Bu model Y'nin bir dönemdeki değerini, X'in o döneme ve geçmiş dönemlere ait değerlerinin bir fonksiyonu olarak açıklamaktadır. b_1, b_2, b_3, ... b_p tahmincileri, X'in çeşitli gecikmelerinin, Y'nin belli bir dönemdeki değerini etkileme düzeyleridir. Çoğu ekonomik uygulamalarda gecikme arttıkça X'in Y üzerindeki etkisinin azalması beklenir.

Zaman gecikmeli modellerin tahmininde, EKK yöntemi kullanılabilir. Ancak bunun bazı sakıncaları vardır:

- Gecikmeli değişkenler arasında çoklu doğrusal bağlantı olabilir.
- b tahmincileri, gecikme arttıkça giderek azalan değerler almayabilir.
- Modele alınan her gecikmeli değişken serbestlik derecesini düşürür. Aynı şekilde, gecikme sayısı kadar gözlem örnek dışı kalacağından, serbestlik derecesi ikinci kez azalır.

Bu sakıncalar nedeniyle, dağıtılmış zaman gecikmeli modeller, Koyck modellerinden yola çıkılarak elde edilir.

Örnek 10-2

Yıl	Tüketim	Gelir	Yıl	Tüketim	Gelir
1959	7876	8641	1977	12846	14095
1960	7926	8660	1978	13258	14662
1961	7954	8794	1979	13417	14899
1962	8220	9077	1980	13216	14813
1963	8434	9274	1981	13245	15009
1964	8817	9805	1982	13270	14999
1965	9257	10292	1983	13829	15277
1966	9674	10715	1984	14415	16252
1967	9854	11061	1985	14954	16597
1968	10313	11448	1986	15409	16981
1969	10593	11708	1987	15740	17106
1970	10717	12022	1988	16211	17621

1971	10975	12345	1989	16430	17801
1972	11508	12770	1990	16532	17942
1973	11950	13539	1991	16249	17755
1974	11756	13310	1992	16520	18062
1975	11899	13404	1993	16810	18075
1976	12446	13793	1994	17152	18320

Bağımlı değişken: tuketim

	Katsayı	*St.Hata*	*t*	*p*	
sabit	-384.105	151.33	-2.5382	0.015892	**
gelir	0.932738	0.0106966	87.1992	0.00001	***

R^2 = 0.995548
Düzeltilmiş R^2 = 0.995417
Akaike bilgi kriteri = 485.232
Schwarz Bayesian kriteri = 488.399

Bağımlı değişken: tuketim

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	-156.798	143.15	-1.0953	0.281543	
tuketi_1	0.424408	0.0976641	4.3456	0.000131	***
Gelir	0.540095	0.0915616	5.8987	0.00001	***

R^2 = 0.997082
Düzeltilmiş R^2 = 0.9969
F-istatistiği (2, 32) = 5467.95 ($p < 0.00001$)
Akaike bilgi kriteri = 457.432
Schwarz Bayesian kriteri = 462.098

Bağımlı değişken: tuketim

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	-122.312	119.199	-1.0261	0.312780	
tuketi_1	0.799533	0.125352	6.3783	0.00001	***
Gelir	0.763857	0.0950361	8.0375	0.00001	***
Gelir_1	-0.569815	0.145186	-3.9247	0.000450	***

R^2 = 0.998051
Düzeltilmiş R^2 = 0.997862
F-istatistiği (3, 31) = 5291.21 ($p < 0.00001$)
Akaike bilgi kriteri = 445.314
Schwarz Bayesian kriteri = 451.535

Örnek 10-3

Yıl	Tüketim	Gelir	Yıl	Tüketim	Gelir
1967	61.284	78.221	1981	120.477	152.846

1968	68.814	83.326	1982	133.868	164.318
1969	76.766	90.49	1983	148.004	172.414
1970	73.576	92.692	1984	149.735	178.433
1971	73.256	94.814	1985	155.2	185.753
1972	67.502	92.59	1986	154.165	192.059
1973	78.832	101.419	1987	155.445	191.288
1974	80.24	105.267	1988	157.199	196.055
1975	84.477	112.149	1989	158.576	202.477
1976	86.038	116.078	1990	169.238	223.225
1977	96.275	122.04	1991	179.001	233.231
1978	101.292	128.578	1992	183.687	242.762
1979	105.448	136.851	1993	198.273	259.555
1980	114.57	144.734			

Bağımlı değişken: tuketim

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	1.41986	3.00025	0.4732	0.640142	
Gelir	0.779962	0.0186895	41.7326	< 0.00001	***

R^2 = 0.985849
Düzeltilmiş R^2 = 0.985283
Degrees of freedom = 25
Akaike kriteri = 166.763
Schwarz Bayesian kriteri = 169.355

Bağımlı değişken: tuketim

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	1.0384	2.50146	0.4151	0.681901	
tuk_1	0.50092	0.121307	4.1293	0.000408	***
Gelir	0.40436	0.0919424	4.3980	0.000209	***

R^2 = 0.991221
Düzeltilmiş R^2 = 0.990458
F-istatistiği (2, 23) = 1298.47 ($p < 0.00001$)
Akaike bilgi kriteri = 149.236
Schwarz Bayesian kriteri = 153.011

Bağımlı değişken: tuketim

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	1.28846	2.22055	0.5802	0.567646	
gelir	0.861178	0.18836	4.5720	0.000149	***
tuk_1	0.796478	0.153765	5.1798	0.000034	***
gelir_1	-0.713919	0.265354	-2.6904	0.013362	**

R^2 = 0.993395
Düzeltilmiş R^2 = 0.992494
F-istatistiği (3, 22) = 1102.85 ($p < 0.00001$)
Akaike bilgi kriteri = 143.841
Schwarz Bayesian kriteri = 148.873

Örnek 10-4

Bu örnek aylık satış ve reklam harcamalarını ele almaktadır.

Ay (t)	$Satış_t$	$Satış_{t-1}$	Reklam	Ay (t)	$Satış_t$	$Satış_{t-1}$	$Reklam_t$
1	12	-	15	19	30.5	29.6	33
2	20.5	12	16	20	28	30.5	62
3	21	20.5	18	21	26	28	22
4	15.5	21	27	22	21.5	26	12
5	15.3	15.5	21	23	19.7	21.5	24
6	23.5	15.3	49	24	19	19.7	3
7	24.5	23.5	21	25	16	19	5
8	21.3	24.5	22	26	20.7	16	14
9	23.5	21.3	28	27	26.5	20.7	36
10	28	23.5	36	28	30.6	26.5	40
11	24	28	40	29	32.3	30.6	49
12	15.5	24	3	30	29.5	32.3	7
13	17.3	15.5	21	31	28.3	29.5	52
14	25.3	17.3	29	32	31.3	28.3	65
15	25	25.3	62	33	32.3	31.3	17
16	36.5	25	65	34	26.4	32.3	5
17	36.5	36.5	46	35	23.4	26.4	17
18	29.6	36.5	44	36	16.4	23.4	1

Modelimiz:

$$Satış_t = \gamma + \alpha Satış_{t-1} + \beta\, Reklam_t + e_t \quad \text{for} \quad t = 2,...,T$$

şeklindedir. Amacımız, reklam harcamalarının satışlar üzerindeki etkisini belirlemektir.

γ, β, and α katsayılarını EKK ile elde edeceğiz. Eğer $|\alpha| < 1$ ise, bu bize zamana göre üssel azalma desenini verir. O halde:

Şu anki ve gelecekteki dönemlerin tümünde meydana gelecek toplam satış artışı:

$$\beta(1 + \alpha + \alpha^2 + \alpha^3 + \ldots) = \beta / (1 - \alpha)$$

Yalnızca k dönem sonraki etki:

$$\beta(1 + \alpha + \ldots + \alpha^k) = \beta(1 - \alpha^{k+1}) / (1 - \alpha)$$

k zamanındaki toplam reklam etkisinin yüzdesi:

$$100(1 - \alpha^{k+1})\ \%$$

Belli bir yüzde (p) etkinin ortaya çıkacağı dönem sayısı (örneğin %95 etkinin ortaya çıkacağı dönem sayısı):

$$k = \log(1 - p) / \log(\alpha) - 1$$

Satış-Reklam örneğimize ilişkin tahmin sonucu aşağıda verilmiştir:

$$SALES_t = 7.45 + 0.528\ SALES_{t-1} + 0.146\ ADVERT_t + \hat{e}_t$$

Elde edilen denkleme göre, mevcut ayda reklam harcamalarındaki 1 birim artış, satışı 0.146 birim artıracaktır. Yanısıra:

Mevcut ve gelecekteki aylarda satışta meydana gelecek toplam artış:

$\beta / (1 - \alpha) = 0.146 / (1 - 0.528) = 0.310$

%95 etkinin ortaya çıkacağı dönem sayısı:

$\log(1 - 0.95)/\log(0.528) - 1 = 3.69$

Buna göre, yaklaşık 4 ay sonra reklamın satış üzerine etkisinin %95'i ortaya çıkacaktır.

Yalnızca 2 dönem sonraki etki:

$\beta (1 + \alpha + \ldots + \alpha^k) = 0.146 (1 - 0.528^{2+1}) / (1 - 0.528) = 0.263$

2 zamanındaki toplam reklam etkisinin yüzdesi:

$100 (1 - \alpha^{k+1}) \% = 85 \%$

10.4. Koyck Dağıtılmış Gecikme Modelleri

Koyck modellerinde, gecikme sayısı arttıkça gecikmeli değişkenlerin katsayıları giderek azalır. Amacımız:

$Y_t = b_0 + b_1X_t + b_2X_{t-1} + b_3X_{t-2} + b_4X_{t-3} + \ldots + b_pX_{t-p}$

denklemini elde etmek ise:

Modelimiz:

$Y_t = a_0 + a_1X_t + \lambda Y_{t-1}$

şeklinde tahmin edilir ve gecikmeli değişkenler:

Denklem 10-2: $b_i = a_1\lambda^i$

formülüyle hesaplanır. Burada:

i = gecikme sayısı, 1, 2,..,p

λ = 0 ile 1 arasında bir değerdir.

Örneğin:

$b_3 = a_1\lambda^3$

olur.

$b_0 = a_0/(1-\lambda)$

formülüyle hesaplanır.

Denklem 10-2'yi, Denklem 10-1'da yerine koyarsak:

Denklem 10-3: $Y_t = b_0 + b_1 (X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \lambda^3 X_{t-3} + \ldots + \lambda^p X_{t-p})$

elde edilir.

λ, sıfır ile bir arasında bir değer olduğundan, λ'nın (n+1)'inci kuvveti, n'inci kuvvetinden daha küçük olacaktır. Örneğin:

λ=0.25 ise,

0.25^2=0.0625,

0.25^3= 0.015625

olur. Görüldüğü gibi, X_t değişkeninin gecikme sayısı büyüdükçe, Y'nin o dönemki değerini, daha küçük ağırlıklarla etkilemektedir.

Koyck modeli, doğrusal olmayan bir yapıya sahip olduğundan EKK ile tahmin edilemez. Ancak bazı düzenlemelerden sonra, Denklem 10-3, tahmincileri doğrusal bir forma çevrilebilir:

Denklem 10-4 Denklem 10-5: $Y_t = b_0 + b_1X_t + \lambda Y_{t-1}$

Ancak bu model, dağıtılmış gecikme modeli değil, otoregresif bir modeldir. Denklem 10-4 Denklem 10-5'ü tahmin etmek için örnek gözlem sayısının yeterince büyük olması gerekir. Gözlem sayısı 25'ten az ise kesinlikle kullanılmamalıdır.

Örnek 10-5

Toplam tüketim harcamalarını (C), gelirin Y bir fonksiyonu olarak tanımlayıp, Koyck modelini elde edelim:

Çizelge 10-1: Tüketim Harcamaları ve Gelir

Yıl	Tüketim Harcamaları C_t	Harcanabilir Gelir Y_t	Y_{t-1}	Y_{t-2}	Y_{t-3}	Y_{t-4}
1964	1417.2	1562.2	*	*	*	*
1965	1497.0	1653.5	1562.2	*	*	*
1966	1573.8	1734.3	1653.5	1562.2	*	*
1967	1622.4	1811.4	1734.3	1653.5	1562.2	*
1968	1707.5	1886.8	1811.4	1734.3	1653.5	1562.2
1969	1771.2	1947.4	1886.8	1811.4	1734.3	1653.5
1970	1813.5	2025.3	1947.4	1886.8	1811.4	1734.3
1971	1873.7	2099.9	2025.3	1947.4	1886.8	1811.4
1972	1978.4	2186.2	2099.9	2025.3	1947.4	1886.8
1973	2066.7	2334.1	2186.2	2099.9	2025.3	1947.4
1974	2053.8	2317	2334.1	2186.2	2099.9	2025.3
1975	2097.5	2355.4	2317.0	2334.1	2186.2	2099.9
1976	2207.3	2440.9	2355.4	2317.0	2334.1	2186.2
1977	2296.6	2512.6	2440.9	2355.4	2317.0	2334.1
1978	2391.8	2638.4	2512.6	2440.9	2355.4	2317.0
1979	2448.4	2710.1	2638.4	2512.6	2440.9	2355.4
1980	2447.1	2733.6	2710.1	2638.4	2512.6	2440.9
1981	2476.9	2795.8	2733.6	2710.1	2638.4	2512.6
1982	2503.7	2820.4	2795.8	2733.6	2710.1	2638.4
1983	2619.4	2893.6	2820.4	2795.8	2733.6	2710.1
1984	2746.1	3080.1	2893.6	2820.4	2795.8	2733.6
1985	2865.8	3162.1	3080.1	2893.6	2820.4	2795.8
1986	2969.1	3261.9	3162.1	3080.1	2893.6	2820.4
1987	3052.2	3289.5	3261.9	3162.1	3080.1	2893.6
1988	3162.4	3404.3	3289.5	3261.9	3162.1	3080.1
1989	3223.3	3464.9	3404.3	3289.5	3261.9	3162.1
1990	3272.6	3524.5	3464.9	3404.3	3289.5	3261.9
1991	3259.4	3538.5	3524.5	3464.9	3404.3	3289.5
1992	3349.5	3648.1	3538.5	3524.5	3464.9	3404.3
1993	3458.7	3704.1	3648.1	3538.5	3524.5	3464.9

1994	3578.5	3835.4	3704.1	3648.1	3538.5	3524.5

Koyck modelimiz:

$$C_t = a_0 + b_1Y_t + \lambda C_{t-1}$$

Tahmin sonuçlarımız:

	Tahminci	StHata	t	p
Sabit	-49.44	31.59	-1.57	0.129
Y_t	0.5348	0.1172	4.56	0.000
C_{t-1}	0.4446	0.124	3.59	0.001
S = 30.25	R^2 = 0.998	Düz.R^2 = 0.998	F=6151.6	p=0.000

Tahmin sonuçlarını denklem formunda yazalım:

$$C_t = a_0 + b_0Y_t + \lambda C_{t-1}$$

$$C_t = -49.44 + 0.5348\ Y_t + 0.4446\ C_{t-1}$$

Bu denklemi, dağıtılmış gecikmeli modele çevirebiliriz. b_1=0.5348 ve λ=0.4446'dır. O halde:

$$b_i = b_1\lambda^i$$

$$b_1 = 0.5348(0.4446)^1 = 0.2377$$

$$b_2 = 0.5348(0.4446)^2 = 0.1057$$

$$b_3 = 0.5348(0.4446)^3 = 0.0470$$

Bu katsayıları, dağıtılmış gecikmeli modelimizde yerine koymadan önce, b_0'ı da hesaplayalım:

$$\beta_0 = a_0/(1-\lambda)$$

$$\beta_0 = -49.44/(1-0.4446)$$

$$\beta_0 = -89.0169$$

Son olarak denklemimiz:

$$C_t = -89.9169 + 0.5348Y_t + 0.2377Y_{t-1} + 0.1057Y_{t-2} + 0.0470Y_{t-3}$$

olur. Görüldüğü gibi, gelir değişkeninin gecikme sayısı arttıkça, tüketim harcamaları üzerinde etkisi azalmaktadır. Eğer, dağıtılmış gecikme modelini, EKK ile tahmin etseydik:

	Tahminci	StHata	t	p
Sabit	-152.8	40.91	-3.74	0.001
Y_t	0.9186	0.1695	5.42	0.000
Y_{t-1}	0.0984	0.2245	0.44	0.665
Y_{t-2}	-0.0536	0.2298	-0.23	0.818
Y_{t-3}	0.0002	0.1747	0	0.999
S = 37.09	R^2 = 0.997	Düz.R^2 = 0.996	F=1697.4	p=0.000

EKK sonuçlarını incelersek, gelir değişkeninin (Y_t) katsayısı beklendiği gibi pozitiftir. İlk gecikmeli gelir değişkeninin katsayısı da, istatistiki açıdan önemsiz olmakla birlikte, aynı şekilde pozitif ve daha düşüktür. Ancak ikinci gecikmede katsayı, negatiftir. Kuşkusuz bu sonuçlar çelişkilidir ve yararlanmak mümkün değildir.

Koyck modelinin bize sunduğu önemli bir bilgi de, uzun dönem etkisidir. Gelirin tüketim harcamalarına kısa dönemde yaptığı etki, gecikmesiz gelir değişkeninin tahmincisidir. Koyck modelinin tahmin sonuçları:

$$C_t = -49.44 + 0.5348\, Y_t + 0.4446\, C_{t-1}$$

şeklindeydi. Buna göre, gelirin kısa dönem etkisi 0.5348'dir. Bir başka anlatımla, kısa dönemde gelir artışının % 53.48'i tüketim harcamalarına gitmektedir. Uzun dönem etkisini hesaplamak için:

$$b_1[1/(1-\lambda)]$$

formülünü kullanacağız. İlgili katsayıları yerine koyarsak:

$$b_1[1/(1-\lambda)] = 0.5348[1/(1-0.4446)] = 0.96$$

elde ederiz. O halde, uzun dönemde gelir artışının %96'sı tüketim harcamalarına gitmektedir.

Harcama ile gelir ilişkisini, çift logaritmik formda elde etseydik, kısa ve uzun dönem esnekliklerini elde edebilirdik.

Koyck modelimiz:

$$\ln C_t = a_0 + b_1 \ln Y_t + \lambda \ln C_{t-1}$$

ise, tahmin sonuçlarımız:

	Tahminci	StHata	T	p
Sabit	-0.1687	0.1117	-1.51	0.143
$\ln Y_t$	0.6385	0.1458	4.38	0.000
$\ln C_{t-1}$	0.3767	0.137	2.75	0.011
S = 0.012	R^2 = 0.998	Düz.R^2 = 0.998	F=6294.4	p=0.000

Denklem olarak sunumumuz:

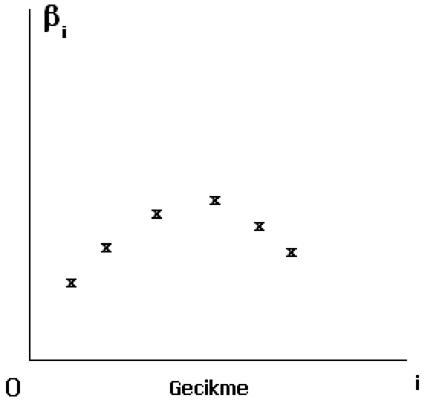
$$\ln C_t = -0.1687 + 0.6385 \ln Y_t + 0.3767 \ln C_{t-1}$$

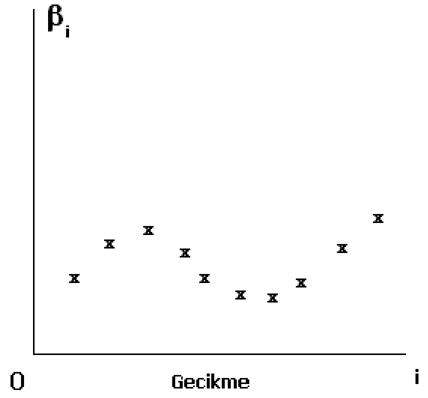
Kısa dönem gelir esnekliği, 0.6385'tir. Uzun dönem esnekliği ise:

$$b_1[1/(1-\lambda)] = 0.6385[1/(1-0.3767)] = 1.02$$

10.5. Almon Dağıtılmış Gecikme Modeli

Koyck β katsayılarının geometrik olarak azaldığını varsayar. Ancak bazı durumlarda β'nın zaman içinde aşağıdaki grafiklerdekine benzer bir seyir göstermesi mümkündür:





(A) (B)

Shirley Almon bu sorunun çözümü için Almon modelini önermiştir. Modelimiz:

Denklem 10-6: $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_k X_{t-k} + u_t$

olsun. Bunu toplam notasyonuyla:

Denklem 10-7: $Y_t = \alpha + \sum_{i=0}^{k} \beta_i X_{t-i} + u_t$

şeklinde yazabiliriz. Almon'a göre Weierstrass teoremi gereği, β_i'nin gecikme uzunluğu i'nin uygun dereceden bir çok terimlisi ile yaklaşık olarak bulunabilir. β_i, grafik (A)'daki gibi ise:

Denklem 10-8: $\beta_i = a_0 + a_1 i + a_2 i^2$

ikinci dereceden bir çok terimli; β_i, grafik (B)'deki gibi ise:

Denklem 10-9: $\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3$

şeklinde ifade edilebilir. Genel yazımı:

Denklem 10-10: $\beta_i = a_0 + a_1 i + a_2 i^2 + .. + a_m i^m$

olur. β'ların seyri grafik (A)'daki gibi ise:

Denklem 10-8'yı, Denklem 10-7'te yerine koyarsak:

$$Y_t = \alpha + \sum_{i=0}^{k} (a_0 + a_1 i + a_2 i^2) X_{i-1} + u_t$$

olur. Ardından:

$$Z_{0t} = \sum_{i=0}^{k} X_{t-i}$$

$$Z_{1t} = \sum_{i=0}^{k} i X_{t-i}$$

$$Z_{2t} = \sum_{i=0}^{k} i^2 X_{t-i}$$

tanımlamalarını yaptıktan sonra:

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + u_t$$

elde edilir. Almon modelde Z değişkenlerine göre regresyon modellemesi yapılır. Bu durumda:

$$\hat{\beta}_0 = \hat{a}_0$$
$$\hat{\beta}_1 = \hat{a}_0 + \hat{a}_1 + \hat{a}_2$$
$$\hat{\beta}_2 = \hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2$$
$$\hat{\beta}_3 = \hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2$$
$$\hat{\beta}_4 = \hat{a}_0 + 4\hat{a}_1 + 16\hat{a}_2$$
....
$$\hat{\beta}_m = \hat{a}_0 + k\hat{a}_1 + k^2\hat{a}_2$$

olur.

Almon modellemede yapılması gerekenler:

1. En büyük k gecikmesi tesbit edilir.
2. m derecesi (Z sayısı) belirlenir (2 ya da 3 olabilir)

Aşağıda m=2 ve gecikme uzunluğu k=5 için Z'ler oluşturulmuştur:

$$Z_{0t} = \sum_{i=0}^{5} X_{t-i} = X_t + X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4} + X_{t-5}$$

$$Z_{1t} = \sum_{i=0}^{5} iX_{t-i} = X_{t-1} + 2X_{t-2} + 3X_{t-3} + 4X_{t-4} + 5X_{t-5}$$

$$Z_{2t} = \sum_{i=0}^{5} i^2 X_{t-i} = X_{t-1} + 4X_{t-2} + 9X_{t-3} + 16X_{t-4} + 25X_{t-5}$$

Z'ler X'lerin bir fonksiyonudur. Z'ler arasında çoklu doğrusallık olabileceği unutulmamalıdır.

Örnek 10-6

Stok, satışın bir fonksiyonu olarak Almon modelle tahmin edilecektir.

k=3 ve m=2 olsun:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$$

Almon modelini tahmin edelim. Z'ler:

$$Z_{0t} = \sum_{i=0}^{3} X_{t-i} = X_t + X_{t-1} + X_{t-2} + X_{t-3}$$

$$Z_{1t} = \sum_{i=0}^{3} iX_{t-i} = X_{t-1} + 2X_{t-2} + 3X_{t-3}$$

$$Z_{2t} = \sum_{i=0}^{3} i^2 X_{t-i} = X_{t-1} + 4X_{t-2} + 9X_{t-3}$$

Yıl	Y	X	X_{t-1}	X_{t-2}	X_{t-3}	Z_{0t} $X_t+X_{t-1}+X_{t-2}+X_{t-3}$	Z_{1t} $X_{t-1}+2X_{t-2}+3X_{t-3}$	Z_{2t} $X_{t-1}+4X_{t-2}+9X_{t-3}$
1955	45069	26480		*	*	*	*	*
1956	50642	27740	26480	*	*	*	*	*
1957	51871	28736	27740	26480	*	*	*	*

1958	50070	27280	28736	27740	26480	110236	163656	378016
1959	52707	30219	27280	28736	27740	113975	167972	391884
1960	53814	30796	30219	27280	28736	117031	170987	397963
1961	54939	30896	30796	30219	27280	119191	173074	397192
1962	58213	33113	30896	30796	30219	125024	183145	426051
1963	60043	35032	33113	30896	30796	129837	187293	433861
1964	63383	37335	35032	33113	30896	136376	193946	445548
1965	68221	41003	37335	35032	33113	146483	206738	475480
1966	77965	44869	41003	37335	35032	158239	220769	505631
1967	84655	46449	44869	41003	37335	169656	238880	544896
1968	90875	50282	46449	44869	41003	182603	259196	594952
1969	97074	53555	50282	46449	44869	195155	277787	639899
1970	101645	52859	53555	50282	46449	203145	293466	672724
1971	102445	55917	52859	53555	50282	212613	310815	719617
1972	107719	62017	55917	52859	53555	224348	322300	749348
1973	120870	71398	62017	55917	52859	242191	332428	761416
1974	147135	82078	71398	62017	55917	271410	363183	822719

$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + u_t$ model tahmin edildiğinde:

	Tahminci	StHata	t	P
α	-7140.75	1992.988	-3.58294	0.003
Z_{0t}	0.661248	0.16548	3.995947	0.002
Z_{1t}	0.902049	0.483131	1.86709	0.084
Z_{2t}	-0.43216	0.166464	-2.59608	0.022
	R^2=0.998	Düz R^2=0.997	Sd=13	

elde edilir. Z'ler elde edildiğine göre dağıtılmış gecikme modelimizin katsayılarını hesaplayabiliriz. Modelimiz:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$$

idi. β tahmincilerinin hesaplanması aşağıdaki formüllerle yapılıyordu:

$$\hat{\beta}_0 = \hat{a}_0 = 0.661248$$

$$\hat{\beta}_1 = \hat{a}_0 + \hat{a}_1 + \hat{a}_2 = 0.661248+0.902049+(-0.43216)=1.1310$$

$$\hat{\beta}_2 = \hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2 = 0.661248+2*0.902049+4*(-0.43216)= 0.7364$$

$$\hat{\beta}_3 = \hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2 = 0.661248+3*0.902049+9*(-0.43216)= -0.5226$$

Şimdi tahmincileri yerine koyalım:

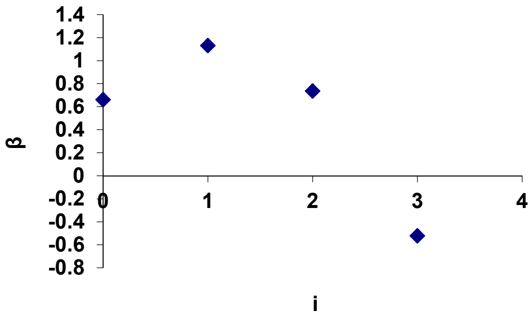
$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$$

$$Y_t=-7140.75+0.661248X_t+1.1310X_{t-1}+0.7364X_{t-2}-0.5226X_{t-3}$$

β tahmincilerinin grafiğini çizelim:

i	**β_i**
0	0.661248

1	1.1310
2	0.7364
3	-0.5226



10.6. Kısmi Düzeltme Modeli

Y_t^* bir şirketin arzu ettiği stok mal düzeyi, Y_t gerçek stok mal düzeyi ve X_t satış miktarı olsun. Arzu edilen stok mal düzeyinin satışlara bağlı olduğunu varsayarsak:

Denklem 10-11 $\quad Y_t^* = \alpha + \beta X_t$

Pazardaki belirsizliklerden dolayı, arzu edilen ve gerçek stok mal düzeyleri arasındaki açık, bir anda kapatılamaz. Ancak her dönemde açığın belli bir kısmı kapatılabilir. Bu durumda t zamanındaki stok mal düzeyi; t-1 zamanındaki stok mal düzeyine, düzeltme faktörü ve hata teriminin eklenmesine eşit olacaktır:

Denklem 10-12 $\quad Y_t = Y_{t-1} + \lambda\,(Y_t^* - Y_{t-1}) + u_t, \qquad 0 < \lambda < 1,$

Bu model, kısmî düzeltme modeli olarak bilinir. λ parametresi, kısmî düzeltme katsayısı; $1/\lambda$, düzeltme hızıdır. Düzeltme katsayısı, açığın bir dönemde kapatılacak oransal miktarını; düzeltme hızı ise, açığın tamamen kapatılabilmesi için geçmesi gereken dönem sayısını verir. Örneğin; $\lambda = 0.25$ ise, bir dönemde açığın %25'i kapatılabilecektir; açığın tamamen kapanması için geçecek süre ise, $1/\lambda=1/0.25=4$ yıldır.

Denklem 10-11'u Denklem 10-12'de yerine koyarsak:

Denklem 10-13 $\quad Y_t = \alpha\lambda + (1-\lambda)Y_{t-1} + \beta\lambda X_t + u_t$

Denklem 10-14 $\quad Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_t + u_t$

Denklem 10-14 tahmin edildikten sonra, β değerleri Denklem 10-13'deki parametrelere eşitlenerek, herbir katsayının değeri hesaplanır.

$\beta_1 = \alpha\lambda,\ \beta_2 = (1-\lambda),\ \beta_3 = \beta\lambda$

Örnek 10-7

Aşağıdaki veriler, 1950-1991 arasındaki stok ve satış miktarlarını içermektedir. Bu verilere uygun bir kısmi düzeltme modeli tahmin edip, sonuçlarını yorumlayalım.

Yıl	Satış (X)	Stok (Y)	Yıl	Satış (X)	Stok (Y)
1950	38596	59822	**1971**	117023	188991
1951	43356	70242	**1972**	131227	203227
1952	44840	72377	**1973**	153881	234406
1953	47987	76122	**1974**	178201	287144
1954	46443	73175	**1975**	182412	288992
1955	51694	79516	**1976**	204386	318345
1956	54063	87304	**1977**	229786	350706
1957	55879	89052	**1978**	260755	400929
1958	54021	87055	**1979**	298328	452636

1959	59729	92097	**1980**	328112	510124
1960	60827	94719	**1981**	356909	547169
1961	61159	95580	**1982**	348771	575486
1962	65662	101049	**1983**	370501	591858
1963	68995	105463	**1984**	411427	651527
1964	73682	111504	**1985**	423940	665837
1965	80283	120929	**1986**	431786	664654
1966	87187	136824	**1987**	459107	711745
1967	90918	145681	**1988**	496334	767387
1968	98794	156611	**1989**	522344	813018
1969	105812	170400	**1990**	540788	835985
1970	108352	178594	**1991**	533838	828184

PAM modelimiz:

$$STOK_t = \alpha\lambda + (1-\lambda)STOK_{t-1} + \beta\lambda SATIS_t + u_t$$

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_t + u_t$$

Model 1: EKK tahminleri 41 gözlem 1951-1991
Bağımlı değişken: STOK

	Tahminci	StHata	t	P	
sabit	2449.3	1717.87	1.4258	0.162096	
$STOK_{t-1}$	0.17949	0.0594496	3.0192	0.004511	***
$SATIS_t$	1.28675	0.0887745	14.4946	< 0.00001	***

R^2 = 0.999334
Düzeltilmiş R^2 = 0.999299
F-istatistiği (2, 38) = 28524.8 ($p < 0.00001$)
Durbin-Watson istatistiği = 1.53123
Durbin's h stat. 1.45894

$$Y_t = 2449.3 + 0.17949Y_{t-1} + 1.28675X_t$$

$\beta_1 = \alpha\lambda$, $0.17949 = (1-\lambda)$ **λ=0.82051**

$\alpha = 2449.3/0.82051$ **α=2985.09**

$1.28675 = \beta * 0.82051$ **β=1.5682**

Hesaplama sonuçlarını:

$$Y_t^* = \alpha + \beta X_t$$

denkleminde yerine koyarsak:

$$\mathbf{Y_t^* = 2985.09 + 1.5682X_t}$$

olur. Böylece mümkün değil gibi görünen arzu edilen stok mal düzeyi ile satış arasındaki ilişkiyi tahmin etmiş olduk.

λ = 0.82 olduğuna göre, satış ve stok arasındaki açığın %82'si bir dönemde kapatılabilecektir; açığın tamamen kapanması için geçecek süre ise, 1/0.82=1.3 yıldır

10.7. Uyumcu Beklentiler Modeli

Y_t tüketim, X_t^* beklenen gelir ve X_t mevcut gelir olsun. Bu durumda tüketimin mevcut gelire değil beklenen gelire bağlı olması beklenir:

$$Y_t = \alpha + \beta X_t^* + u_t$$

β beklenen gelir üzerinden marjinal tüketim eğilimidir. X_t^* gözlemlenebilir bir değişken olmadığından, bu denklem tahmin edilemez. Bu durumda modelde yeni bir yapılanma yoluna gitmemiz gerekir. Buna göre, tüketicilerin daha önceki dönemlerde gelire ilişkin beklentilerinin ne denli gerçekleştiğini kontrol ettiğini düşünelim. Beklentideki değişimin ($X_t^* - X^*_{t-1}$), X_{t-1} ve X^*_{t-1} arasındaki sapmaya bağlı olduğu varsayılırsa:

$$X_t^* - X^*_{t-1} = \lambda(X_{t-1} - X^*_{t-1}) \qquad 0 < \lambda < 1$$

Eğer t-1 zamanında gerçekleşen gelir, beklentilerden daha fazla olursa, tüketicilerin beklentilerini yükseltmesini umarız. Bu durumda:

$$X_t^* = \lambda X_{t-1} + (1-\lambda)X^*_{t-1}$$

Denklemlerde yapılan düzenlemelerden sonra:

$$Y_t = \alpha\lambda + (1- \lambda)Y_{t-1} + \beta\lambda X_{t-1} + u_t + (1-\lambda)u_{t-1}$$

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_{t-1} + v_t$$

Burada:

$$\beta_1 = \alpha\lambda,\ \beta_2 = (1- \lambda),\ \beta_3 = \beta\lambda,\ v_t = u_t + (1-\lambda)u_{t-1}$$

formülleriyle hesaplanır.

Örnek 10-8

YIL	Tüketim	Gelir
1967	61284	78221
1968	68814	83326
1969	76766	90490
1970	73576	92692
1971	73256	94814
1972	67502	92590
1973	78832	101419
1974	80240	105267
1975	84477	112149
1976	86038	116078
1977	96275	122040
1978	101292	128578
1979	105448	136851
1980	114570	144734
1981	120477	152846
1982	133868	164318
1983	148004	172414
1984	149735	178433
1985	155200	185753
1986	154165	192059
1987	155445	191288
1988	157199	196055
1989	158576	202477
1990	169238	223225

1991	179001	233231
1992	183687	242762
1993	198273	259555

$Tüketim_t = \beta_1 + \beta_2 Tüketim_{t-1} + \beta_3 Gelir_{t-1} + v_t$

$Y_t = \alpha\lambda + (1-\lambda)Y_{t-1} + \beta\lambda X_{t-1} + u_t + (1-\lambda)u_{t-1}$

$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_{t-1} + v_t$

Model 2: EKK tahminleri 26 gözlem 1968-1993
Bağımlı değişken: Y_t

	Tahminci	StHata	t	P	
Sabit	1338.88	3032.74	0.4415	0.662991	
Y_{t-1}	0.553707	0.19709	2.8094	0.009952	***
X_{t-1}	0.379689	0.156899	2.4200	0.023828	**

Bağımlı değişkenin ortalaması = 121921
Bağımlı değişkenin st sapması = 41385.1
Sum of kare residuals = 5.51568e+008
Hataların st hatası= 4897.06
$R^2 = 0.987118$
Düzeltilmiş $R^2 = 0.985998$
F-istatistiği (2, 23) = 881.244 (p < 0.00001)
Durbin-Watson istatistiği = 1.66286
First-order autocorrelation coeff. = 0.149269
Durbin's h stat. 4.39105
(Using 3 for h stat, with T' = 25)
Log-likelihood = -256.205
Akaike bilgi kriteri = 518.409
Schwarz Bayesian kriteri = 522.184

$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_{t-1}$

$Y_t = 1338.88 + 0.553707Y_{t-1} + 0.379689X_{t-1}$

$\beta_2 = (1-\lambda)$, $0.553707=(1-\lambda)$ $\boldsymbol{\lambda=0.446293}$

$\beta_1 = \alpha\lambda$ $1338.88=\alpha*0.446293$ $\boldsymbol{\alpha=3000}$

$\beta_3 = \beta\lambda$ $0.379689=\beta*0.446293$ $\boldsymbol{\beta=0.85076}$

$Y_t = \alpha + \beta X_t^*$

$\mathbf{Y_t = 3000 + 0.85076X_t^*}$

Böylece, X_t^* gözlemlenebilir bir değişken olmamasına rağmen, beklenen gelir üzerinden marjinal tüketim eğilimini hesaplayabileceğimiz bir model geliştirdik.

11. ÖNGÖRÜMLEME

Herhangi bir organizasyonu yönetmenin en kritik yönü gelecek için plan yapmaktır.Bir organizasyonun uzun süreli başarısı yöneticilerin geleceği önceden sezmeleri ve uygun strateji geliştirmeleri ile yakından ilgilidir. Uygun yargı, önsezi ve ekonomik durumdan haberdar olma yöneticiye gelecekte olacabilecek olaylar hakkında fikir ve düşünce verebilmektedir. Bununla birlikte, bu düşünceyi gelecek üç aydaki satış miktarı veya bir sonraki yılın hammadde masrafı olarak kullanılabilecek bir rakama dönüştürmek çok zordur.

ZAMAN SERİLERİNİN BİLEŞENLERİ

Zaman serilerindeki verilerin gidişatını (seyrini) açıklamak için zaman serilerinin çeşitli bileşenlerden oluştuklarını düşünmek yardımcı olacaktır. Genel varsayım dört ayrı bileşendir; trend (eğilim), dönemsel, mevsimlik ve düzensiz.

TREND BİLEŞENİ

Zaman serileri analizinde ölçüler saatlik, günlük, haftalık, aylık, yıllık veya herhangi bir düzenli aralık olarak alınabilir. Genellikle zaman serileri verilerinin düzensiz iniş çıkış göstermelerine rağmen, daha geniş bir periyotta daha yüksek veya daha düşük değerlere aşamalı değişiklik veya hareket göstermeye devam edebilirler. Zaman serilerinin bu aşamalı değişikliğine “trend” adı verilmektedir ve genellikle nüfustaki, nüfusun demografik özelliklerindeki, teknolojideki ve/veya tüketici tercihlerindeki değişimlerin uzun dönemdeki sonucudur.

DÖNEMSEL BİLEŞENİ

Zaman serilerinin eğilim gösterebilmesine rağmen, zaman serilerinin gelecek değerlerinin hepsi olmamakla beraber trend çizgisinde yer alabilmektedir. Bu durumda, zaman serileri sık sık trend çizgisinin altında veya üzerinde alternatif ardışık noktalar göstermektedir. Trend çizgisinin altında veya üzerinde tekrarlanan ve bir yıldan fazla süren ardışık noktalar zaman serilerinin “dönemsel bileşeni” olarak gösterilmektedir.

MEVSİMLİK BİLEŞEN

Zaman serileri bir yıl içerisinde genellikle düzgün gidişatlı değişkenlik gösterirler. Örneğin, bir yüzme havuzu yapımcısı sonbahar ve kış aylarında düşük satış rakamları beklerken ilkbahar ve yaz aylarında en yüksek noktaya ulaşacağını düşünür. Kar temizleme aleti ve kalın kıyafetler üreten firmalar ise tam tersi mevsimlik satışları hedeflemektedir. Genellikle mevsimlik hareketlerin bir yıl içerisinde oluştuğunu düşünmemize rağmen, mevsimlik bileşen bir yıldan daha kısa bir zaman içerisinde tekrarlanabilir. Örneğin, günlük trafik gün içerisinde mevsimlik hareket, yoğun zamanlarında en yüksek seviyesini, akşam ve kalan diğer zamanlarda orta derece yoğunluk ve geceyarısından sabaha kadar yavaş akış gösterebilir.

DÜZENSİZ BİLEŞEN

Düzensiz bileşen zaman serilerini etkileyen beklenmeyen ve tekrarlanmayan faktörler nedeniyle kısa dönemde oluşmaktadır. Zaman serilerindeki ilerleyişini önceden tahmin edemeyiz.

11.1. TAHMİNLEME YÖNTEMLERİ

11.1.1. HAREKETLİ ORTALAMALAR

Haftalık Pirinç Satışı

Hafta (t)	Satış(1000 kg) (Y_t)
1	17
2	21
3	19
4	23
5	18
6	16
7	20
8	18
9	22
10	20
11	15
12	22

Üç Dönemlik Hareketli Ortalama

İlk üç hafta için hareketli ortalama (4. hafta tahmini)

Hareketli Ortalama (1-3. hafta) = (17+21+19) / 3 = 19

Bu ortalamayı 4. haftanın tahminidir. Bu durumda 4. haftanın değeri 23 olduğu için 4. haftada tahmin hatası 4 tür.

İkinci üç hafta için hareketli ortalama (5. hafta tahmini)

Hareketli Ortalama (2-4. hafta) = (21+19+23) / 3 = 21

5. haftanın tahmini 21'dir. Tahmini hata ise 18-21= -3'tür.

Tahminin doğruluğu

Tahminleme yöntemini seçerken dikkat edilmesi gereken en önemli nokta tahminin kesin olmasıdır. Açıkçası tahmin hatasının en küçük olmasını isteriz.

Bu amaçla kullanılan ölçütler:

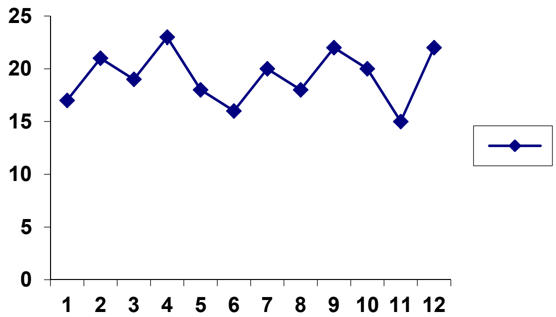
Ortalama hata kareleri: $MSE = \dfrac{\sum e_t^2}{n}$

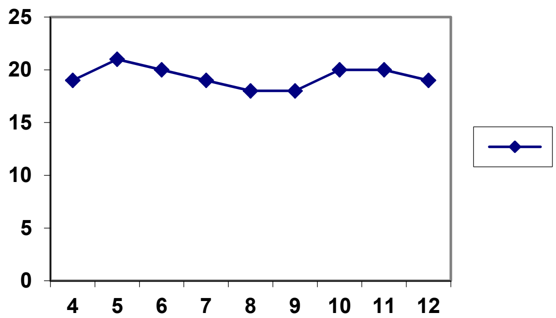
Ortalama mutlak yüzde hata: $MAPE = \dfrac{\sum |e_t \; / Y_t|}{n}.100$

Ortalama mutlak sapma: $MAD = \dfrac{\sum |e_t|}{n}$

Üç Haftalık Hareketli Ortalama Hesaplarının Özeti

Hafta	Y_t	$\hat{Y}_t$	e_t	Hatanın Karesi	$\lvert e_t/Y_t \rvert$
1	17				
2	21				
3	19				
4	23	19	4	16	4/23=0.173913
5	18	21	-3	9	3/18=0.166667
6	16	20	-4	16	0.25
7	20	19	1	1	0.05
8	18	18	0	0	0
9	22	18	4	16	0.181818
10	20	20	0	0	0
11	15	20	-5	25	0.333333
12	22	19	3	9	0.136364
13		19			
			Toplam	92	1.292095





Hata Karelerinin Ortalaması = MSE = 92 / 9 = 10.22

Ortalama Mutlak Yüzde Hata= 1.292095 / 9 *100= %14.4

Ortalama Mutlak Sapma = MAD = 24 / 9 = 2.67

13. Haftanın Tahmin Değeri

13. hafta tahmini

Hareketli Ortalama (10-12. hafta) = (20+15+22) / 3 = 19

Dört Haftalık Hareketli Ortalama

Dört Haftalık Hareketli Ortalama

Hafta	Y_t	$\hat{Y}_t$	e_t	Hatanın Karesi	e_t/Y_t
1	17				
2	21				
3	19				
4	23				
5	18	20	-2	4	0.111111
6	16	20.25	-4.25	18.0625	0.265625
7	20	19	1	1	0.05
8	18	19.25	-1.25	1.5625	0.069444
9	22	18	4	16	0.181818
10	20	19	1	1	0.05
11	15	20	-5	25	0.333333
12	22	18.75	3.25	10.5625	0.147727
13		19.75			
			Toplam	77.1875	1.209

Hata Karelerinin Ortalaması = MSE = 77.1875 / 8 = 9.65

Ortalama Mutlak Yüzde Hata= MAPE=1.209 / 8 *100= %15.11

Ortalama Mutlak Sapma = MAD = 21.75 / 8 = 2.72

11.1.2. AĞIRLIKLI HAREKETLİ ORTALAMALAR

3 haftalık ağırlıklı hareketli ortalama:

1+2+3 = 6

1/6, 2/6, 3/6

4. hafta için tahmin= 1/6 (17)+ 2/6 (21)+ 3/6 (19)=19.33

5. hafta için tahmin= 1/6 (21)+ 2/6 (19)+ 3/6 (23)=21.33

6. hafta için tahmin= 1/6 (19)+ 2/6 (23)+ 3/6 (18)=19.83

13. hafta için tahmin= 1/6 (20)+ 2/6 (15)+ 3/6 (22)=19.75

Ağırlıklı Üç Haftalık Hareketli Ortalama

Hafta	Y_t	$\hat{Y}_t$	e_t	Hatanın Karesi	e_t/Y_t
1	17				
2	21				
3	19				
4	23	19.33333	3.66667	13.44447	0.15942
5	18	21.33333	-3.33333	11.11109	0.185185
6	16	19.83333	-3.83333	14.69442	0.239583
7	20	17.83333	2.16667	4.694459	0.108334
8	18	18.33333	-0.33333	0.111109	0.018518
9	22	18.33333	3.66667	13.44447	0.166667
10	20	20.33333	-0.33333	0.111109	0.016667
11	15	20.33333	-5.33333	28.44441	0.355555
12	22	17.83333	4.16667	17.36114	0.189394
13		19.75			
			Toplam	103.4167	1.439

Hata Karelerinin Ortalaması = 103.4167 / 9 = 11.49

Hata Karelerinin Ortalaması = MSE = 103.4167 / 9 = 11.49
Ortalama Mutlak Yüzde Hata= MAPE=1.439 / 9 *100= %16
Ortalama Mutlak Sapma = MAD = 26.83 / 9 = 2.98

11.1.3.ÜSTEL YUMUŞATMA

Üstel yumuşatma, tahmin olarak geçmiş zaman serilerinin ağırlıklı ortalamalarını kullanmaktadır. Sadece son gözlem için bir ağırlık seçtiğimiz ağırlıklı hareketli ortalamalar yönteminin özel bir durumudur.

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

F_{t+1} = Zaman serisinin t+1 periyodu için tahmin

Y_t = Zaman serisinin t periyodu için gözlenen değer

F_t = Zaman serisinin t periyodu için tahmin

α = Yumuşatma katsayısı ($0 \leq \alpha \leq 1$)

Hesapları başlatmak için F_1 in 1. periyotta gözlenen değere eşit olduğunu varsayarsak, $F_1 = Y_1$ olur. Böylece ikinci periyotta tahmin,

$$F_2 = \alpha Y_1 + (1-\alpha)F_1 \qquad = \alpha Y_1 + (1-\alpha)F_1 \qquad = Y_1$$

$$F_3 = .2Y_2 + .8F_2 = .2(21) + .8(17) = 17.8$$

$$F_4 = .2Y_3 + .8F_3 = .2(19) + .8(17.8) = 18.04$$

α =0.2 için Pirinç Satış Tahminlerinin HKT Hesaplamaları				
Hafta	Zaman Serisi Değeri	Tahmin	Tahmin Hatası	Tah.Hata.Karesi
(t)	Y_t	F_t	$Y_t - F_t$	$(Y_t - F_t)^2$
1	17			
2	21	17	4	16
3	19	17.8	1.2	1.44
4	23	18.04	4.96	24.6
5	18	19.03	-1.03	1.061
6	16	18.83	-2.83	8.009
7	20	18.26	1.74	3.028
8	18	18.61	-0.61	0.372
9	22	18.49	3.51	12.32
10	20	19.19	0.81	0.656
11	15	19.35	-4.35	18.92
12	22	18.48	3.52	12.39
13		19.18		
			Toplam	98.8

MSE=98.80/11=8.98

MAPE= 1.47 / 11 *100= %13

MAD = 28.56 / 11 = 2.59

$$F_{13} = .2Y_{12} + .8F_{12} = .2(22) + .8(18.48) = 19.18$$

α =0.3 için Pirinç Satış Tahminlerinin HKT Hesaplamaları				
Hafta	Zaman Serisi Değeri	Tahmin	Tahmin Hatası	Tah.Hata.Karesi
(t)	Y_t	F_t	$Y_t - F_t$	$(Y_t - F_t)^2$
1	17			
2	21	17.00	4.00	16.00
3	19	18.20	.80	.64
4	23	18.44	4.56	20.79
5	18	19.81	-1.81	3.28
6	16	19.27	-3.27	10.69
7	20	18.29	1.71	2.92
8	18	18.80	-.80	.64
9	22	18.56	3.44	11.83
10	20	19.59	.41	.17
11	15	19.71	-4.71	22.18
12	22	18.30	3.70	13.69
13		19.41		
			Toplam	102.83

MSE=102.83/11=9.35

MAPE= 1.52 / 11 *100= %13.8

MAD = 29.21 / 11 = 2.66

$$F_{13} = 0.3Y_{12} + 0.7F_{12} = 0.3(22) + 0.7(18.30) = 19.41$$

11.1.4.TREND

Trend, bir zaman serisi değişkeninin geçmişteki eğilimine bakıp, gelecekte de bunun devam edeceği varsayılır.

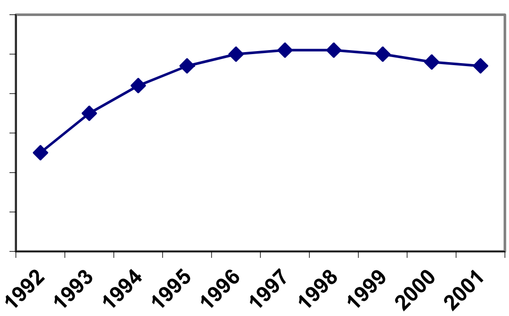
Y = f (zaman) Y=f(t)

Doğrusal trend: $Y_t=b_o+b_1t$

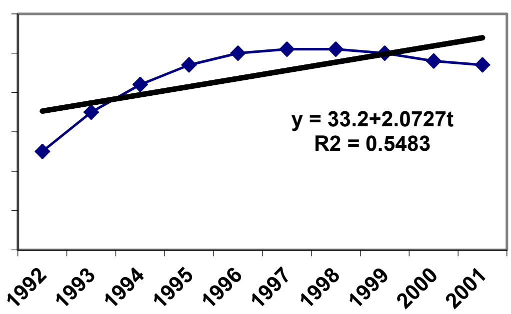
Kuadratik trend: $Y_t=b_o+b_1t + b_2t^2$

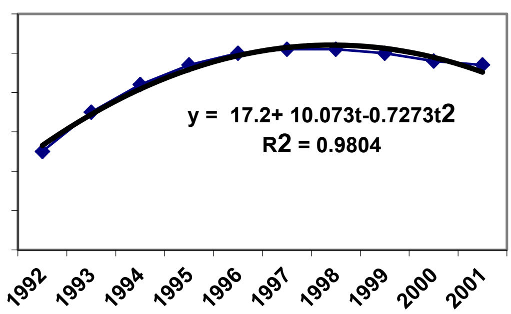
3. dereceden trend: $Y_t= b_o+b_1t + b_2t^2+ b_3t^3$

YIL	t	Yt
1992	1	25
1993	2	35
1994	3	42
1995	4	47
1996	5	50
1997	6	51
1998	7	51
1999	8	50
2000	9	48
2001	10	47



Y= f(t)





Doğrusal (1992 = 1, Δt=1)

$Y = 33.2 + 2.07\, t$

Değişken	Katsayı	StHata	T	P
Sabit	33.200	4.127	8.04	0.000
t	2.0727	0.6651	3.12	0.014

S = 6.041   R-Sq = 54.8%   R-Sq(adj) = 49.2%

Kuadratik (1992 = 1, Δt=1)

$Y = 17.2 + 10.1\, t - 0.727\, t^2$

Değişken	Katsayı	StHata	T	P
Sabit	17.200	1.584	10.86	0.000
t	10.0727	0.6614	15.23	0.000
t^2	-0.72727	0.05860	-12.41	0.000

S = 1.346   R-Sq = 98.0%   R-Sq(adj) = 97.5%

Tahminler:

Doğrusal

YIL	t	Yt	$\hat{Y}_t$ =bo+bıt	$\hat{Y}_t$	et	e2
1992	1	25	Y_1=33.2+2.0727(1)	35.2727	-10.2727	105.529
1993	2	35	Y_2=33.2+2.0727(2)	37.3455	-2.3455	5.501
1994	3	42	Y_3=33.2+2.0727(3)	39.4182	2.5818	6.666
1995	4	47	Y_4=33.2+2.0727(4)	41.4909	5.5091	30.350
1996	5	50	Y_5=33.2+2.0727(5)	43.5636	6.4364	41.427
1997	6	51	Y_6=33.2+2.0727(6)	45.6364	5.3636	28.769
1998	7	51	Y_7=33.2+2.0727(7)	47.7091	3.2909	10.830
1999	8	50	Y_8=33.2+2.0727(8)	49.7818	0.2182	0.048
2000	9	48	Y_9=33.2+2.0727(9)	51.8545	-3.8545	14.858
2001	10	47	Y_{10}=33.2+2.0727(10)	53.9273	-6.9273	47.987
					HKT	291.965

MSE=29.1965 MAPE=%11.8 MAD=4.68

YIL	t	$\hat{Y}_t$ =bo+bıt	$\hat{Y}_t$
2002	11	Y_{11}=33.2+2.0727(11)	56.00
2003	12	Y_{12}=33.2+2.0727(12)	58.07
2004	13	Y_{13}=33.2+2.0727(13)	60.15

Kuadratik

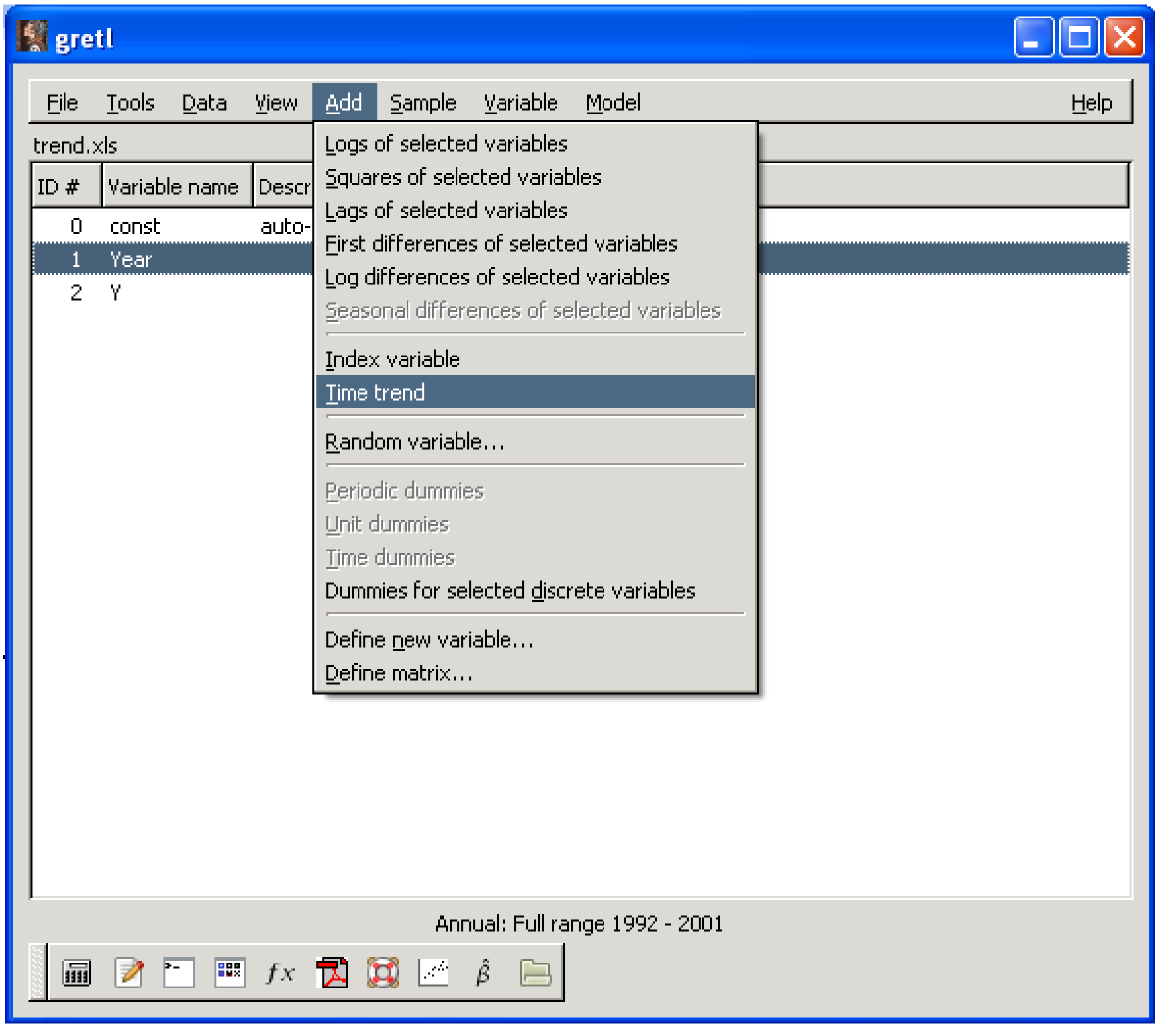
YIL	t	Y_t	$\hat{Y}_t$ =bo+bıt+b2t2	$\hat{Y}_t$	et	e2
1992	1	25	Y_1=17.2+10.0727(1)-0.72727(1)^2	26.5455	-1.54545	2.38843
1993	2	35	Y_2=17.2+10.0727(2)-0.72727(2)^2	34.4364	0.56364	0.31769
1994	3	42	Y_3=17.2+10.0727(3)-0.72727(3)^2	40.8727	1.12727	1.27074
1995	4	47	Y_4=17.2+10.0727(4)-0.72727(4)^2	45.8545	1.14545	1.31207
1996	5	50	Y_5=17.2+10.0727(5)-0.72727(5)^2	49.3818	0.61818	0.38215
1997	6	51	Y_6=17.2+10.0727(6)-0.72727(6)^2	51.4545	-0.45455	0.20661
1998	7	51	Y_7=17.2+10.0727(7)-0.72727(7)^2	52.0727	-1.07273	1.15074
1999	8	50	Y_8=17.2+10.0727(8)-0.72727(8)^2	51.2364	-1.23636	1.52860
2000	9	48	Y_9=17.2+10.0727(9)-0.72727(9)^2	48.9455	-0.94545	0.89388
2001	10	47	Y_{10}=17.2+10.0727(10)-0.72727(10)^2	45.2000	1.80000	3.24000
					HKT	12.69091

MSE=1.269091 MAPE=%2.54 MAD=1.05

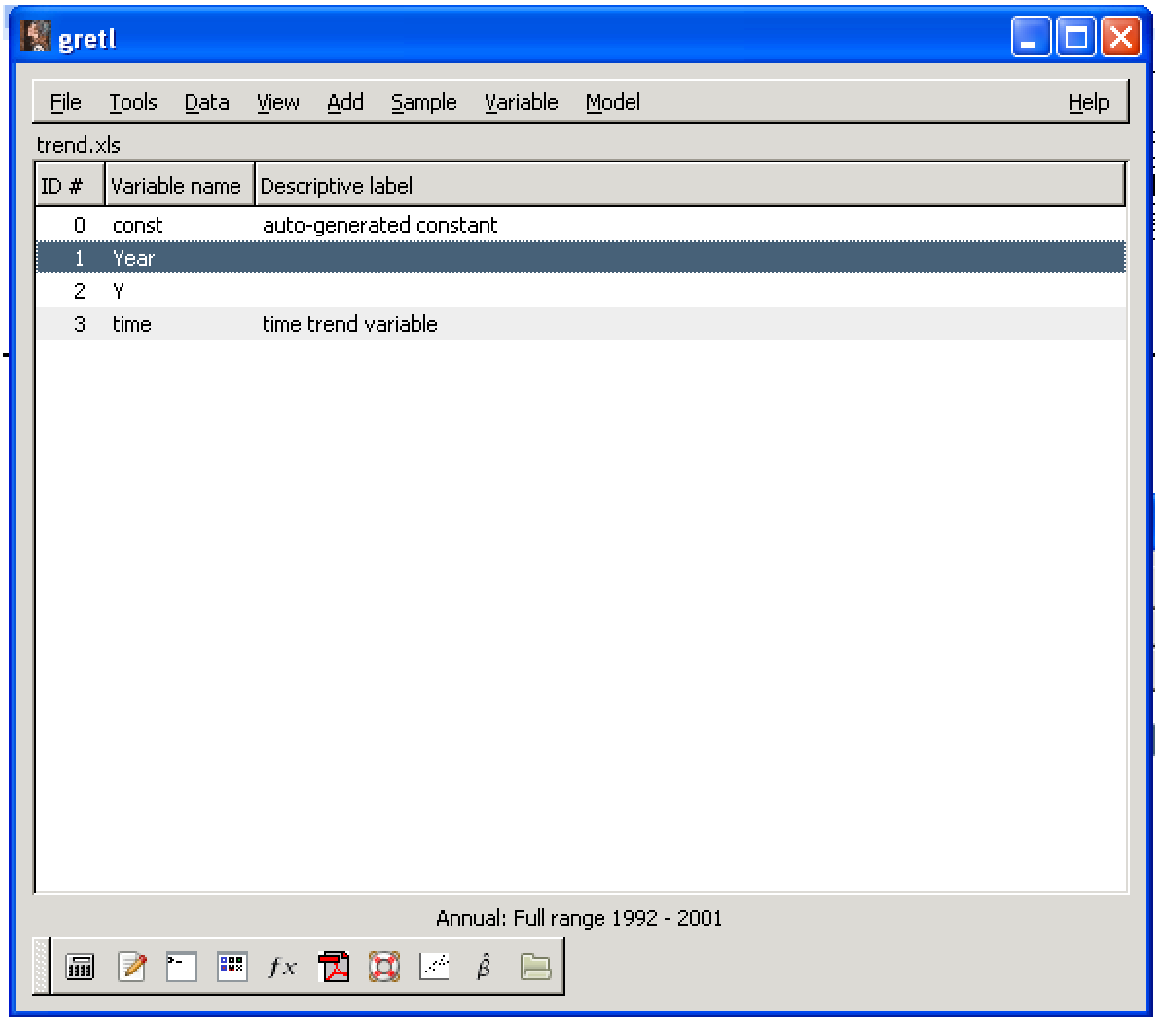
YIL	t	$\hat{Y}_t$ =bo+bıt+b2t2	$\hat{Y}_t$
2002	11	Y_{11}=17.2+10.0727(11)-0.72727(11)^2	44.00
2003	12	Y_{12}=17.2+10.0727(12)-0.72727(12)^2	33.345
2004	13	Y_{13}=17.2+10.0727(13)-0.72727(13)^2	25.236

Şimdi trend tahminlemesini GRETL yardımı ile yapalım:

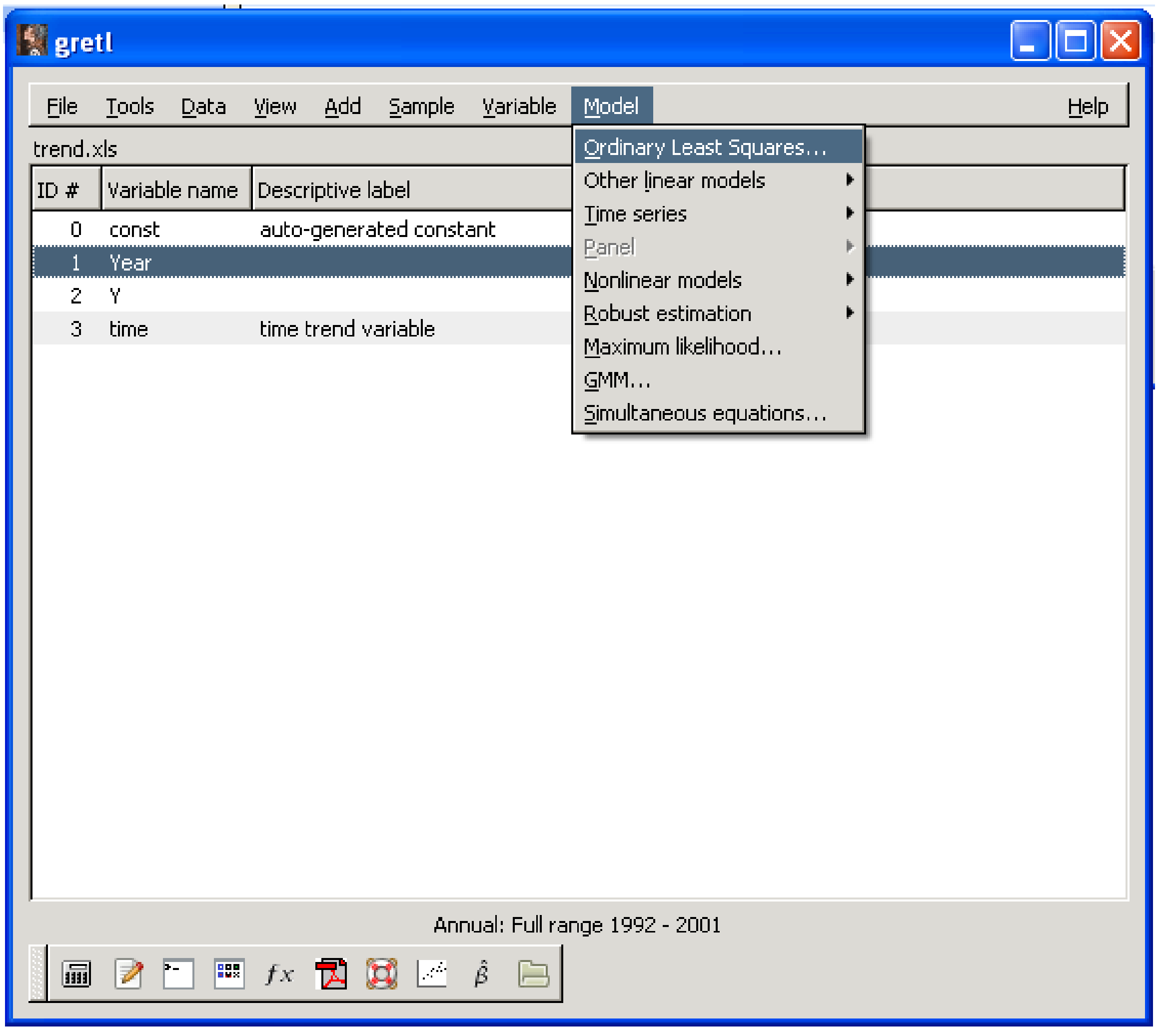
Önce trend değişkenini oluşturalım:

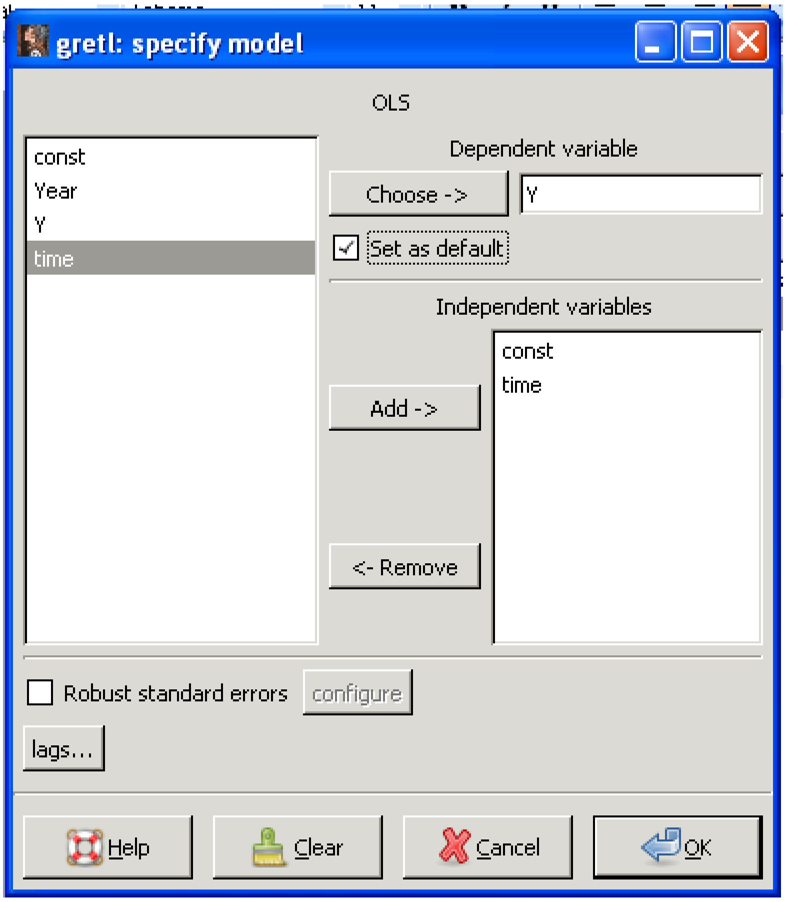


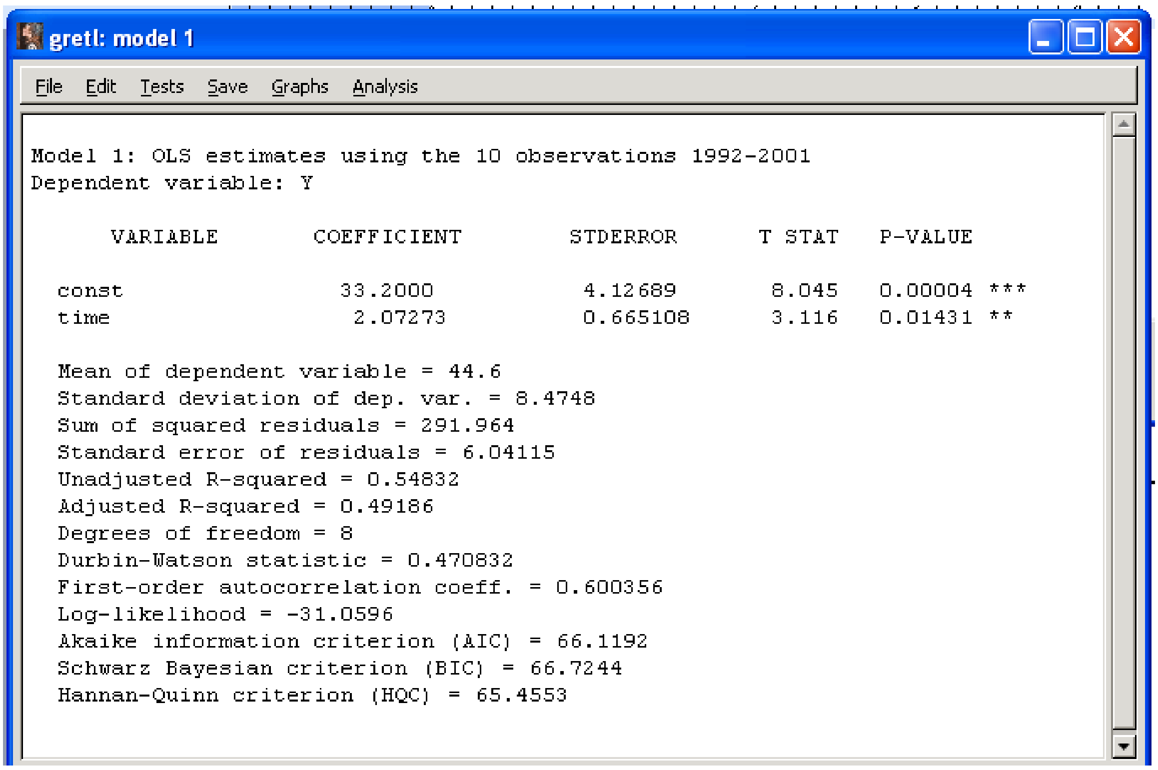
Trend değişkeni, time adıyla değişken listesine kaydedilmiş olmalıdır:



Doğrusal trend denklemi: $Y_t = b_o + b_1 t$

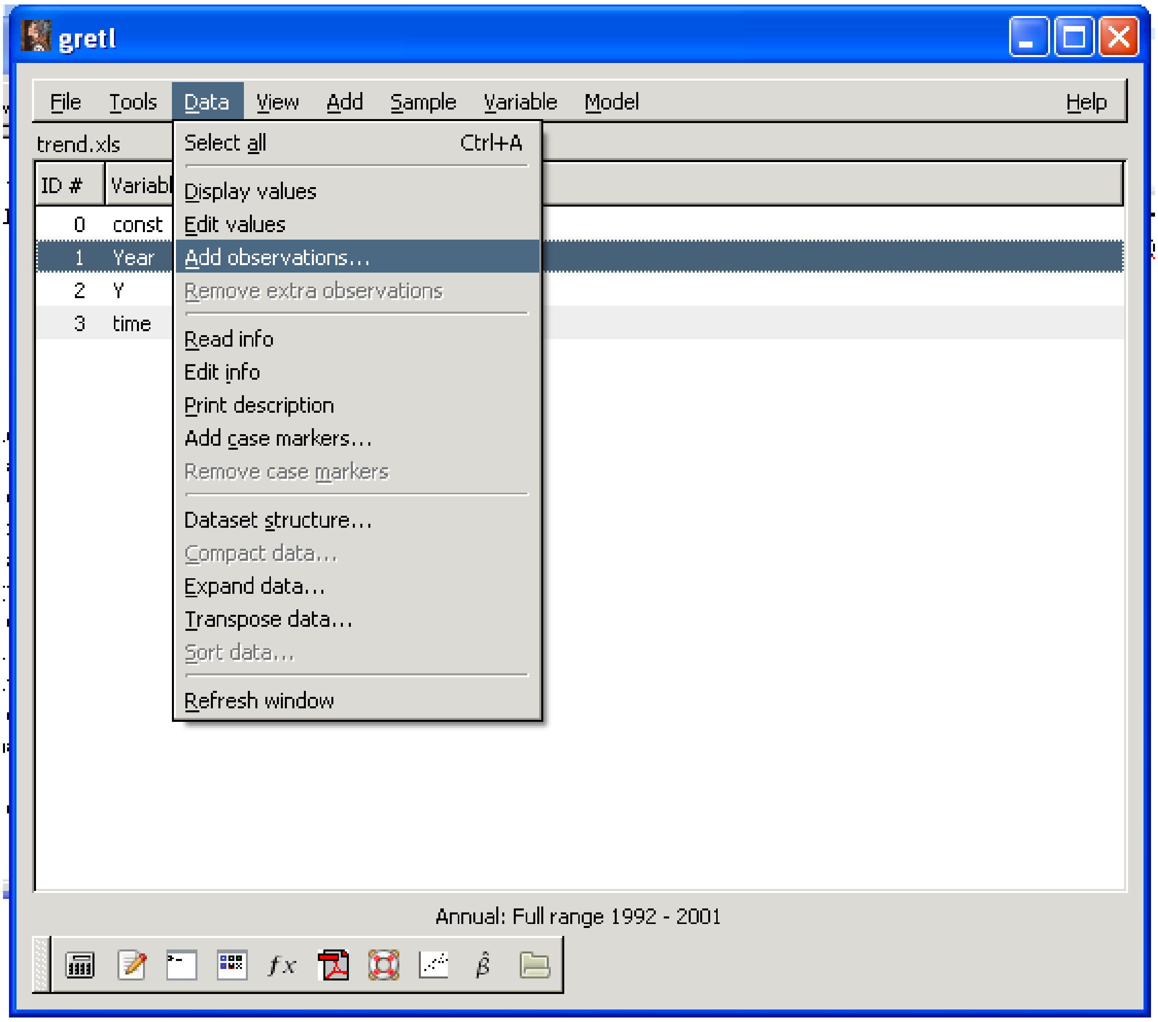

gretl
File Tools Data View Add Sample Variable Model Help
trend.xls
ID # Variable name Descriptive label
0 const auto-generated constant
1 Year
2 Y
3 time time trend variable
Ordinary Least Squares...
Other linear models
Time series
Panel
Nonlinear models
Robust estimation
Maximum likelihood...
GMM...
Simultaneous equations...
Annual: Full range 1992 - 2001



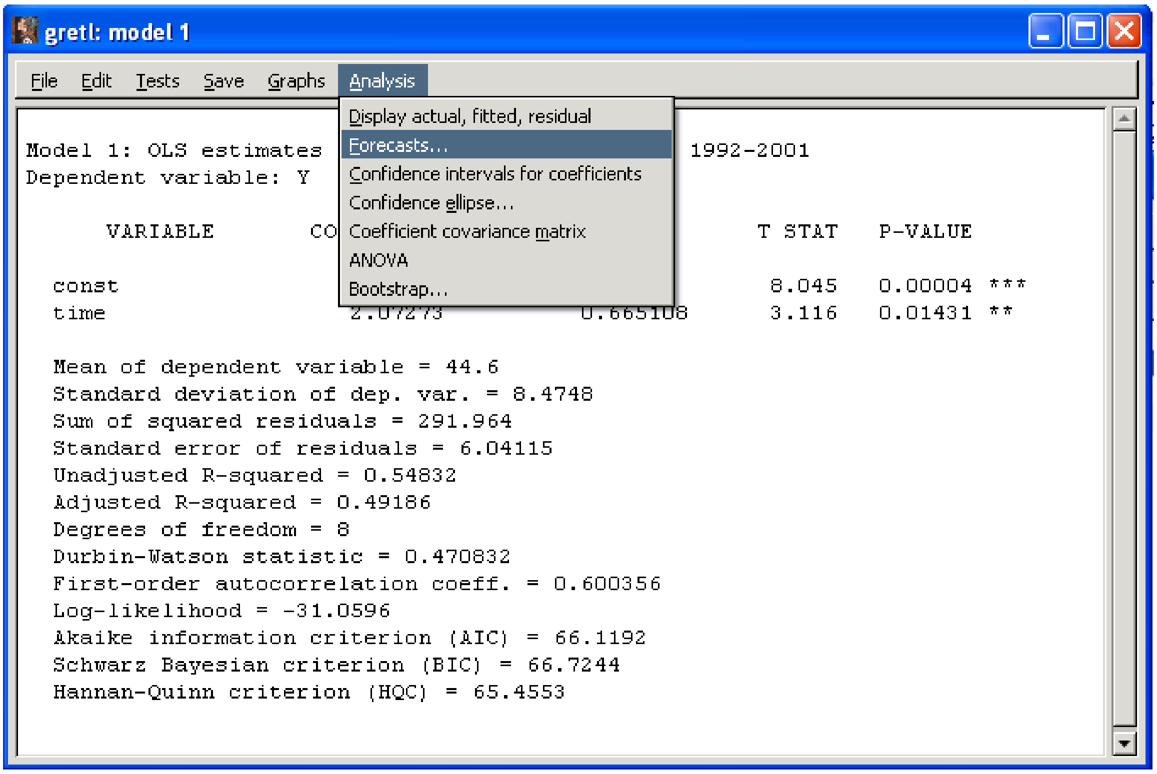


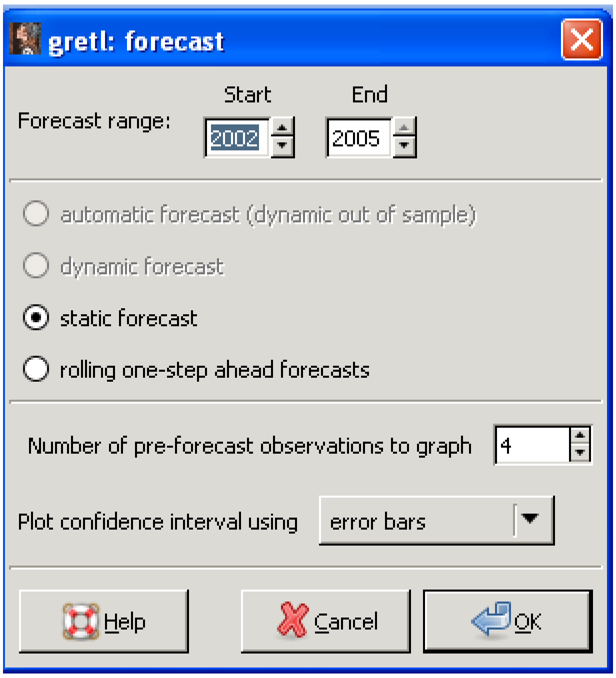

```
Model 1: OLS estimates using the 10 observations 1992-2001
Dependent variable: Y

      VARIABLE        COEFFICIENT        STDERROR      T STAT    P-VALUE

  const                 33.2000          4.12689       8.045     0.00004 ***
  time                   2.07273         0.665108      3.116     0.01431 **

  Mean of dependent variable = 44.6
  Standard deviation of dep. var. = 8.4748
  Sum of squared residuals = 291.964
  Standard error of residuals = 6.04115
  Unadjusted R-squared = 0.54832
  Adjusted R-squared = 0.49186
  Degrees of freedom = 8
  Durbin-Watson statistic = 0.470832
  First-order autocorrelation coeff. = 0.600356
  Log-likelihood = -31.0596
  Akaike information criterion (AIC) = 66.1192
  Schwarz Bayesian criterion (BIC) = 66.7244
  Hannan-Quinn criterion (HQC) = 65.4553
```

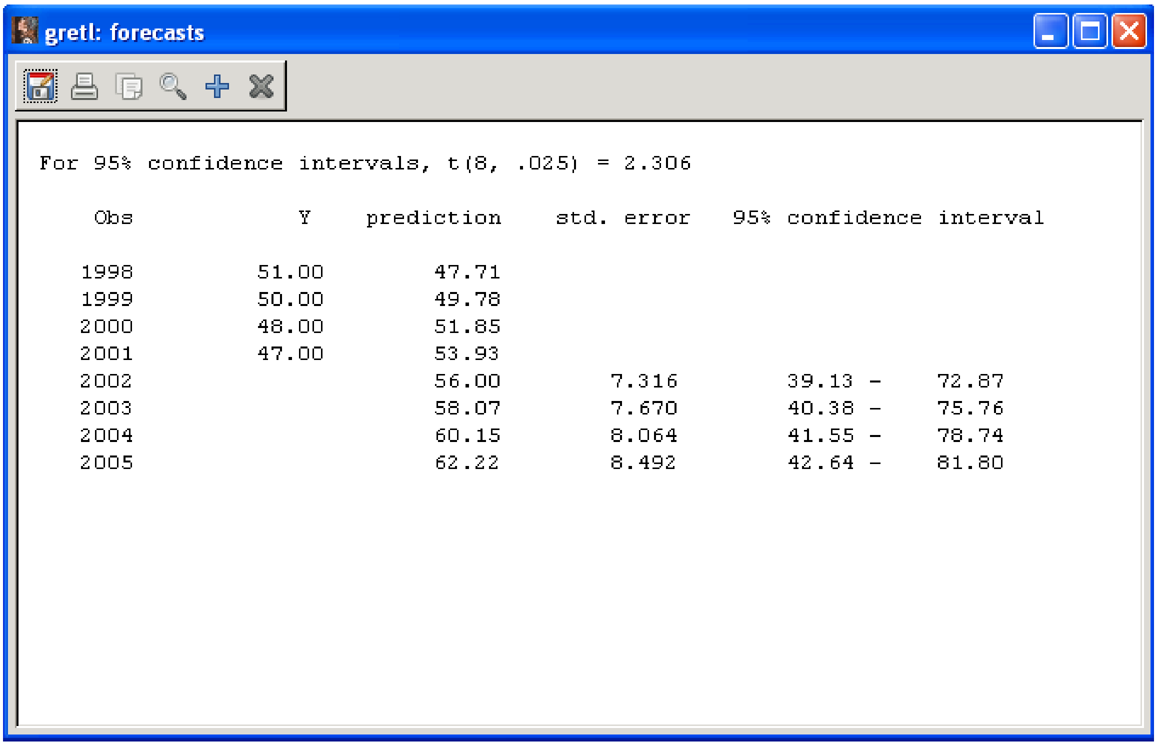
Şimdi 1992-2001 yılları arasındaki veri setimizi, 2005 yılına kadar genişletelim:


gretl
File Tools Data View Add Sample Variable Model Help
trend.xls
ID # Variabl
0 const
1 Year
2 Y
3 time
Select all Ctrl+A
Display values
Edit values
Add observations...
Remove extra observations
Read info
Edit info
Print description
Add case markers...
Remove case markers
Dataset structure...
Compact data...
Expand data...
Transpose data...
Sort data...
Refresh window
Annual: Full range 1992 - 2001


Add observations
Number of observations to add:
4
Cancel
OK


gretl: model 1
File Edit Tests Save Graphs Analysis
Display actual, fitted, residual
Forecasts...
Confidence intervals for coefficients
Confidence ellipse...
Coefficient covariance matrix
ANOVA
Bootstrap...
Model 1: OLS estimates
1992-2001
Dependent variable: Y
VARIABLE
T STAT
P-VALUE
const
8.045
0.00004 ***
time
3.116
0.01431 **
Mean of dependent variable = 44.6
Standard deviation of dep. var. = 8.4748
Sum of squared residuals = 291.964
Standard error of residuals = 6.04115
Unadjusted R-squared = 0.54832
Adjusted R-squared = 0.49186
Degrees of freedom = 8
Durbin-Watson statistic = 0.470832
First-order autocorrelation coeff. = 0.600356
Log-likelihood = -31.0596
Akaike information criterion (AIC) = 66.1192
Schwarz Bayesian criterion (BIC) = 66.7244
Hannan-Quinn criterion (HQC) = 65.4553


gretl: forecast
Start
End
Forecast range:
2002
2005
automatic forecast (dynamic out of sample)
dynamic forecast
static forecast
rolling one-step ahead forecasts
Number of pre-forecast observations to graph
4
Plot confidence interval using
error bars
Help
Cancel
OK

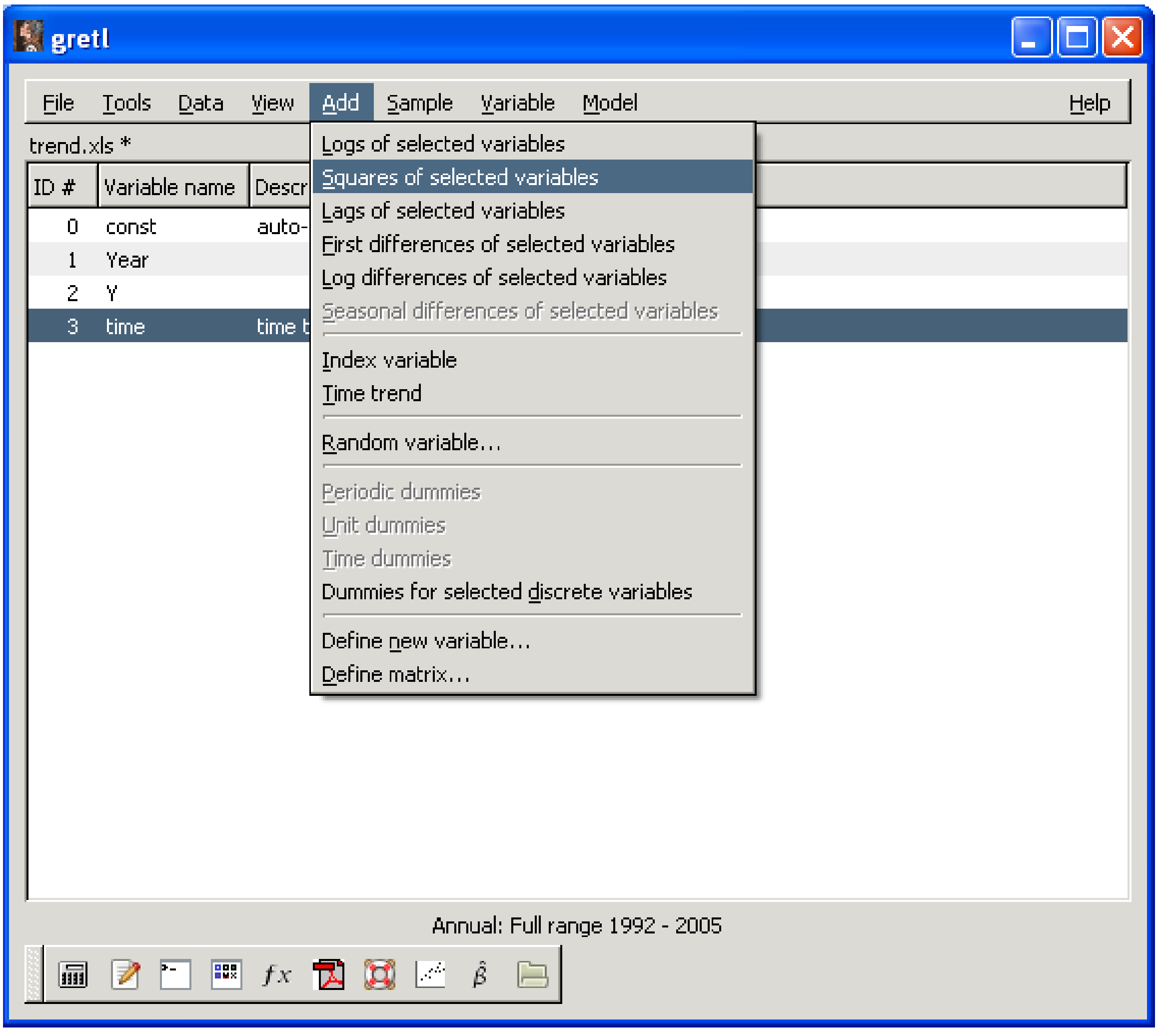


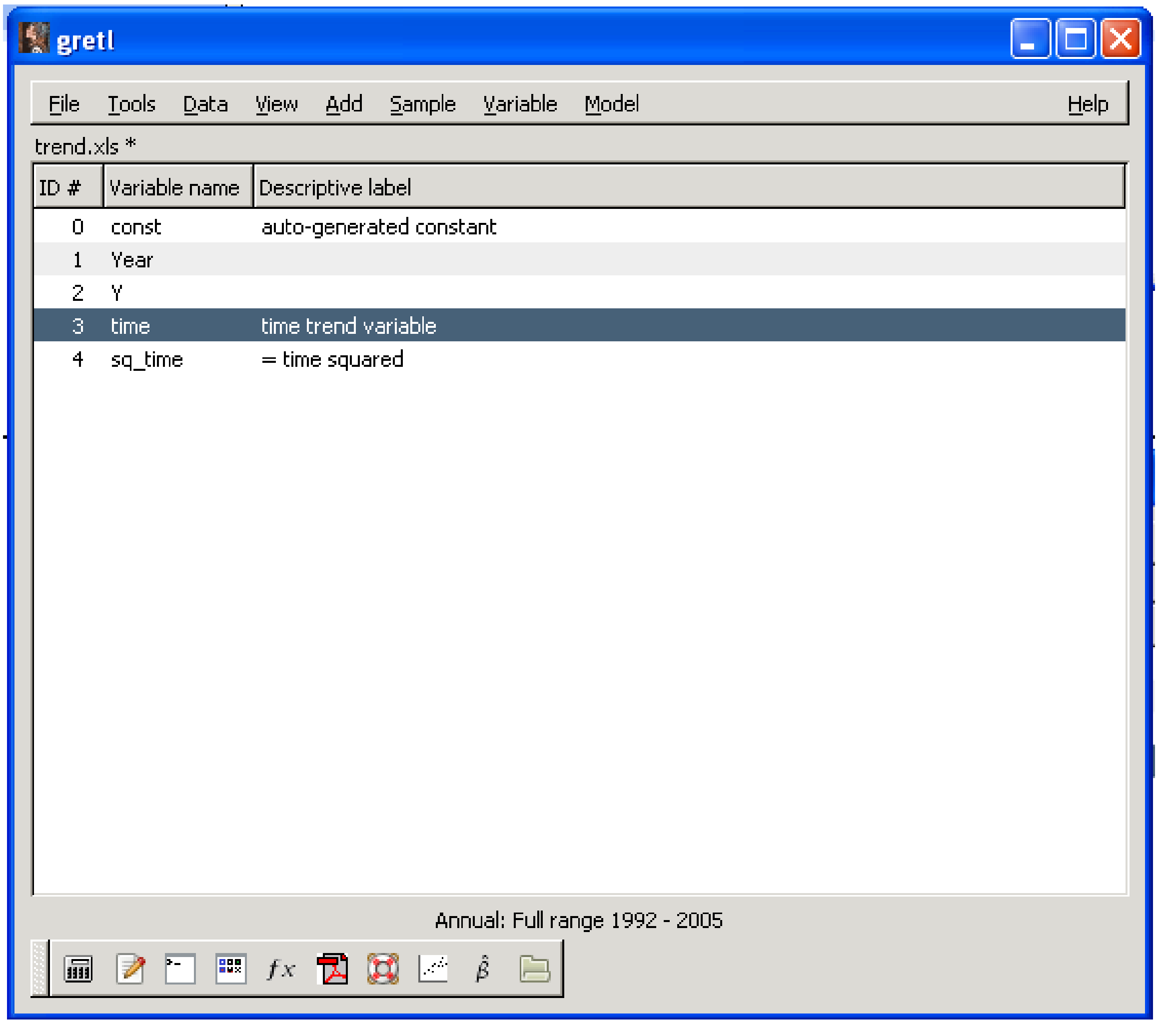
For 95% confidence intervals, t(8, .025) = 2.306

Obs	Y	prediction	std. error	95% confidence interval
1998	51.00	47.71		
1999	50.00	49.78		
2000	48.00	51.85		
2001	47.00	53.93		
2002		56.00	7.316	39.13 - 72.87
2003		58.07	7.670	40.38 - 75.76
2004		60.15	8.064	41.55 - 78.74
2005		62.22	8.492	42.64 - 81.80

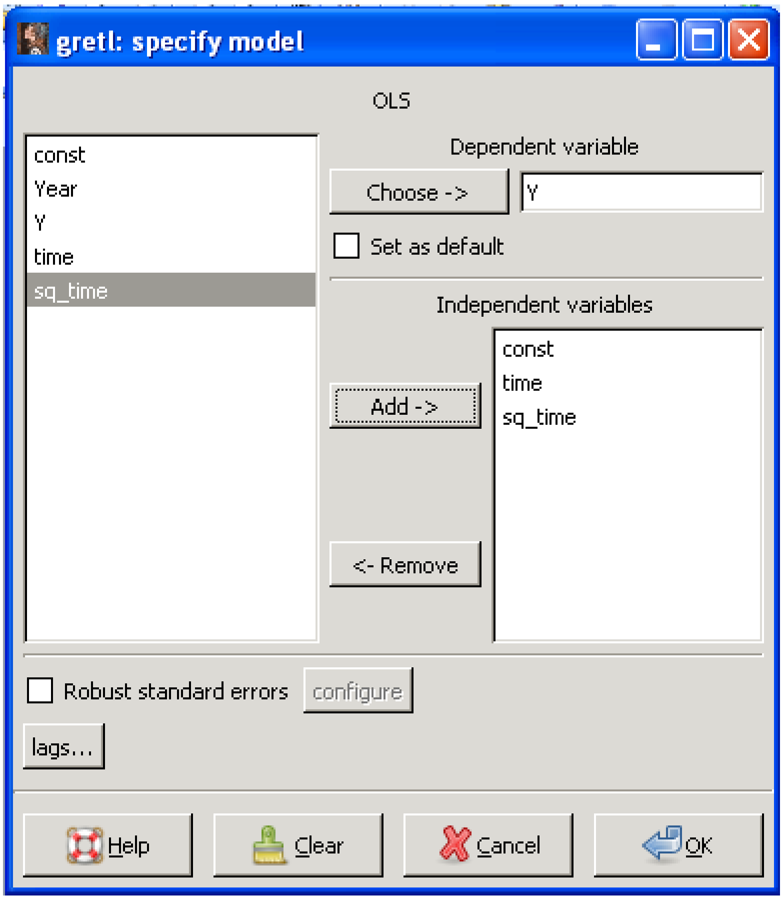
Kuadratik trend denklemi: $Y_t=b_0+b_1t+b_2t^2$

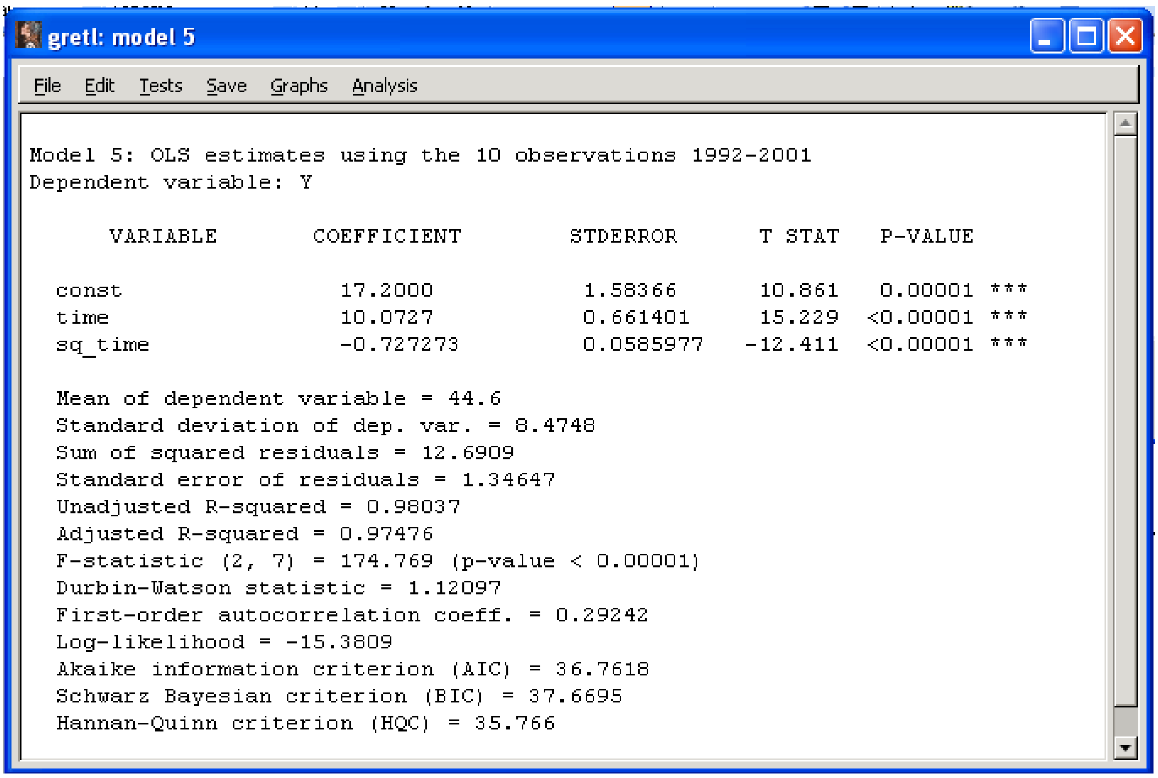
Önce trend değişkeninin karesini alalım:


gretl
File Tools Data View Add Sample Variable Model Help
trend.xls *
ID # Variable name Descr
0 const auto-
1 Year
2 Y
3 time time t
Logs of selected variables
Squares of selected variables
Lags of selected variables
First differences of selected variables
Log differences of selected variables
Seasonal differences of selected variables
Index variable
Time trend
Random variable...
Periodic dummies
Unit dummies
Time dummies
Dummies for selected discrete variables
Define new variable...
Define matrix...
Annual: Full range 1992 - 2005

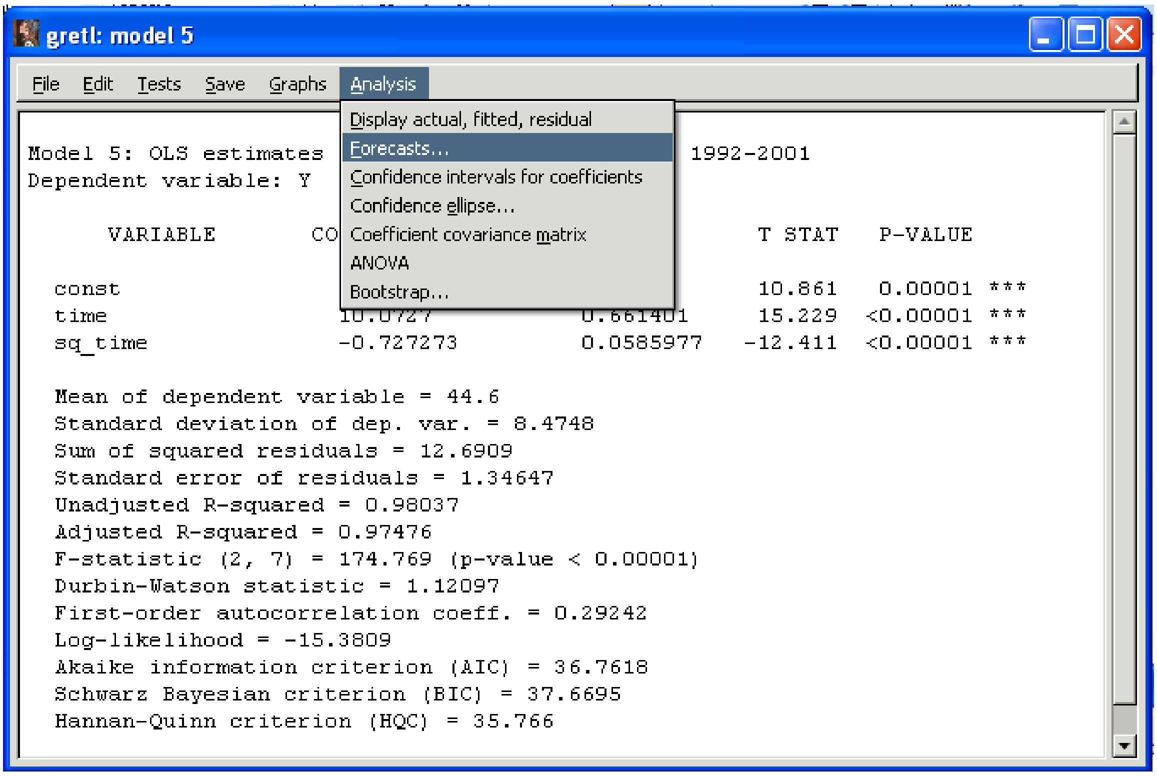


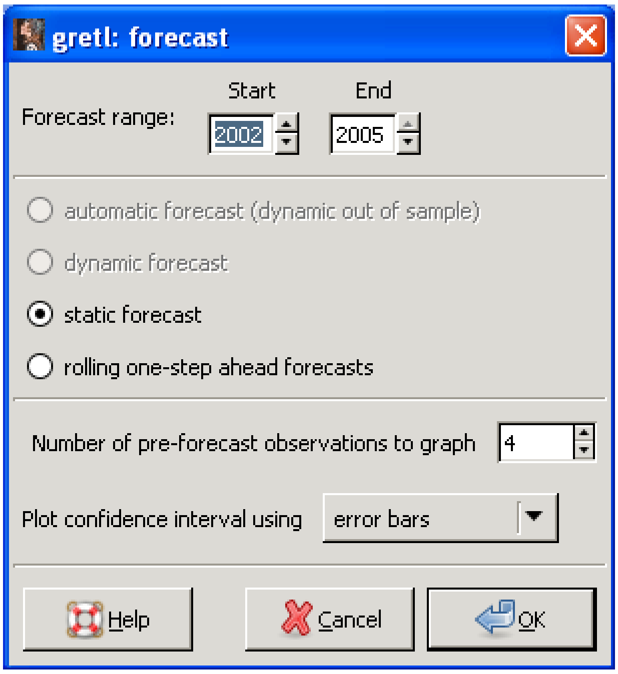
Kuadratik trend denklemini tahmin edelim:

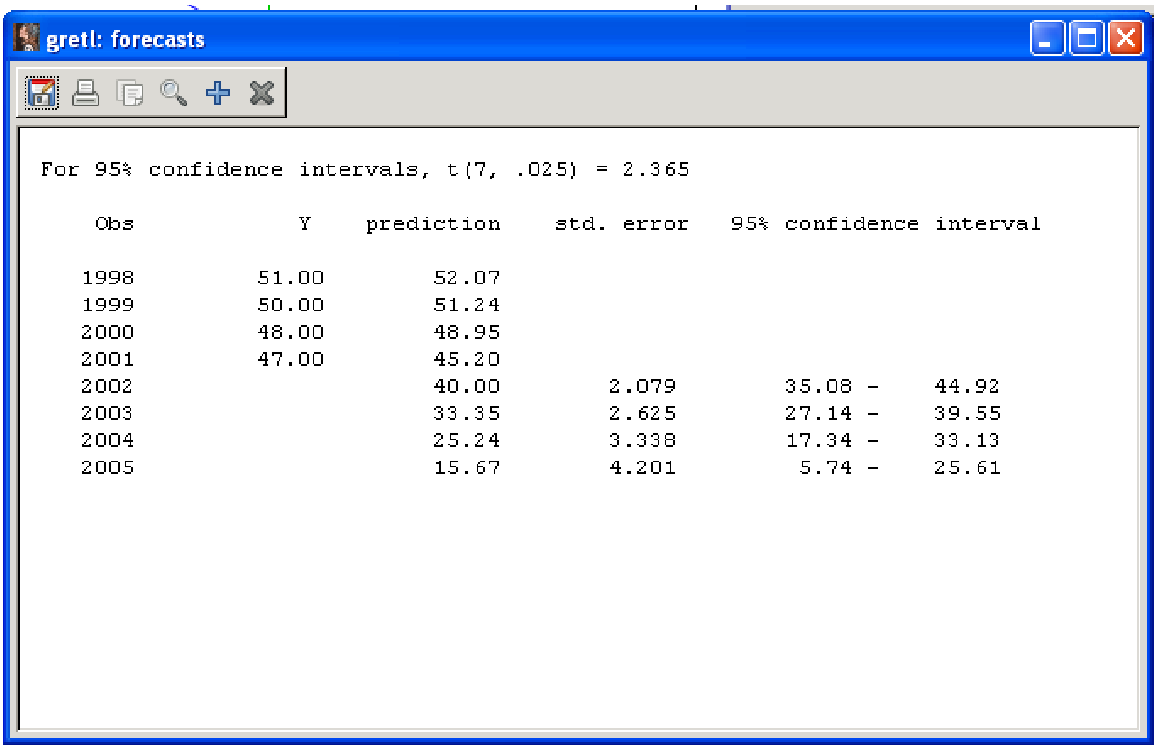





```
Model 5: OLS estimates using the 10 observations 1992-2001
Dependent variable: Y

     VARIABLE       COEFFICIENT        STDERROR      T STAT    P-VALUE

  const               17.2000          1.58366       10.861    0.00001 ***
  time                10.0727          0.661401      15.229   <0.00001 ***
  sq_time             -0.727273        0.0585977    -12.411   <0.00001 ***

  Mean of dependent variable = 44.6
  Standard deviation of dep. var. = 8.4748
  Sum of squared residuals = 12.6909
  Standard error of residuals = 1.34647
  Unadjusted R-squared = 0.98037
  Adjusted R-squared = 0.97476
  F-statistic (2, 7) = 174.769 (p-value < 0.00001)
  Durbin-Watson statistic = 1.12097
  First-order autocorrelation coeff. = 0.29242
  Log-likelihood = -15.3809
  Akaike information criterion (AIC) = 36.7618
  Schwarz Bayesian criterion (BIC) = 37.6695
  Hannan-Quinn criterion (HQC) = 35.766
```


gretl: model 5
File Edit Tests Save Graphs Analysis
Display actual, fitted, residual
Forecasts...
Confidence intervals for coefficients
Confidence ellipse...
Coefficient covariance matrix
ANOVA
Bootstrap...
Model 5: OLS estimates
1992-2001
Dependent variable: Y
VARIABLE
CO
T STAT
P-VALUE
const 10.861 0.00001 ***
time 10.0727 0.661401 15.229 <0.00001 ***
sq_time -0.727273 0.0585977 -12.411 <0.00001 ***
Mean of dependent variable = 44.6
Standard deviation of dep. var. = 8.4748
Sum of squared residuals = 12.6909
Standard error of residuals = 1.34647
Unadjusted R-squared = 0.98037
Adjusted R-squared = 0.97476
F-statistic (2, 7) = 174.769 (p-value < 0.00001)
Durbin-Watson statistic = 1.12097
First-order autocorrelation coeff. = 0.29242
Log-likelihood = -15.3809
Akaike information criterion (AIC) = 36.7618
Schwarz Bayesian criterion (BIC) = 37.6695
Hannan-Quinn criterion (HQC) = 35.766


gretl: forecast
Forecast range:
Start
End
2002
2005
automatic forecast (dynamic out of sample)
dynamic forecast
static forecast
rolling one-step ahead forecasts
Number of pre-forecast observations to graph
4
Plot confidence interval using
error bars
Help
Cancel
OK



For 95% confidence intervals, t(7, .025) = 2.365

Obs	Y	prediction	std. error	95% confidence interval
1998	51.00	52.07		
1999	50.00	51.24		
2000	48.00	48.95		
2001	47.00	45.20		
2002		40.00	2.079	35.08 - 44.92
2003		33.35	2.625	27.14 - 39.55
2004		25.24	3.338	17.34 - 33.13
2005		15.67	4.201	5.74 - 25.61

11.2. Trend Değişkenli Modeller

X ve Y gibi yıllık gözlemleri olan değişkenlerimiz olsun. Trend değişkeni içeren bir regresyon denklemi:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 t + e_t$$

Şeklinde olabilir. β_2 katsayısı, diğer değişkenler sabitken Y'deki yıllık değişimi ölçer. Denklemimizi:

$$\ln(Y_t) = \alpha_0 + \alpha_1 \ln(X_t) + \alpha_2 t + u_t$$

olarak ta ifade edebiliriz. Bu durumda, trend değişkenine ait katsayı($\alpha 2$), büyüme oranı olarak yorumlanır. Örneğimizde, X'in belli bir değeri için:

$\alpha_2 > 0$ ise $100(\alpha_2)$ Y'deki yüzde büyüme oranıdır.

$\alpha_2 < 0$ ise $100(\alpha_2)$ Y'deki yüzde küçülme oranıdır

Hesaplanan bu değerler anlık büyüme oranlarıdır. Y'deki yıllık bileşik büyüme oranı, diğer değişkenler sabitken:

$$g = 100\,[\text{antilog}_e\,(\alpha_2) - 1]$$

formülüyle hesaplanır.

Ya da:

Eğer sorumuz "trend değişkenindeki bir birim (yıl) değişmenin, bağımlı değişkende meydana getireceği oransal değişme nedir?" ise, bir başka açıdan yaklaşmak mümkündür.

Modelimiz:

$$\ln Y = b_0 + b_1 T + b_2 X$$

olsun. T, trend değişkeni; X ise ölçülebilir bir değişkendir. Bu denklemden yararlanarak, trend değişkenindeki 1 birim (yıl) değişmenin Y'de meydana getirdiği oransal değişimi hesaplayalım:

OD = (Bağımlı değişkendeki oransal değişme) / (Trendde bir yıllık değişme)

OD = (ΔY/Y) / dT

Gerekli düzenlemeleri yaptıktan sonra:

OD = (dY/Y) / dT = (∂y/∂T).(1/Y)

olur. ∂Y/∂T, kısmi türevdir. Logaritmik fonksiyonlarda türev alma kurallarını uyguladığımızda:

$$OD = \frac{\partial y}{\partial T}.\frac{1}{y} = b_1 y \frac{1}{y} = b_1$$

olur.

Örnek 11-1

Bu örnekte kırmızı et talebini ele alacağız. Bağımsız değişkenlerimiz: kırmızı et fiyatı, tavuk eti fiyatı ve gelir olarak belirlenmiştir (**EtTuk.xls**).

ETTUK : Kişi başına kırmızı et tüketimi

ETFIY : Reel Taze ve dondurulmuş et fiyatı – tüketici fiyat indeksi (1986=100)

TAVFIY : Reel Taze ve dondurulmuş tavuk eti fiyatı – tüketici fiyat indeksi (1986=100)

GELIR : Reel Kişi başına harcanabilir gelir

TUFE : Tüketici fiyat indeksi (1986=100)

Yıl	*ETTUK*	*ETFIY*	*TAVFIY*	*GELIR*	*TUFE*
1975	79.2	42	51.3	4883	44
1976	84.2	39.3	51.3	5453	45.2
1977	80.1	41.8	51.6	5941	48.9
1978	75	61.1	59.9	6634	56.5
1979	63.7	80.3	66	7408	63.9
1980	63.1	87.2	71	8281	70.8
1981	64.8	89.5	84.5	9545	78.9
1982	64.2	88.9	86.8	10430	84.6
1983	63.9	89.6	89.9	10843	87.7
1984	60.9	95.5	95.8	11686	92.6
1985	61.7	97.9	91.8	12387	95.2
1986	61.2	100	100	12902	100
1987	58.2	109.1	106.1	13613	104.4
1988	58.2	110.8	107.6	14658	107.2
1989	56.6	113.2	120	15783	111.1
1990	54.5	117.6	126.1	16263	115.7
1991	53.4	118.3	123.5	16570	121.2
1992	51.7	116.8	123	16753	120.8
1993	49.6	123.1	125.6	16874	122.8
1994	50.6	124.1	119.3	17003	123.3

Yukarıdaki komutlarda öncelikle fiyatlar ve gelir reelleştirilmiştir. Bunun için bu değişkenler TUFE indeksine bölünmüştür. Ekonomik teoriye göre kırmızı et fiyatına ait katsayı negatif olmalıdır. Gelir katsayısı ise normal mal için pozitif, adi mal için negatif olmalıdır. Tavuk fiyatı katsayısının pozitif olması ikame, negatif olması tamamlayıcı, sıfır olması ise ilişkisiz olduğu anlamına gelir. Regresyon analizi sonuçları:

Bağımlı değişken: ETTUK

	Katsayı	*St.Hata*	*t*	*p*	
sabit	97.0947	19.3892	5.0077	0.00016	***
REtFiy	-31.448	4.7108	-6.6757	<0.00001	***
RTavFiy	-0.109865	9.95679	-0.0110	0.99134	
RGelir	0.13361	0.131758	1.0140	0.32664	
time	-1.8059	0.217735	-8.2940	<0.00001	***

Bağımlı değişkenin ortalaması = 62.74
Bağımlı değişkenin st sapması = 9.93327
Hata kareleri toplamı = 39.6386
Hataların st hatası= 1.6256
R^2 = 0.97886
Düzeltilmiş R^2 = 0.97322
F-istatistiği (4, 15) = 173.608 (p < 0.00001)
Durbin-Watson istatistiği = 1.75767
First-order autocorrelation coeff. = 0.00437018
Log-likelihood = -35.2195
Akaike bilgi kriteri = 80.439
Schwarz Bayesian kriteri = 85.4176
Hannan-Quinn kriteri = 81.4108

1975/94 döneminde trend değişkeni negatif ve anlamlıdır. Yani, Herşey sabitken k.et tüketimi her yıl ortalama 1.8059 birim düşmektedir.

Log-log denklemi:

Model 2: EKK tahminleri 20 gözlem 1975-1994
Bağımlı değişken: l_ETTUK

	Katsayı	*St.Hata*	*t*	*p*	
sabit	2.55765	1.0463	2.4445	0.02734	**
l_REtFiy	-0.410051	0.0660422	-6.2089	0.00002	***
l_RTavFiy	-0.0833765	0.137436	-0.6067	0.55315	
l_RGelir	0.390741	0.22159	1.7634	0.09819	*
time	-0.0296351	0.00285677	-10.3736	<0.00001	***

Bağımlı değişkenin ortalaması = 4.12772
Bağımlı değişkenin st sapması = 0.152246
Hata kareleri toplamı = 0.00701967

Hataların st hatası= 0.0216328
R^2 = 0.98406
Düzeltilmiş R^2 = 0.97981
F-istatistiği (4, 15) = 231.516 (p < 0.00001)
Durbin-Watson istatistiği = 2.36191
First-order autocorrelation coeff. = -0.314861
Log-likelihood = 51.1689
Akaike bilgi kriteri = -92.3379
Schwarz Bayesian kriteri = -87.3592
Hannan-Quinn kriteri = -91.366

log-log denklemi tahmin ettiğimizde, trend değişkenine ait katsayıyı -0.0296 buluruz. Buna göre, diğer tüm değişkenler sabit tutulduğunda, kişi başına ortalama yıllık azalış:

$g=[\text{antilog}_e(-0.0296)-1]*100=-2.96\ \%$

olarak hesaplanır.

Diğer yaklaşımla, b_1 katsayısının bize doğrudan yıllık değişim oranını %-2.96 olarak verdiğini görebiliriz.

11.3. Otoregresif (AR) Modeller

Bir değişkenin geleceğe dönük kestirimi, başka değşkenlere baş vurmaksızın sadece değişkenin kendisinden yararlanılarak yapılabilir. Bu kestirimler belli bir teorik modele dayanmaz. Değişkenin geçmişte gösterdiği hareket, gelecekte göstereceği hareketi kestirmek için kullanılır. Yüksek frekanslı veriler (aylık, günlük gibi) uygun tahmin yöntemini değiştirebilecek karmaşık zaman serisi özellikleri gösterebilir.

Otoregresif model (AR), önemli zaman serisi yaklaşımlarından biridir. AR modeli, basit olarak bir zaman serisinin halihazır değerinin, bu serinin bir veya daha önceki değerlerine karşı doğrusal regresyonudur. Bir başka ifadeyle y_t gibi bir zaman serisi değişkeninin belli bir oranının gelecek döneme, y_{t+1}'in belli bir oranının daha sonraki döneme aktarıldığı bir süreçtir. Her dönemin bir sonraki döneme aktarılması nedeniyle, otoregresyon uzun solukludur. AR modelleri aşağıdaki gibi gösterilir:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + \ldots + \phi_1 Y_{t-p} + \varepsilon_t$$

Burada Y_t zaman serisi, ε_t hata terimi (beyaz gürültü=white noise), Φ_1 , ..., Φ_p modelin tahmincileri, δ ise aşağıda gösterildiği şekilde zaman serisinin ortalamasıdır:

$$\delta = \sum_{i=1}^{p}(1-\phi_i)$$

p, AR modelinin derecesidir. AR modelleri, bu amaçla geliştirilmiş yöntemlerin yanısıra, EKK ile de tahmin edilebilir.

Hareketli ortalama (MA) sürecinde, hata terimi gelecek dönemlere taşınır. MA modelleri:

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

şeklinde gösterilir. Burada Y_t zaman serisi, μ zaman serisinin ortalaması, θ_1, ..., θ_q modelin tahmincileridir. q, MA modelinin derecesidir. MA modeli, zaman serisinin halihazır değerinin, bu serinin bir veya daha önceki değerlerinin tesadüfi şoklarına karşı doğrusal regresyonudur. Tesadüfi şokların, tipik olarak normal dağılış gösterdiği varsayılır. Bu modelin altında yatan düşünce, tesadüfi şokların zaman serisine ait gelecekteki değerleri yönlendirdiğidir.

MA tahminlemesi, AR tahminlemesinden daha zordur. Zira hata terimleri uydurulan modele bağlıdır. Bu, MA modellerinin EKK yerine tekrarlamalı doğrusal olmayan yöntemlerle tahmin edilebileceği anlamına gelmektedir. MA modellerinin yorumu biraz daha karmaşıktır.

Bazan tek başına AR veya MA, bazan da her ikisi birden aynı modelde bulunabilir.

11.4. Durağanlık

Çoğu zaman serisi yöntemi, verilerin durağan olmasını gerektirir. Durağanlık; ortalama, varyans ve otokorelasyon zaman içinde değişme göstermediği, yani sabit kaldığı anlamına gelir. Bir başka ifadeyle durağan bir zaman serisi; ortalaması, varyansı ve otokorelasyonu zaman içinde değişmeyen ve trendsiz düz görüntülü bir seridir. Durağan olmayan bir seri:

$$Y_t = \theta Y_{t-1} + \varepsilon_t$$

modelinde $\theta = 1$ durumunu (birim kök) gösterir. Eğer zaman serisi durağan değilse, seriyi durağanlaştırmak için yaygın olarak fark alma işlemine başvurulur:

1. *Dereceden fark alma:*

   Z_t zaman serisinin 1. dereceden farkı için yeni bir seri tanımlanır:

   $\Delta Z_t = Z_t - Z_{t-1}$

2. *Dereceden fark alma:*

   Yine Z_t zaman serisinin 2. dereceden farkı için yeni bir seri tanımlanır:

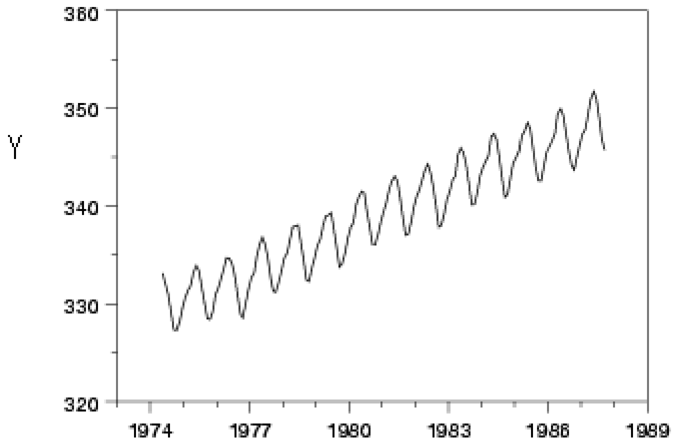
   $\Delta_2 Z_t = Z_t - Z_{t-2}$

Farkı alınmış veriler, başlangıçtakinden daha az sayıda olacaktır. Birden daha çok fark alınabilir ancak ekonomik serilerde genellikle bir fark yeterlidir.

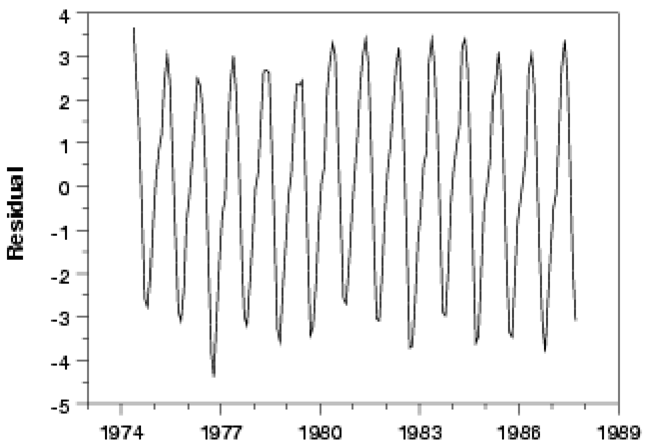
Eğer zaman serisi değişkeni trend gösteriyorsa, verilere basit bir doğru veya 2. dereceden bir eğri eğri uydurulabilir. Bu modelden elde edilen hataları kullanarak modelleme yapılır. Amaç uzun dönem trendini uzaklaştırmak olduğundan, basit bir doğru yeterli olabilir.

Varyans sabit değilse, serinin karekökü veya logaritması alınır. Böylece varyans kararlı bir hale gelir. Negatif veri problemi varsa, tüm verileri pozitif yapacak makul bir sabit sayı verilere eklenir. Daha sonra eklenen değer, modele göre tahmin edilen değerlerden çıkarılır.

Mevsimsellik durağanlığı bozmakla birlikte, bu genellikle zaman serisi modeline kolayca dahil edilebilir.



Grafikten, verilerin artan bir trend gösterdiği anlaşılmaktadır. Basit doğrusal bir tahmin, bu yukarı doğru trendi yok edebilecektir. Ayrıca grafik dönemsel bir davranışı da sergilemektedir.



Yukarıdaki grafik, orijinal verilerle tahmin edilmiş doğrusal modelin hatalarına aittir. Hataların gösterdiği desen sistematik bir dalgalanma gösterse de, doğrusal trend giderildikten sonra verilerin ortalaması ve varyansı sabitlenmiş görünmektedir.

11.4.1.Birim Kök Durağanlık Testi

Bu test, herhangi bir zaman serisinin durağanlık gösterip göstermediğini belirler. Modelimiz:

$$Y_t = \rho\, Y_{t-1} + u_t$$

olsun. Burada u_t, ortalaması sıfır, varyansı sabit, otokorelasyon göstermeyen hata terimidir. Bu özelliğe sahip hata terimine "beyaz gürültü" (white noise) adı verilir. Eğer Y_{t-1}'nin tahmincisi olan ρ 1'e eşitse birim kök vardır denir. Bunun anlamı, Y_t zaman serisinin durağanlık göstermediğidir. Yukarıdaki denklemi fark denklemi olarak yazarsak yeni durum:

$$\Delta Y_t = (\rho-1)\, Y_{t-1} + u_t$$

$$\Delta Y_t = \delta Y_{t-1} + u_t$$

$$\delta = (\rho-1)$$

Yeni duruma göre zaman serisinde durağanlık söz konusu değilse δ'nın sıfır olması gerekir.

Dickey Fuller veya Philips Peron birim kök testlerinde aşağıdaki regresyonlar uygulanır:

1) $\Delta Y_t = \delta Y_{t-1} + u_t$

2) $\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$

3) $\Delta Y_t = \beta_0 + \beta_1 t + \delta Y_{t-1} + u_t$

Denklemlerdeki t, trend değişkenidir.

Dickey Fuller veya Philips Peron birim kök testlerinde hipotezlerimiz:

H_0: Seri durağan değil [Birim kök var ($\delta = 0$ veya $\rho = 1$)]

H_1: Seri durağan (Birim kök yok)

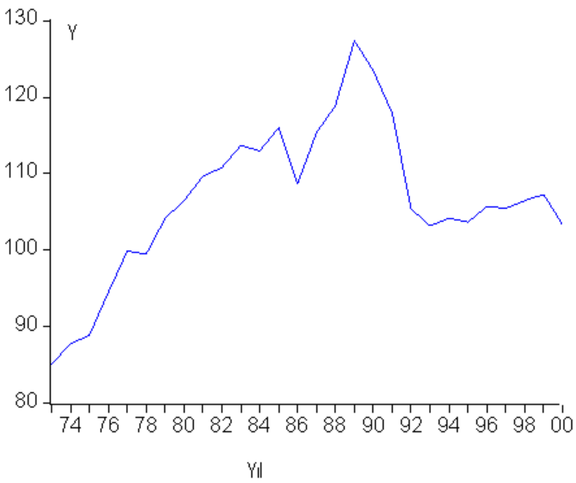
Hesaplanan DF τ (tau) değerinin mutlak değeri, tablo değerinin mutlak değerinden büyük olması durumunda H_0 hipotezi reddedilir. Yani seri durağveır denir.

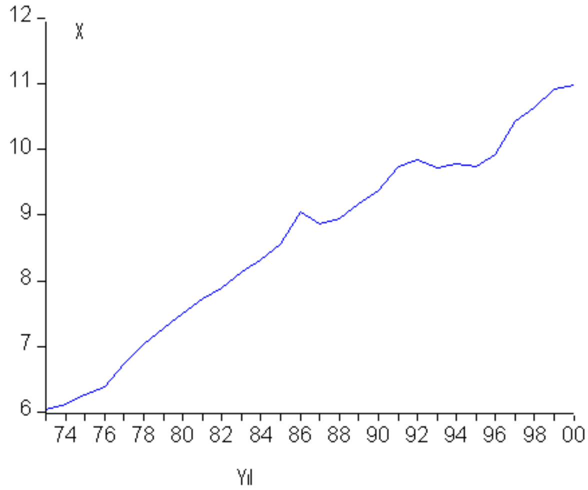
Örnek 11-2

Aşağıda verilen Y ve X gibi iki zaman serisi için durağanlık testi yapalım. Y, Y ülkesinin tüketim harcamaları, X, X ülkesinin borsa kapanış değerleridir (**BirimKok.gdt**).

Yıl	Y	X	ΔY	ΔX	Yıl	Y	X	ΔY	ΔX
1973	85.1	6.04	-	-	**1987**	115.4	8.87	-	-
1974	87.8	6.11	2.7	0.07	**1988**	118.9	8.94	3.5	0.07
1975	88.9	6.27	1.1	0.16	**1989**	127.4	9.18	8.5	0.24
1976	94.5	6.38	5.6	0.11	**1990**	123.5	9.38	-3.9	0.2
1977	99.9	6.73	5.4	0.35	**1991**	117.9	9.74	-5.6	0.36
1978	99.5	7.03	-0.4	0.3	**1992**	105.4	9.83	-12.5	0.09
1979	104.2	7.28	4.7	0.25	**1993**	103.2	9.72	-2.2	-0.11
1980	106.5	7.51	2.3	0.23	**1994**	104.2	9.77	1	0.05
1981	109.7	7.73	3.2	0.22	**1995**	103.7	9.73	-0.5	-0.04
1982	110.8	7.89	1.1	0.16	**1996**	105.7	9.93	2	0.2
1983	113.7	8.13	2.9	0.24	**1997**	105.5	10.42	-0.2	0.49
1984	113	8.32	-0.7	0.19	**1998**	106.5	10.63	1	0.21
1985	116	8.56	3	0.24	**1999**	107.3	10.91	0.8	0.28
1986	108.7	9.04	-7.3	0.48	**2000**	103.3	10.97	-4	0.06

Önce Y ve X'in grafiklerini çizeceğiz:



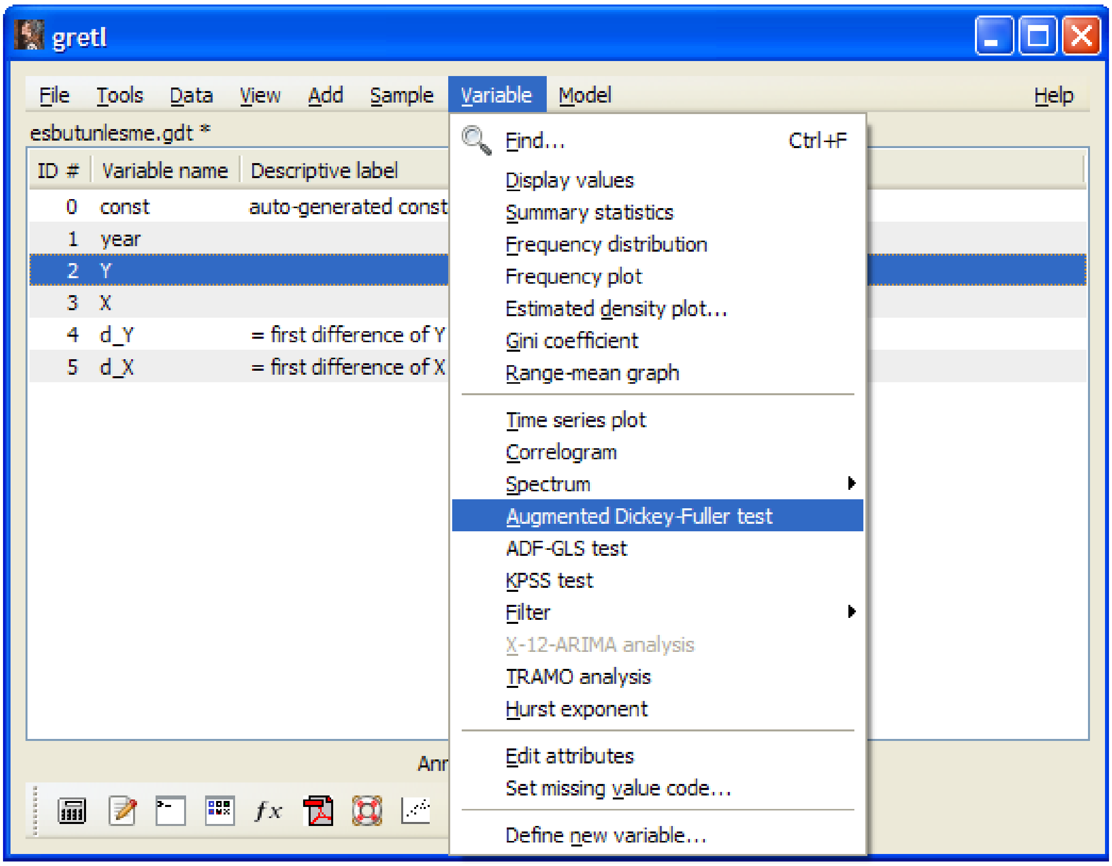


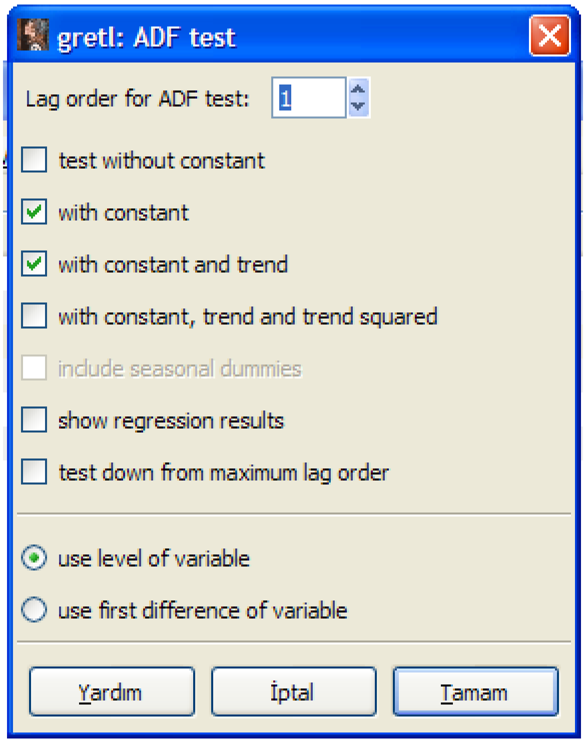
Önce Y ile X arasındaki ilişkiyi, Y'nin X'e göre regresyonuyla inceleyelim:

$$Y=77.3293+3.40599X \qquad R^2=0.26$$
$$t: \quad (7.958) \quad (3.062)$$

R^2=0.26 olmakla birlikte, X'e ait tahminci istatistiki olarak anlamlıdır. Buna göre X, Y'yi açıklamaktadır. Ancak bu ilişkiyi bir de Y ve X'in durağanlıklarını test ettikten sonra inceleyeceğiz.

Y'nin durağanlığını belirlemek için, a) sabit b) Sabit ve trendli ADF modeline göre ADF testini uygulayalım.





Augmented Dickey-Fuller tests, order 1, for Y
sample size 26
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.087
Tahmin değeri (a - 1): -0.206303
test istatistiği: tau_c(1) = -2.30174
asimtotik p 0.1714

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.070
Tahmin değeri (a - 1): -0.171511
test istatistiği: tau_ct(1) = -1.71828
asimtotik p 0.7434

Görüldüğü gibi gerek sabitli, gerekse sabit ve trendli ADF istatistiği için p değerleri %1, %5 veya %10'dan büyük olduğundan sıfır hipotezi reddedilmez. Yani Y serisi durağan değildir.

Aynı şekilde X'in durağanlığını belirlemek için, a) sabit b) Sabit ve trendli ADF modeline göre ADF testini uygulayalım.

Augmented Dickey-Fuller tests, order 1, for X
sample size 26
birim-kök sıfır hipotezi: a = 1

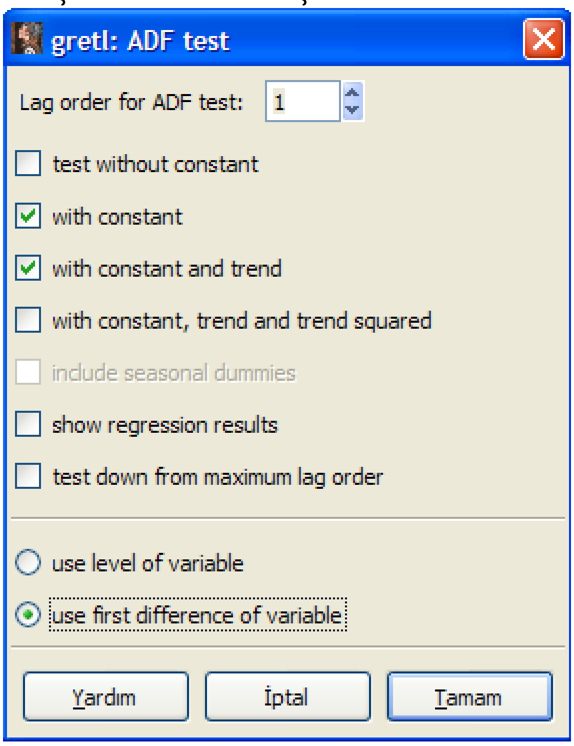
Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.012
Tahmin değeri (a - 1): -0.0228789
test istatistiği: tau_c(1) = -1.01976
asimtotik p 0.7485

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.057
Tahmin değeri (a - 1): -0.311646
test istatistiği: tau_ct(1) = -2.2346
asimtotik p 0.4696

Görüldüğü gibi ADF istatistiği için p değeri %1, %5 veya %10'dan büyük olduğundan sıfır hipotezi reddedilmez. Yani X serisi de durağan değildir.

Y ve X serilerini durağanlaştırmak için her iki değişkenin de birinci farklarını alalım ve durağanlıklarını test edelim:

ΔY için ADF testi sonuçları:



Augmented Dickey-Fuller tests, order 1, for d_Y
sample size 25

birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: 0.099
Tahmin değeri (a - 1): -0.266283
test istatistiği: tau_c(1) = -2.70891
asimtotik p 0.07244

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: 0.081
Tahmin değeri (a - 1): -0.240621
test istatistiği: tau_ct(1) = -2.21921
asimtotik p 0.4782

ΔX için ADF testi sonuçları:

Augmented Dickey-Fuller tests, order 1, for d_X
sample size 25
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.024
Tahmin değeri (a - 1): -1.02268
test istatistiği: tau_c(1) = -3.568
asimtotik p 0.006434

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.043
Tahmin değeri (a - 1): -1.0526
test istatistiği: tau_ct(1) = -3.61442
asimtotik p 0.02855

Görüleceği üzere Y ve X'in birinci farkları durağveır. Artık Y ve X arasndaki ilişkiyi, durağanlığını belirlediğimiz ΔY ve ΔX üzerinden ele alalım:

$$\Delta Y = 1.31755 - 3.52408\ \Delta Y \qquad R^2 = 0.01$$
$$t: \quad (0.958) \quad (-0.608)$$

R^2 sıfıra çok yakın ve ΔY'nin tahmincisi istatistiki açıdan önemsizdir. Bir başka ifadeyle X, Y'yi açıklamamaktadır. Bu durumda durağan olmayan Y ve X serileri arasında yukarıda bulduğumuz ilişki sahtedir diyebiliriz. Gerçekten de Y ülkesinin tüketim harcamaları ile X ülkesinin borsa kapanış değerleri arasında herhangi bir ilişki olmaması gerekir. Bu, durağan olmayan serilerle yapılacak regresyon modellemelerinin, sahte regresyona yolaçabileceğini göstermektedir.

11.4.2. Korelogram Yardımıyla Durağanlığın Belirlenmesi

Korelogram örnek otokorelasyon fonksiyonunun grafiğidir. Otokorelasyon katsayılarının oluşturduğu bir kümenin yorumlanmasında kullanılan bir araçtır. k gecikme için hazırlanan bir grafiktir. Bir başka ifadeyle herbir k gecikme için hesaplanmış korelasyon katsayıları dizisinin, k'ya göre çizilmiş grafiğidir.

Korelogram çoğunlukla veri setinin tesadüfiliğin belirlemede kullanılır. Eğer tesadüfilik varsa, otokorelasyonlar yaklaşık sıfırdır. Tesadüfiliğin olmadığı durumlarda otokorelasyonlardan en az biri veya daha fazlası sıfırdan farklı olacaktır.

Korelogram Box-Jenkins modellerinin tanımlama aşamasında da kullanılır. Bir zaman serisinin otokorelasyon fonksiyonu (ACF) geometrik azalan bir grafik gösteriyorsa, durağan olmadığı anlaşılır. Bu serinin kısmi otokorelasyon fonksiyonu (PACF) incelendiğinde 1. gecikme için istatistiki açıdan önemli bir sıçrama görülmektedir. Buna göre, durağanlığı gidermek için serinin 1. farkı alınmalıdır.

```
Autocorrelation    Partial               AC     PAC    Q-Stat  Prob
                   Correlation
      . |*******|        . |*******|  1  0.969  0.969  85.459  0.000
      . |*******|        . |  .    |  2  0.936  -0.053 166.06  0.000
      . |*******|        . |  .    |  3  0.902  -0.025 241.80  0.000
      . |*******|        . |  .    |  4  0.866  -0.045 312.51  0.000
      . |****** |        . |  .    |  5  0.830  -0.024 378.24  0.000
      . |****** |        .*|  .    |  6  0.791  -0.062 438.73  0.000
      . |****** |        . |  .    |  7  0.752  -0.029 494.04  0.000
      . |*****  |        . |  .    |  8  0.713  -0.024 544.31  0.000
      . |*****  |        . |  .    |  9  0.675  0.009  589.99  0.000
      . |*****  |        . |  .    |  10 0.638  -0.010 631.35  0.000
      . |*****  |        . |  .    |  11 0.602  -0.020 668.57  0.000
      . |****   |        . |  .    |  12 0.566  -0.012 701.90  0.000
      . |****   |        . |  .    |  13 0.532  0.020  731.81  0.000
      . |****   |        . |  .    |  14 0.500  -0.012 758.55  0.000
      . |****   |        . |  .    |  15 0.468  -0.021 782.28  0.000
      . |***    |        . |  .    |  16 0.437  -0.001 803.29  0.000
      . |***    |        . |  .    |  17 0.405  -0.041 821.62  0.000
      . |***    |        . |  .    |  18 0.375  -0.005 837.51  0.000
      . |***    |        . |  .    |  19 0.344  -0.038 851.06  0.000
      . |**     |        . |  .    |  20 0.313  -0.018 862.45  0.000
      . |**     |        .*|  .    |  21 0.279  -0.066 871.67  0.000
      . |**     |        . |  .    |  22 0.246  -0.019 878.92  0.000
      . |**     |        . |  .    |  23 0.214  -0.007 884.49  0.000
      . |*.     |        . |  .    |  24 0.182  -0.018 888.59  0.000
      . |*.     |        . |  .    |  25 0.153  0.017  891.52  0.000
      . |*.     |        . |  .    |  26 0.123  -0.024 893.47  0.000
      . |*.     |        . |  .    |  27 0.095  -0.007 894.65  0.000
      . |*.     |        . |  .    |  28 0.068  -0.011 895.26  0.000
      . | .     |        . |  .    |  29 0.043  -0.007 895.51  0.000
      . | .     |        . |  .    |  30 0.019  -0.005 895.56  0.000
      . | .     |        . |  .    |  31 -0.004 -0.002 895.56  0.000
      . | .     |        . |  .    |  32 -0.026 -0.028 895.65  0.000
      . | .     |        . |  .    |  33 -0.046 0.007  895.96  0.000
      .*| .     |        . |  .    |  34 -0.061 0.047  896.51  0.000
      .*| .     |        . |  .    |  35 -0.075 0.004  897.35  0.000
      .*| .     |        . |  .    |  36 -0.085 0.036  898.45  0.000
```

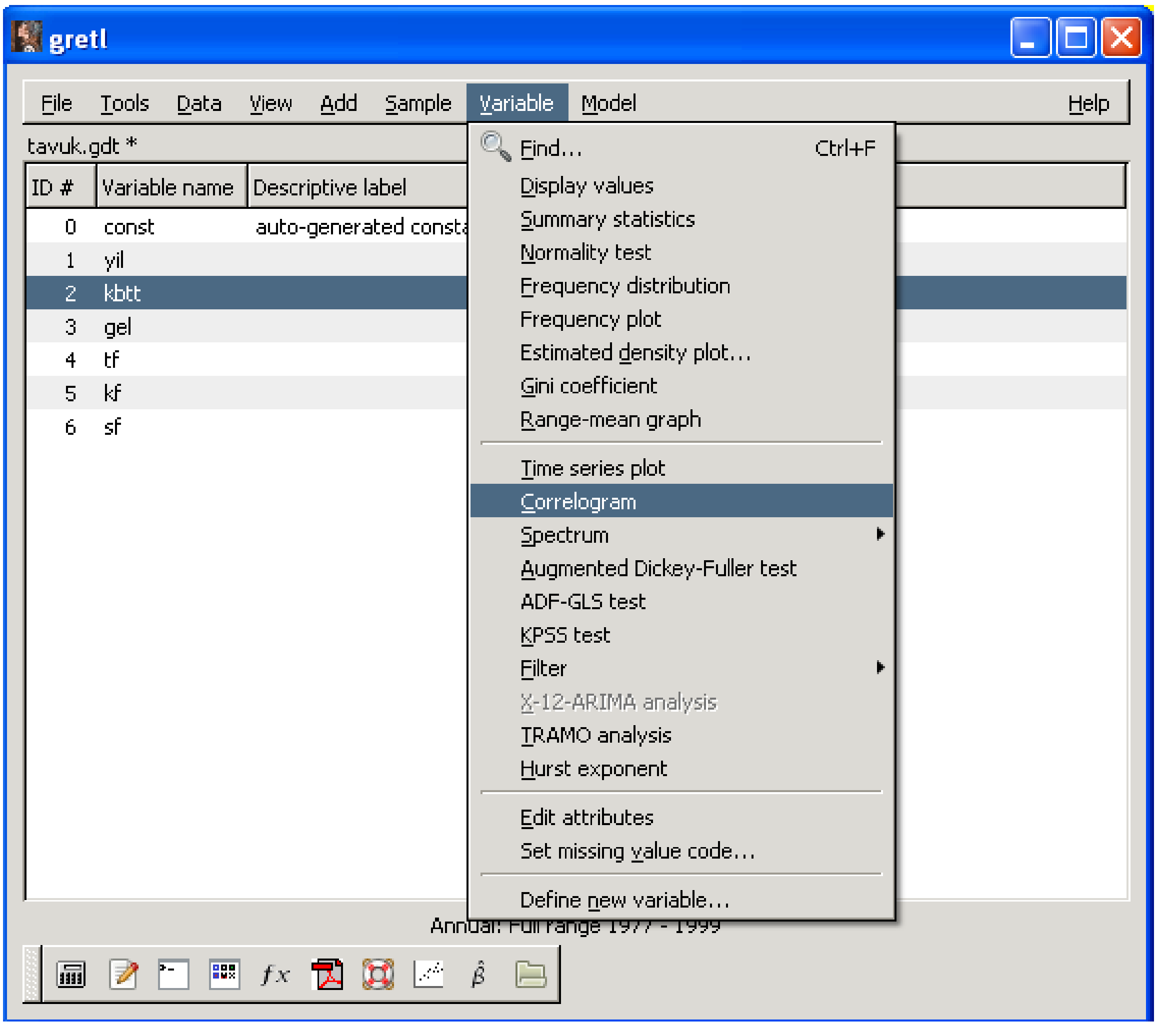
Örnek

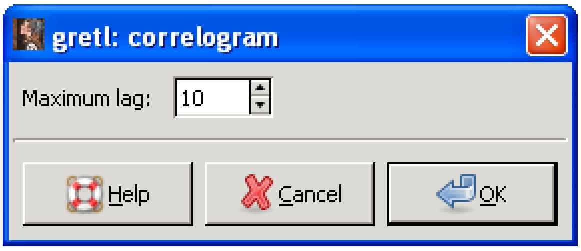
Durağanlığı belirlemek üzere kişi başına tavuk eti tüketim (KBTT) değişkenini ele alalım:

Çizelge 2: KBTT zaman serisi (KBTT.gdt)

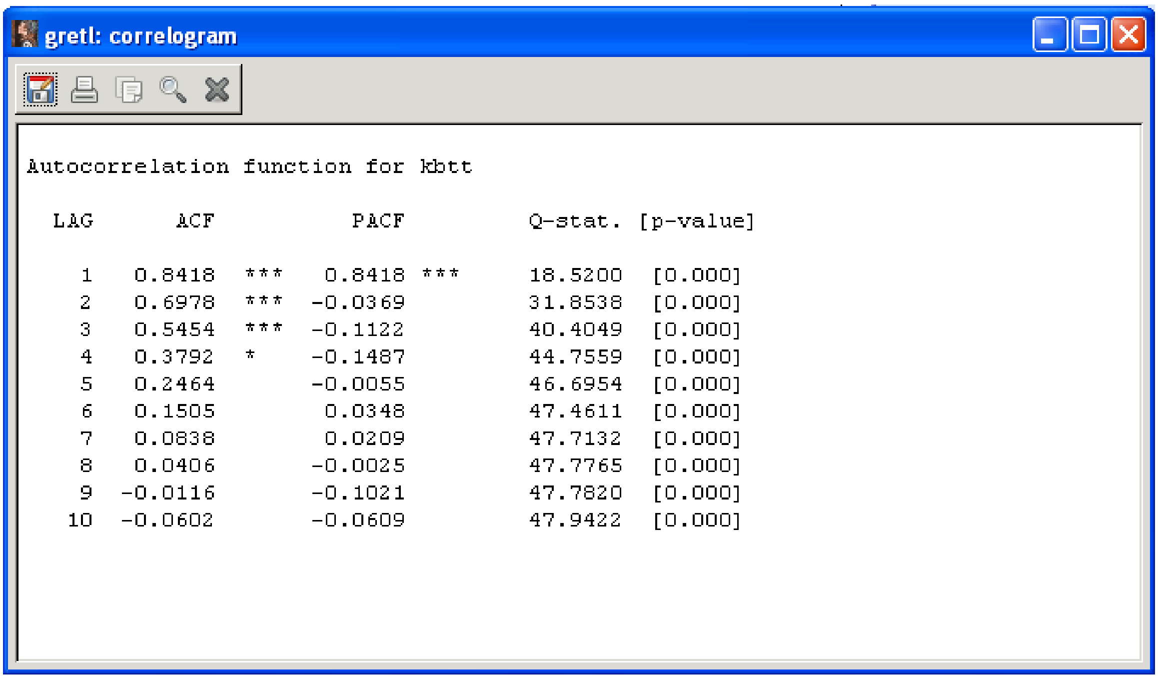
Yıl	KBTT	Yıl	KBTT
1977	27.8	1989	41.8
1978	29.9	1990	40.4
1979	29.8	1991	40.7
1980	30.8	1992	40.1
1981	31.2	1993	42.7
1982	33.3	1994	44.1
1983	35.6	1995	46.7
1984	36.4	1996	50.6
1985	36.7	1997	50.1
1986	38.4	1998	51.7
1987	40.4	1999	52.9
1988	40.3		

GRETL'da Korelogram hazırlama süreci:

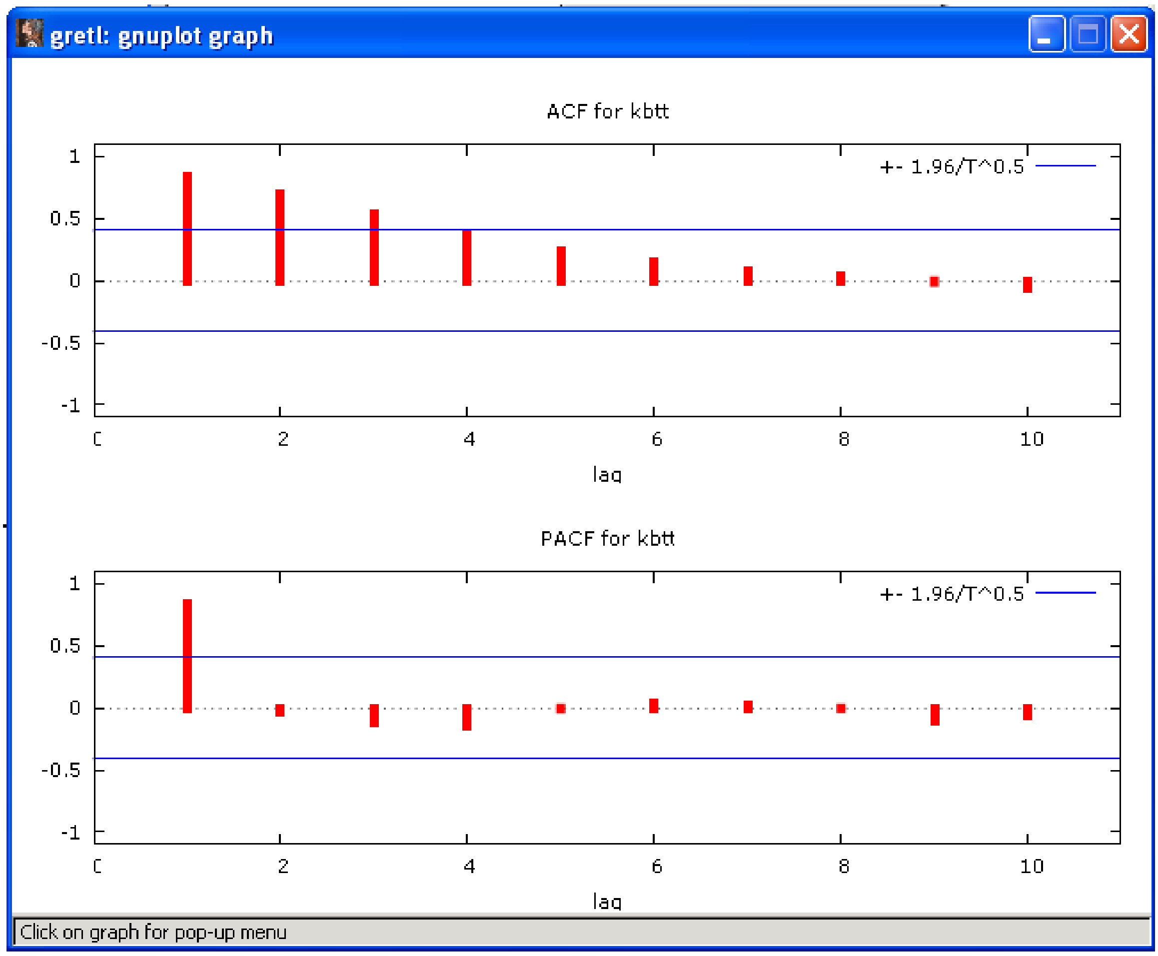




Aşağıda sunulan korelogramda 1-10 arasındaki gecikmeler için otokorelasyon fonksiyonuna (ACF) ait değerler verilmektedir. Sağ taraftaki Q istatistikleri, ACF değerlerinin tamamının sıfırdan farklı olduğu sonucunu vermektedir. Bu durumda tesadüfilik söz konusu değildir. Aynı zamanda durağanlık olmadığı anlamına gelmektedir.

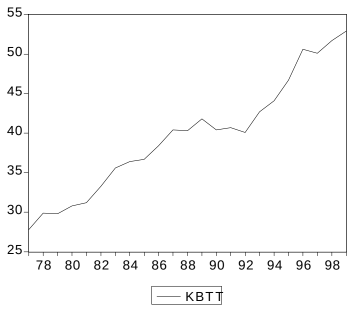



```
Autocorrelation function for kbtt

  LAG       ACF          PACF         Q-stat. [p-value]

    1   0.8418  ***   0.8418  ***    18.5200  [0.000]
    2   0.6978  ***  -0.0369         31.8538  [0.000]
    3   0.5454  ***  -0.1122         40.4049  [0.000]
    4   0.3792  *    -0.1487         44.7559  [0.000]
    5   0.2464       -0.0055         46.6954  [0.000]
    6   0.1505        0.0348         47.4611  [0.000]
    7   0.0838        0.0209         47.7132  [0.000]
    8   0.0406       -0.0025         47.7765  [0.000]
    9  -0.0116       -0.1021         47.7820  [0.000]
   10  -0.0602       -0.0609         47.9422  [0.000]
```



Otokorelasyonun geometrik olarak azalması, durağanlığın olmadığına işaret etmektedir. Q istatistiği korelogramda hazır olarak verilmiştir. Q=47.94 için p değeri 0.000 olduğuna göre H_0 reddedilir. Yani seri durağan değildir.

Şimdi serinin grafiğini görelim:



KBTT değişkeninin durağanlığını test etmek üzere bu kez birim kök testi yapalım. Grafik, seride trend olduğunu göstermektedir. O halde modelimizde trend mutlaka olmalıdır. Biz önce sabitin, daha sonra da hem sabit hem de trendin bulunduğu modelleri uygulayacağız:

Augmented Dickey-Fuller tests, order 1, for kbtt
sample size 21
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: 0.045
Tahmin değeri (a - 1): 0.0173766
test istatistiği: tau_c(1) = 0.376214
asimtotik p 0.982

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.003
Tahmin değeri (a - 1): -0.347708
test istatistiği: tau_ct(1) = -1.70681
asimtotik p 0.7486

Her iki modeldeki p değeri "birim kök vardır" hipotezini kabul ettiğine göre, seri durağan değildir.

Hata! Başvuru kaynağı bulunamadı.'den farkedileceği üzere 1. gecikmede PACF güven sınırları dışına sıçramaktadır. Buna göre, serimizin birinci farkını almamız, durağanlaştırma için yeterli olacaktır:

$$\Delta(KBTT)_t = KBTT_t - KBTT_{t-1}$$

Çizelge 3: KBTT 1. farkı

Yıl	$KBTT_t$	$\Delta(KBTT)_t$
1977	27.8	*
1978	29.9	2.1
1979	29.8	-0.1
1980	30.8	1
1981	31.2	0.4
1982	33.3	2.1
1983	35.6	2.3
1984	36.4	0.8
1985	36.7	0.3
1986	38.4	1.7
1987	40.4	2
1988	40.3	-0.1
1989	41.8	1.5
1990	40.4	-1.4
1991	40.7	0.3
1992	40.1	-0.6
1993	42.7	2.6
1994	44.1	1.4
1995	46.7	2.6
1996	50.6	3.9
1997	50.1	-0.5
1998	51.7	1.6
1999	52.9	1.2

Şimdi KBTT'nin birinci farkına birim kök testi uygulayalım:

Augmented Dickey-Fuller tests, order 1, for d_kbtt
sample size 20

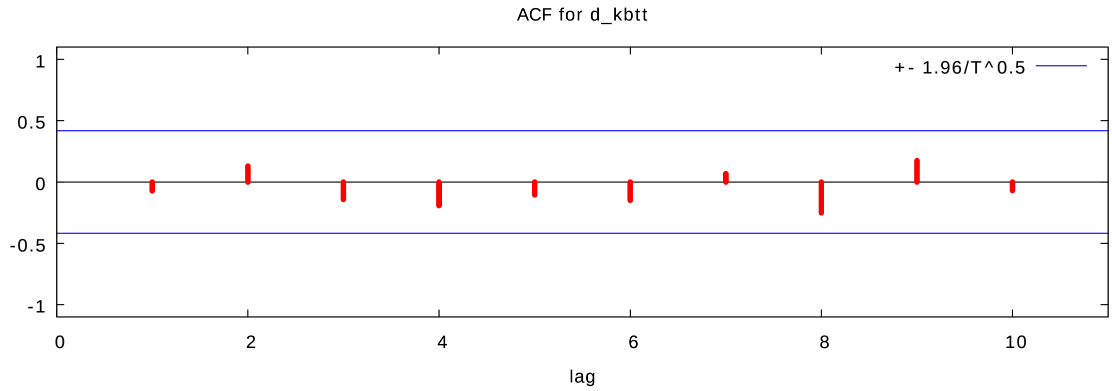
birim-kök sıfır hipotezi: a = 1

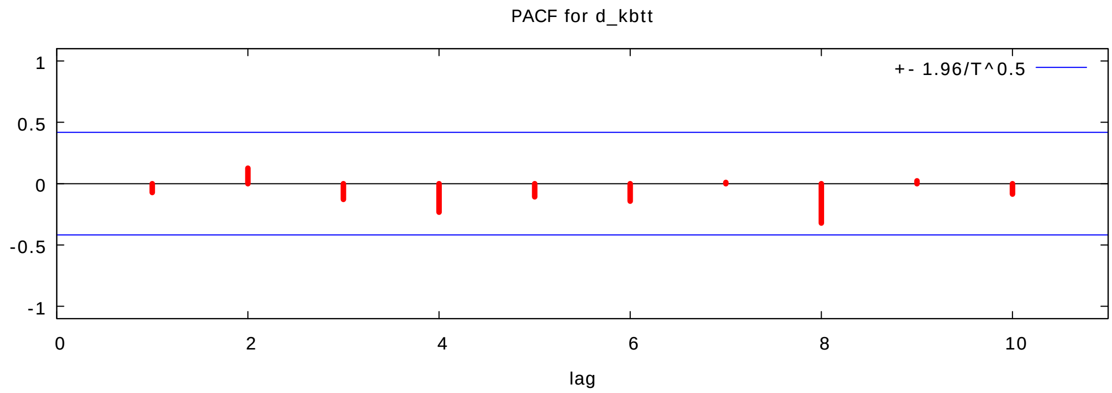
Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: 0.006
Tahmin değeri (a - 1): -0.898754
test istatistiği: tau_c(1) = -2.61883
asimtotik p 0.08908

Görüldüğü gibi p değeri sıfır hipotezinin reddedilmesini gerektirmektedir. O halde serinin durağan olduğuna karar verebiliriz. Bunu doğrulamak üzere, kbtt değişkeninin birinci farkına korelogram uygulaması yapabiliriz:

Autocorrelation function for d_kbtt

LAG	ACF	PACF	Q-stat. [p]
1	-0.0721	-0.0721	0.1308[0.718]
2	0.1301	0.1255	0.5775[0.749]
3	-0.1421	-0.1276	1.1391[0.768]
4	-0.1928	-0.2325	2.2294[0.694]
5	-0.1042	-0.1067	2.5668[0.766]
6	-0.1494	-0.1430	3.3033[0.770]
7	0.0693	0.0097	3.4723[0.838]
8	-0.2508	-0.3214	5.8455[0.665]
9	0.1740	0.0225	7.0748[0.629]
10	-0.0698	-0.0855	7.2891[0.698]





Görüldüğü gibi PACF grafiğinde güven sınırlarını aşan bir sıçrama yoktur. Bu da serimizin durağan olduğunu doğrulamaktadır.

11.5. ARIMA ile Öngörümleme

Genel ARIMA modeli 1976'da Box ve Jenkins tarafından önerilmiştir. Modelde otoregresif parametrenin yanısıra, hareketli ortalama ve fark parametreleri bulunur. Bunlar: Otoregresif parametre (p), fark alma sayısı (d) ve hareketli ortalama parametresidir. Box Jenkins gösterimi:

ARIMA (*p, d, q*)

şeklindedir. Örneğin ARIMA (0,1,2)'de sıfır otoregresif parametre, iki hareketli ortalama parametre ve bir farkı alınmış seri bulunmaktadır.

ARIMA; bir zaman serisi değişkeninin mevcut değerlerini:

- Geçmiş değerleriyle
- Mevcut ve geçmiş tesadüfi hatalarıyla

ilişkilendirir.

Box-Jenkins Modeli

B-J ARMA modeli, AR ve MA modellerinin bir karmasıdır:

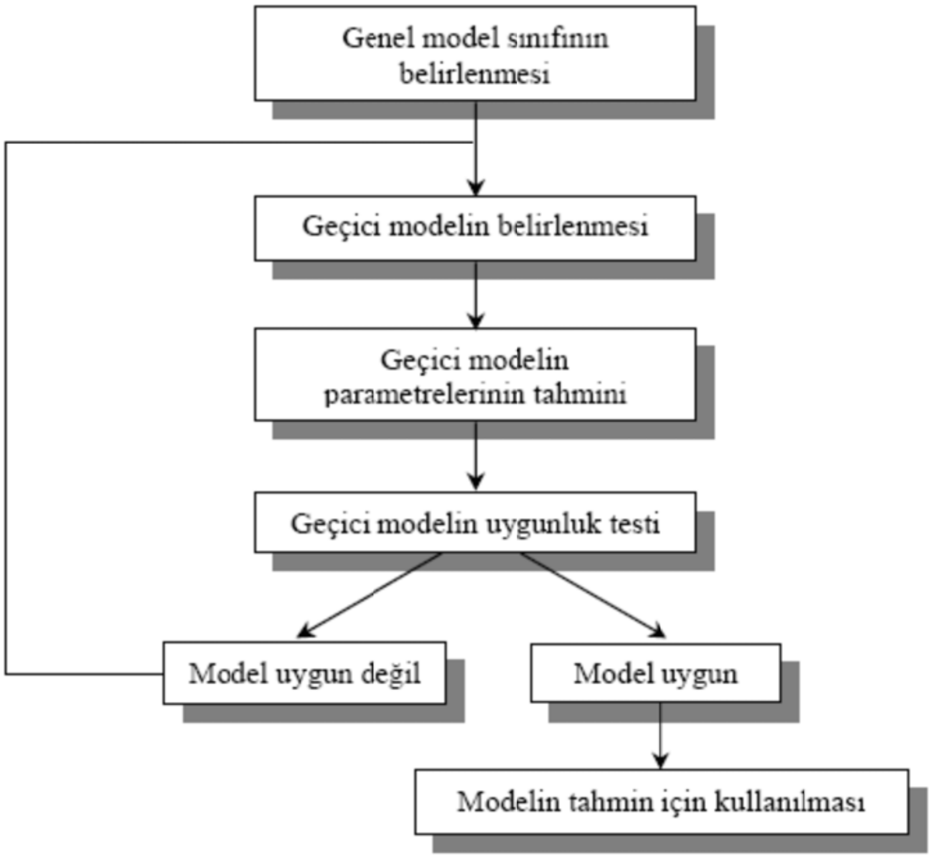
$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + \ldots + \phi_1 Y_{t-p} + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \ldots - \theta_q A_{t-q}$$

B-J modeli zaman serisinin durağan olduğunu kabul eder. B-J, durağanlığı yakalayana kadar bir veya daha fazla fark alınmasını önerir. Fark alma işlemi I ile ifade edilir ve ARIMA'ya ulaşılır.

Logaritma ve karekök alma gibi bazı dönüştürme işlemleriyle, ortalaması sıfır olan seriler elde edilebilir.

B-J zaman serisi modelinin kurulmasında üç ana aşama vardır:

1. **Tanımlama:** Uygun p ve q değerleri bulunur. Bunun için korelogram ve kısmi korelogramdan yararlanılır.
2. **Modelin tahhminlenmesi:** p ve q'lu parametrelerin tahmincilerinin hesaplanması.
3. **Modelin geçerliliğini doğrulama:** ARIMA modelden elde edilen hata terimi eğer durağansa, modelin geçerli olduğu kabul edilir.



Box-Jenkins yöntemiyle model belirleme aşamaları

Otokortelasyon fonksiyonu (ACF)

ACF geçmiş gecikmelerin y_t üzerindeki etkisini sabit tutmadan, (y_t ve y_{t-1}), (y_t ve y_{t-2}), (y_t ve y_{t-3}) arasındaki korelasyonlar dizisidir. ACF, MA için q değerlerini belirlemede kullanılır.

k gecikmeli ACF ρ_k ile gösterilir ve:

$$\rho_k = \gamma_k / \gamma_o$$

formülüyle hesaplanır. Burada γ_k k gecikmesi için kovaryans, γ_o ise varyansı göstermektedir.

Kısmi Otokorelasyon Fonksiyonu (PACF)

y_t'nin geçmiş gecikmelerinin etkilerini sabit tutarak sunulan, (y_t ve y_{t-1}), (y_t ve y_{t-2}), (y_t ve y_{t-3}) arasındaki korelasyonlar dizisidir. PACF, AR için p değerlerini belirlemede kullanılır.

ACF ve PACF'nin teorik desenleri:

Model Tipi	ACF deseni	PACF deseni
AR(p)	Üssel olarak azalır veya sinüs eğrisi gösterir ya da herikisi birden	p gecikmesinden sonra katsayı aniden düşerek istatistiksel olarak anlamsız olur.
MA(q)	Q gecikmelerde önemli sıçramalar	Üssel veya sinüzodial olarak azalır
ARMA(p,q)	Üssel olarak azalır	Üssel olarak azalır

ARIMA Tahminlemesi

Örnek

Q değişkenine Box-Jenkins yaklaşımını uygulayalım (**Q.gdt**):

Yıl	Q	dQ
1970	77	
1971	86	9
1972	87	1
1973	92	5
1974	84	-8
1975	93	9
1976	92	-1
1977	100	8
1978	102	2
1979	113	11
1980	101	-12
1981	117	16
1982	117	0
1983	88	-29
1984	111	23
1985	118	7
1986	109	-9
1987	108	-1
1988	92	-16
1989	107	15
1990	114	7
1991	111	-3

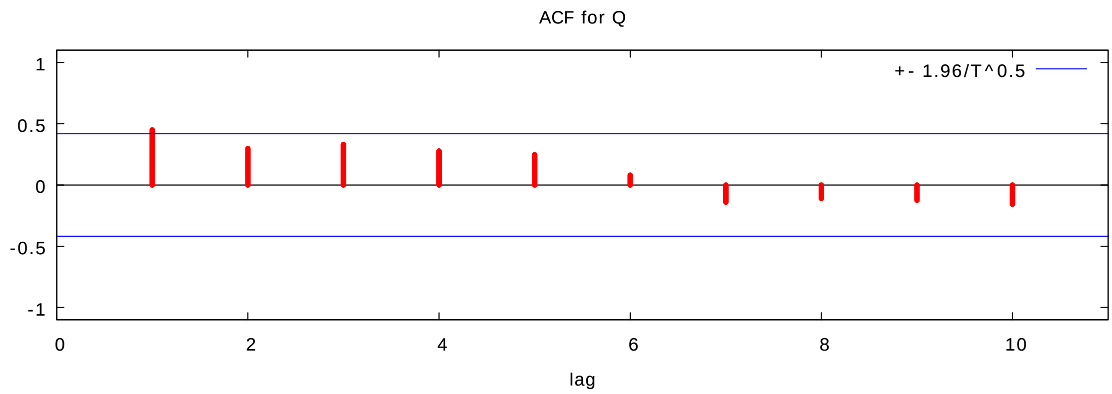
Box-Jenkins, serinin durağanlaştırılmasını istemektedir:

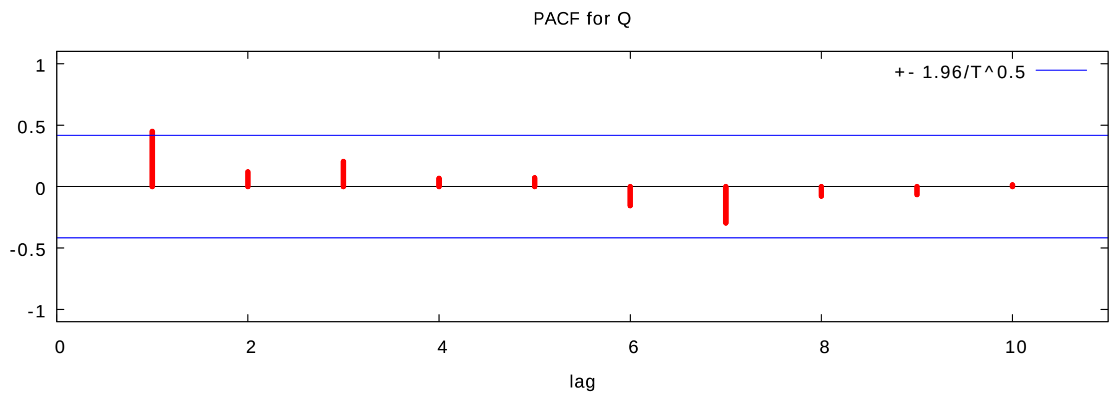
Q için durağanlık testi

Korelogram:

Autocorrelation function for Q

LAG	ACF	PACF	Q-stat.[p]
1	0.4492**	0.4492**	5.0740[0.024]
2	0.2968	0.1190	7.4000[0.025]
3	0.3316	0.2041	10.4550[0.015]
4	0.2771	0.0675	12.7081[0.013]
5	0.2478	0.0711	14.6154[0.012]
6	0.0808	-0.1553	14.8306[0.022]
7	-0.1396	-0.2958	15.5170[0.030]
8	-0.1093	-0.0770	15.9678[0.043]
9	-0.1250	-0.0655	16.6019[0.055]
10	-0.1571	0.0135	17.6883[0.060]



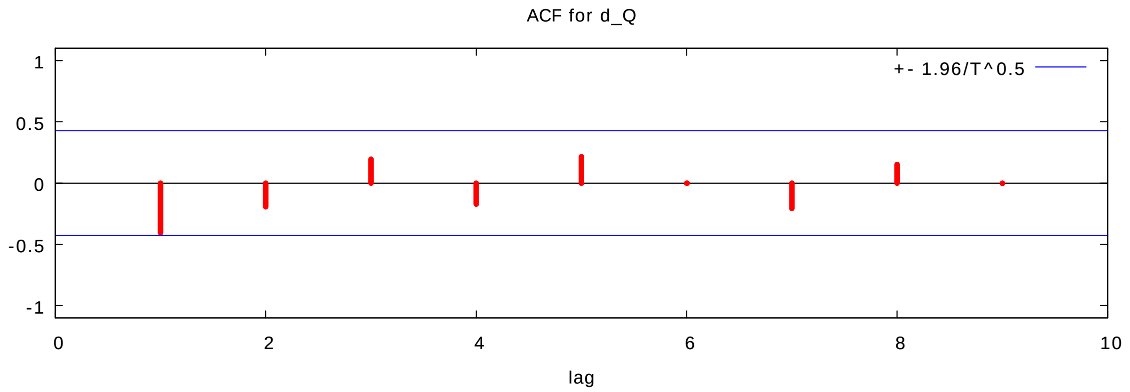


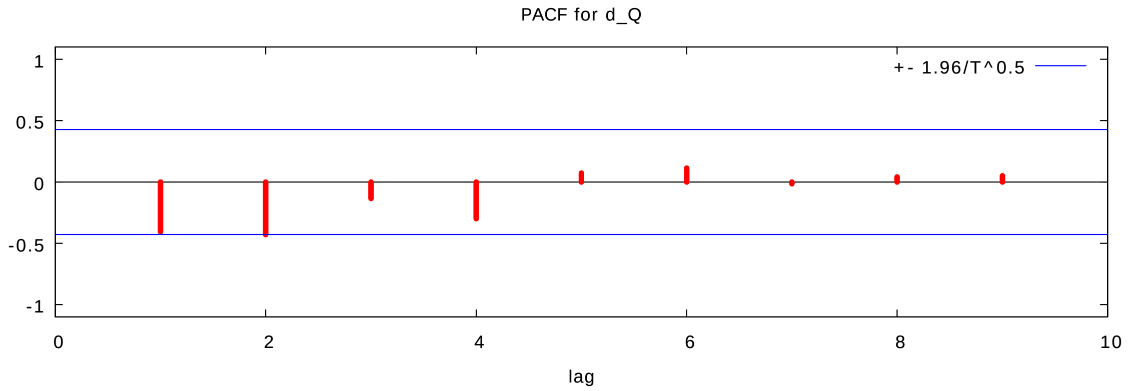
PACF'de 1. gecikmede güven sınırları dışına sıçrama söz konusudur. O halde 1. farkın alınması durağanlaştırma için yeterli olacaktır.

Bu kez 1. fark için korelogram alalım:

Autocorrelation function for d_Q

LAG	ACF	PACF	Q-stat.[p]
1	-0.4064*	-0.4064*	3.9892[0.046]
2	-0.1934	-0.4295**	4.9397[0.085]
3	0.1950	-0.1350	5.9596[0.114]
4	-0.1710	-0.2985	6.7901[0.147]
5	0.2158	0.0721	8.1956[0.146]
6	-0.0003	0.1126	8.1956[0.224]
7	-0.2063	-0.0147	9.6637[0.208]
8	0.1524	0.0420	10.5261[0.230]
9	-0.0029	0.0498	10.5265[0.310]





Görüldüğü gibi ACF'de üssel bir azalma ve PACF'de ani bir düşme söz konusu değildir. O halde seri durağandır. Bunu birim kök testi ile de doğrulayabiliriz:

Augmented Dickey-Fuller tests, order 1, for d_Q
sample size 19
birim-kök sıfır hipotezi: a = 1

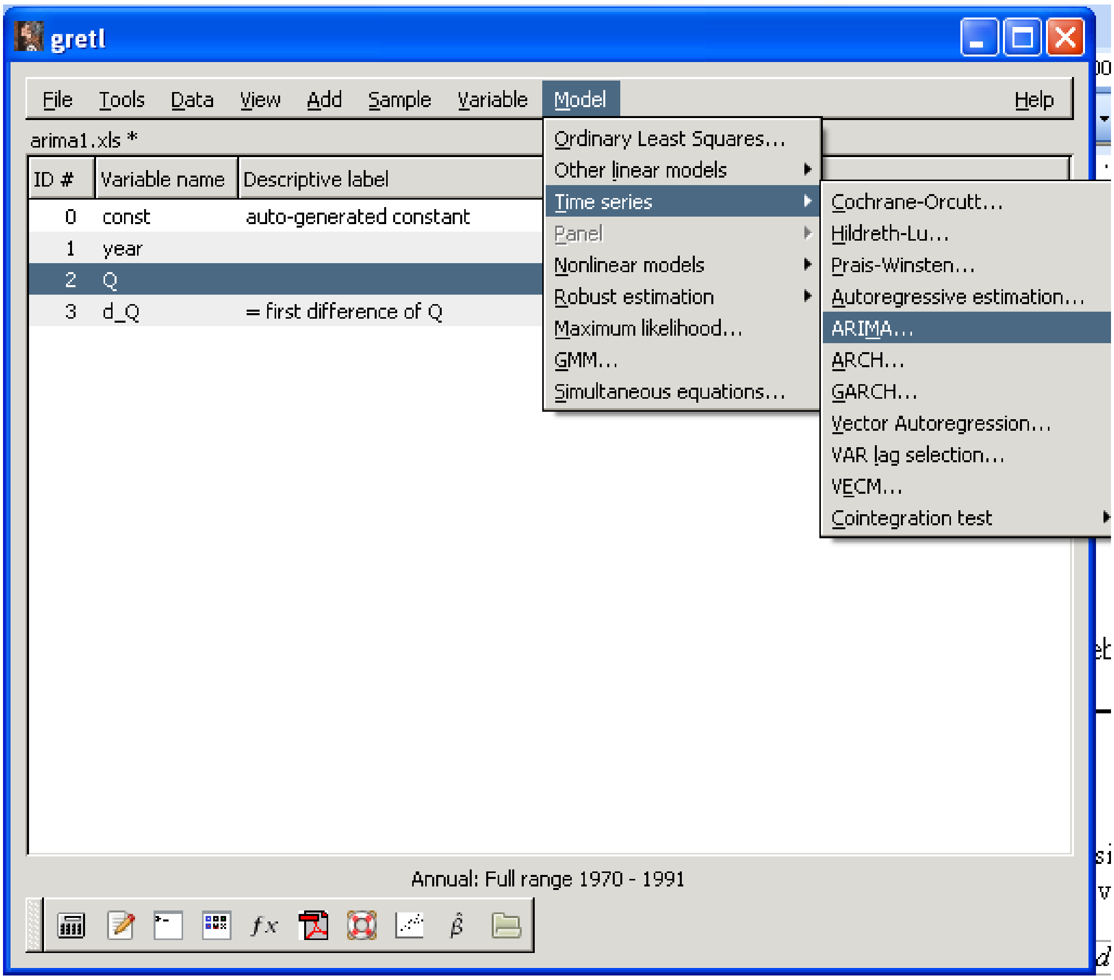
Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.080
Tahmin değeri (a - 1): -2.03344
test istatistiği: tau_c(1) = -5.41693
asimtotik p 2.652e-006

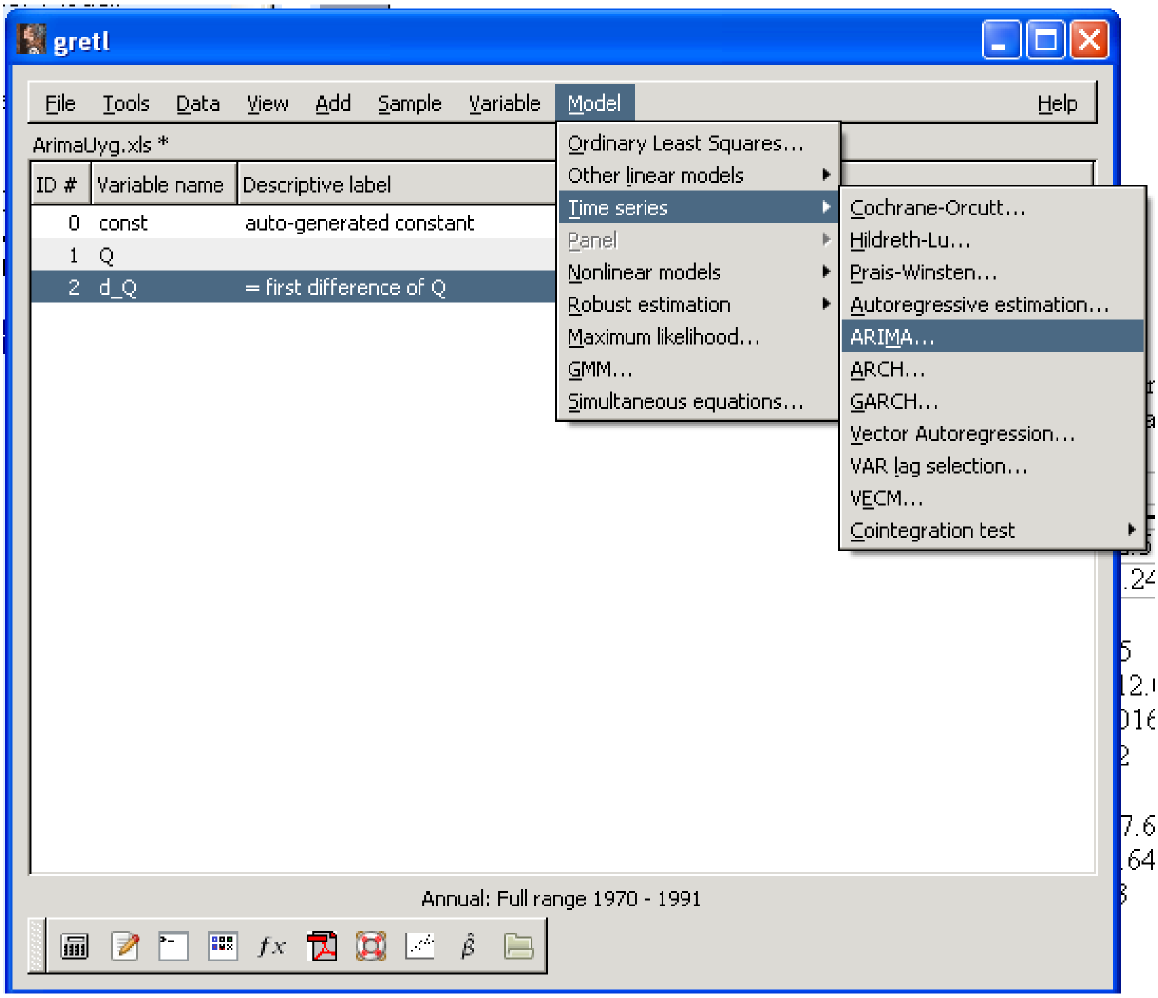
Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.104
Tahmin değeri (a - 1): -2.06613

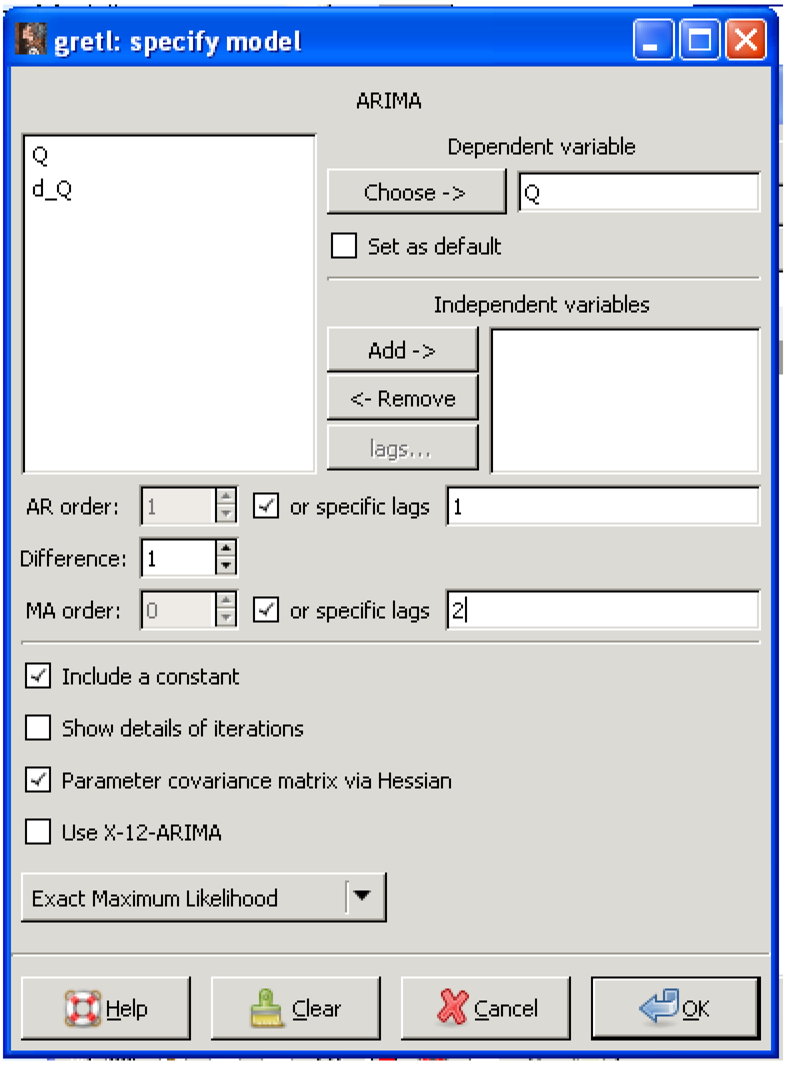
test istatistiği: tau_ct(1) = -5.34873
asimtotik p 3.539e-005

Sonuç: Seri durağandır.

Artık ARIMA tahminlemesi yapabiliriz. AR için p, MA için q değerlerinin ne olduğunu, yine ACF ve PACF üzerinden belirleyeceğiz. Hatırlanacağı üzere PACF, AR için p değerlerini; ACF ise MA için q değerlerini verecektir. Yukarıdaki grafikte PACF, AR için 1'inci gecikmeye; ACF, MA için 2. gecikmeye işaret etmektedir. O halde modelimizi; AR(1) ve MA(2) ile kurabiliriz.




gretl
File Tools Data View Add Sample Variable Model Help
ArimaUyg.xls *
ID # Variable name Descriptive label
0 const auto-generated constant
1 Q
2 d_Q = first difference of Q
Ordinary Least Squares...
Other linear models
Time series
Panel
Nonlinear models
Robust estimation
Maximum likelihood...
GMM...
Simultaneous equations...
Cochrane-Orcutt...
Hildreth-Lu...
Prais-Winsten...
Autoregressive estimation...
ARIMA...
ARCH...
GARCH...
Vector Autoregression...
VAR lag selection...
VECM...
Cointegration test
Annual: Full range 1970 - 1991

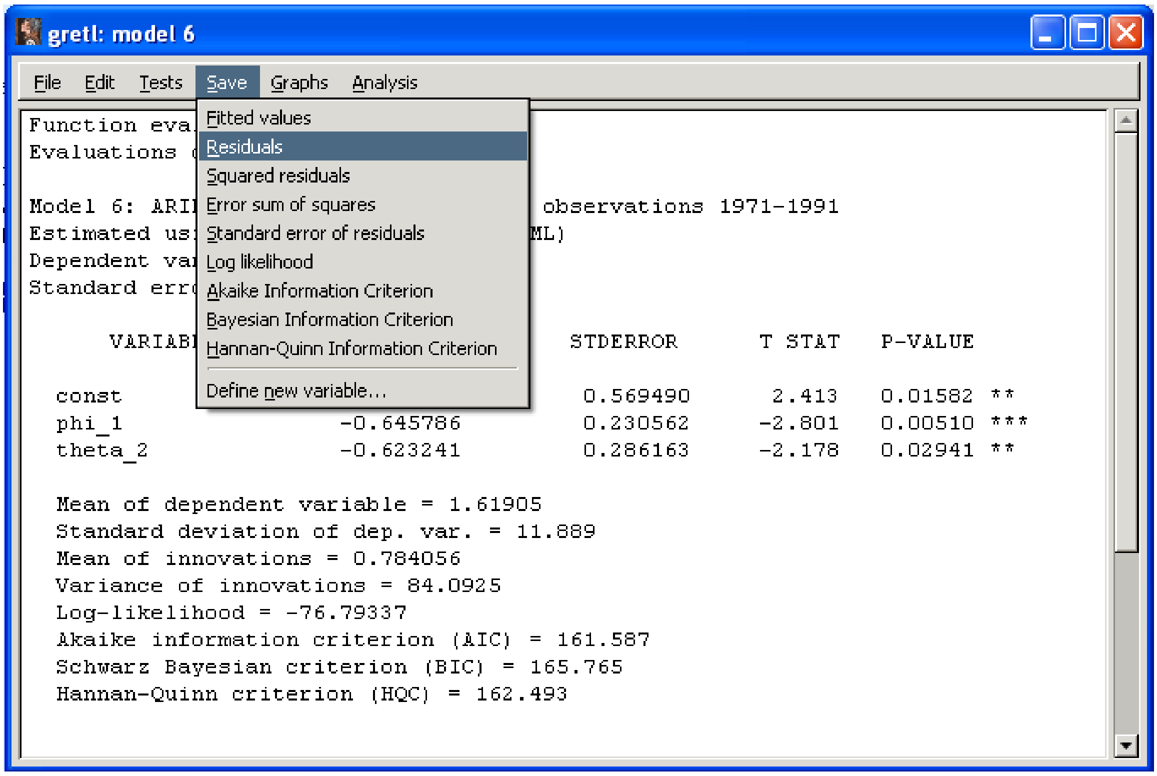


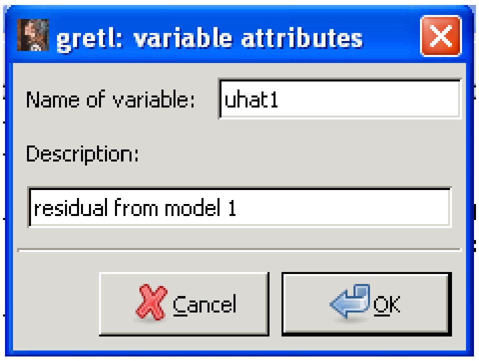
Model 6: ARIMA estimates using the 21 gözlem 1971-1991
Bağımlı değişken: (1-L) Q
Stveard errors based on Hessian

	Katsayı	*St.Hata*	*t*	*p*	
sabit	1.37421	0.56949	2.4130	0.01582	**
phi_1	-0.645786	0.230562	-2.8009	0.00510	***
theta_2	-0.623241	0.286163	-2.1779	0.02941	**

Bağımlı değişkenin ortalaması = 1.61905
Bağımlı değişkenin st sapması = 11.889
Mean of innovations = 0.784056
Variance of innovations = 84.0925
Log-likelihood = -76.7934
Akaike bilgi kriteri = 161.587
Schwarz Bayesian kriteri = 165.765
Hannan-Quinn kriteri = 162.493

Görüldüğü gibi, phi_1 [=AR(1)] ve theta_2 [=MA(2)] istatistiki açıdan anlamlıdır. Bu, seçilen p ve q değerlerinin isabetli olduğunu göstermektedir. Ancak modelin doğruluğunu, ARIMA modelinden elde edilen hata değerinin durağan olup olmadığını test ettikten sonra anlayabileceğiz. Bunun için tahminlemeden hemen sonra hata terimini değişken olarak saklamamız gerekir:





Uhat1 değişkenine birim kök uygularsak:

Augmented Dickey-Fuller tests, order 1, for uhat1
sample size 19
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.008
Tahmin değeri (a - 1): -1.0779
test istatistiği: tau_c(1) = -2.98947
asimtotik p 0.0359

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.020
Tahmin değeri (a - 1): -1.42035
test istatistiği: tau_ct(1) = -3.57143
asimtotik p 0.03225

Görüldüğü üzere sıfır hipotezi reddedilmektedir. O halde hata terimi durağandır. Bu, modelimizin doğru seçildiğini göstermektedir.

ARIMA ile öngörümlemeyi kolayca anlayabilmek için, aynı seri için sadece AR(1) ve AR(2) uygulamaları üzerinde duracağız.

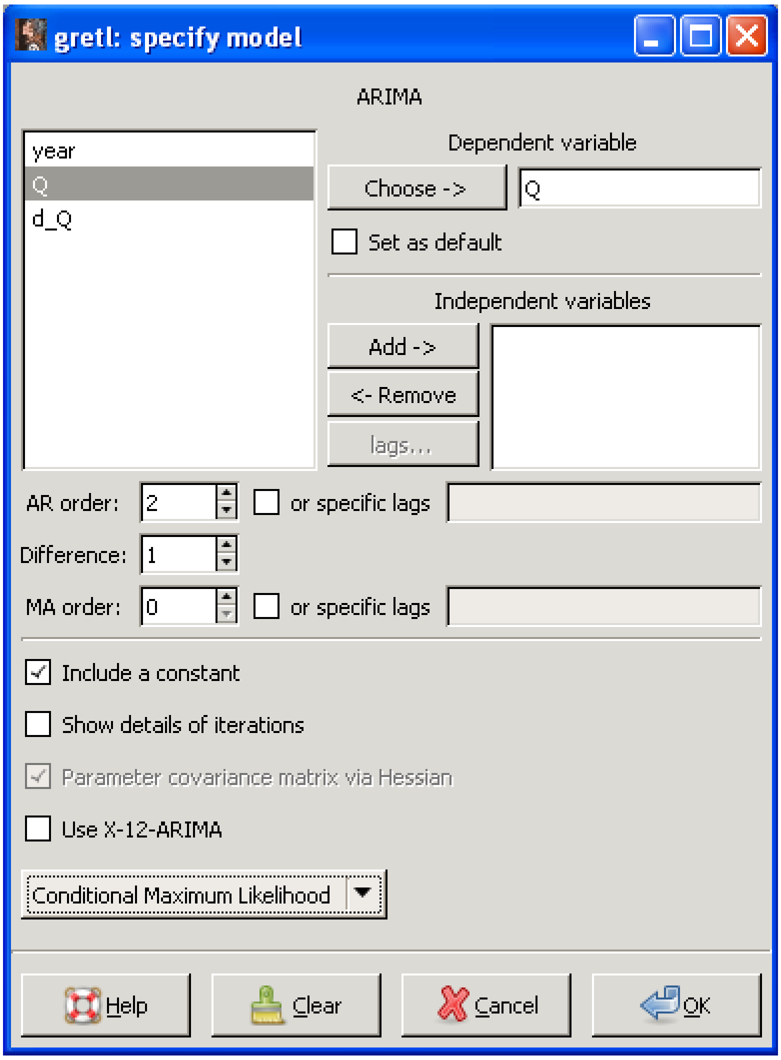
Önce AR(1) Modelini tahmin edelim:

Model 4: ARIMA estimates using the 20 gözlem 1972-1991
Bağımlı değişken: (1-L) Q

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	2.00677	2.56566	0.7822	0.43412	
phi_1 [(AR(1)]	-0.409067	0.242236	-1.6887	0.09127	*

Bağımlı değişkenin ortalaması = 1.25
Bağımlı değişkenin st sapması = 12.0738
Mean of innovations = 1.77636e-016
Variance of innovations = 115.022
Log-likelihood = -75.83
Akaike bilgi kriteri = 157.66
Schwarz Bayesian kriteri = 160.647
Hannan-Quinn kriteri = 158.243

AR(2) modeli



Model 5: ARIMA estimates using the 19 gözlem 1973-1991
Bağımlı değişken: (1-L) Q

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	2.83223	2.6553	1.0666	0.28614	
phi_1	-0.595408	0.435087	-1.3685	0.17116	
phi_2	-0.438034	0.22295	-1.9647	0.04945	**

Bağımlı değişkenin ortalaması = 1.26316
Bağımlı değişkenin st sapması = 12.4045
Mean of innovations = 9.34925e-017
Variance of innovations = 97.2718
Log-likelihood = -70.4462
Akaike bilgi kriteri = 148.892
Schwarz Bayesian kriteri = 152.67
Hannan-Quinn kriteri = 149.532

AR(1) mi yoksa AR(2)'mi sorusunun cevabını AIC veya Schwarz'a bakarak verebiliriz. AR(2)'nin bu 3 kriter açısından daha iyi olduğu açıktır.

Şimdi önce AR(1) ile kestirim yapalım. Modelimiz:

$$\Delta\hat{Q} = 2.00677\text{-}0.40907\ (\Delta Q)_{t-1}$$

şeklindedir. Cebirsel işlemlerle denklemi açarsak:

$$Q_t - Q_{t-1} = 2.00677\text{-}0.40907\ (\Delta Q)_{t-1}$$

$$Q_t = 1.\ 2.00677\text{-}0.40907\ \Delta Q_{t-1} + Q_{t-1}$$

haline gelir. O halde örneğin 1973 yılının tahmini Q değerini hesaplayabilmek için, 1972'ye ait ΔQ ve 1972'ye ait Q değerini kullanmamız gerekecek. Yani:

$$Q_{1973} = 2.00677\text{-}0.40907\ (\Delta Q)_{1972} + Q_{1972}$$

$$Q_{1973} = 2.00677\text{-}0.40907\ (1) + (87)$$

$$Q_{1973} = 88.5977$$

Şimdi bu modele göre fark değerlerini tahmin edelim:

Yıl	t	Q	ΔQ_t	$\Delta\hat{Q}$	$\hat{Q}_t$
1970	1	77			
1971	2	86	9	1.58051	
1972	3	87	1	-1.37563	84.32517
1973	4	92	5	1.811803	88.59771
1974	5	84	-8	0.218086	91.96144
1975	6	93	9	5.397667	89.27931
1976	7	92	-1	-1.37563	91.32517
1977	8	100	8	2.608661	94.41584
1978	9	102	2	-0.9772	98.73424
1979	10	113	11	1.413374	103.1886
1980	11	101	-12	-2.17249	110.507
1981	12	117	16	6.991384	107.9156
1982	13	117	0	-4.16464	112.4617
1983	14	88	-29	2.210232	119.0068
1984	15	111	23	13.76468	101.8697
1985	16	118	7	-6.95364	103.5982
1986	17	109	-9	-0.57877	117.1433
1987	18	108	-1	5.796096	114.6884
1988	19	92	-16	2.608661	110.4158
1989	20	107	15	8.585101	100.5519
1990	21	114	7	-3.76621	102.8708
1991	22	111	-3	-0.57877	113.1433

Bunun için fark alma işlemini geri döndürmemiz gerekir. Bir başka ifade ile integrasyon uygulaması yapacağız.

$$Q_t - Q_{t-1} = 1.424186 - 0.40907\ (Q_{t-1} - Q_{t-2})$$

$$Q_t = 2.00677 - 0.40907\ (Q_{t-1} - Q_{t-2}) + Q_{t-1}$$

$$Q_t = 2.00677 - 0.40907\ Q_{t-1} + 0.40907\ Q_{t-2} + Q_{t-1}$$

$$\mathbf{Q_t = 2.00677 + 0.59093\ Q_{t-1} + 0.40907\ Q_{t-2}}$$

olur. Örneğin 1970 yılını kestirelim:

$$\mathbf{Q_{1970} = 2.00677 + 0.59093\, Q_{1969} + 0.40907\, Q_{1968}}$$

Görüldüğü gibi 1968 ve 1969 yıllarına ait veri bulunmamaktadır. O yüzden 1970 için kestirim yapamayız. 1971 için kestirim:

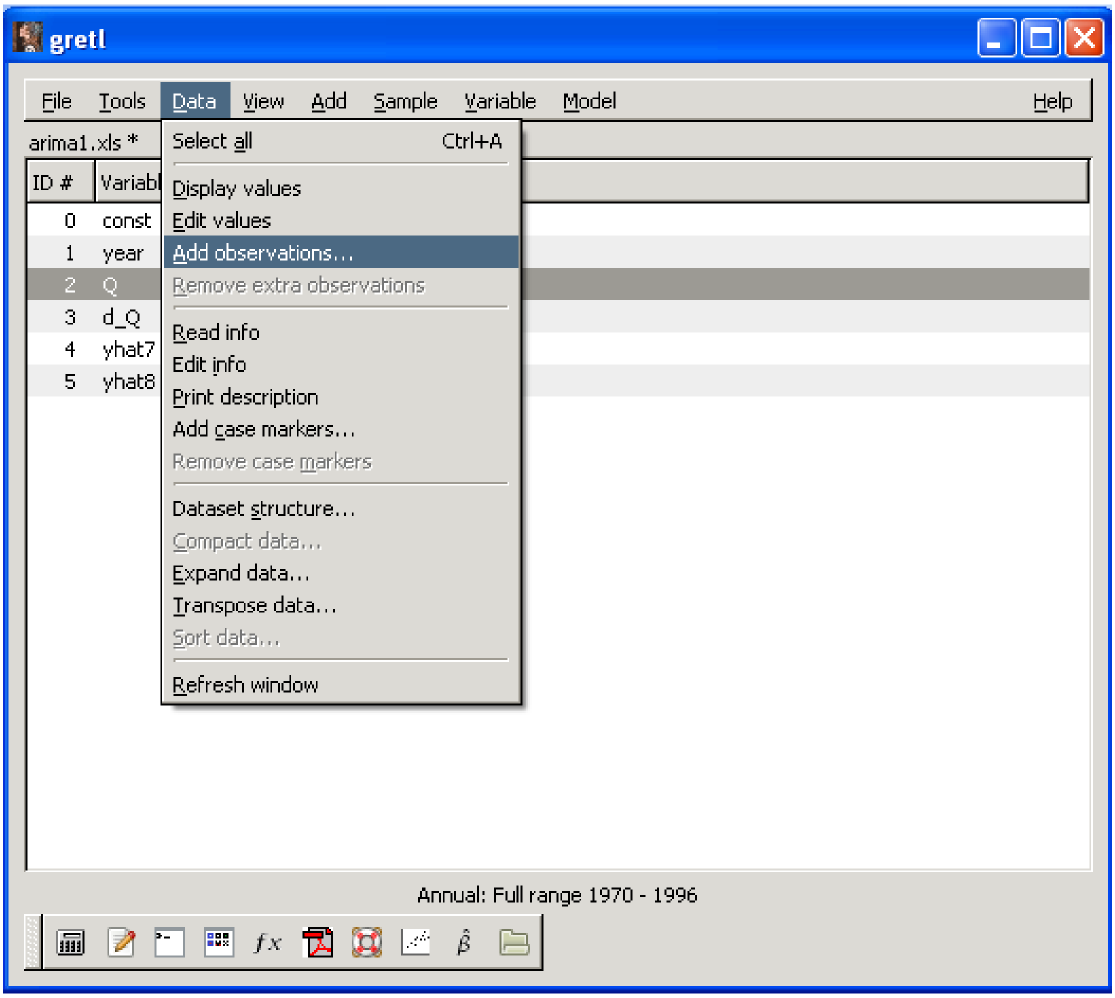
$$\mathbf{Q_{1971} = 2.00677 + 0.59093\, Q_{1970} + 0.40907\, Q_{1969}}$$

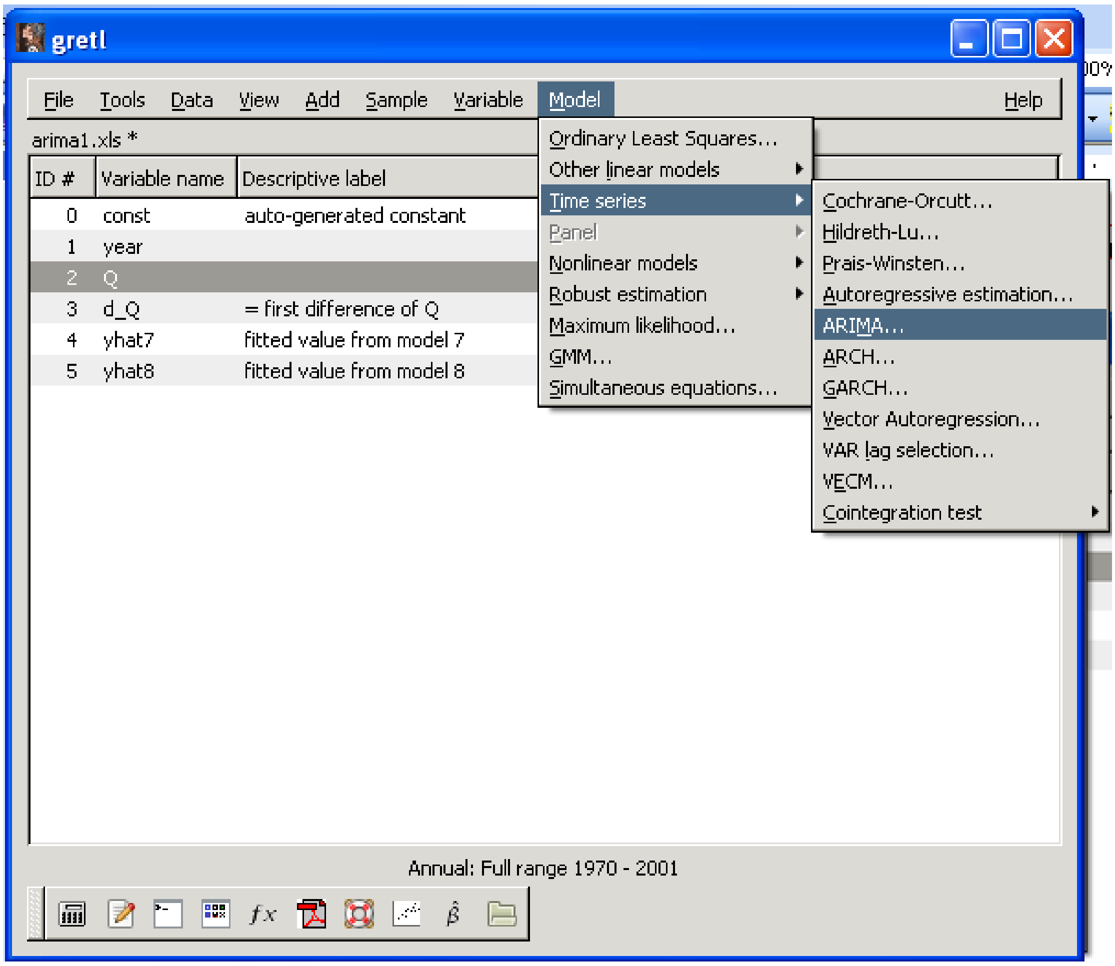
1969 yıllarına ait veri bulunmamaktadır. O halde 1971 için de kestirim yapılamaz. Şimdi sırasıyla takip eden yılları hesaplayalım:

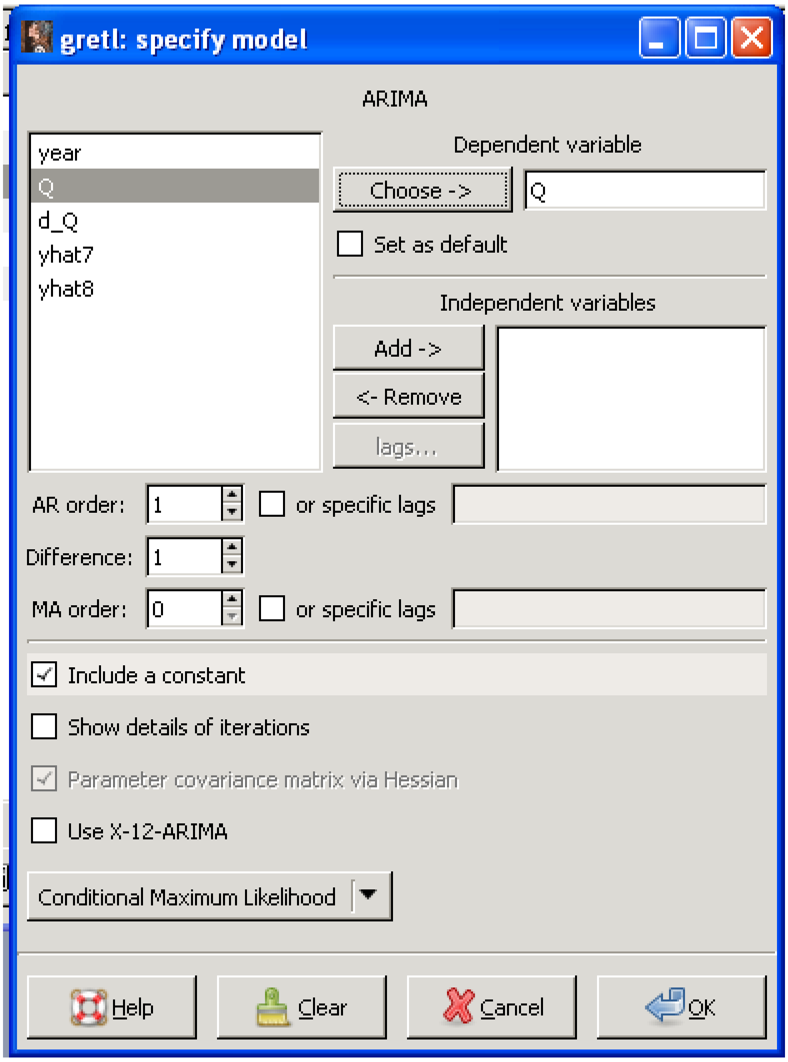
$\mathbf{Q_{1972} = 2.00677 + 0.59093\, Q_{1971} + 0.40907\, Q_{1970}}$		
$Q_{1972} = 2.00677 + 0.59093 *(86) + 0.40907 *(77)$	=	84.32517
$\mathbf{Q_{1973} = 2.00677 + 0.59093\, Q_{1972} + 0.40907\, Q_{1971}}$		
$Q_{1972} = 2.00677 + 0.59093 *(87) + 0.40907 *(86)$	=	88.59771
$\mathbf{Q_{1974} = 2.00677 + 0.59093\, Q_{1973} + 0.40907\, Q_{1972}}$	=	91.96144
$\mathbf{Q_{1975} = 2.00677 + 0.59093\, Q_{1974} + 0.40907\, Q_{1973}}$	=	89.27931
$\mathbf{Q_{1976} = 2.00677 + 0.59093\, Q_{1975} + 0.40907\, Q_{1974}}$	=	91.32517
$\mathbf{Q_{1977} = 2.00677 + 0.59093\, Q_{1976} + 0.40907\, Q_{1975}}$	=	94.41584
$\mathbf{Q_{1978} = 2.00677 + 0.59093\, Q_{1977} + 0.40907\, Q_{1976}}$	=	98.73424
$\mathbf{Q_{1979} = 2.00677 + 0.59093\, Q_{1978} + 0.40907\, Q_{1977}}$	=	103.1886
$\mathbf{Q_{1980} = 2.00677 + 0.59093\, Q_{1979} + 0.40907\, Q_{1978}}$	=	110.507
$\mathbf{Q_{1981} = 2.00677 + 0.59093\, Q_{1980} + 0.40907\, Q_{1979}}$	=	107.9156
$\mathbf{Q_{1982} = 2.00677 + 0.59093\, Q_{1981} + 0.40907\, Q_{1980}}$	=	112.4617
$\mathbf{Q_{1983} = 2.00677 + 0.59093\, Q_{1982} + 0.40907\, Q_{1981}}$	=	119.0068
$\mathbf{Q_{1984} = 2.00677 + 0.59093\, Q_{1983} + 0.40907\, Q_{1982}}$	=	101.8697
$\mathbf{Q_{1985} = 2.00677 + 0.59093\, Q_{1984} + 0.40907\, Q_{1983}}$	=	103.5982
$\mathbf{Q_{1986} = 2.00677 + 0.59093\, Q_{1985} + 0.40907\, Q_{1984}}$	=	117.1433
$\mathbf{Q_{1987} = 2.00677 + 0.59093\, Q_{1986} + 0.40907\, Q_{1985}}$	=	114.6884
$\mathbf{Q_{1988} = 2.00677 + 0.59093\, Q_{1987} + 0.40907\, Q_{1986}}$	=	110.4158
$\mathbf{Q_{1989} = 2.00677 + 0.59093\, Q_{1988} + 0.40907\, Q_{1987}}$	=	100.5519
$\mathbf{Q_{1990} = 2.00677 + 0.59093\, Q_{1989} + 0.40907\, Q_{1988}}$	=	102.8708
$\mathbf{Q_{1991} = 2.00677 + 0.59093\, Q_{1990} + 0.40907\, Q_{1989}}$	=	113.1433

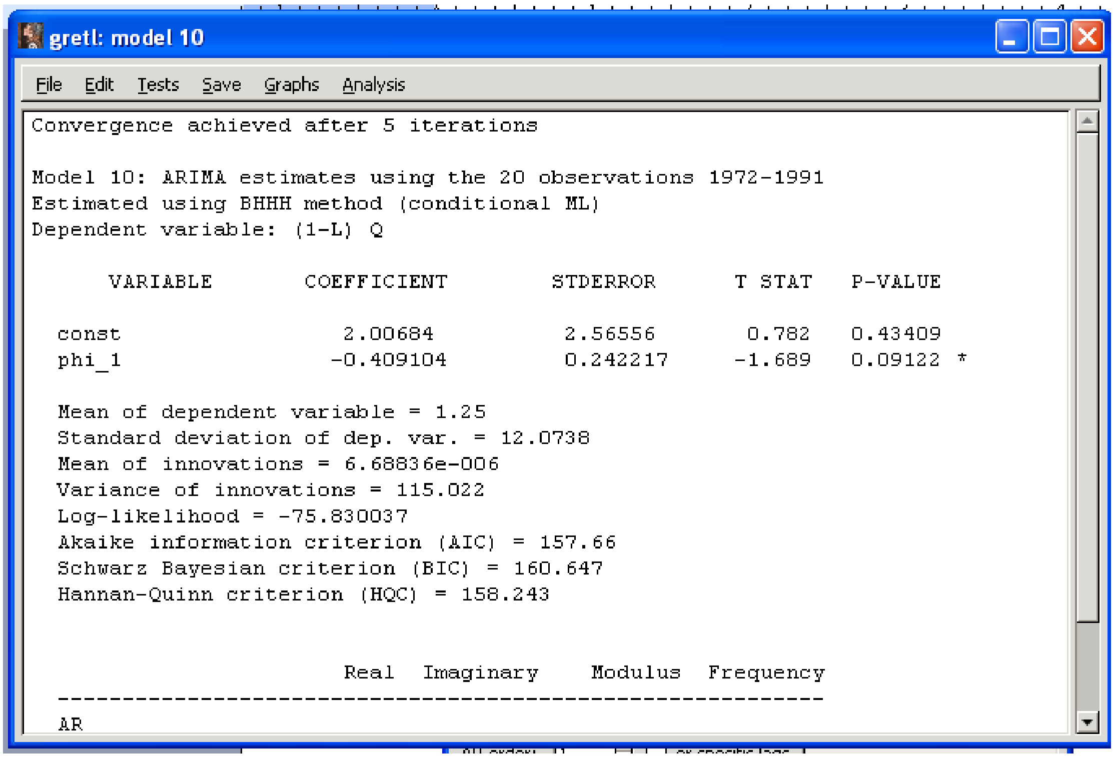
AR(1) ile 1991'den sonrasının kestirimi:

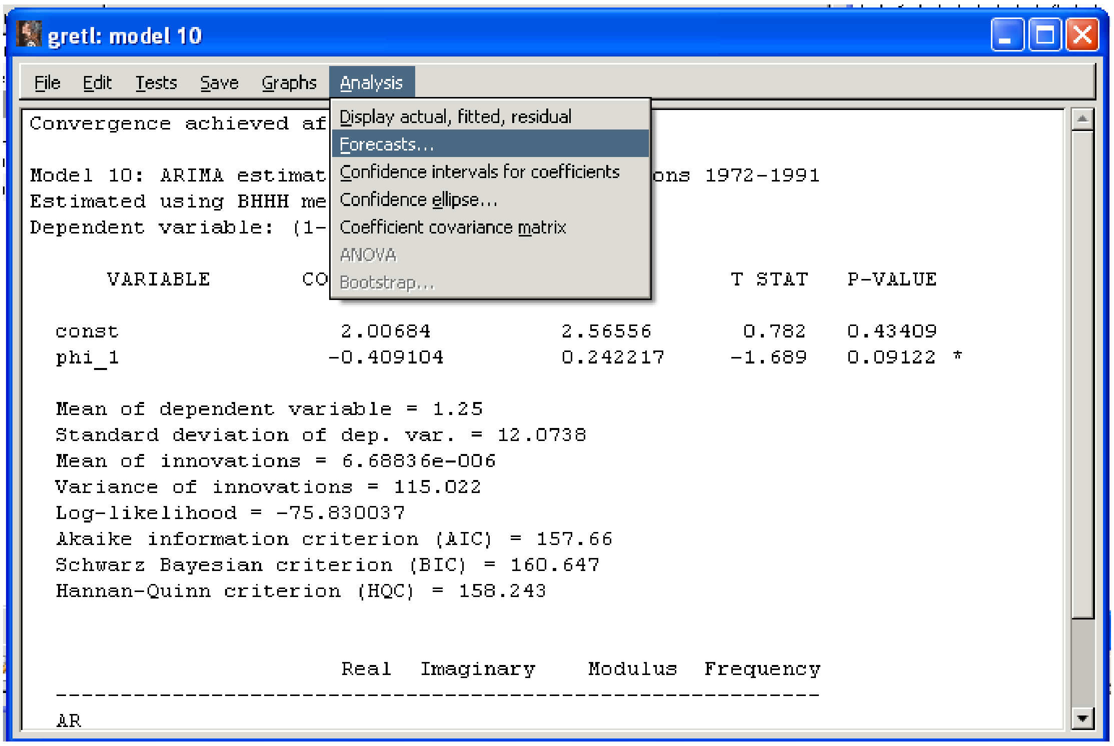
1992 ile 1996 yılları için kestirim yapmak istediğimizi varsayalım. Bunun için once gözlemlerimizi 1991'den 1996'ya kadar genişletmemiz gerekiyor:

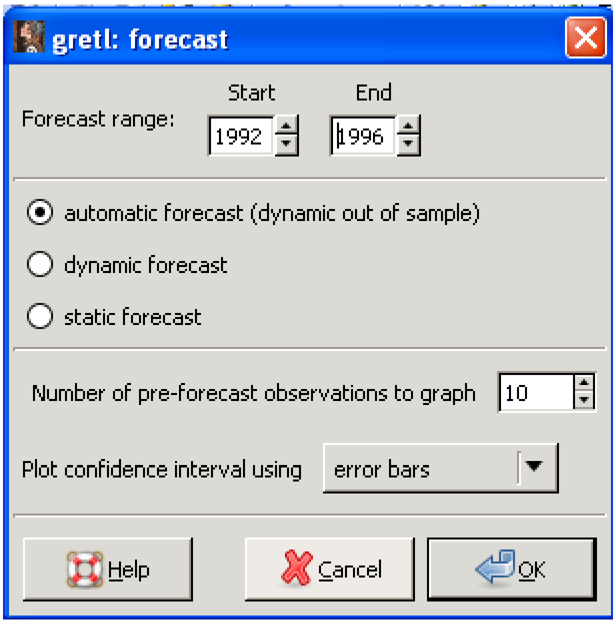


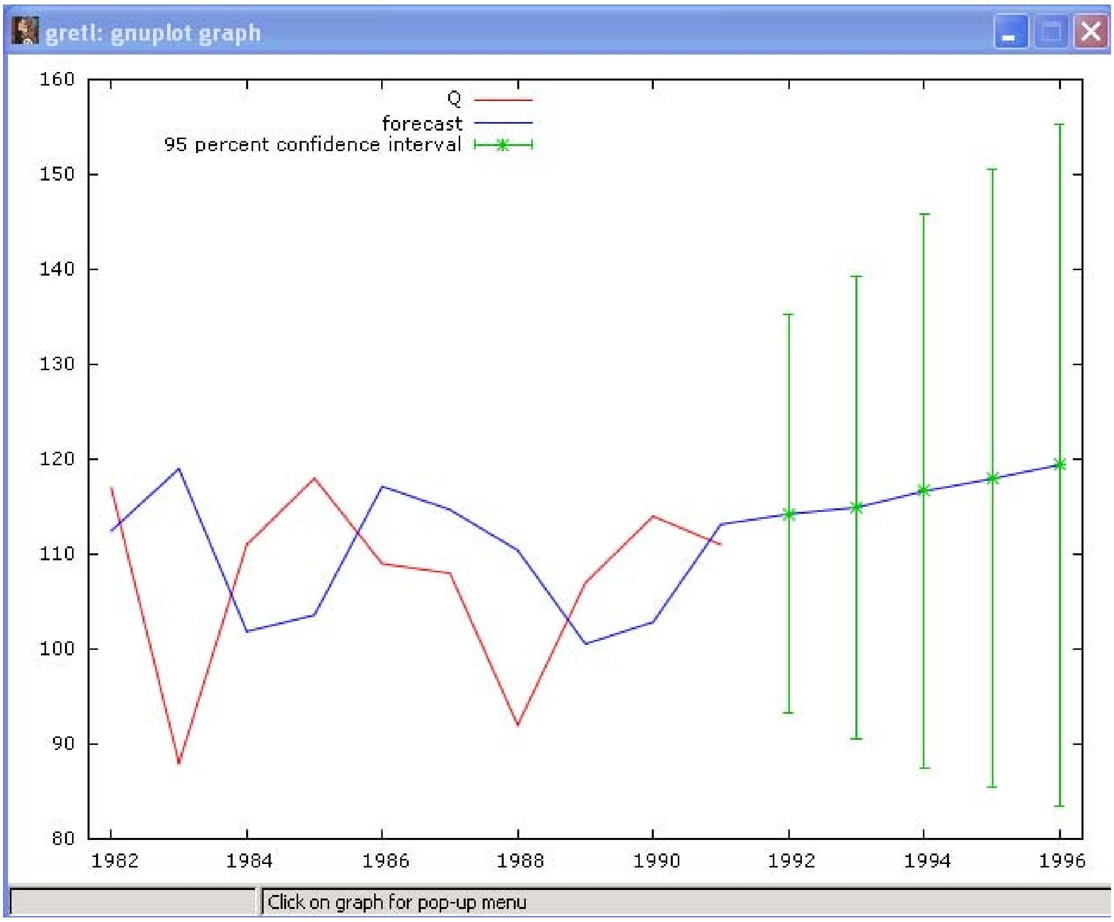

gretl
File Tools Data View Add Sample Variable Model Help
arima1.xls *
ID # Variable name Descriptive label
0 const auto-generated constant
1 year
2 Q
3 d_Q = first difference of Q
4 yhat7 fitted value from model 7
5 yhat8 fitted value from model 8
Ordinary Least Squares...
Other linear models
Time series
Panel
Nonlinear models
Robust estimation
Maximum likelihood...
GMM...
Simultaneous equations...
Cochrane-Orcutt...
Hildreth-Lu...
Prais-Winsten...
Autoregressive estimation...
ARIMA...
ARCH...
GARCH...
Vector Autoregression...
VAR lag selection...
VECM...
Cointegration test
Annual: Full range 1970 - 2001


gretl: specify model
ARIMA
year
Q
d_Q
yhat7
yhat8
Dependent variable
Choose ->
Q
Set as default
Independent variables
Add ->
<- Remove
lags...
AR order: 1
or specific lags
Difference: 1
MA order: 0
or specific lags
Include a constant
Show details of iterations
Parameter covariance matrix via Hessian
Use X-12-ARIMA
Conditional Maximum Likelihood
Help
Clear
Cancel
OK


gretl: model 10
File Edit Tests Save Graphs Analysis
Convergence achieved after 5 iterations
Model 10: ARIMA estimates using the 20 observations 1972-1991
Estimated using BHHH method (conditional ML)
Dependent variable: (1-L) Q
VARIABLE COEFFICIENT STDERROR T STAT P-VALUE
const 2.00684 2.56556 0.782 0.43409
phi_1 -0.409104 0.242217 -1.689 0.09122 *
Mean of dependent variable = 1.25
Standard deviation of dep. var. = 12.0738
Mean of innovations = 6.68836e-006
Variance of innovations = 115.022
Log-likelihood = -75.830037
Akaike information criterion (AIC) = 157.66
Schwarz Bayesian criterion (BIC) = 160.647
Hannan-Quinn criterion (HQC) = 158.243
Real Imaginary Modulus Frequency
AR


gretl: model 10
File Edit Tests Save Graphs Analysis
Display actual, fitted, residual
Forecasts...
Confidence intervals for coefficients
Confidence ellipse...
Coefficient covariance matrix
ANOVA
Bootstrap...
Convergence achieved af
Model 10: ARIMA estimat ons 1972-1991
Estimated using BHHH me
Dependent variable: (1-
VARIABLE CO T STAT P-VALUE
const 2.00684 2.56556 0.782 0.43409
phi_1 -0.409104 0.242217 -1.689 0.09122 *
Mean of dependent variable = 1.25
Standard deviation of dep. var. = 12.0738
Mean of innovations = 6.68836e-006
Variance of innovations = 115.022
Log-likelihood = -75.830037
Akaike information criterion (AIC) = 157.66
Schwarz Bayesian criterion (BIC) = 160.647
Hannan-Quinn criterion (HQC) = 158.243
Real Imaginary Modulus Frequency
AR


gretl: forecast
Start End
Forecast range: 1992 1996
automatic forecast (dynamic out of sample)
dynamic forecast
static forecast
Number of pre-forecast observations to graph 10
Plot confidence interval using error bars
Help Cancel OK



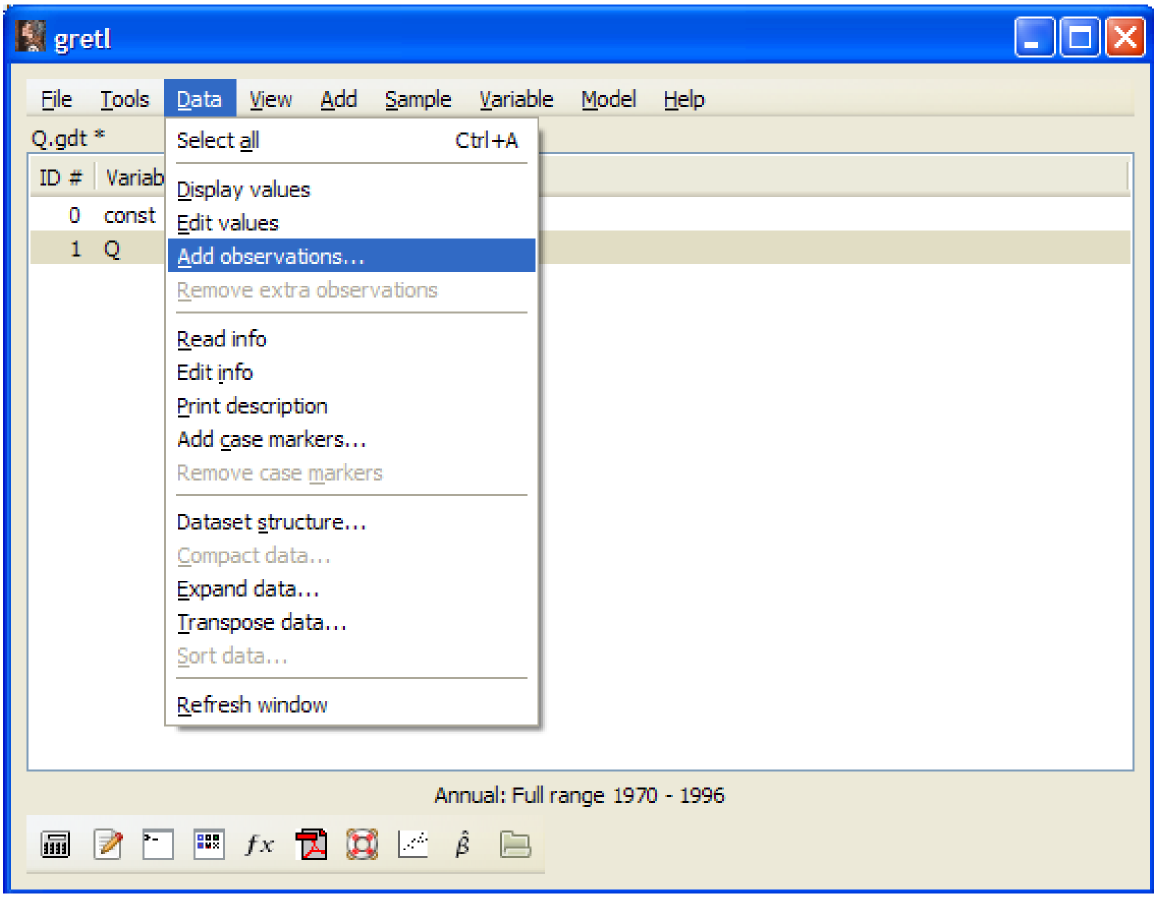
gretl: forecasts

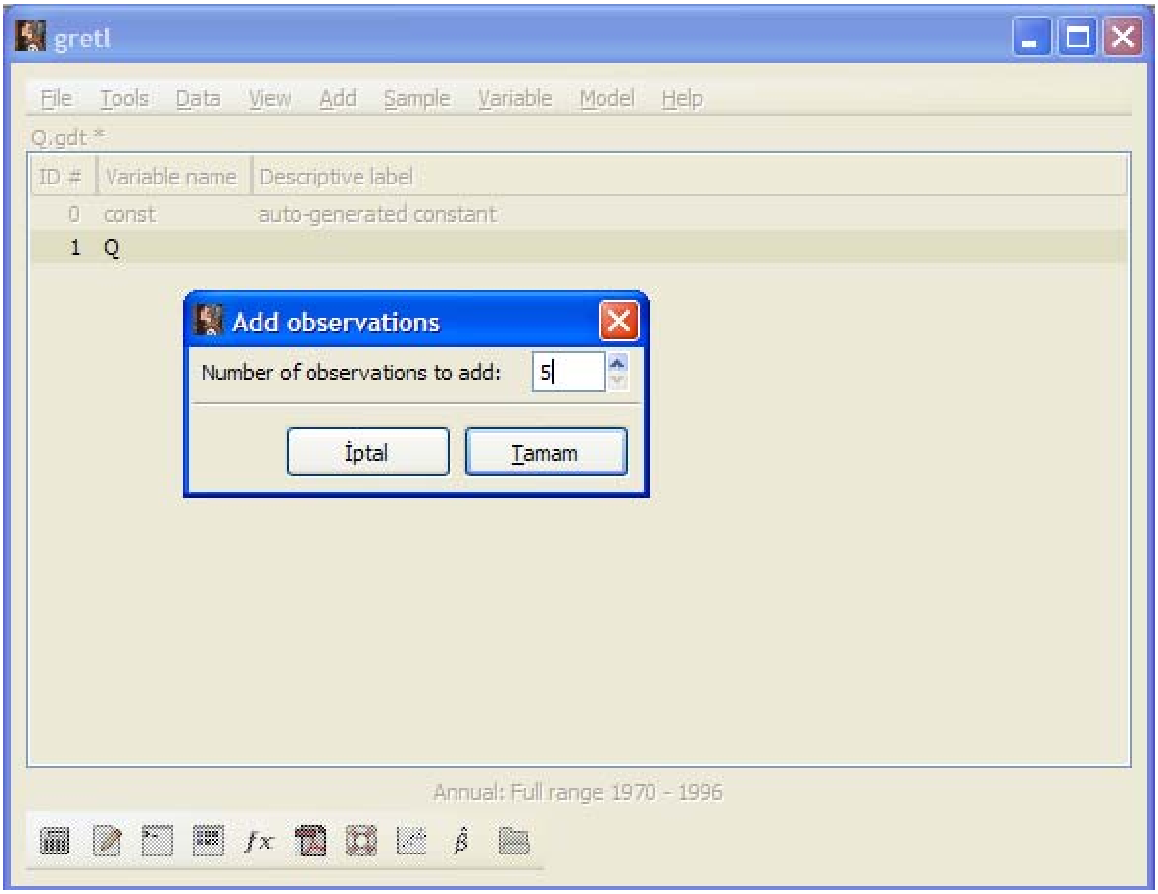
For 95% confidence intervals, z(.025) = 1.96

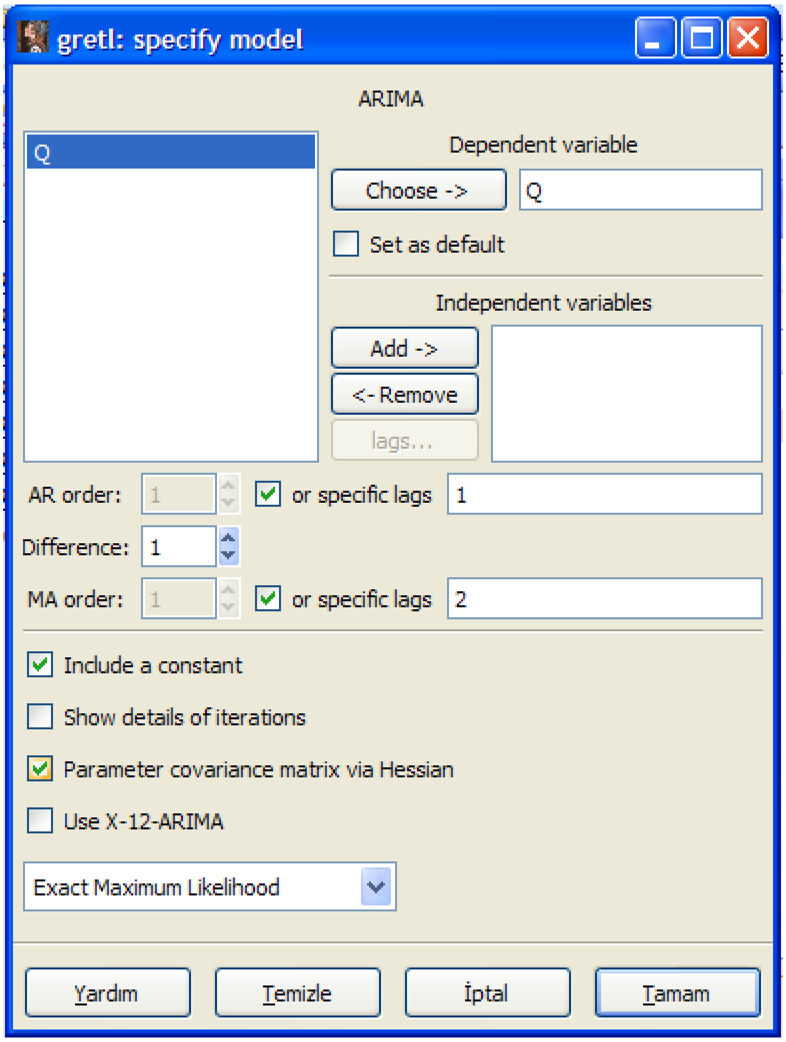
Obs	Q	prediction	std. error	95% confidence interval
1982	117.00	112.46		
1983	88.00	119.01		
1984	111.00	101.87		
1985	118.00	103.60		
1986	109.00	117.14		
1987	108.00	114.69		
1988	92.00	110.42		
1989	107.00	100.55		
1990	114.00	102.87		
1991	111.00	113.14		
1992		114.23	10.725	93.21 - 135.25
1993		114.92	12.457	90.50 - 139.33
1994		116.64	14.877	87.49 - 145.80
1995		117.95	16.615	85.38 - 150.51
1996		119.42	18.311	83.53 - 155.31

Şimdi AR(1) MA(2) modelinin öngörümleme için kullanımını gösterelim:

Önce kaç yıl ötesi için öngörüde bulunacağımıza karar vermeliyiz. Biz 5 yıl için için öngörüde bulunalım:



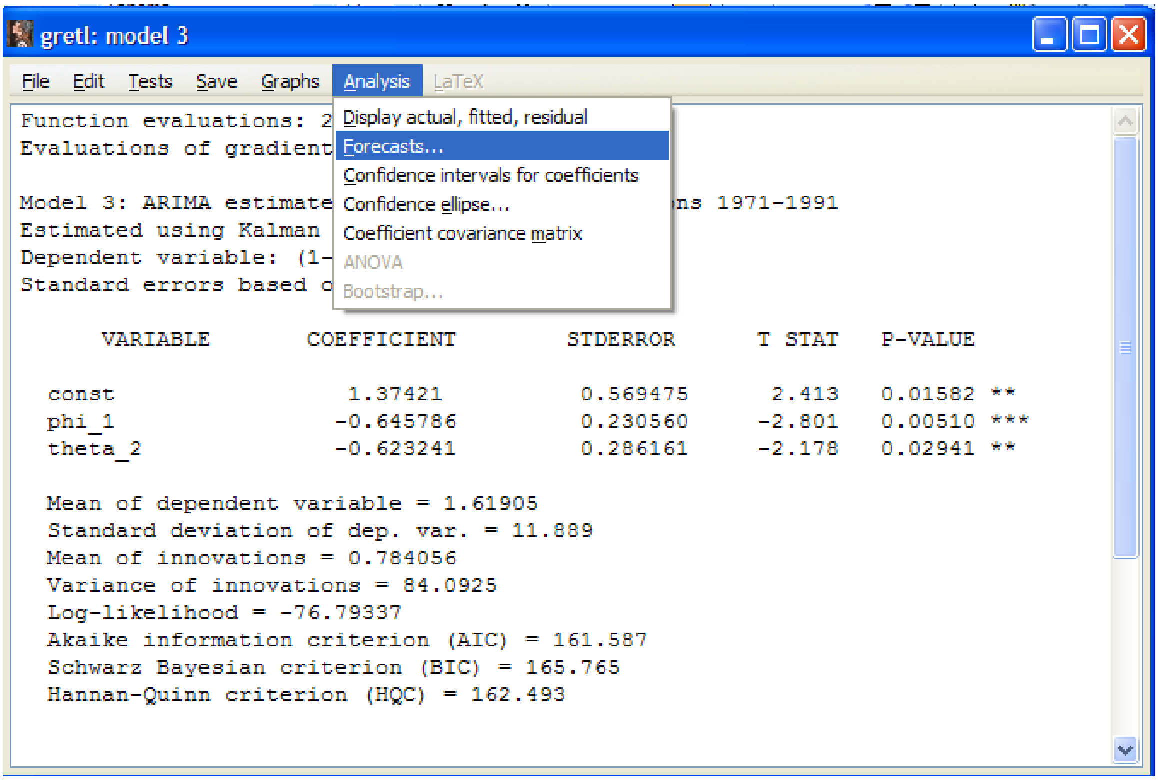

gretl
File Tools Data View Add Sample Variable Model Help
Q.gdt *
ID # Variable name Descriptive label
0 const auto-generated constant
1 Q
Add observations
Number of observations to add: 5
İptal
Tamam
Annual: Full range 1970 - 1996

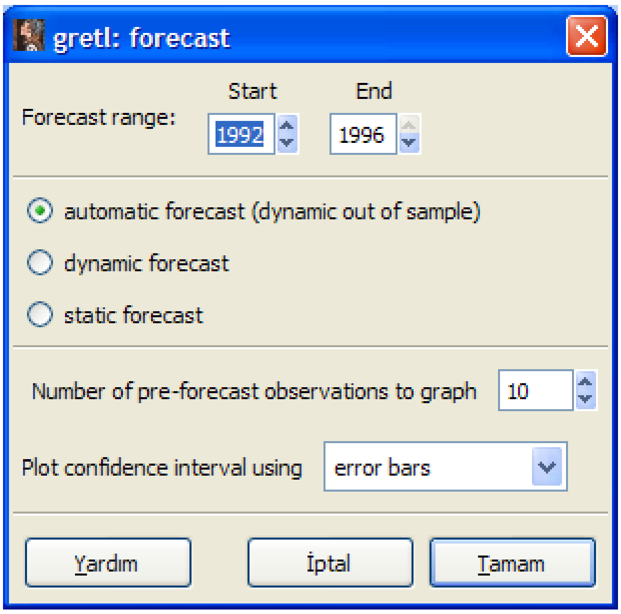


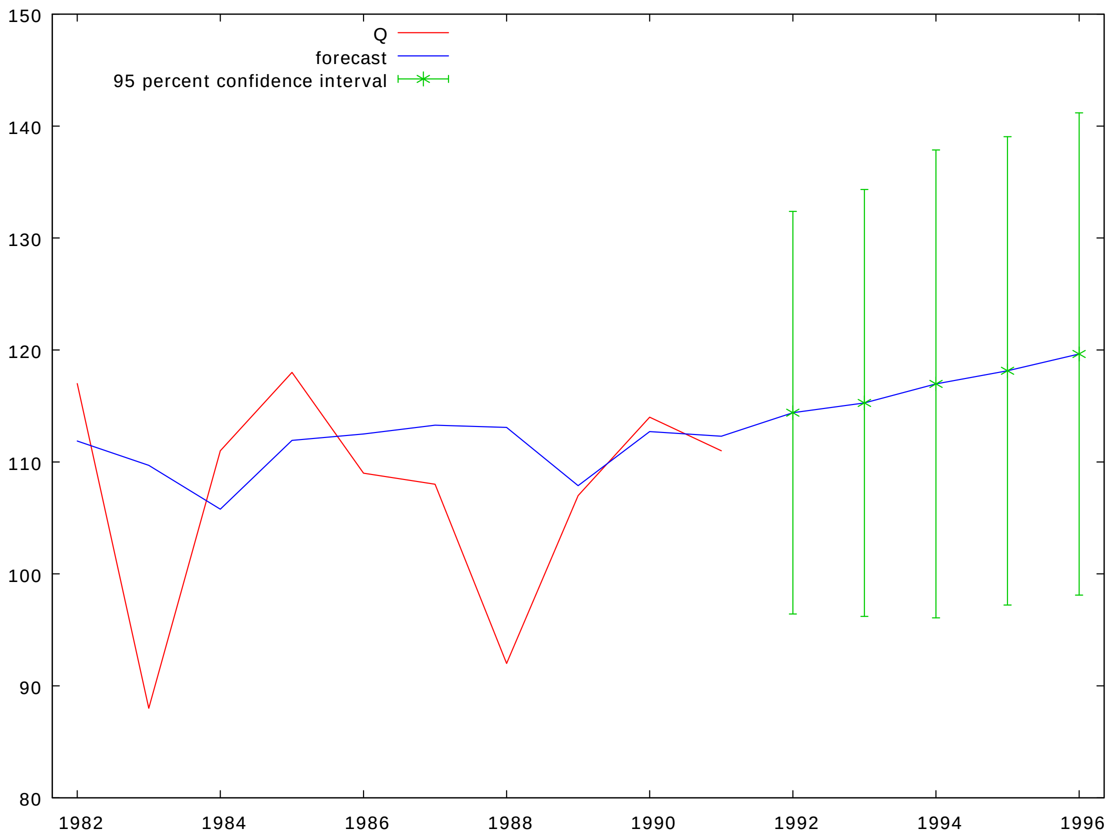
Model 2: ARIMA estimates using the 21 gözlem 1971-1991
Bağımlı değişken: (1-L) Q
Standard errors based on Hessian

	Katsayı	*St.Hata*	*t*	*p*	
Sabit	1.37421	0.56949	2.4130	0.01582	**
phi_1	-0.645786	0.230562	-2.8009	0.00510	***
theta_2	-0.623241	0.286163	-2.1779	0.02941	**

Bağımlı değişkenin ortalaması = 1.61905
Bağımlı değişkenin st sapması = 11.889
Mean of innovations = 0.784056
Variance of innovations = 84.0925
Log-likelihood = -76.7934
Akaike bilgi kriteri = 161.587
Schwarz Bayesian kriteri = 165.765
Hannan-Quinn kriteri = 162.493


gretl: model 3
File Edit Tests Save Graphs Analysis LaTeX
Display actual, fitted, residual
Forecasts...
Confidence intervals for coefficients
Confidence ellipse...
Coefficient covariance matrix
ANOVA
Bootstrap...
Function evaluations: 2
Evaluations of gradient
Model 3: ARIMA estimate
ns 1971-1991
Estimated using Kalman
Dependent variable: (1-
Standard errors based o
VARIABLE COEFFICIENT STDERROR T STAT P-VALUE
const 1.37421 0.569475 2.413 0.01582 **
phi_1 -0.645786 0.230560 -2.801 0.00510 ***
theta_2 -0.623241 0.286161 -2.178 0.02941 **
Mean of dependent variable = 1.61905
Standard deviation of dep. var. = 11.889
Mean of innovations = 0.784056
Variance of innovations = 84.0925
Log-likelihood = -76.79337
Akaike information criterion (AIC) = 161.587
Schwarz Bayesian criterion (BIC) = 165.765
Hannan-Quinn criterion (HQC) = 162.493


gretl: forecast
Forecast range:
Start
End
1992
1996
automatic forecast (dynamic out of sample)
dynamic forecast
static forecast
Number of pre-forecast observations to graph
10
Plot confidence interval using
error bars
Yardım
İptal
Tamam



11.6. Vektör Otoregresyon (VAR)

VAR modelleri yaygın olarak makro ekonomide kullanılmaktadır. Kısıtsız VAR modellerinin, çok denklemli yapısal modellerden daha iyi tahminleme sağladığı belirtilmektedir (Greene,1998). Ayrıca, dinamik eşanlı denklem sistemlerinde, değişkenlerin endojen ve egzojen olarak ayrımlanmasındaki güçlüğün, VAR yaklaşımı ile aşılabildiği de ifade edilmektedir (Maddala, 1992).

Ekonometrik modellerde genellikle, bağımsız değişkenlerle bağımlı değişken arasındaki ilişkinin tek yönlü olduğu kabul edilir. Buna göre, bağımsız değişkenler dışsaldır ve bağımsız değişkenler bağımlı değişken tarafından belirlenmez. Ancak, eğer bir fonksiyonda iki yönlü açıklayıcılık varsa tek denklemli bir modelden sözedilemez. Değişkenlerin tümü için ayrı birer denklemin tanımlandığı bir sistemin belirlenmesi gerekir. Böyle bir sistemin eşanlı denklemler sistemi olarak adlandırıldığı bilinmektedir (Koutsoyiannis, 1989).

Eşanlı denklem sistemlerinde bazı değişkenler içsel (endogenous), bazıları da dışsal (exogenous) veya önceden belirlenmiş (predetermined) değişkenler olarak ele alınır. Bu modeller tahmin edilmeden önce sistemdeki denklemlerin eksiksiz mi yoksa fazladan mı tanımlandıkları bilinmek zorundadır. Ancak tanımlama, önceden belirlenmiş değişkenlerden bazılarının sadece denklemlerin bir kısmında bulunduğu varsayılarak yapılır. Bu karar, çoğu zaman öznel olması nedeniyle eleştirilmektedir (Gujarati, 1995).

Eğer değişkenler arasında gerçek bir eşanlılık varsa, tümüne eşit muamele yapılması gerekir. İçsel ve dışsal değişkenler arasında herhangi bir üstünlük ayrımı yapılmamalıdır. Bu amaçla Sims tarafından VAR (Vector Autoregression) geliştirilmiştir (Gujarati, 1995).

VAR, teoriye uygun olmadığı yönünde eleştirler almış olmasına rağmen (Cooley, Leroy,1985; Darnel, Evans, 1990; Robertson, Orden, 1990), tarım ve makro ekonomi ilişkilerini inceleyen birçok araştırmada bu yöntem kullanılmıştır (Bessler, 1984; Orden, 1986; Orden, Fackler, 1989; Miran, Güler, 1996; Bayaner, 1998).

Bir VAR modelinde; birbiriyle ilişkisi olduğu düşünülen değişkenlerin ve en fazla kaç gecikme alınacağının belirlenmesi yeterlidir (Pindyck, Rubinfeld, 1991). VAR yönteminin üç özelliği üzerinde durulmaktadır:

- Yöntem basittir; hangi değişkenin içsel hangi değişkenin dışsal olduğunu belirlemek gerekmez. Tüm değişkenler içseldir.
- Tahmin edilmesi kolaydır: En küçük kareler yöntemi herbir denkleme ayrı ayrı uygulanır.
- Bu yöntemle yapılan tahminler, çok daha kapsamlı eşanlı modellerin tahminlerinden daha iyidir.

VAR yaklaşımı, pek çok konuda tahminleme ve politika analizi amaçlarıyla kullanılmaktadır (Darnell, Evans, 1990). Bunun yanında VAR, etki-tepki fonksiyonlarının (impulse-response function = IRF) belirlenmesinde başarıyla uygulanmaktadır. IRF, VAR sistemindeki bağımlı değişkenlerin, denklemlerin hata terimlerinde meydana gelen şoklara tepkisini ölçer. Bir başka anlatımla IRF, içsel değişkenlerden herbirinin, diğer içsel değişkenlerdeki şok değişmelere zaman içinde nasıl bir duyarlılık göstereceğini belirler. Bu amaçla şok değişme olarak hata terimlerinde 1 standart sapmalık değişmeler esas alınır.

VAR modelinin genel ifadesi (Denklem 11-1'de verilmiştir:

(Denklem 11-1)

$$y_t = \mu + \Delta_1 y_{t-1} + \ldots + \Delta_p y_{t-p} + v_t$$

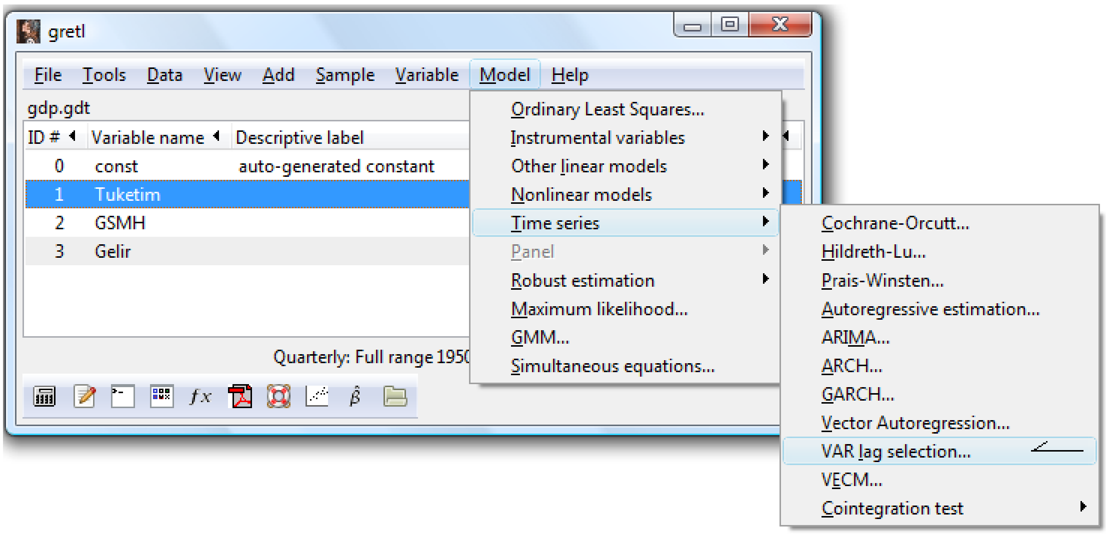
VAR Modelindeki denklemlerin bireysel ifadeleri ise, (Denklem 11-2'de sunulmuştur.

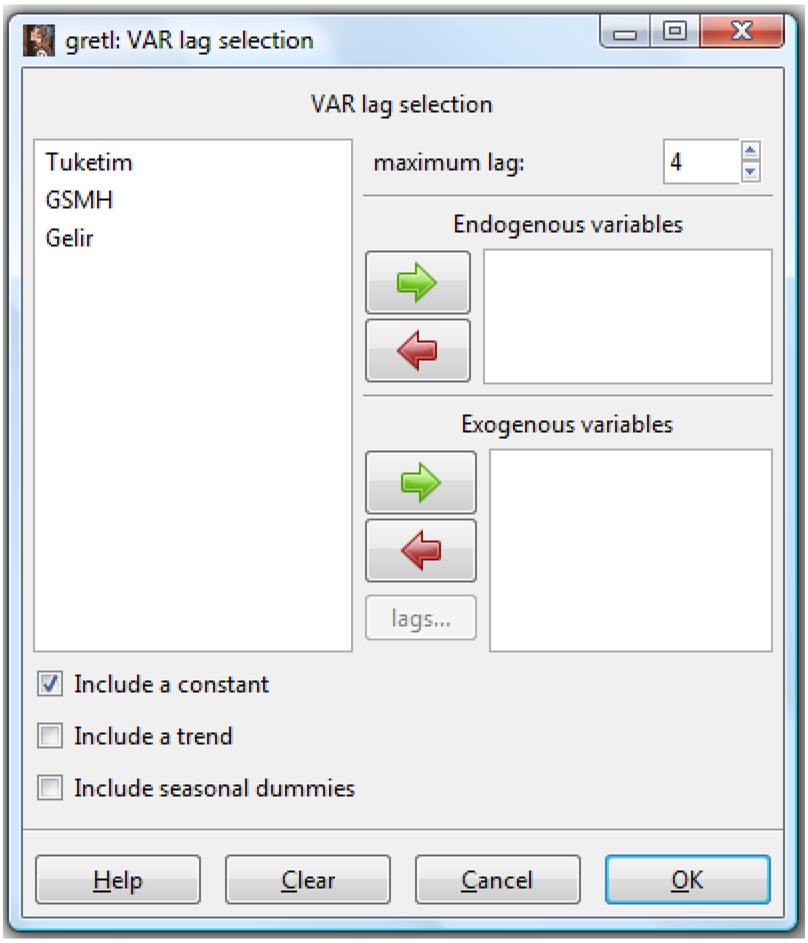
(Denklem 11-2)

$$y_{mt} = \mu_m + \sum_{j=1}^{p} (\Delta_j)_{m1} y_{1,t-j} + \sum_{j=1}^{p} (\Delta_j)_{m2} y_{2,t-j} + .. + \sum_{j=1}^{p} (\Delta_j)_{mM} y_{M,t-j} + \in_{mt}$$

VAR uygulamasını tüketim harcamaları (tuketim), GSMH ve kişi başına gelir (gelir) değişkenleri için uygulayacağız (**GDP.gdt**)

Bunun için önce en uygun gecikme uzunluğunu belirlememiz gerekiyor. Gretl bu konuda bize yardımcı olacaktır:





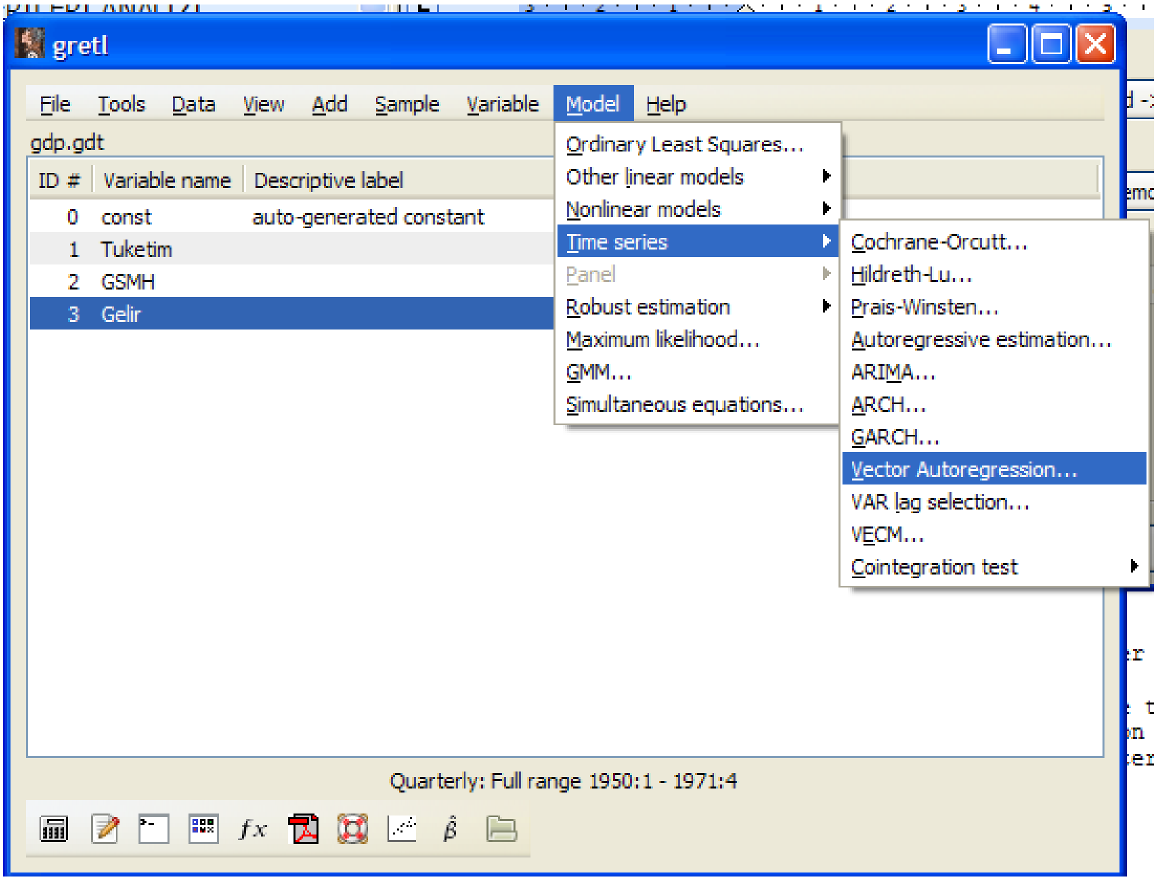
VAR system, maximum lag order 4

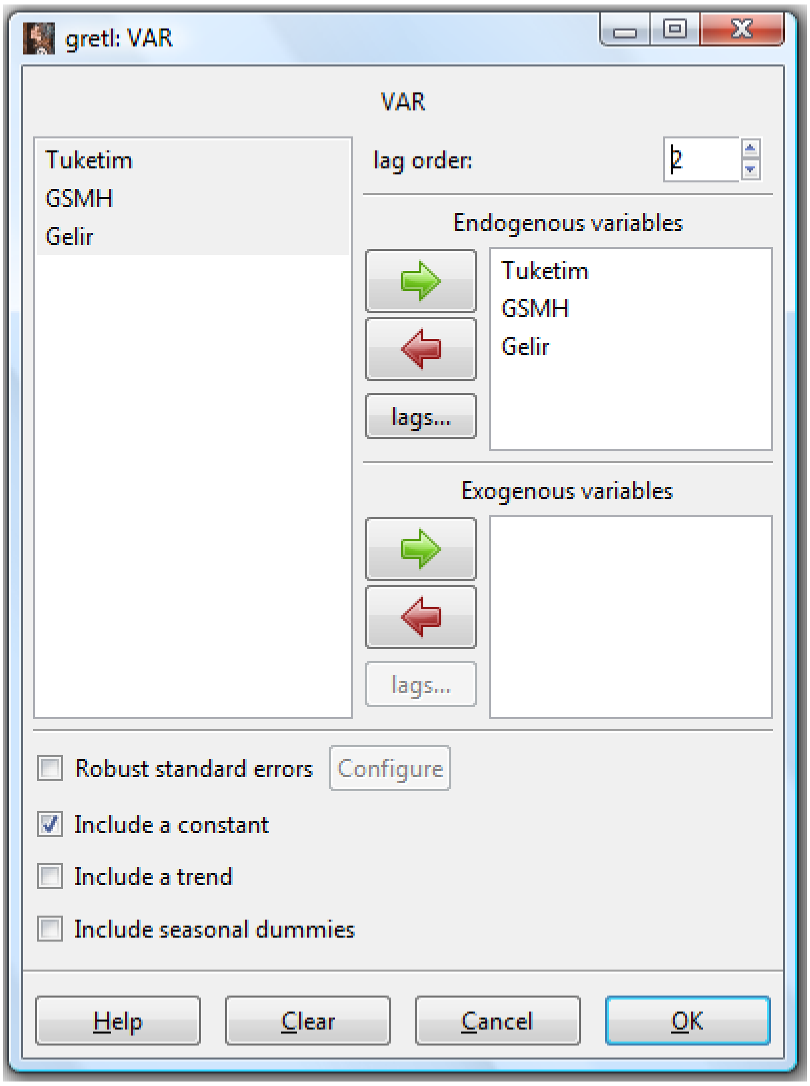
The asterisks below indicate the best (that is, minimized) values
of the respective bilgi criteria, AIC = Akaike criterion,
BIC = Schwartz Bayesian kriteri and HQC = Hannan-Quinn criterion.

lags	loglik	p(LR)	AIC	BIC	HQC
1	-1199.70949		28.850226	29.197485*	28.989821

2 -1184.19615 0.00029 28.695146* 29.302851 28.939438*
3 -1178.59385 0.26194 28.776044 29.644193 29.125032
4 -1172.18431 0.17097 28.837722 29.966315 29.291407

Görüldüğü gibi AIC ve HQC kriterlerine göre en uygun gecikme uzunluğu 2'dir.





VAR system, lag order 2
EKK estimates, gözlem 1950:3-1971:4 (T = 86)
Log-likelihood = -1212.1895
Determinant of covariance matrix = 3.5118485e+008
AIC = 28.6788
BIC = 29.2781
HQC = 28.9200
Portmanteau test: LB(21) = 157.587 (df = 171, p 0.760790)

Equation 1: Tuketim

	Katsayı	*St.Hata*	*t*	*p*	
sabit	156.163	100.496	1.5539	0.12420	
Tuketim_1	0.238142	0.116439	2.0452	0.04416	**
Tuketim_2	0.119559	0.117435	1.0181	0.31174	
GSMH_1	-0.356925	0.284661	-1.2539	0.21359	
GSMH_2	0.314342	0.272176	1.1549	0.25160	
Gelir_1	1.34272	0.52872	2.5396	0.01306	**
Gelir_2	-0.622416	0.604715	-1.0293	0.30649	

Bağımlı değişkenin ortalaması = 2823.97
Bağımlı değişkenin st sapması = 474.65
Hata kareleri toplamı = 308268
Hataların st hatası= 62.467
R^2 = 0.98390
F-istatistiği (6, 79) = 804.758 (p < 0.00001)
Durbin-Watson istatistiği = 1.95929
First-order autocorrelation coeff. = 0.0186302

F-tests of zero restrictions:
All lags of Tuketim F(2, 79) = 3.3973, p 0.0384
All lags of GSMH F(2, 79) = 0.83701, p 0.4368
All lags of Gelir F(2, 79) = 5.3219, p 0.0068
All vars, lag 2 F(3, 79) = 0.88956, p 0.4503

Equation 2: GSMH

	Katsayı	*St.Hata*	*t*	*p*	
sabit	254.175	44.6859	5.6880	<0.00001	***
Tuketim_1	-0.0965999	0.051775	-1.8658	0.06579	*
Tuketim_2	-0.108238	0.0522179	-2.0728	0.04145	**
GSMH_1	0.704323	0.126576	5.5644	<0.00001	***
GSMH_2	-0.178629	0.121024	-1.4760	0.14392	
Gelir_1	0.876588	0.235098	3.7286	0.00036	***
Gelir_2	-0.0209499	0.268889	-0.0779	0.93809	

Bağımlı değişkenin ortalaması = 3888.72
Bağımlı değişkenin st sapması = 618.081
Hata kareleri toplamı = 60950
Hataların st hatası= 27.7762
R^2 = 0.99812
F-istatistiği (6, 79) = 7001.57 (p < 0.00001)
Durbin-Watson istatistiği = 2.08729
First-order autocorrelation coeff. = -0.0455121

F-tests of zero restrictions:
All lags of Tuketim F(2, 79) = 5.2797, p 0.0070
All lags of GSMH F(2, 79) = 24.043, p 0.0000
All lags of Gelir F(2, 79) = 22.958, p 0.0000
All vars, lag 2 F(3, 79) = 4.3944, p 0.0065

Equation 3: Gelir

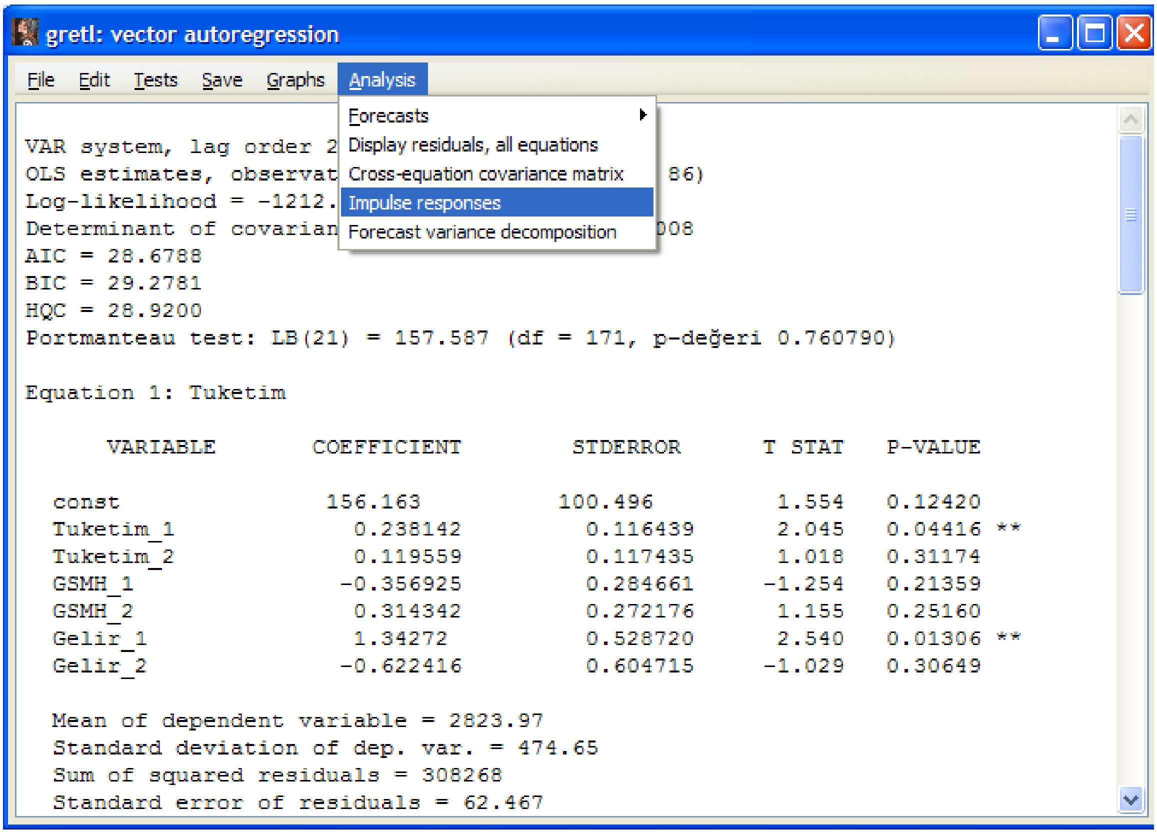
	Katsayı	*St.Hata*	*t*	*p*	
sabit	138.515	26.0431	5.3187	<0.00001	***
Tuketim_1	0.0142053	0.0301746	0.4708	0.63910	
Tuketim_2	-0.0646349	0.0304327	-2.1239	0.03681	**
GSMH_1	-0.114172	0.0737686	-1.5477	0.12569	
GSMH_2	-0.143144	0.0705332	-2.0295	0.04578	**
Gelir_1	1.05176	0.137015	7.6762	<0.00001	***
Gelir_2	0.348727	0.156709	2.2253	0.02891	**

Bağımlı değişkenin ortalaması = 2554.11
Bağımlı değişkenin st sapması = 454.531
Hata kareleri toplamı = 20702.2
Hataların st hatası= 16.1881
R^2 = 0.99882
F-istatistiği (6, 79) = 11155.6 (p < 0.00001)
Durbin-Watson istatistiği = 2.10834
First-order autocorrelation coeff. = -0.0645239

F-tests of zero restrictions:

All lags of Tuketim	F(2, 79) =	2.2598, p 0.1111
All lags of GSMH	F(2, 79) =	13.029, p 0.0000
All lags of Gelir	F(2, 79) =	158.45, p 0.0000
All vars, lag 2	F(3, 79) =	3.2656, p 0.0256

For the system as a whole
Null hypothesis: the longest lag is 1
Alternative hypothesis: the longest lag is 2
Likelihood ratio test: Chi-square(9) = 29.8783 (p 0.000460)



Responses to a one-standard error shock in Tuketim
(Tüketimdeki 1 st hatalık şoka tepkiler)

period Tuketim GSMH Gelir

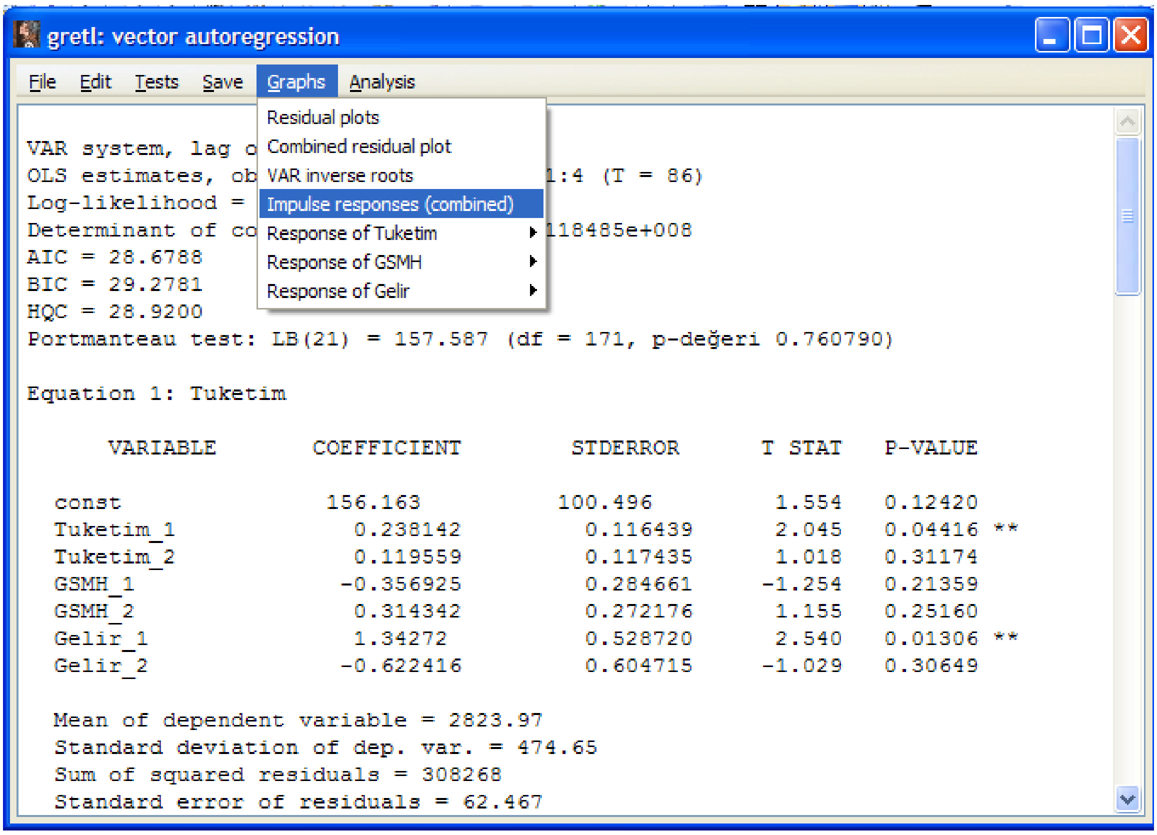
1	59.871	4.136	5.9132
2	20.721	2.313	6.5976
3	17.745	-1.9322	4.5696
4	10.149	-1.8636	5.9092
5	9.6866	1.2156	7.2952
6	8.618	5.4259	9.3431
7	9.6604	9.7607	11.074
8	10.606	13.551	12.594
9	11.93	16.538	13.69
10	13.008	18.664	14.446
11	13.938	20.02	14.883
12	14.589	20.756	15.092
13	15.025	21.042	15.134
14	15.264	21.033	15.077
15	15.364	20.854	14.968
16	15.365	20.599	14.84
17	15.309	20.331	14.716
18	15.224	20.088	14.608
19	15.132	19.885	14.519
20	15.043	19.729	14.45

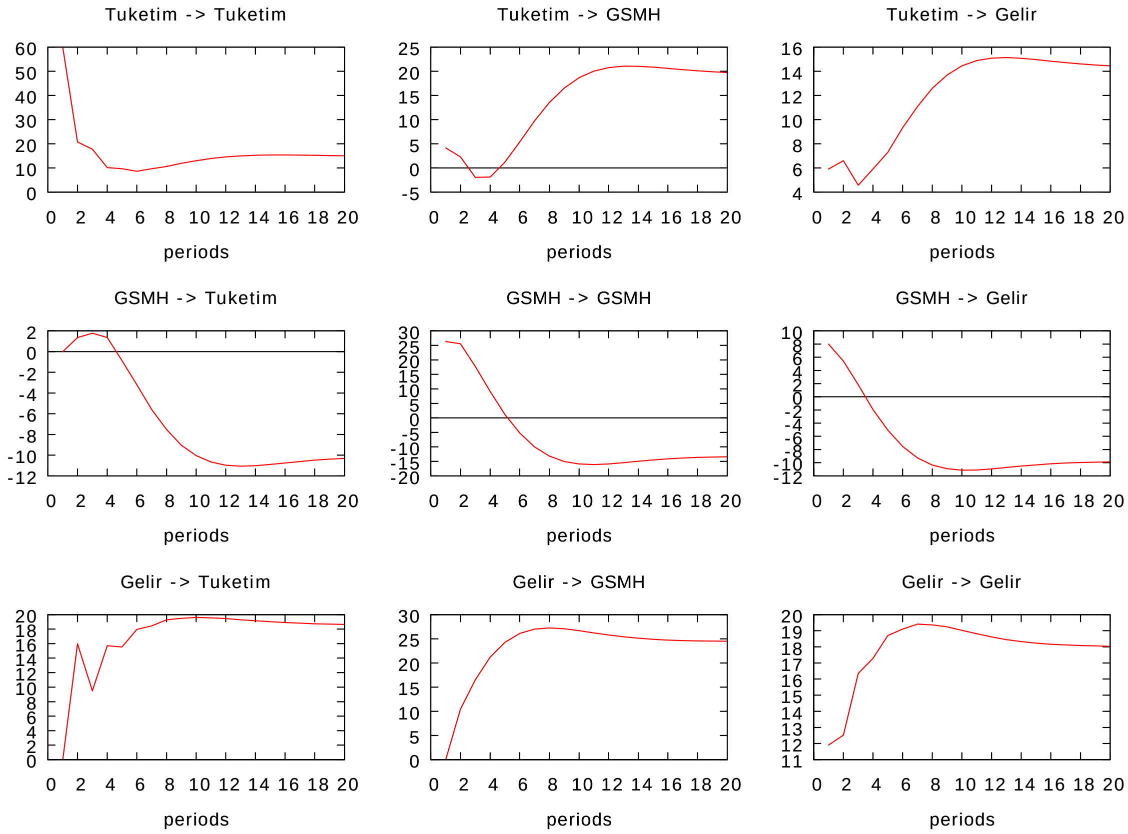
Responses to a one-standard error shock in GSMH
(GSMH'deki 1 st hatalık şoka tepkiler)

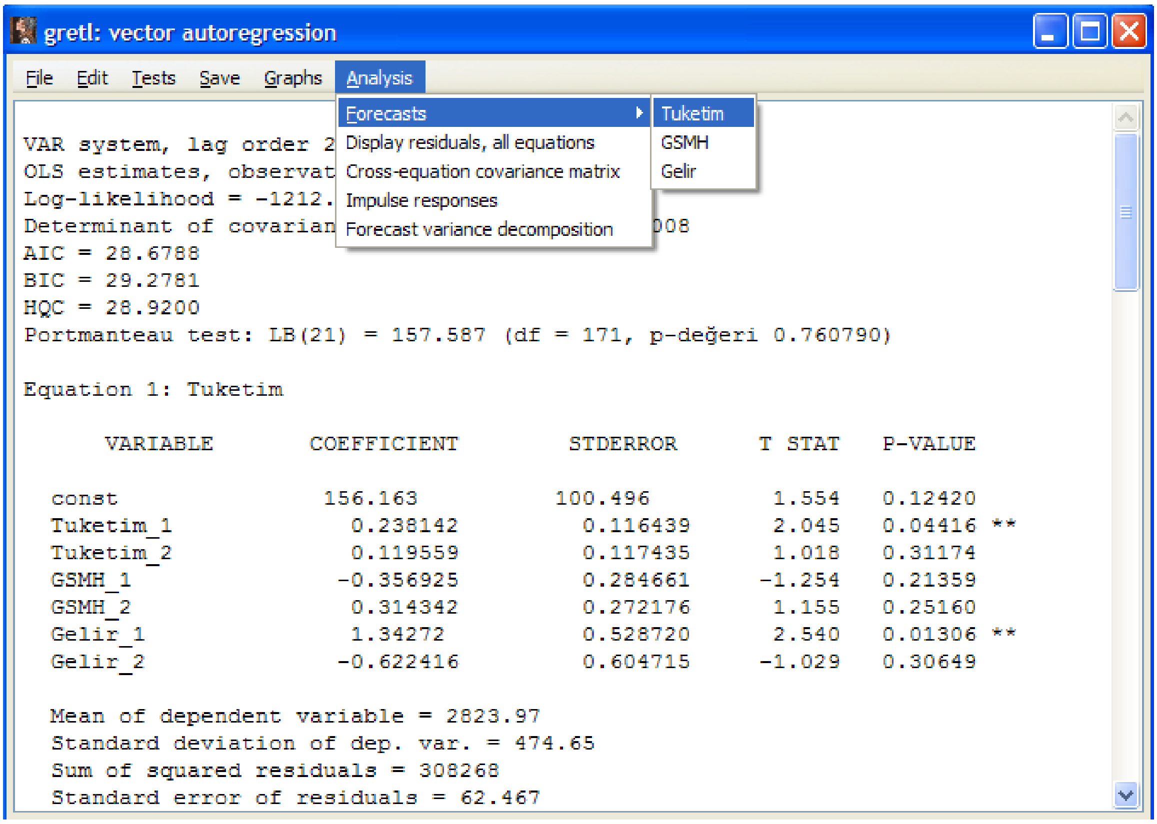
period	Tuketim	GSMH	Gelir
1	0	26.299	8.0062
2	1.3635	25.541	5.418
3	1.7671	17.741	1.8293
4	1.3641	9.105	-1.9311
5	-0.86845	1.1896	-5.0671
6	-3.2079	-5.2536	-7.5424
7	-5.5923	-10.014	-9.2598
8	-7.5311	-13.186	-10.346
9	-9.0319	-15.041	-10.917
10	-10.047	-15.904	-11.127
11	-10.669	-16.092	-11.1
12	-10.977	-15.872	-10.944
13	-11.068	-15.45	-10.732
14	-11.021	-14.967	-10.516
15	-10.902	-14.512	-10.323
16	-10.755	-14.131	-10.168
17	-10.61	-13.838	-10.051
18	-10.483	-13.631	-9.9704
19	-10.381	-13.497	-9.9177
20	-10.304	-13.419	-9.8858

Responses to a one-standard error shock in Gelir
(Gelirdeki 1 st hatalık şoka tepkiler)

period	Tuketim	GSMH	Gelir
1	0	0	11.902
2	15.981	10.433	12.518
3	9.4822	16.528	16.352
4	15.714	21.204	17.286
5	15.535	24.247	18.706
6	17.948	26.124	19.104
7	18.437	27.008	19.414
8	19.285	27.25	19.359
9	19.471	27.073	19.237
10	19.625	26.689	19.022
11	19.554	26.23	18.813
12	19.451	25.786	18.614
13	19.296	25.404	18.452
14	19.149	25.102	18.324
15	19.012	24.879	18.229
16	18.9	24.726	18.162
17	18.809	24.626	18.114
18	18.74	24.563	18.08
19	18.687	24.525	18.054
20	18.647	24.499	18.032





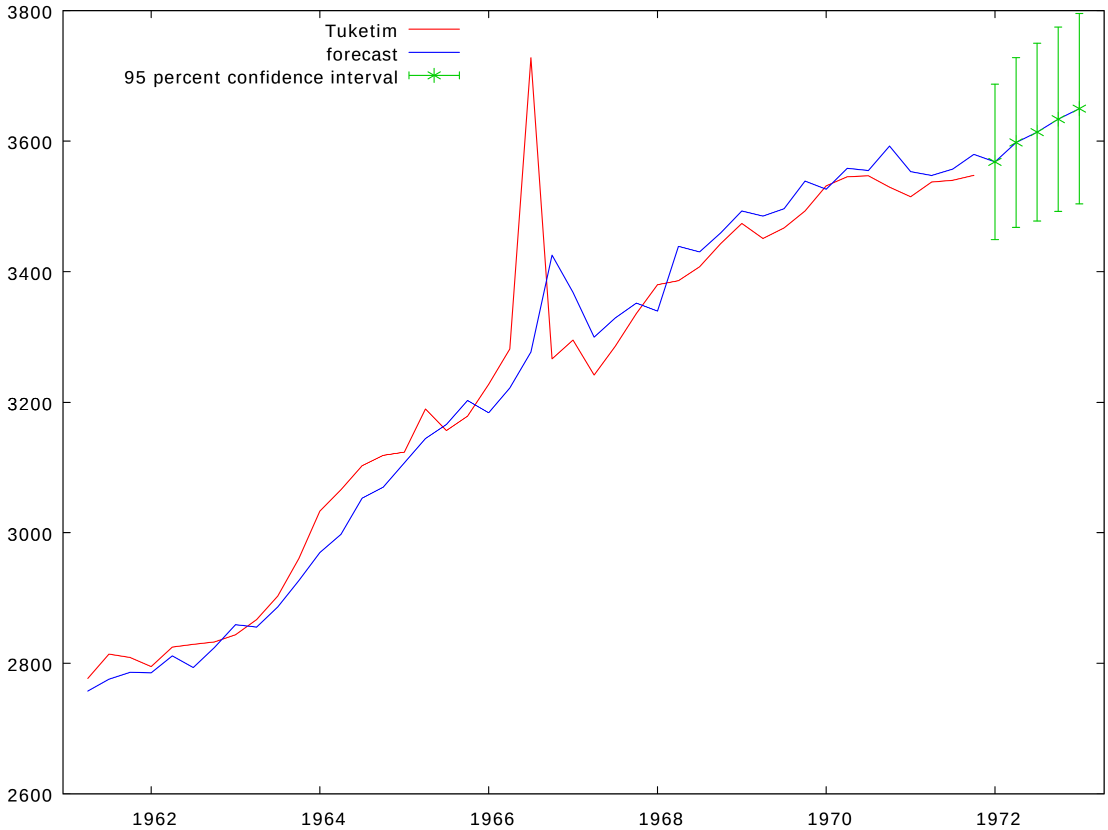



```
VAR system, lag order 2
OLS estimates, observat
Log-likelihood = -1212.
Determinant of covarian                              008
AIC = 28.6788
BIC = 29.2781
HQC = 28.9200
Portmanteau test: LB(21) = 157.587 (df = 171, p-değeri 0.760790)

Equation 1: Tuketim

      VARIABLE        COEFFICIENT        STDERROR      T STAT    P-VALUE

  const                156.163         100.496         1.554    0.12420
  Tuketim_1              0.238142        0.116439      2.045    0.04416 **
  Tuketim_2              0.119559        0.117435      1.018    0.31174
  GSMH_1                -0.356925        0.284661     -1.254    0.21359
  GSMH_2                 0.314342        0.272176      1.155    0.25160
  Gelir_1                1.34272         0.528720      2.540    0.01306 **
  Gelir_2               -0.622416        0.604715     -1.029    0.30649

  Mean of dependent variable = 2823.97
  Standard deviation of dep. var. = 474.65
  Sum of squared residuals = 308268
  Standard error of residuals = 62.467
```

95% güven aralıkları, t(79, .025) = 1.990

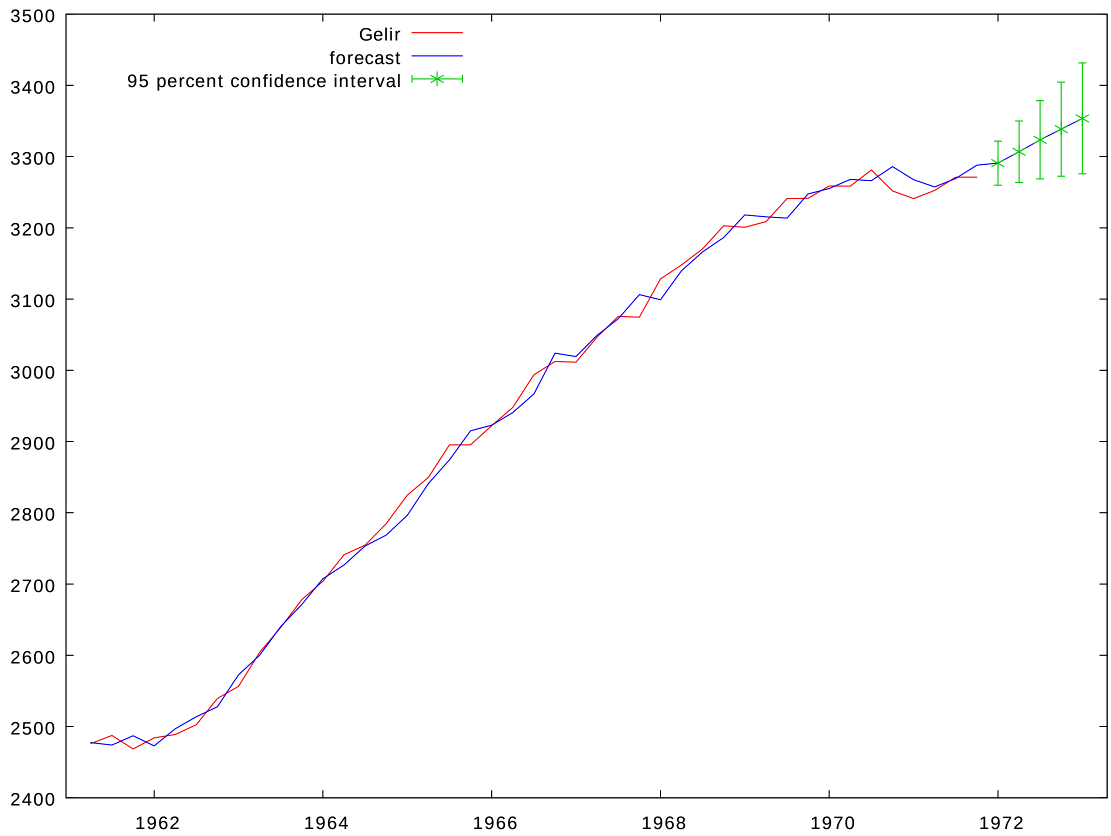
Obs	*Tuketim*	*prediction*	*St.Hata*	*95% confidence interval*

1972:1	undefined	3568.22	59.8708	(3449.05, 3687.39)
1972:2	undefined	3597.97	65.3539	(3467.89, 3728.06)
1972:3	undefined	3613.80	68.4037	(3477.64, 3749.95)
1972:4	undefined	3633.63	70.9287	(3492.45, 3774.81)
1973:1	undefined	3649.70	73.2585	(3503.89, 3795.52)



95% güven aralıkları, t(79, .025) = 1.990

Obs	*Gelir*	*prediction*	*St.Hata*	*95% confidence interval*
1972:1	undefined	3290.86	15.5153	(3259.98, 3321.75)
1972:2	undefined	3306.83	21.6866	(3263.66, 3350.00)
1972:3	undefined	3323.55	27.6033	(3268.61, 3378.50)
1972:4	undefined	3338.53	33.1569	(3272.54, 3404.53)
1973:1	undefined	3353.59	39.0922	(3275.78, 3431.41)



11.7. Eşbütünleşme ve Hata Düzeltme Modeli

Bazı durumlarda Y ve X gibi iki seri durağan olmadığı halde uzun dönemde birlikte hareket edebilirler. Eğer:

$$Y_t = b_0 + b_1X_t + e_t$$

modelinin tahmininden elde edilen e_t hata teriminin durağan olduğu belirlenirse, Y ve X'in eşbütünleşme gösterdiğinden söz edilir. Eğer eşbütünleşme varsa, bağımlı değişkeni açıklamak için uzun dönem süreci kullanılmalıdır. Y, uzun dönem dengesinin altında (ya da üstünde) kalıyorsa, takip eden dönemde Y'nin azalması (ya da artması) beklenir.

Eğer Y ve X eşbütünleşme gösteriyorsa, sadece değişkenlerin farkını alarak regresyon modellemesi yapmak yeterli olmayacaktır. Değişkenler arasındaki uzun dönem ilişkisinin de dikkate alınması gerekecektir. Bu amaçla hata düzeltme modelinden (error correction model=ECM) yararlanılır. ECM'de açıklayıcı değişken olarak uzun dönem dengesinden sapmalar da yer alır.

Uzun dönem ilişkisini:

$$Y_t = b_0 + b_1X_t + e_t$$

şeklinde tahmin edeceğimizi varsayalım. Bu durumda uzun dönem ilişkisinden sapmalar:

$$e_t = Y_t - b_0 - b_1X_t$$

olur. Bu sapmalar aşağıdaki ECM denkleminde yer alır:

$$\Delta Y_t = c_0 + c_1\Delta X_t + c_2e_{t-1} + u_t$$

modeldeki tüm değişkenler durağan olduğundan, ECM modeli EKK ile tahmin edilebilir.

Örnek 11-3

Bir yatırımcı tüketim ve gelir arasındaki ilişkiyi öğrenmek istiyor. 1978-2002 arasındaki veriler aşağıdaki gibidir (**ECM.gdt**).

YIL	C	Y	YIL	C	Y
1969	1393.6	2212.3	**1987**	2829.8	4279.3
1970	1432.6	2261.7	**1988**	2951.6	4493.7
1971	1461.5	2309.08	**1989**	3020.2	4624
1972	1533.8	2449.1	**1990**	3009.7	4611.9
1973	1596.6	2554	**1991**	3046.4	4724.9
1974	1692.3	2702.9	**1992**	3081.5	4623.6
1975	1799.1	2874.8	**1993**	3240.6	4810
1976	1902	3060.2	**1994**	3407.6	5138.2
1977	1958.6	3140.2	**1995**	3565.5	5329.5
1978	2070.2	3288.6	**1996**	3708.7	5489.9
1979	2147.5	3388	**1997**	3822.3	5648.4
1980	2197.8	3388.2	**1998**	3972.7	5862.9
1981	2279.5	3500.1	**1999**	4064.6	6060.4
1982	2415.9	3690.3	**2000**	4132.2	6138.7
1983	2532.6	3902.3	**2001**	4105.8	6079
1984	2514.7	3888.2	**2002**	4219.8	6244.4
1985	2570	3865.1	**2003**	4339.7	6383.8
1986	2714.3	4081.1	**2004**	4471.1	6604.2

Tahminleme doğrusal bir model ile yapılacaktır:

$$C_t = b_0 + b_1Y_t + e_t$$

Önce lnTuketim ve lnGelir değişkenlerinin durağanlıklarını test edelim:

C için ADF test sonuçları

Augmented Dickey-Fuller tests, order 2, for C
sample size 33
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.061
Tahmin değeri (a - 1): 0.00622487
test istatistiği: tau_c(1) = 0.619542
asimtotik p 0.9903

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.014
Tahmin değeri (a - 1): -0.243856
test istatistiği: tau_ct(1) = -2.13187
asimtotik p 0.5273

Y için ADF test sonuçları

Augmented Dickey-Fuller tests, order 2, for Y
sample size 33

birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.069
Tahmin değeri (a - 1): 0.00513383
test istatistiği: tau_c(1) = 0.377717
asimtotik p 0.9821

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.005
Tahmin değeri (a - 1): -0.375617
test istatistiği: tau_ct(1) = -2.30952
asimtotik p 0.4282

Görüldüğü gibi her iki değişken de durağan değildir. Şimdi

$$C_t = b_0 + b_1 Y_t + e_t$$

modelini tahmin edelim:

Model 1: EKK tahminleri 36 gözlem 1969-2004
Bağımlı değişken: C

	Katsayı	*St.Hata*	*t*	*p*	
sabit	-221.425	19.2949	-11.4759	<0.00001	***
Y	0.71029	0.00431991	164.4223	<0.00001	***

R^2 = 0.99874
Düzeltilmiş R^2 = 0.99871

Bu modelden elde edilen e_t değerleri aşağıda sunulmuştur.

YIL	e_t	e_{t-1}	YIL	e_t	e_{t-1}
1969	43.65031		**1987**	11.68072	36.96021
1970	47.56198	43.65031	**1988**	-18.8055	11.68072
1971	42.80844	47.56198	**1989**	-42.7563	-18.8055
1972	15.65362	42.80844	**1990**	-44.6618	-42.7563
1973	3.944191	15.65362	**1991**	-88.2245	-44.6618
1974	-6.118	3.944191	**1992**	18.82784	-88.2245
1975	-21.4169	-6.118	**1993**	45.52977	18.82784
1976	-50.2046	-21.4169	**1994**	-20.5874	45.52977
1977	-50.4279	-50.2046	**1995**	1.434073	-20.5874
1978	-44.2349	-50.4279	**1996**	30.70354	1.434073
1979	-37.5377	-44.2349	**1997**	31.72257	30.70354
1980	12.62021	-37.5377	**1998**	29.76534	31.72257
1981	14.83875	12.62021	**1999**	-18.6169	29.76534
1982	16.14157	14.83875	**2000**	-6.63266	-18.6169
1983	-17.7399	16.14157	**2001**	9.371657	-6.63266
1984	-25.6248	-17.7399	**2002**	5.889678	9.371657
1985	46.08287	-25.6248	**2003**	26.77524	5.889678
1986	36.96021	46.08287	**2004**	1.627307	26.77524

Şimdi e_t'nin durağanlığını test edelim:

e_t için ADF durağanlık testi

Augmented Dickey-Fuller tests, order 2, for e_t
sample size 33
birim-kök sıfır hipotezi: a = 1

Sadece sabit
model: (1 - L)y = b0 + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: 0.004
Tahmin değeri (a - 1): -0.652918
test istatistiği: tau_c(1) = -3.17436
asimtotik p 0.02155

Sabit ve trend
model: (1 - L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
e'nin 1. sıra otokorelasyon katsayısı: -0.017
Tahmin değeri (a - 1): -0.675437
test istatistiği: tau_ct(1) = -3.28392
asimtotik p 0.06887

ADF testi %10 için e_t'nin durağan olduğunu göstermektedir. O halde tüketim ve gelir arasında eş bütünleşme olduğundan söz edilebilir. Bir başka ifadeyle tüketim ve gelir arasında uzun dönem ilişkisi vardır. Buna göre uzun dönem marjinal tüketim eğilimi 0.71'dir.

Şimdi uzun dönemde düzeltme gerekip gerekmediğini, ECM model ile belirleyelim:

$$\Delta C_t = a_0 + a_1 \Delta Y_t + a_2 e_{t-1} + u_t$$

olmalıdır. Bu modeli tahmin ettiğimizde:

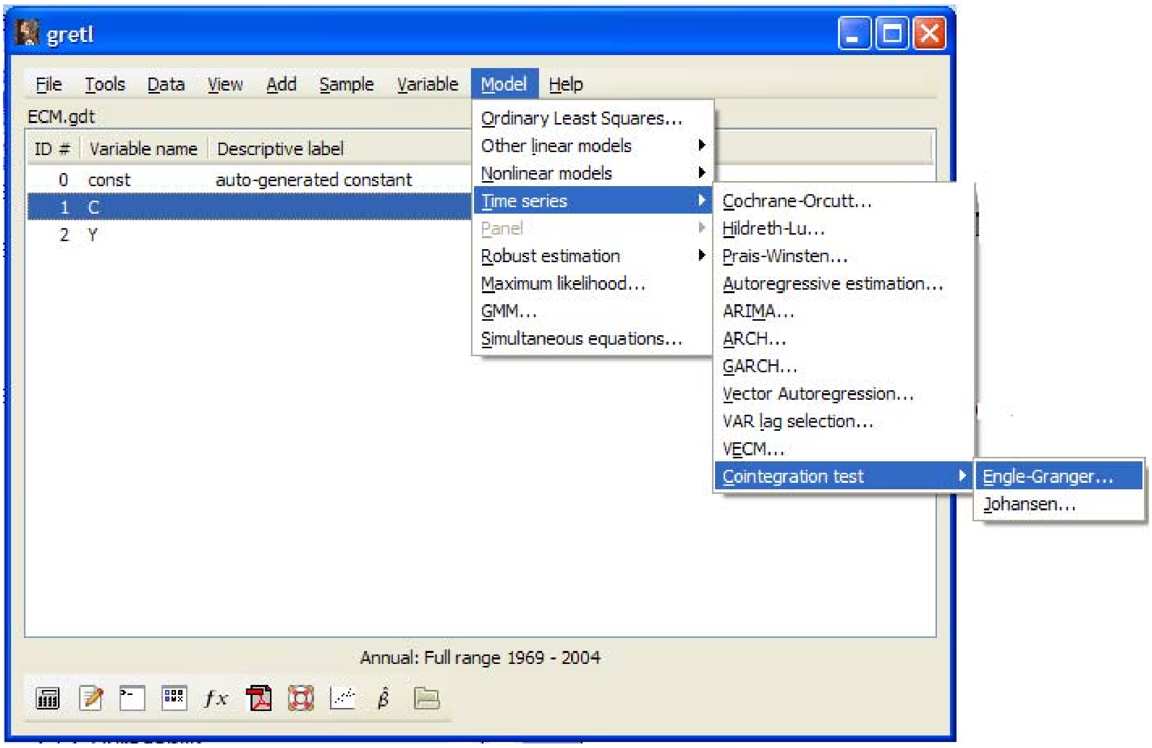
Model 3: EKK tahminleri 35 gözlem 1970-2004
Bağımlı değişken: d_C

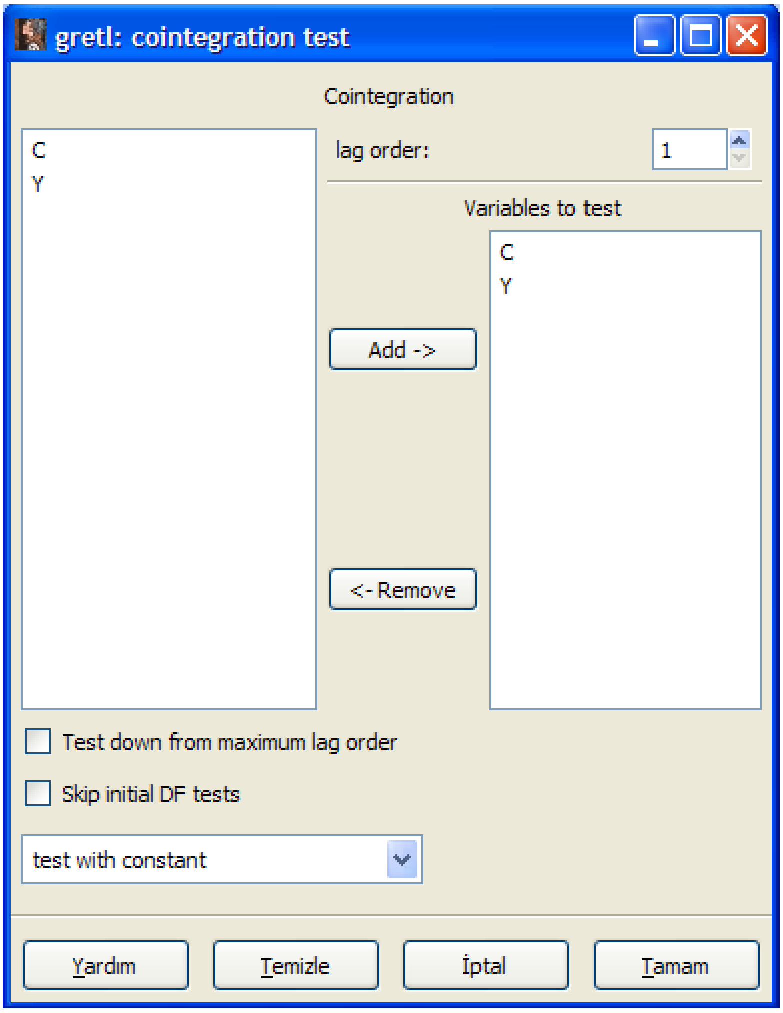
	Katsayı	*St.Hata*	*t*	*p*	
Sabit	23.2987	8.1554	2.8568	0.00746	***
d_Y	0.51498	0.0557108	9.2438	<0.00001	***
e_1	-0.18697	0.152632	-1.2250	0.22953	

R^2 = 0.77476
Düzeltilmiş R^2 = 0.76068
F-istatistiği (2, 32) = 55.0343 (p < 0.00001)
Durbin-Watson istatistiği = 1.68792
First-order autocorrelation coeff. = 0.155973
Log-likelihood = -160.485

elde edilir. Modeldeki ΔY_t'ye ait tahminci (0.51), kısa dönemde dönemde gelirin tüketim üzerindeki etkisini göstermektedir. Yani kısa dönem marjinal tüketim eğilimi 0.51'dir. e_1'nin istatistiki açıdan önemli olmaması, tüketim ile gelir arasındaki dengesizliğin hemen o yılda düzeltildiğini, uzun dönemdeki dengenin korunduğunu göstermektedir.

Şimdi eşbütünleşme analizini bu süreci Gretl yardımıyla yapalım:





Step 1: testing for a unit root in C

Augmented Dickey-Fuller test, order 1, for C
sample size 34
birim-kök sıfır hipotezi: a = 1

  Sadece sabit
  Tahmin değeri (a - 1): 0.00570085
  test istatistiği: tau_c(1) = 0.587917
  asimtotik p 0.9895

Step 2: testing for a unit root in Y

Augmented Dickey-Fuller test, order 1, for Y
sample size 34
birim-kök sıfır hipotezi: a = 1

  Sadece sabit
  Tahmin değeri (a - 1): 0.0044091
  test istatistiği: tau_c(1) = 0.332468
  asimtotik p 0.98

Step 3: cointegrating regression

Cointegrating regression -
EKK tahminleri 36 gözlem 1969-2004
Bağımlı değişken: C

VARIABLE	KATSAYI	STDERROR	T STAT	P
sabit	-221.425	19.2949	-11.476	<0.00001 ***
Y	0.710290	0.00431991	164.422	<0.00001 ***

R-squared = 0.99874
Düzeltilmiş R-squared = 0.99871
Durbin-Watson istatistiği = 0.938149
First-order autocorrelation coeff. = 0.506688
Akaike bilgi kriteri (AIC) = 358.004
Schwarz Bayesian kriteri (BIC) = 361.171
Hannan-Quinn kriteri (HQC) = 359.11

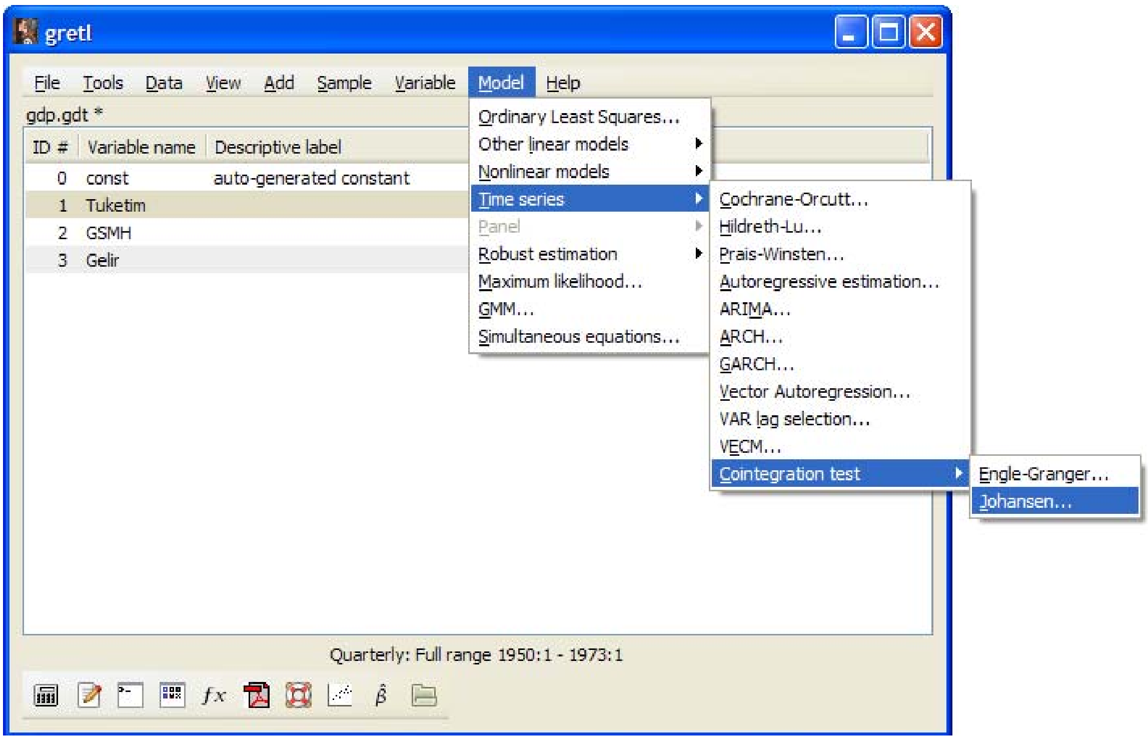
Step 4: Dickey-Fuller test on residuals

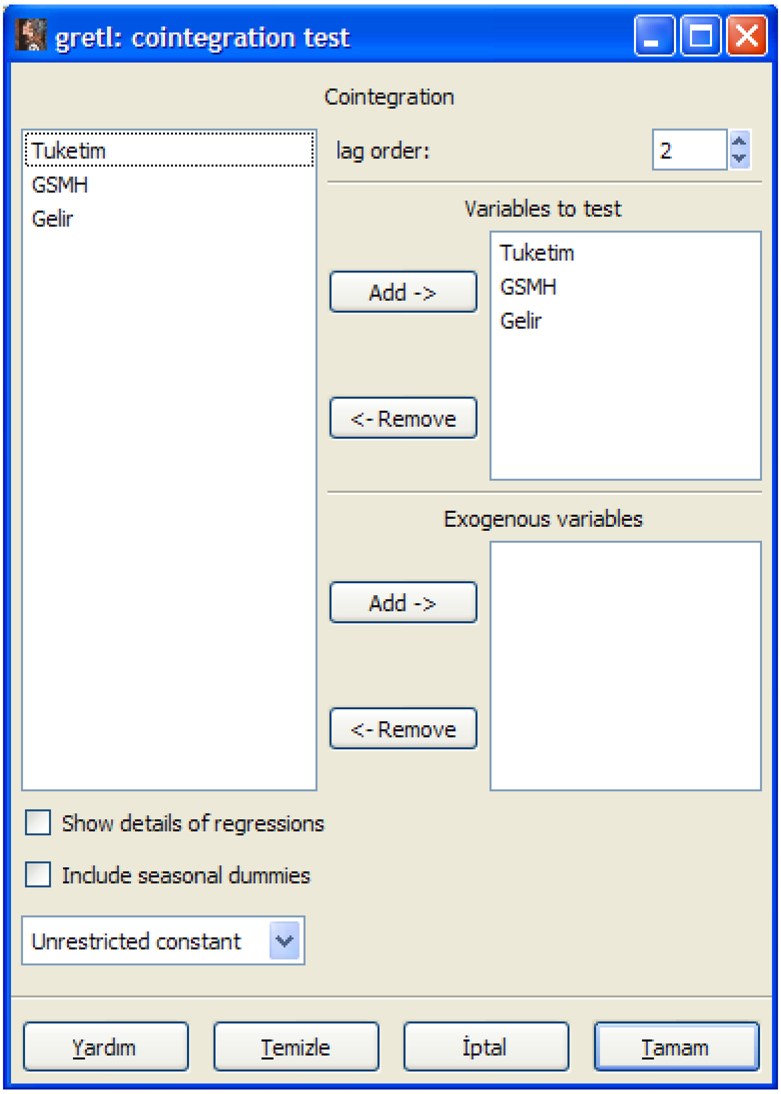
lag order 1
sample size 34
birim-kök sıfır hipotezi: a = 1

Tahmin değeri (a - 1): -0.629466
test istatistiği: tau_c(2) = -3.81594
asimtotik p 0.0129

There is evidence for a cointegrating relationship if:
(a) The unit-root hypothesis is not rejected for the individual variables.
(b) The unit-root hypothesis is rejected for the residuals (uhat) from the cointegrating regression.

Johansen Metodu ile Eşbütünleşme:





Johansen test:
Number of equations = 3
Lag order = 2
Estimation period: 1950:3 - 1971:4 (T = 86)

Case 3: Unrestricted Sabitant

Rank	Eigenvalue	Trace test	p-deÄŸeri	Lmax test	p-deÄŸeri
0	0.32851	62.490	[0.0000]	34.251	[0.0002]
1	0.27925	28.239	[0.0003]	28.162	[0.0001]
2	0.00089541	0.077040	[0.7813]	0.077040	[0.7814]

eigenvalue	0.32851	0.27925	0.00089541

beta (cointegrating vectors)

Tuketim	0.0090303	0.018971	0.0014202
GSMH	0.028053	-0.0074570	0.0012070
Gelir	-0.047383	-0.0095261	-0.0053288

alpha (adjustment vectors)

Tuketim	-9.3934	-29.462	1.0213
GSMH	-17.543	-2.4291	-0.24037
Gelir	-8.7840	1.5033	0.26300

renormalized beta

Tuketim	1.0000	-2.5440	-0.26652
GSMH	3.1066	1.0000	-0.22651
Gelir	-5.2471	1.2775	1.0000

renormalized alpha

Tuketim	-0.084826	0.21970	-0.0054421
GSMH	-0.15842	0.018114	0.0012809
Gelir	-0.079323	-0.011210	-0.0014014

long-run matrix (alpha * beta')

	Tuketim	GSMH	Gelir
Tuketim	-0.64230	-0.042583	0.72031
GSMH	-0.20484	-0.47431	0.85564
Gelir	-0.050430	-0.25732	0.40049