Uludağ Üniversitesi
İktisadi ve İdari Bilimler Fakültesi Dergisi
Cilt XII, Sayı 1-2, Mart-Kasım 1991

KUKLA DEĞİŞKENLERİN DEĞİŞİK KULLANIMLARI

Mustafa SEVÜKTEKİN*

1. GİRİŞ

Ekonometrik modellerde kukla değişkenler daha çok birer açıklayıcı değişken olarak yer almaktadır. Nitel vasıftaki bu değişkenlerin ekonometrik modellerde kullanımı genellikle iki temel biçimde ele alınmaktadır: Varyans analizi modelleri ve kovaryans analizi modelleri. Yalnızca kukla değişkenlerin yer aldığı regresyon modellerine "Varyans Analizi Modelleri" ve nicel diğer değişkenlerin yanında kukla değişkenlerin yer aldığı regresyon modellerine ise "Kovaryans Analizi Modelleri" denilmektedir. Bu çalışmamızda bu iki tür model üzerinde pek fazla durmadan regresyon denkleminde bağımlı değişken yerine nitel vasıftaki kukla değişkenlerin nasıl kullanıldığı üzerinde durulacaktır. Kukla bağımlı değişkenlerin yer aldığı regresyon modellerine ise "Doğrusal Olasılık Modelleri" denilmektedir. Doğrusal olasılık modellerinin tanımını kısaca açıkladıktan sonra onların tahminleri ve tahmin sorunları üzerinde durulmaya çalışılacaktır.

Bağımlı değişken niteliğindeki kukla değişkenlerin regresyon modelleri üzerinde kullanımı ile ilgili bilgilerin sunulmasının hemen arkasından varyans analizi modelinin farklı bir kullanımı açıklanmaya çalışılmaktadır. Kukla değişkenlerin çok sayıda yer alabileceği durumlar üzerinde kısaca durularak bir kovaryans analizi modeli ile buradaki problem ele alınarak incelenmektedir. Kukla değişkenlerin çokça başvurulduğu diğer bir kullanım alanı ise mevsimlik ayarlamalardır. Farazi bir örnek ile kukla değişkenlerin mevsimlik ayarlamalarda

---

* *Yard. Doç. Dr.; U.Ü. İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü Ekonometri Anabilim Dalı Öğretim Üyesi.*

nasıl kullanıldığı gösterilmektedir. Diğer önemli bir kullanım alanı ise parçalı doğrusal regresyon modellerinde görülmektedir. Son olarak önemli bir kullanımı bulunan ve sıkça başvurulan yapısal farklılıkların test edilmesinde kukla değişkenlerin kullanımı, önce yöntem olarak açıklanmaya çalışılmakta ve arkasından sayısal bir örnekle izah edilmektedir. Burada özellikle Chow Testinden yararlanılmaktadır.

2. DOĞRUSAL OLASILIK MODELLERİ (LPM)

Bir regresyon modelinde kukla değişkenler yalnızca açıklayıcı değişkenler olarak yer almazlar. Ayrıca bunları bazı hallerde bağımlı değişken olarak da görmemiz mümkündür. Örneğin, işsizlik oranının, ortalama ücret hadlerinin, aile gelirlerinin veya eğitim düzeylerinin bir fonksiyonu olarak, orta yaştaki erkeklerin işgücüne katılımı ile ilgili bir fonksiyon da, bağımlı değişken olarak işgücüne katılma iki şekilde ele alınabilir: Eğer kişi işgücüne katılmış ise 1, katılmamış ise 0. İkili yapıdaki bu örnekleri çoğaltmamız mümkündür. Örneğin bir ailenin ev, hastalık sigortası ya vardır ya da yoktur; karı-koca birlikte işgücüne ya dahildirler, ya da dahil değillerdir vs. Bütün bu örneklerin biricik çehresi evet ve hayır gibi iki tipde cevapların çıkarılabileceği, yani ikili yapıda bir bağımlı değişken ile tanımlanabileceğidir[1].

Kukla bağımlı değişkenlerin regresyon modellerinde nasıl kullanıldığını görebilmek için aşağıdaki modeli incelemeye çalışalım[2]:

(2.1) $$Y_i = \alpha + \beta X_i + u_i$$

burada Y_i = 1 aile eğer ev sahibi ise
= 0 aile eğer ev sahibi değil ise,
X_i = ailenin gelirini göstermektedir.

Bağımlı değişkenin bir kukla değişken olduğu bu tür modellere "Doğrusal Olasılık Modelleri" (LPM) denilmektedir. Çünkü E $(Y_i \mid X_i)$, X_i veri iken Y_i nin koşullu beklentisi, koşullu olasılık olarak yorumlanır. Bu ise ancak X_i veri iken sözkonusu olabilir, yani Pr $(Y_i = 1 \mid X_i)$ dir. Böylece yukarıdaki örnekte bir ailenin sahip olduğu bir evin olasılığı ve gelir X_i veri iken E $(Y_i \mid X_i)$ dir. Bu modellere doğrusal olasılık modelleri denilmesinin daha açık bir izahını şöyle verebiliriz: Modeldeki kalıntı terimi u_i nin klasik doğrusal regresyon varsayımlarını sağladığını kabul edelim. Y_i için beklenen değer:

---

1 Gujarati (1979, s. 312-13).

2 Bu tür modeller için bakınız Gujarati (1979, s. 312), Kmenta (1971, s. 425-26), Stewart ve Wallis (1981, s. 185-86) ve Pindyck ve Rubinfeld (1981, s. 275-76).

(2.2) $$E(Y_i \mid X_i) = \alpha + \beta X_i$$

olacaktır. Bilindiği gibi Y_i 0 ve 1 değerlerini almaktadır. $Y_i = 1$ ise P_i olasılığını alsın (yani olay vuku bulduğunda), $Y_i = 0$ ise $1 - P_i$ olasılığını alsın (yani olay vuku bulmadığında). Bu durumu şematik olarak kısaca şöyle gösterebiliriz:

Y_i	Olasılık
1	P_i
0	$1 - P_i$
	1

Dolayısıyla buradan yararlanılarak matematiksel bekleti tanımı şöyle verebiliriz:

$$E(Y_i) = 0 \,.\, (1 - P_i) + 1 \,.\, (P_i)$$

(2.3) $$E(Y_i) = P_i$$

Yukarıdaki ikinci ve üçüncü eşitlikleri karşılaştırıp birlikte ele aldığımızda

(2.4) $$E(Y_i \mid X_i) = \alpha + \beta X_i = P_i$$

sonucunu buluruz, yani (2.1) in koşulu beklentisi gerçekte Y_i nin koşullu olasılığı olarak yorumlanabilir. Çünkü olasılık 0 ve 1 arasında yer almakta ve dolayısıyla sınırlandırma şu şekilde ifade edilmektedir:

(2.5) $$0 \leq E(Y_i \mid X_i) \leq 1$$

3. DOĞRUSAL OLASILIK MODELLERİNİN TAHMİNİ

Yukarıda ele aldığımız Model (2.1) in herhangi bir regresyon modelinden farkı yoktur, dolayısıyla bu modelin parametrelerine olağan en küçük kareler yöntemini uygulayarak tahminlerini yapabiliriz[3]. Ancak tahmin aşamasında karşımıza önemli sayılabilecek bazı problemler çıkmaktadır. Bu problemler üzerinde kısaca durmamızda yarar vardır.

---

3 Tahmin yöntemi hakkında daha fazla bilgi için bakınız Gujarati (1979, s. 314) ve Dutta (1975, s. 168-69).

1. u_i nin Normal Dağılım Göstermemesi: Olağan en küçük karelerin tahmin yöntemi kalıntıların (u ların) normal dağıldığını varsaymasına rağmen, istatistiki vardama amacıyla normal dağılımın gerekli olduğunu kabul ediyoruz veya en azından varsayıyoruz. Fakat u_i nin normal dağıldığı varsayımı olasılık modelleri için uzun süre savunulmamaktadır. Çünkü u_i yalnızca iki değer ihtiva ettiğinden Y_i ye benzer. Model (2.1) i şimdi şöyle yazabiliriz:

$$u_i = Y_i - \alpha - \beta X_i \tag{3.1}$$

buradan $Y_i = 1$ için $\quad u_i = 1 - \alpha - \beta X_i$ ve

$Y_i = 0$ için $\quad u_i = - \alpha - \beta X_i$

olacaktır. Dolayısıyla u_i nin artık burada normal dağıldığını ileri süremeyiz. Kesikli bir dağılıma sahip olduğu açıktır.

Normal dağılım varsayımının ihlal edilmesi göründüğü kadar kritik bir sorun teşkil etmeyebilir. Zira biliyoruz ki, olağan en küçük karelerin nokta tahminleri böyle durumlarda eğilimsizliklerini hala devam ettirmektedirler, fakat amaç eğer nokta tahmini ise normallik varsayımı ancak o zaman önemsiz olmaktadır. Bundan başka örneklem büyüklüğü artarken tanımsız olarak olağan en küçük kareler tahminleri genellikle normal dağılım gösterme eğilimindedirler. Bu açıdan büyük örneklemlerde doğrusal olasılık modellerinin istatistiki vardaması normal dağılım varsayımı altında olağan en küçük kareler yöntemi takip edilerek yapılır[4].

2. Kalıntıların Heteroskedastik (Dağılan) Varyansları: İkinci önemli problem kalıntıların heteroskedastik (dağılan) varyans özelliği göstermeleridir. $E(u_i) = 0$ ve $E(u_iu_j) = 0$, $i \neq j$ için olsa bile (yani serisel korrelasyon olmadığında) kalıntılar u_i ler homoskedastik olduğundan bu uzun süre devam etmeyecektir. u_i nin olasılık dağılımları aşağıdaki gibi ifade edilebilir:

Y_i	u_i	$f(u_i)$	Olasılık
0	$-\alpha - \beta X_i$	$1 - \alpha - \beta X_i$	$1 - P_i$
1	$1 - \alpha - \beta X_i$	$\alpha + \beta X_i$	P_i
			1

Burada verilen olasılık dağılımı daha önce Y_i için verilen olasılık dağılımına benzemektedir.

---

4 Daha fazla bilgi için bakınız Malinvaud (1980, s. 195-97).

Goldberger bu durumda temel olarak u_i nin dağılan varyans varsayımının geçerli olamıyacağını ileri sürer[5]. Aşağıdaki tanımdan hareketle

$$\begin{aligned} Var(u_i) &= E(u_i - E(u_i))^2 \\ &= E(u^2_i) \qquad\qquad E(u_i) = 0 \text{ varsayımından} \end{aligned}$$

ve dolayısıyla u_i nin önceki olasılık tanımını kullanarak

$$\begin{aligned} Var(u_i) &= E(u_i^2) \\ &= (-\alpha - \beta X_i)^2 (1 - P_i) + (1 - \alpha - \beta X_i)^2 (P_i) \\ &= (-\alpha - \beta X_i)^2 (1 - \alpha - \beta X_i) \\ &\quad + (1 - \alpha - \beta X_i)^2 (\alpha + \beta X_i) \end{aligned}$$

$$(3.2) \qquad Var(u_i) = (\alpha + \beta X_i)(1 - \alpha - \beta X_i)$$

veya

$$\begin{aligned} Var(u_i) &= E(Y_i \text{ I } X_i)(1 - \alpha - \beta X_i) \\ &= E(Y_i \text{ I } X_i)(1 - E(Y_i \text{ I } X_i)) \end{aligned}$$

$$(3.3) \qquad Var(u_i) = P_i (1 - P_i)$$

Yukarıda elde edilen sonuçtan ötürü u_i heteroskedastiktir ve onun varyansı $E(Y_i \text{ I } X_i)$ ve X_i üzerine sistematik olarak bağımlıdır ya da onun bağımlı olduğu gösterilebilir. Heteroskedastisite durumunda α ve β nın olağan en küçük kareler tahminlerine daha az güven duyulur. Tahminler eğilimsiz olsalar bile etkin değillerdir. Heteroskedastik bir yapıdaki u_i için tahmin metodu seçmek gereklidir. Ayrıca dikkat edilirse u_i bu modelde kesikli bir dağılıma sahiptir. Dolayısıyla bu durum normal dağılım varsayımının kuralı dışındadır. Ayrıca tahminler için anlamlılık testlerinin uygulanması da kural dışıdır[6].

Heteroskedastisite problemini yeniden çözmenin bir yolu Model (2.1) in her iki tarafını dönüştürülmüş verilere bölmektir.

$$\begin{aligned} \sqrt{E(X_i \text{ I } X_i)(1 - E(Y_i \text{ I } X_i)} &= \sqrt{P_i (1 - P_i)} \quad \text{diyelim ki,} \\ &= \sqrt{w_i} \end{aligned}$$

5 Bu sürecin gerekçeleri için bakınız Goldberger (1964, s. 249-50) ve Goldberger (1968, s. 112-18).
6 Bakınız Dutta (1975, s. 69).

$$(3.4) \qquad \frac{Y_i}{\sqrt{w_i}} = \frac{\alpha}{\sqrt{w_i}} + \beta \frac{X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}}$$

Denklem (3.4) deki kalıntı terim artık homoskedastik bir yapıda olacaktır. Olağan en küçük kareler yöntemini bu durumda (3.4) denklemine uygulamamız mümkün olacaktır.

Gerçek E $(Y_i \mid X_i)$ bilinmediğinden, w_i de bilinmemektedir. Bu açıdan w_i nin tahmininde aşağıdaki iki adımdan yararlanılacaktır[7].

Adım I: Kural; Model (2.1) deki regresyona heteroskedastisite problemine rağmen olağan en küçük kareler yöntemini uygulamaktır ve böylece gerçek E $(Y_i \mid X_i)$ nin tahmini $\hat{Y}$ yi bulmaktır. Daha sonra w_i nin bir tahmini $\hat{w}_i = \hat{Y}_i (1 - \hat{Y}_i)$ den elde etmektir.

Adım II: Model (3.4) deki gibi dönüştürülmüş verileri kullanarak w_i nin tahmin edilmesi ve uygulanacak kural; dönüştürülmüş veriler üzerine olağan en küçük kareler yöntemini uygulamaktır.

3. $0 \leq$ E $(Y_i \mid X_i) \leq 1$ in Gerçekleşmemesi: Doğrusal olasılık modellerinde E $(Y_i \mid X_i)$ ve X_i veri iken Y_i nin koşullu olasılık ölçüsü 0 ile 1 arasındadır. Bu önsel olarak doğru olmasına (veya kabul edilmesine) rağmen E $(Y_i \mid X_i)$ nin tahmincisi Y_i bunu garanti etmez. Bu doğrusal olasılık modellerinin en küçük kareler tahminleriyle ortaya çıkan gerçek bir problemdir. Fakat bu kısıtlamanın yerine getirilmesi elzemdir. Tahmin edilen Y_i nin 0 ve 1 arasında olmasını bilmenin iki yolu vardır: Birisi doğrusal olasılık modellerine olağan en küçük kareleri uygulayarak tahmin etmek ve tahmin edilen Y_i nin 0 ile 1 arasında olup olmadığını bulmaktır. Eğer tahminlerden bazıları 0 dan küçük ise (yani negatif ise) Y_i bu durumlar için sıfır olarak varsayılacaktır. Eğer tahminler 1 den büyük ise 1 olarak varsayılacaktır. İkinci yöntem, tahmin edilen koşullu olasılıkların Y_i lerin 0 ve 1 arasında olmasını garanti edecek bir yöntem bulmaktır. Bunun için birkaç yöntem mevcuttur fakat buradaki amacımız dışına taşdığı için değinmeden geçeceğiz[8].

4. KUKLA DEĞİŞKENLERE KEYFİ ÖLÇÜLERİN VERİLMESİ

Bazı durumlarda kukla değişkenler için keyfi ölçülerin verilmesi başvurulan kurallardan biridir. Örneğin aşağıdaki bölgeler arası çay tüketimi üzerine verilen farazi örnek üzerinde duralım:

---

7 Bakınız Gujarati (1979, s. 315) ve Goldberger (1964, s. 250).

8 Yöntemler hakkında daha fazla bilgi için bakınız Goldfeld ve Quandt (1972, Bölüm 4) ve Pindyck ve Rubinfeld (1981, Bölüm 10).

Kuzey : 1
Güney: 2
Batı : 3
Doğu : 4

Tüketim için davranış kalıpları üzerine bölgesel yerleşimlerin etkilerini ölçmek istiyoruz. Tahmin edilen davranış kalıplarını yorumlarken bu bölgesel ayırımları da dikkate almak ve karşılaştırmak istiyoruz.

Bu amaçla aşağıdaki modeli dikkate alalım:

$$C_i = \alpha + \beta R_i + u_i \tag{4.1}$$

burada C_i = i ninci ailenin çay tüketimi

R_i = i ninci ailenin ikamet ettiği bölge, bölgeler yukarıda vurgulandığı gibi yerleşim bölgelerine göre 1, 2, 3 ve 4 değerlerini alacaktır.

(4.1) nolu denklemin kalıntı terimi u_i klasik doğrusal regresyon varsayımlarını sağladığını düşündüğümüzde, C_i için birbirinden farklı dört değer beklenecektir, bunlar:

(4.2)

$E(C_i \mathrm{I} R_i) = \alpha + \beta$ i ninci aile R = 1 de ikamet etmektedir,

$E(C_i \mathrm{I} R_i) = \alpha + 2\beta$ i ninci aile R = 2 de ikamet etmektedir,

$E(C_i \mathrm{I} R_i) = \alpha + 3\beta$ i ninci aile R = 3 de ikamet etmektedir,

$E(C_i \mathrm{I} R_i) = \alpha + 4\beta$ i ninci aile R = 4 de ikamet etmektedir.

Yukarıdaki sonuçların vurguladığı nokta (veya gerçek) çay tüketimi üzerine bölgesel etkinin biricik oranının ve bu orantıyı gösteren faktör β yı belirlemektir. Bu faktör bir ve iki bölgeleri arasında, iki ve üç bölgeleri arasında ve tekrar üç veya dört bölgeleri arasında aynıdır. Bir önsel bilgi olmaksızın bu bir keyfi varsayımın getirisi, ya da başka bir ifadenin sonucudur.

Buradaki problemi tek ikili kukla değişken kullanımından kaçınarak tanımlamakta yarar vardır. Bundan ötürü modeli aşağıdaki gibi yeniden yazalım:

$$C_i = \alpha + \beta_1 R_i(1) + \beta_2 R_i(2) + \beta_3 R_i(3) + u_i \tag{4.3}$$

burada $R_i(1)$ = 1 eğer i ninci aile birinci bölgede ikamet ediyorsa
= 0 diğer durumlar,

$R_i(2)$ = 1 eğer i ninci aile ikinci bölgede ikamet ediyorsa
= 0 diğer durumlar,

$R_i(3)$ = 1 eğer i ninci aile üçüncü bölgede ikamet ediyorsa
= 0 diğer durumlar.

(4.3) deki modelde kalıntı u_i klasik doğrusal regresyon modelinin ileri sürülen varsayımlarını sağlıyorsa Y_i için beklenen değerler şöyle olacaktır:

$E(C_i \mid R_i) = \alpha$	$R_i(1) = R_i(2) = R_i(3) = 0$ olduğunda, yani i ninci aile dördüncü bölgede ikamet ediyorsa,
$E(C_i \mid R_i) = \alpha + \beta_1$	i ninci aile birinci bölgede ikamet ediyorsa,
$E(C_i \mid R_i) = \alpha + \beta_2$	i ninci aile ikinci bölgede ikamet ediyorsa,
$E(C_i \mid R_i) = \alpha + \beta_3$	i ninci aile üçüncü bölgede ikamet ediyorsa.

Yukarıdaki sonuçların yorumunda çay tüketimi ile ilgili olarak bölgeler arası ilişkiyi yansıtan keyfi bir varsayıma yer verilmemektedir[9].

5. KUKLA DEĞİŞKENLERİN ÇOK SAYIDA BULUNDUĞU HAL

Regresyon modellerinde kukla değişkenlerin çok sayıda kullanılması veya yer alması beraberinde önemli birçok problemi getirmektedir. Sayısallaştırılamayan nitel türdeki değişkenleri sayısallaştırmak için son zamanlarda yoğun çabalar ve gayretler gözlenmektedir. Modellerde yer alan kukla değişkenlerin sağladığı bilgiden çok ortaya çıkardığı sorunların sayısının daha fazla olması, onların birer "başbelası" olarak nitelendirilmesine yol açmaktadır. Örneğin bir tüketim fonksiyonunun yatay kesit çalışmasında j ninci ailenin tüketimi, j ninci ailenin yalnızca gelirine Y_j ye bağlı olmayabilir. Ayrıca ırk Z_{2j}, din Z_{3j}, eğitim Z_{4j}, evlilik durumu Z_{5j}, aile büyüklüğü Z_{6j}, ve i1h değişkenlerine de bağlı olabilir. Dolayısıyla etkileyici bu diğer faktörler modelde birer kukla değişken olarak yer alacaklardır.

Dikkat edilirse, yukarıda tüketim fonksiyonu ile ilişkisi ileri sürülen diğer faktörler her ne şekilde olursa olsun önemsiz faktörler değillerdir. Bu nedenle, bu tür değişkenlerin etkileri tesadüfi ise, bunlar bir ortalama üzerinde toplanarak birçoğu (veya tamamı) dışlanarak indirgeme yapılabilir. Testler yapıldığında görülecektir ki, bu değişkenlerin modelden düşürülmesiyle fazla bir şey kaybedilmemektedir. Çünkü bağımlı değişkendeki sistematik değişmenin kaynağı hakkında herhangi bir bilgi kaybı sözkonusu olmayacaktır. Öte yandan ekonometrik çalışmaların birçoğunda yatay kesit verileri ile tüketim fonksiyonunun açıklanmasında aile geliri tek önemli bir değişken olarak yeterli görülmektedir.

---

9 Bakınız Dutta (1975, s. 169-70) ve Intriligator (1978, s. 61-2).

İleri sürülen diğer mülahazaları aşağıdaki gibi ifade edebiliriz:

$$(5.1) \quad C_{jt} = \alpha + \beta_1 Y_{jt} + \beta_2 Z_{2jt} + \beta_3 Z_{3jt} + \beta_4 Z_{4jt} + ... + \beta_K Z_{Kjt} + u_{jt}$$

$$C_{jt+1} = \alpha + \beta_1 Y_{jt+1} + \beta_2 Z_{2jt+1} + \beta_3 Z_{3jt+1} + ... + \beta_K Z_{Kjt+1} + u_{jt+1}$$

Fark alma işlemi uygulanırsa

$$(5.2) \quad \Delta C_j = \beta_1 \Delta Y_j + \beta_2 \Delta Z_{2j} + \beta_3 \Delta Z_{3j} + ... + \beta_K \Delta Z_{Kj} + v_j$$

burada $v_j = u_{jt+1} - u_{jt}$ dir. Belirli varsayımlar altında ırk, din, eğitim, evlilik durumu ve ailenin büyüklüğü simgeleri j ninci aile için belirli bir zamanda değişmediği kabul edilir. Dolayısıyla değişkenler:

$$\begin{aligned} \Delta Z_{2j} &= Z_{2jt+1} - Z_{2jt} \\ \Delta Z_{3j} &= Z_{3jt+1} - Z_{3jt} \\ \dots\dots &\quad \dots\dots\dots \quad \dots\dots\dots \\ \Delta Z_{Kj} &= Z_{Kjt+1} - Z_{Kjt} \end{aligned}$$

sıfır olacaklardır. Daha sonra denklemi aşağıdaki duruma indirgeyebiliriz:

$$(5.3) \quad \Delta C_j = \beta_1 \Delta Y_j + v_j$$

Gerçek bir durumda bu değişkenlerin tümü ihtiyaçtan dolayı değişmeden aynen kalacaktır. Kukla değişkenler seti Z ler ekseriye sayısallaştırılamayan ve nitel iki gruba ayrılabilir. Geçici yatay kesit verilerinden oluşan örneklemler aldığımızda bu alt setlerden birinin üzerine gözlemler değişmeden kalacaktır. Dolayısıyla regresyon denklemlerinde ilk farklara göre açıklanacak ve kukla değişkenlerin sayısı birkaç tane olacaktır. Veri bir model için veri bir aralık boyunca t ve t+1 arasında ırk, din, hatta eğitim bile değişmeden aynen kalacaktır ancak evlilik durumu ile ailenin büyüklüğü zamanla değişebilir[10].

6. MEVSİMLİK AYARLAMALARDA KUKLA DEĞİŞKENLERİN KULLANIMI

Ekonomik zaman dizilerinin birçoğu mevsimlik kalıpları göstermek için aylık veya çeyrek yıllık verilerden yararlanılır. Örneğin yılbaşılarında dükkanların

10 Bakınız Dutta (1975, s. 170-71).

satış hacimlerinde, tatil dönemlerinde hane halklarının para taleplerinde, yaz mevsimlerinde soğuk içecekler ve dondurma tüketiminde, hasat mevsiminin sonunda ekin fiyatlarının yükselmesi vs. çoğunlukla zaman dizilerinin mevsimlik faktörlerini veya unsurlarını ortadan kaldırmak için bazı yöntemlere başvurulabilir. Bir zaman dizisinden mevsimlik unsuru ortadan kaldırma süreci mevsimsizleştirme veya mevsimlik ayarlama olarak bilinmektedir[11].

Bir zaman dizisini mevsimsizleştirmenin birkaç yöntemi vardır. Burada yalnızca kukla değişkenlerin kullanıldığı yönteme başvuracağız. Zaman dizisi verilerinin mevsimlik ayarlamalarında kukla değişken tekniği (sıfır-bir ölçeği) sıkça başvurulan önemli tekniklerden birisidir. Kukla değişkenler üzerine doğrusallık ve toplanabilirlik varsayımı veri iken, mevsimlik ayarlama metodu geleneksel kurallardan daha az keyiflilik taşır. Örneğin çeyrek yıllık verileri kullanarak aşağıdaki tablo ile ilgili bir model kurmaya çalışalım[12]:

$$
\begin{aligned}
Q_1 &= 1 \text{ eğer birinci çeyrek yıllık gözlem ise} \\
&= 0 \text{ diğer durumlar,} \\
Q_2 &= 1 \text{ eğer ikinci çeyrek yıllık gözlem ise} \\
&= 0 \text{ diğer durumlar,} \\
Q_3 &= 1 \text{ eğer üçüncü çeyrek yıllık gözlem ise} \\
&= 0 \text{ diğer durumlar,} \\
Q_4 &= 1 \text{ eğer dördüncü çeyrek yıllık gözlem ise} \\
&= 0 \text{ diğer durumlar.}
\end{aligned}
$$

Q_1	Q_2	Q_3	Q_4
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Kârlar ve satışlar ile ilgili aşağıdaki modeli dikkate alalım:

$$P_t = \alpha_0 + \alpha_1 S_t + \beta_1 Q_{1t} + \beta_2 Q_{2t} + \beta_3 Q_{3t} + \beta_4 Q_{4t} + u_t \tag{6.1}$$

burada P_t = t dönemindeki kârlar

S_t = t dönemindeki satışlar

11 Bir zaman dizisi genellikle dört temel unsuru bünyesinde ihtiva eder, ya da etkisi altında kalır. Bunlar: 1. Uzun dönem eğilimi (trend), 2. Mevsimlik dalgalanmalar, 3. Konjonktürel dalgalanmalar ve 4. Düzensiz hareketler veya tam tesadüfilik. Bu konuda daha fazla bilgi için bakınız Serper (1986, s. 208-16).

12 Gujarati (1979, s. 300); Stewart ve Wallis (1981, s. 178-80); Jonhston (1972, s. 186-89) ve İşyar (1989, s. 206-8).

Q_{1t}, Q_{2t}, Q_{3t} ve Q_{4t} = yukarıda tanımlandığı gibidir.

Dikkat edilirse mevsim ile ilgili nitel değişken dört kategoriyi içermektedir. Dolayısıyla kukla değişken tuzağına düşmemek için kukla değişken sayılarından bir eksiğini modelde ele almak gerekecektir. Yani dört kategori olduğundan bir eksiği üç kukla değişkene yer verilecektir. Bu amaçla modelimiz aşağıdaki gibi olacaktır:

(6.2) $$P_t = \alpha_0 + \alpha_1 S_t + \beta_1 Q_{1t} + \beta_2 Q_{2t} + \beta_3 Q_{3t} + u_t$$

Modelden Q_{4t} nin düşürülmesi tamamen tesadüfidir. Dolayısıyla sonuçta (6.2) denklemi tahmin edilecektir. β_1, β_2 ve β_3 diferansiyel kesme terimleridir. Modelde u_t için klasik doğrusal regresyon modelinin varsayımları geçerli ise P_t için aşağıda birbirinden farklı beklenen değerler elde edebiliriz:

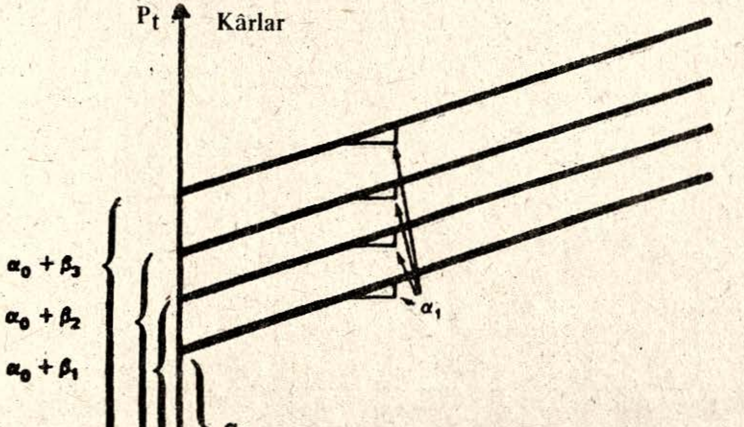
$E(P_t \mid S_t) = (\alpha_0 + \beta_1) + \alpha_1 S_t$ — Q_1 çeyrek yıllık gözlem için

$E(P_t \mid S_t) = (\alpha_0 + \beta_2) + \alpha_1 S_t$ — Q_2 çeyrek yıllık gözlem için

$E(P_t \mid S_t) = (\alpha_0 + \beta_3) + \alpha_1 S_t$ — Q_3 çeyrek yıllık gözlem için

$E(P_t \mid S_t) = \alpha_0 + \alpha_1 S_t$ — Q_4 çeyrek yıllık gözlem için

Dikkat edilirse eğim katsayısı α_1 sabittir. Bu durumu aşağıdaki gibi geometrik olarak gösterebiliriz.

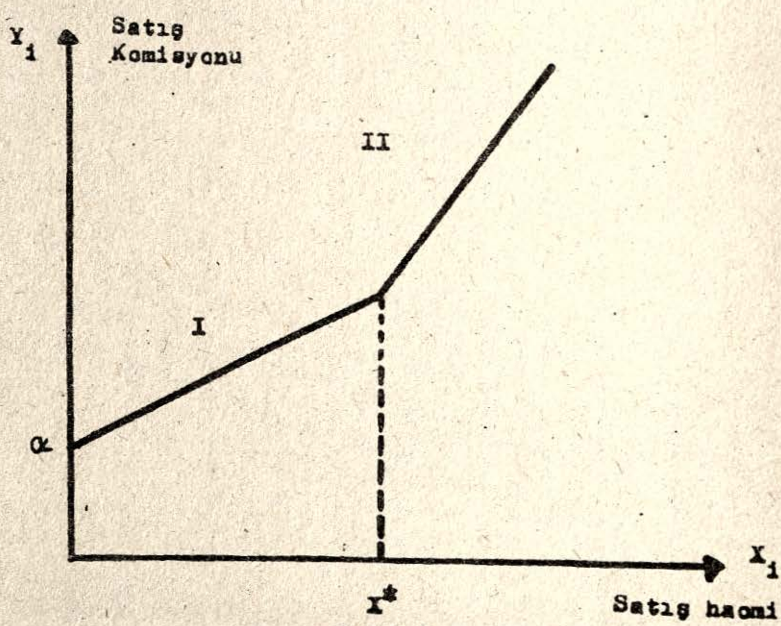


Şekil: 1
Kârlar ve satışlar ile ilgili regreson

7. PARÇALI DOĞRUSAL REGRESYONLARDA KUKLA DEĞİŞKEN KULLANIMI

Kukla değişkenlerin diğer bir kullanım alanı aşağıda farazi bir örnek üzerinde durularak anlatılmaya çalışılacaktır. Farazi örneğimiz herhangi bir şirketin ürün satışı kâr payı ile ilgilidir. Şirketin sattığı ürünlerden belli bir miktar komisyon payı almaktadır. Ancak aldığı bu komisyon payı satış oranları ile yakından ilişkilidir. Yani belli bir satış miktarından sonra ancak bu komisyon payı geçerli olmaktadır. Örneğin X^* düzeyi olarak kabul edeceğimiz bir eşik değerinden sonra bu komisyon payını alabilecektir. Daha spesifik olması açısından varsayalım ki, satış komisyonları satışlar ile doğrusal olarak bağımlıdır. Fakat bu bağımlılık X^* eşik değerine kadar daha yavaş, eşik değerinden sonra daha hızlıdır. Bu durumu geometrik olarak aşağıdaki I ve II olarak işaretlenen iki parçadan oluşmuş bir parçalı doğrusal regresyona sahip oluruz, dolayısıyla eğim eşik değerinden sonra değişmektedir.

Komisyon, satışlar ve eşik değeri X^* veri iken kukla değişkenler yardımıyla aşağıdaki şekilde gösterilen parçalı doğrusal regresyonun iki parçasının eğimleri (ki bu eğimler farklıdır) tahmin edebiliriz. Bunun için aşağıdaki modelden yararlanabiliriz:



Şekil: 2

Satış komisyonu ve satış hacmi arasındaki farazi ilişkiler (Dikkat edilirse Y ekseni üzerindeki kesme terimi α minimum noktayı göstermektedir)

$$(7.1) \qquad Y_i = \alpha + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i$$

burada Y_i = satış komisyonları,
X_i = satış personeli tarafından yapılan satış hacmi,
D_i = 1 eğer $X_i > X^*$ ise
= 0 eğer $X_i < X^*$ ise.

Modelde yer alan kalıntı değişken u_i nin klasik doğrusal regresyon varsayımlarını sağladığını kabul edelim. Dolayısıyla Y_i için beklenen değerler şöyle olacaktır:

$$(7.2) \qquad E(Y_i \mid D_i = 0, X_i, X^*) = \alpha + \beta_1 X_i$$

(7.2) denklemi eşik X^* düzeyine kadar, aşağıdaki denklem ise

$$(7.3) \qquad E(Y_i \mid D_i = 1, X_i, X^*) = \alpha - \beta_2 X^* + (\beta_1 + \beta_2) X_i$$

eşik X^* düzeyinin ötesindeki ortalama satış komisyonlarını vermektedir.

Böylece β_1 I inci parçadaki regresyon doğrusunun eğimini vermekte ve $(\beta_1 + \beta_2)$ yukarıdaki şekilde gösterilen parçalı doğrusal regresyonun II inci parçasındaki eğimi vermektedir. Başvurulacak bir hipotez testi ile tahmin edilen diferansiyel eğim katsayısı β_2 nin istatistiki anlamlılığını, eşik değer X^* de kırılma olup olmadığını kolaylıkla öğrenmek mümkündür[13].

8. YAPISAL FARKLILIKLARIN TEST EDİLMESİNDE KUKLA DEĞİŞKEN KULLANIMI

Ekonometrik araştırmalarda örneklem gözlemleri bireysel birimler için gerek zaman, gerek yatay kesit olarak verilerin biriktirilmesi veya derlenmesi herhangi bir modeli çözümlemek, test etmek amacıyla kullanılabilir. Farzedelim ki, Y ve X gibi iki değişken ile ilgili olarak 5 yıllık zaman dizisi verilerinin derlendiğini kabul edelim. Dizinin uzun olması belki serbestlik derecesi açısından bazı avantajlar taşıdığı söylenebilir. Fakat bu zaman mesafesi boyunca temel ekonomik yapıda birçok önemli ekonomik değişimlerin vuku bulacağı muhakkaktır ve bu değişimler de model ile açıklanmaya çalışılmaktadır. Bu tür yapısal değişmeler kısa bir dönem sürse bile önemli etkileri sözkonusu olabilir. Örneğin bir savaş halinin yaşanması, bir devrimin gerçekleşmesi ve doğal felaketlerin yaşanma-

13 Bakınız Gujarati (1979, s. 302-3); Pindyck ve Rubinfeld (1981, s. 126-27) ve İşyar (1989, s. 224-26).

sı gibi vs. Yatay kesit çalışmalarında örneklem gözlemleri veri bir zaman noktası ile ilgili olsa bile aynı yapı içerisine düşmeyebilir. Kırsal alanlarda çiftçilerin tüketim kalıpları muhtemelen şehirde çalışanların tüketim kalıplarından yapısal olarak farklılık gösterebilir. Örneğin Türkiye'deki ve Almanya'daki ailelerin geniş bir yatay kesit örneklemine sahip olduğumuzu düşünelim. Bu örneklemdeki veriler hane halklarının tasarruf davranışlarıyla ilgili gözlemlerden oluşsun. İki farklı ekonomik yapıya ait olan veriler gözlenecektir. Burada ortaya çıkan bu tür yapısal farklılıklar problemi nasıl test edilecektir? Çünkü farklı yapılar için doğru regresyon katsayıları muhtemelen farklı olacaktır ve birleştirilmiş örneklemlerden tahmin edilen katsayılar güvenilir olmayacaktır. Yapısal farklılıkların test edilmesinde en çok başvurulan testlerin başında Chow testi gelmektedir. Kukla değişken için bir test ise yine Chow testinden yararlanılarak yapılmaktadır[14].

8.1. Chow Testi

Chow Testi gerçekte aynı yapıya ait iki veya daha fazla farklı yapılara dair gözlemlerin bir veri setini alt setlere ayırarak regresyon katsayılarının tahminlerinin testini yapar. Bu test kısaca şu adımlarda yapılmaktadır:

Adım I: Tüm örneklemden $\hat{\alpha}$ ve $\hat{\beta}$ nın en küçük kareler tahminleri bulunur, Y ve X üzerine örneklem gözlemleri kullanılarak modelden bulunan parametreler yazılır, yani

$$Y_t = \alpha + \beta X_t + u_t \qquad (8.1)$$

denklemi tahmin edilerek bulunan tahminler aşağıdaki sonucun hesaplanmasında yararlanılır:

$$\Sigma \hat{u}^2_t = \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2 \qquad (8.2)$$

hesaplanan bu sonuç kalıntıların kareleri toplamını T – K serbestlik derecesiyle vermektedir (burada K= 2 dir).

Adım II: İki farklı yapının olduğu farzedilerek örneklem gözlemleri ayrı ayrı alt örneklem gruplarına ayrılır. İlk alt örneklem gruplarının büyüklükleri aynı olmayabilir. Ancak alt grupların büyüklükleri en azından parametrelerin tahmin edilebileceği kafi büyüklükte olmalıdır. $T = T_1 + T_2$ olduğundan, T_1 ve T_2 ile alt örneklemleri tasarımlayabiliriz.

14 Chow Testi için temel ekonometri kalıplarının birçoğuna bakılabilir. Bakınız Dutta (1975, s. 172-73), Stewart ve Wallis (1981, s. 199-200) ve daha detaylı bilgi için bakınız Chow (1960).

T_1 ve T_2 her bir alt örneklem için ayrı ayrı olağan en küçük kareler tahminleri bulunur. Bunun için ayrı ayrı iki regresyon denkleminden yararlanacağız:

(8.3) $$Y_{t1} = \alpha_1 + \beta_1 X_{t1} + u_{t1}$$

(8.4) $$Y_{t2} = \alpha_2 + \beta_2 X_{t2} + u_{t2}$$

burada alt imler birinci ve ikinci alt imleri göstermektedir. Alt örneklemler için kalıntıların kareleri toplamlarır hesaplanır:

(8.5) $$\Sigma \hat{u}^2_{t1} = \Sigma \left(Y_{t1} - \hat{\alpha}_1 - \hat{\beta}_1 X_{t1}\right)^2$$

(8.6) $$\Sigma \hat{u}^2_{t2} = \Sigma \left(Y_{t1} - \hat{\alpha}_2 - \hat{\beta}_2 X_{t2}\right)^2$$

bunlar sırasıyla $T_1 - K$ ve $T_2 - K$ serbestlik dereceleriyle dikkate alınmaktadır. $T_1 + T_2 - 2K$ veya $T - 2K$ serbestlik derecesiyle $\Sigma \hat{u}^2_{t1} + \Sigma \hat{u}^2_{t2}$ hesaplanır. Nihai olarak aşağıdaki kalıntıların kareleri toplamı bulunur:

(8.7) $$\Sigma u_t^{*2} = \Sigma u^2_t - \left(\Sigma \hat{u}^2_{t1} + \Sigma \hat{u}^2_{t2}\right)$$

Adım III: Son adımda aşağıda verilen F testi hesaplanır:

(8.8) $$F = \frac{\Sigma \hat{u}^{*2}_t / K}{\left(\Sigma \hat{u}^2_{t1} + \Sigma \hat{u}^2_{t2}\right)/(T - 2K)}$$

Hesaplanan bu F oranı K ve T – 2K serbestlik dereceleriyle uygun anlamlılık düzeylerinde F tablosundaki teorik değeri ile karşılaştırılır.

Boş hipotez iki farklı alt örneklemin yapılarının tahmin edilen katsayılarının eşit olduklarını vurgular:

$$H_0 \;:\; (\hat{\alpha}_1, \hat{\beta}_1) = (\hat{\alpha}_2, \hat{\beta}_2)$$

$$H_1 \;:\; (\hat{\alpha}_1, \hat{\beta}_1) \neq (\hat{\alpha}_2, \hat{\beta}_2)$$

Hesaplanan gözlem değerinin teorik (tablo) değerden küçük ise, tahmin edilen katsayılar arasındaki fark istatistiki olarak anlamlı değildir ve boş hipotezin kabul edildiği bu tartışmada iki farklı yapıda seçilen anlamlılık düzeyinde aynıdır. Kısaca iki alt örneklemden elde edilen katsayılar arasında fark yoktur. Eğer hesaplanan oran F tablosundaki teorik (tablo) değerini aşarsa, tahmin edilen regresyon katsayıları arasındaki fark seçilen anlamlılık düzeyinde farklı olduğu sonucuna varılır.

8.2. Kukla Değişken Testi

Aşağıda verilen sayısal bir örnek yardımıyla kukla değişken testini açıklamaya çalışalım. Örneğimiz farazi değerlerden oluşmaktadır. Burada önce Chow Testine başvurarak yapısal farklılık için bir test yapacağız, arkasından kukla değişken yardımıyla bu testi teyid etmeyi ele alacağız. Yukarıdaki Model (8.1) i dikkate alalım:

$$Y_t = \alpha + \beta X_t + u_t$$

aşağıdaki farazi veriler yardımıyla bu denklemi tahmin ederek Chow testini hesaplayalım:

$$\Sigma \hat{u}^2_t = \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2$$

kalıntıların kareleri toplamını tablo değerleri yardımıyla hesaplarız, buna göre:

t	Y_t	X_t	t	Y_t	X_t
1966	0.36	8.8	1975	0.59	15.5
1967	0.21	9.4	1976	0.90	16.7
1968	0.08	10.0	1977	0.95	17.7
1969	0.20	10.6	1978	0.82	18.6
1970	0.10	11.0	1979	1.04	19.7
1971	0.12	11.9	1980	1.53	21.1
1972	0.41	12.7	1981	1.94	22.8
1973	0.40	13.5	1982	1.75	23.9
1974	0.43	14.3	1983	1.99	25.2

Adım I:

$$\Sigma \hat{u}^2_t = 0.5722$$
$$\text{s. d.} = T - K = 18 - 2 = 16$$

Adım II:

$$T = T_1 + T_2$$
$$T = 1966 - 1983 = 18$$
$$T_1 = 1966 - 1974 = 9$$
$$T_2 = 1975 - 1983 = 9$$

Her bir alt örneklem için kalıntıların kareleri toplamını bulmaya çalışalım:

$$\Sigma\, \hat{u}^2_{t1} = \Sigma\, (Y_{t1} - \hat{\alpha}_1 - \hat{\beta}_1 X_{t1})^2 = 0.1396$$
$$\text{s. d.} = T_1 - K = 9 - 2 = 7$$

$$\Sigma\, \hat{u}^2_{t2} = \Sigma\, (Y_{t2} - \hat{\alpha}_2 - \hat{\beta}_2 X_{t2})^2 = 0.2167$$
$$\text{s. d.} = T_2 - K = 9 - 2 = 7$$

$$\begin{aligned}\Sigma\, \hat{u}^{*2}_t &= \Sigma\, \hat{u}^2_t - (\Sigma\, \hat{u}^2_{t1} + \Sigma\, \hat{u}^2_{t2}) \\ &= 0.5722 - (0.1306 - 0.2167) \\ &= 0.2159\end{aligned}$$

F gözlem değeri hesaplanarak teorik değerler ile karşılaştırılması yapılır:

Adım III:

$$F = \frac{\Sigma\, \hat{u}^{*2}_t / K}{(\Sigma\, \hat{u}^2_{t1} + \Sigma\, \hat{u}^2_{t2})/(T - 2K)}$$

$$= \frac{(0.2159)/2}{(0.3563)/(18 - 2(2))} = 4.24$$

$K = 2$ ve $T - 2K = 14$ serbestlik dereceleriyle yüzde 95 güven düzeyinde F nin tablo (teorik) değeri 3.74'dür. Her iki F değerini karşılaştırdığımız zaman gözlenen F değeri teorik F değerinden daha büyük olduğu görülür. Dolayısıyla tahmin edilen regresyon katsayıları arasındaki fark seçilen anlamlılık düzeyinde farklıdır, yani T_1 ve T_2 alt örneklemleri % 95 güven düzeyinde iki farklı yapıdan çekildiğini bize test sonucu vermektedir. Bu nedenle genel regresyon medeli içine alan örneklem dönemi durağan bir yapıya sahip değildir ve modelin tahmin edilen katsayılarına pek güven duyulmaz.

Şimdi aynı veri setini kullanarak kukla değişken yardımıyla yapısal farklılığı incelemeye çalışalım. Gujaratinin kurup geliştirdiği modeli burada biz de ele alarak tahmin etmeye çalışalım[15]: Gujaratinin önerdiği model şöyledir:

$$(8.9) \qquad Y_t = \alpha + \beta_1 D + \beta_2 X_t + \beta_3 (DX)_t + u_t \qquad t = 1, 2, 3, \ldots\ldots, T$$

15 Bu konuda daha fazla bilgi için bakınız Dutta (1975, s. 175-7) ve ayrıca bakınız Gujarati (1970), Gujarati (1970a).

burada D = 1 eğer gözlem ikinci alt örnekleme ait ise,
= 0 eğer gözlem birinci alt örnekleme ait ise.

Eğer kukla değişken D için tersi bir durum alınsa idi sonuç yine etkilenmiyecekti, yani birinci alt örneklem için 1, ikinci alt örneklem için 0 değerleri de verilebilirdi. Yukarıda verilen Model (8.9) u tabloda yer alan verileri kullanarak tahmin ettiğimizde

$$(8.10) \qquad Yt = \underset{(0.3333)}{-0.2663} - \underset{(0.4703)}{1.4839D} + \underset{(0.0290)}{0.0470X_t} + \underset{(0.0332)}{0.1034(DX)_t}$$

sonucunu buluruz. (Denklemde parametrelerin altındaki değerler standart hataları göstermektedir). Burada yüzde 95 güven düzeyinde kesme terimi hariç diğerleri istatistiki olarak anlamlıdır.

Kukla değişken tanımından hareket ederek yukarıda bulunan tahmini şöyle yorumlayabiliriz: Örneklem dönemi T_2 (ikinci alt örneklem) için regresyon denklemi

$$(8.11) \qquad E(Y_t) = (-0.2663 - 1.4839(1)) + (0.0475X_t + 0.1034(1.X)_t$$
$$= -1.7502 + 0.1504\,X_t$$

beklenen değeri verecek ve T_1 dönemi (birinci alt örneklem) için

$$(8.12) \qquad E(Y_t) = -0.2663 + 0.0470X_t$$

beklenen değeri verecektir.

Dikkat edilirse son iki regresyon tahmininde hem eğim katsayısı, hem de kesme terimi iki alt örneklem için farklı anlamlılıklar vermektedir. Dolayısıyla buradaki bulgular Chow testinden elde edilen sonuçları teyid etmektedir.

KAYNAKLAR

Chow, G.C.; "Tests of Equality Between Sets of Coefficients in Two Linear Regression", *Econometrica*,July, 28(3), ss. 91-605, 1960.

Dutta, M.; *EconometricMethods*, Cincinati: South-Western Publ. Comp., 1975.

Goldberger, A.S.; *Econometric Theory*, New York, John Wiley, 1964.

————; *Topics in Regression Analysis*, New York: The Macmillan Comp., 1968.

Goldfeld, S.M. ve R.E. Quandt; *Nonlinear Methods in Econometrics*, Amsterdam: North-Holland Publishing, 1972.

Gujarati, D.; "Use of Dummy Variables in Testing for Equality Between Sets of Coefficients in Two Linear Regression: A Note," *The American Statistician*, 24(1), ss. 50-2, 1970.

————; "Use of Dummy Variables in Testing for Equality for Between Sets of Coefficient in Linear Regressions: A Generalization", *The American Statistician*, 24(5), ss. 18-21, 1970a.

————; *BasicEconometrics*, London: International Student edit., 1979.

Intriligator, M.D.; *Econometrics Models, Techniques and Applications*, Amsterdam: North-Holland Publ. Comp., 1978.

İşyar, Y.; *EkonometrikModellerI-II*, Bursa (Basılmamış Ders Notları - teksir), 1989.

Johston, J.; *Econometric Methods*, New York: McGraw-Hill Book Company, 1972.

Kmenta, J.; *ElementsofEconometrics*, New York: The Mcmillian Comp., 1971.

Malinvaud, M.D.; *Statistical Methods of Econometrics*, Third Edition, Amsterdam: North-Holland Publishing Comp., 1980.

Pindyck, S. ve D.L. Rubinfeld; *EconometricModelsandEconomicsForecasts*, London: McGraw-Hill International Book Comp., 1981.

Serper, Ö.; *Uygulamalı İstatistik - 2*, İstanbul: Filiz Kitapevi, 1986.

Stewart, M.B. ve K.F. Wallis; *Introductory Econometrics*, Oxford: Billing and Sons Ltd., 1981.