Bölüm 10

Kukla Değişkenlerle Bağlanım

10.1 Nitel Değişkenlerle Bağlanım

- Bağlanım çözümlemelerinde bağımlı değişken, sayısal büyüklükler yanında nitel değişkenlerden de etkilenebilir.

  Nicel Değişkenler

  Gelir, üretim, fiyat, maliyet, enflasyon, işsizlik oranı, yaş, boy, çocuk sayısı, ...

  Nitel Değişkenler

  Cinsiyet, ırk, din, savaş, coğrafi bölge, hükümet politikalarında değişme, grev, ...

- Görgül çalışmalarda karşılaşılan pek çok sorunu çözmede nitelik bildiren *"kukla"* (dummy) değişkenlerden yararlanılır.
- Bu değişkenler aynı zamanda *"nitel"* (qualitative) değişken, *"ulamsal"* (categorical) değişken veya *"gösterge"* (indicator) değişkeni olarak da adlandırılmaktadırlar.
- Kukla değişkenler bir veri sınıflandırma aracıdır.
- Nitel özellikleri nicel olarak gösterebilmek için, niteliğin varlık ya da yokluğunu gösteren 1 ve 0 değerlerini alırlar.
- Örnek olarak bir kimsenin üniversite mezunu olduğu 1 ile, olmadığı ise 0 ile gösterilebilir.

- Kuklaların mutlaka 0 ya da 1 değerleri almaları gerekmez. Doğrusal ilişkili herhangi bir sayı çifti kullanılabilir.
- Diğer yandan, yorumlamada sağladığı kolaylıktan dolayı uygulamada {0, 1} çifti yeğlenmektedir.
- Bağlanım modellerinde kukla değişkenlerin kullanılması ek bir zorluk getirmemektedir.
- Kukla içeren bağlanımların hesaplanması bilindik şekilde olur. Katsayıların anlamlılığı da t istatistiği ile sınanabilir.

10.1.1 VARÇÖZ Modelleri

- Bir bağlanım modelindeki tüm açıklayıcı değişkenlerin birer kukla değişken olması mümkündür.
- Böyle modellere *"VARÇÖZ"* (ANOVA) modelleri de denir.
- VARÇÖZ modelleri özellikle toplumbilim, psikoloji, eğitim, pazar araştırması gibi alanlarda yaygındır.

Örnek olarak, Şubat-Mart 2011 döneminde Ankara Çankaya'da satılığa çıkartılan ikinci el daire fiyatlarının üç ayrı semtte nasıl farklılık gösterdiğini inceleyen aşağıdaki modeli ele alalım.

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i$$

- Y burada satış ilanı verilen apartman dairesinin 1000 TL olarak fiyatıdır.
- Kukla değişkenler ise D ile gösterilmektedir.
- $D_{2i} = 1$, daire Dikmen'de ise; $D_{2i} = 0$, eğer değilse. $D_{3i} = 1$, daire Kavaklıdere'de ise; $D_{3i} = 0$, eğer değilse.
- *Not:* Eğer $D_{2i} = 0$ ve $D_{3i} = 0$ olursa daire Kavaklıdere ya da Dikmen'de değil, üçüncü seçenek olan Cebeci semtinde bulunuyor demektir.
- Bu modelin bize gösterdiği şey şudur:

$$\begin{aligned} \text{Dikmen'de ortalama fiyat:} \quad & E(Y_i|D_{2i} = 1, D_{3i} = 0) = \beta_1 + \beta_2 \\ \text{Kavaklıdere'de ortalama fiyat:} \quad & E(Y_i|D_{2i} = 0, D_{3i} = 1) = \beta_1 + \beta_3 \\ \text{Cebeci'de ortalama fiyat:} \quad & E(Y_i|D_{2i} = 0, D_{3i} = 0) = \beta_1 \end{aligned}$$

- Bağlanıma ilişkin tahmin sonuçları aşağıdaki gibidir:

$$\begin{array}{lllll} \hat{Y}_i = & 91{,}4615 & +\,47{,}2746\, D_{2i} & +\,94{,}1218\, D_{3i} & \\ \text{öh} & (11{,}6983) & (13{,}6480) & (16{,}8850) & \\ t & (7{,}8184) & (3{,}4638) & (5{,}5743) & R^2 = 0{,}3491 \end{array}$$

- D_2 ve D_3 anlamlı olduğu için, Ankara'nın bu üç semtinde daire fiyatlarının farklı olduğunu söyleyebiliriz.
- Buna göre Cebeci'deki ortalama bir dairenin fiyatı yaklaşık 91 bin TL iken, Dikmen ve Kavaklıdere'deki ortalama fiyat ise sırası ile $91 + 47 \approx 140$ ve $91 + 94 \approx 185$ bin liradır.

Birkaç Önemli Nokta

Kukla değişken kullanımı ile ilgili dikkat edilmesi gereken bazı noktalar şunlardır:

1. Bir nitel değişkende m sayıda sınıf ya da *"ulam"* (category) varsa, $(m - 1)$ kukla değişken kullanılmalıdır. Aksi halde *"kukla değişken tuzağı"* (dummy variable trap) denilen *"tam eşdoğrusallık"* (exact collinearity) oluşur. Ancak, sıfır noktasından geçen bağlanımlarda ulam sayısı kadar kukla değişken koymak mümkündür.
2. "Dikmen" ve "Dikmen değil" gibi seçeneklere hangi değerin atanacağı isteğe bağlıdır. Eğer Dikmen değil $= 1$ olursa β_2 katsayısı da eksi çıkar. Demek ki kukla değişken içeren modelleri yorumlarken, 1 ve 0 değerlerinin nasıl verildiğini bilmek önemlidir.
3. 0 değeri verilen ulam, *"yazın"* (literature) içerisinde farklı adlarla karşımıza çıkabilmektedir:

Türkçe	İngilizce
"Taban ulam"	(Base category)
"Kıyas ulamı"	(Benchmark category)
"Karşılaştırma ulamı"	(Comparison category)
"Denetim ulamı"	(Control category)
"Gönderi ulamı"	(Reference category)
"Atlanan ulam"	(Omitted category)

Örneğimizde taban ulam Cebeci semtidir. Taban ulam, diğerlerini karşılaştırmada kullanılan sınıftır. Taban ulamı gösteren ya da ölçen terim de β_1 sabit terimidir. D_2 ve D_3 kukla değişkenlerine gelen β_2 ve β_3 katsayılarına ise *"sabit terim farkı"* (constant term difference) adı verilir.

10.1.2 KOVÇÖZ Modelleri

- Birçok iktisadi araştırmada, yalnızca kukla değişkenlerin kullanıldığı VARÇÖZ modellerine çok sık rastlanmaz.
- Bunun yerine nitel ve nicel değişkenlerin birlikte olduğu *"kovaryans çözümlemesi"* (analysis of covariance) ya da *"KOVÇÖZ"* (ANCOVA) modelleri yeğlenir.
- Bu modellerde nicel değişkenlere *"denetim değişkeni"* (control variable) de denir.

KOVÇÖZ modellerine bir örnek olarak Ankara'daki daire fiyatları örneğimizi geliştirelim.

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + u_i$$

- Y_i burada satılığa çıkartılan apartman dairesinin fiyatıdır.
- X_i ise dairenin metrekare cinsinden büyüklüğüdür.
- $D_{2i} = 1$, daire Dikmen'de ise; $D_{2i} = 0$, eğer değilse. $D_{3i} = 1$, daire Kavaklıdere'de ise; $D_{3i} = 0$, eğer değilse.
- *Not:* Cebeci'yi taban olarak ulamlandırmayı sürdürüyoruz.
- Bağlanım tahminleri aşağıdaki gibidir:

$$\begin{array}{lllll}
\hat{Y}_i = & -27{,}7639 & +\ 15{,}1179\ D_{2i} & +\ 72{,}4160\ D_{3i} & +\ 1{,}24994\ X_i \\
\text{öh} & (15{,}6787) & (9{,}7259) & (11{,}4113) & (0{,}1431) \\
t & (-1{,}7708) & (1{,}5544) & (6{,}3460) & (8{,}7354) \\
R^2 = & 0{,}7217 & & &
\end{array}$$

- Yeni modeldeki sabit terim fark katsayısının Kavaklıdere için anlamlıyken Dikmen için anlamlı olmadığını görüyoruz.
- Diğer bir deyişle, metrekare sabitken, Dikmen'deki daire fiyatlarının Cebeci ile aynı olduğu reddedilmemekte, Kavaklıdere için ise 72 bin liralık bir fark gözlenmektedir.
- Bulguların Dikmen için farklı çıkmasının nedeni, ilk baştaki modelde X'i hesaba katmamış olmamızdır.
- Sonuçlar Ankara'nın bu üç semtindeki dairelerin metrekare fiyatının yaklaşık 1250 lira olduğunu göstermektedir.
- Burada sabit terimleri farklı olan ancak aynı β_4 eğimini paylaşan 3 farklı bağlanımı ele aldığımıza dikkat ediniz.

10.2 Kukla Değişken Kullanım Şekilleri

10.2.1 Chow Sınamasının Kukla Almaşığı

- Önceki örnekte, nitel değişkenlerin sabit terimi etkilediği ama eğim katsayısını etkilemediği varsayılmıştı.
- Diğer yandan, eğer farklı ulamların eğim katsayısı da farklı ise sabit terim farklarını sınamanın pek anlamı yoktur.
- Birden fazla bağlanımın aynı olup olmadığını sınamak için çok adımlı Chow sınamasının kullanılabildiğini biliyoruz.
- Farklı bağlanımları sabit terimler, eğimler ya da her ikisi yönünden ayırt edebilen daha genel bir sınama yöntemi kukla değişkenler ile olanaklıdır.
- Türkiye için tüketim harcamaları ve milli gelir verilerimizi anımsayalım:

Çizelge: Türkiye'de Tüketim ve GSYH (1987–2006)

Yıl	C	Y	Yıl	C	Y
1987	51.019	74.416	1997	77.620	112.892
1988	51.638	76.143	1998	78.113	116.541
1989	51.105	76.364	1999	76.077	111.083
1990	57.803	83.371	2000	80.774	119.147
1991	59.366	84.271	2001	73.356	110.267
1992	61.282	88.893	2002	74.894	118.923
1993	66.545	96.391	2003	79.862	125.778
1994	62.962	91.600	2004	87.897	137.110
1995	66.011	97.729	2005	95.594	147.200
1996	71.614	104.940	2006	100.584	156.249

Türkiye'deki 1994 krizini anımsayalım. Verileri 1994 öncesi ve sonrası olarak ikiye ayıralım ve şu iki modeli inceleyelim:

$$\begin{aligned}
&\text{1987-1993 dönemi: } Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}, \;\; n_1 = 7 \\
&\text{1994-2006 dönemi: } Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}, \;\; n_2 = 13
\end{aligned}$$

Yukarıdaki iki model dört farklı olasılık sunmaktadır:

1. Eğer $\lambda_1 = \gamma_1$ ve $\lambda_2 = \gamma_2$ ise, iki bağlanım sabit terim ve eğim olarak aynıdır: *"Çakışan"* (coincident) bağlanımlar.
2. Eğer $\lambda_1 \neq \gamma_1$ ve $\lambda_2 = \gamma_2$ ise, iki bağlanım yalnızca sabit terimler yönünden farklıdır: *"Koşut"* (parallel) bağlanımlar.
3. Eğer $\lambda_1 = \gamma_1$ ve $\lambda_2 \neq \gamma_2$ ise, iki bağlanım aynı sabit terimli ama farklı eğimlidir: *"Uyumlu"* (concurrent) bağlanımlar.

4. Eğer $\lambda_1 \neq \gamma_1$ ve $\lambda_2 \neq \gamma_2$ ise, iki bağlanım bütünüyle farklıdır: *"Benzemez"* (dissimilar) bağlanımlar.

- Elimizdeki iki modeli karşılaştırabilmek için tüm n_1 ve n_2 gözlemlerini toplayıp aşağıdaki bağlanımı tahmin edelim:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

- $E(u_t) = 0$ varsayımı ile şu iki bağlanımı buluruz:

$$\begin{aligned} E(Y_t|D_t = 0, X_t) &= \alpha_1 + \beta_1 X_t \\ E(Y_t|D_t = 1, X_t) &= (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_t \end{aligned}$$

- Y_t ve X_t farklı yıllar için tüketim ve geliri göstermektedir.
- $D_t = 0$ 1994 öncesi, $D_t = 1$ ise 1994 ve sonrası dönemdir.
- α_2 sabit terim farkıdır.
- β_2 ise eğim katsayısı farkı olup, ikinci dönem işlevinin eğim katsayısının ilk ya da temel döneme ait eğim katsayısından ne kadar farklı olduğunu gösterir.
- Model tahmini şu sonuçları vermektedir:

$$\begin{array}{lllll} \hat{Y}_t = & -4{,}7884 & +\,16{,}2163\ D_t & +\,0{,}7455\ X_t & -\,0{,}1796\ D_t X_t \\ \text{öh} & (6{,}9547) & (7{,}5961) & (0{,}0836) & (0{,}0874) \\ t & (-0{,}6885) & (2{,}1348) & (8{,}9146) & (-2{,}0556) \\ R^2 = & 0{,}9887 & & & \end{array}$$

- Buna göre 1987-94 dönemi tasarruf-gelir bağlanımı şudur:

$$\hat{Y}_t = -4{,}7884 + 0{,}7455\ X_t$$

- 1994-2006 dönemi tasarruf-gelir bağlanımı ise şöyledir:

$$\begin{aligned} \hat{Y}_t &= (-4{,}7884 + 16{,}2163) + (0{,}7455 - 0{,}1796)\ X_t \\ &= 11{,}4279 + 0{,}5659\ X_t \end{aligned}$$

- Sabit terim farkı ve eğim farkının her ikisinin de istatistiksel olarak anlamlı bulunması, bu iki bağlanımın "benzemez" olduğunu göstermektedir.

Kukla değişken yönteminin Chow sınamasına üstünlükleri şunlardır:

1. Kukla değişken yaklaşımı, tek bir bağlanım tahmini içerdiği için uygulama yönünden basittir.
2. Kukla değişkenler, iki bağlanımın farklı olup olmadığının yanı sıra farkın sabit terimden mi yoksa eğimden mi kaynaklandığını da göstermektedir.
3. Tek bağlanım olması önsav sınamalarında kolaylık sağlar.
4. Verilerin bir arada kullanılması serbestlik derecesini arttırır. *Dikkat:* Modele eklenen her kukla değişkenin serbestlik derecesini bir azalttığı unutulmamalıdır.

10.2.2 Karşılıklı Etkileşim

Kukla değişkenlerin bir diğer kullanım alanı da açıklayıcı değişkenler arası karşılıklı etkileşimi incelemektir.
Ankara örneğimize dönelim ve şimdi de şu modeli ele alalım:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i$$

- Y_i burada evin fiyatını, X_i ise m^2 alanını göstermektedir.
- $D_{2i} = 1$, kot daire ise; $D_{2i} = 0$, eğer değilse. $D_{3i} = 1$, su deposu bulunuyorsa; $D_{3i} = 0$, eğer yoksa.
- Model tahmin sonuçları aşağıdaki gibidir.

$$\begin{array}{lccccc} \hat{Y}_i & = & 1{,}2103 & -\ 46{,}2989\ D_{2i} & +\ 20{,}4479\ D_{3i} & +\ 1{,}2023\ X_i \\ \text{öh} & & (19{,}0367) & (12{,}8606) & (9{,}7909) & (0{,}1633) \\ t & & (0{,}0636) & (-3{,}6001) & (2{,}0885) & (7{,}3639) \\ R^2 & = & 0{,}5942 & & & \end{array}$$

- Bulgular kot dairelerin yaklaşık 46 bin lira ucuz olduğunu, apartmanda su deposu bulunmasının ise ortalama daire fiyatını yaklaşık 20 bin TL yükselttiğini göstermektedir.
- Tahmin etmiş olduğumuz modeldeki üstü kapalı varsayım, D_2 ve D_3'ün fark etkilerinin birbirinden bağımsız olduğudur.
- Diğer bir deyişle, su deposu olsa da olmasa da kot dairenin fark etkisinin aynı olduğu kabul edilmektedir.
- Belli bir uygulamada bu varsayım savunulamayabilir.

- D_2 ve D_3 gibi iki ayrı nitel değişken arasında var olabilecek karşılıklı etkileşim şu şekilde ele alınır:

$$\hat{Y}_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i$$

- Burada
  α_2 kot dairenin fark etkisini,
  α_3 su deposu bulunmasının fark etkisini,
  α_4 kot daire ve su deposu olmasının birlikte fark etkisini
  göstermektedir.

- Karşılıklı etkileşimi öneren model tahminleri şöyledir:

$$\begin{array}{llllll}
\hat{Y}_i = & 1{,}1340 & -\,44{,}7608\, D_{2i} & +\,21{,}1225\, D_{3i} & -\,4{,}3179\, (D_{2i}D_{3i}) & +\,1{,}2014\, X_i \\
\text{öh} & (19{,}2074) & (16{,}1315) & (10{,}7339) & (26{,}9206) & (0{,}1648) \\
t & (0{,}0590) & (-2{,}7747) & (1{,}9678) & (-0{,}1604) & (7{,}2912) \\
R^2 = & 0{,}5944 & & & &
\end{array}$$

- *"Etkileşim kuklası"* (interaction dummy) α_4'ün istatistiksel olarak anlamlı olup olmadığı yine t sınamasıyla bulunabilir.

- Sonuçlar, bir apartmanda su deposu bulunmasının kot daire fiyatlarını da diğer daireler ile aynı şekilde artırdığını göstermektedir.

10.2.3 Parça-Yollu Doğrusal Bağlanım

- Kukla değişkenlerin bir diğer kullanım alanı da *"parça-yollu bağlanım"* (piecewise regression) modelleridir.

- Bu modellere yönelik olarak, Ankara'daki satılık daireler örneğimizdeki fiyat-metrekare ilişkisini göz önüne alalım.

- Daire fiyatlarının *"eşik"* (threshold) düzeyi denilen bir X^* değeri öncesinde ve sonrasında farklı şekilde değiştiğini varsayalım.

- Örnek olarak, daire fiyatları metrekareye göre doğrusal olarak artsın ancak X^* eşik düzeyinden sonra daha dik bir eğimle artıyor olsun.

- Buna göre, elimizdeki model iki farklı parçadan oluşan bir doğrusal bağlanım modelidir.

- Bu tür modeller daha genel bir tür olan *"kama işlevleri"* (spline functions) yaklaşımına bir örnektir.

- Parça-yollu bağlanımı açıklamak için şu modele bakalım:

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i$$

- Y_i burada dairenin fiyatını, X_i de metrekare genişliğini göstermektedir.
- X^* değeri genişliğin eşik düzeyidir ve önceden bellidir.
- $D_i = 1$, eğer $X_i \geq X^*$ ise;$D_i = 0$, eğer $X_i < X^*$ ise.
- $E(u_i) = 0$ varsayımı altında şunu görebiliriz:

Eşiğe kadar: $E(Y_i|D_i = 0, X_i, X^*) = \alpha_1 + \beta_1 X_i$
Eşik sonrası: $E(Y_i|D_i = 1, X_i, X^*) = \alpha_1 - \beta_2 X^* + (\beta_1 + \beta_2) X_i$

- Buna göre β_1 parça-yollu bağlanımın birinci parçasının, $(\beta_1 + \beta_2)$ ise ikinci parçasının eğimini vermektedir.
- Kırılma yoktur diyen önsav için $\hat{\beta}_2$'nın p değerine bakılır.
- Verilerden, $X^* = 120m^2$ sonrasında fiyatların değişiyor olabileceğini çıkardığımızı varsayalım.
- Fiyat (Y) ve genişlik (X) verilerini bir parça-yollu doğrusal bağlanım modeline yakıştırırsak şu bulguları elde ederiz:

$\hat{Y}_i =$	$-0{,}4698$	$+\ 1{,}1967\ X_i$	$+\ 0{,}2490\ (X_i - X^*)D_i$	
öh	(33,2861)	(0,3203)	(0,5660)	
t	$(-0{,}0141)$	(3,7365)	(0,4400)	$R^2 = 0{,}4822$

- Dairelerin metrekare fiyatı yaklaşık 1200 TL kadardır.
- 120 metrekare üstünde fiyat $(1200 + 250)$ olmakla birlikte, aradaki fark istatistiksel olarak anlamlı değildir.
- Öyleyse $(X - X^*)D$ değişkeni modelden çıkartılabilir.

10.3 Kukla Değişkenlere İlişkin Konular

10.3.1 Mevsimsel Çözümlemeler

- Günlük, aylık ya da üç aylık verilere dayanan birçok zaman serisi; *"mevsimsel örüntü"* (seasonal pattern) ya da *"düzenli salınımsal hareket"* (regular oscillatory movement) gösterir.
- Buna örnek olarak yeni yıl öncesi mağaza satışlarını ya da bayram öncesi hanelerin artan para talebini gösterebiliriz.
- Bir zaman serisinin çeşitli bileşenleri üzerinde ayrı ayrı yoğunlaşmak için, mevsimsel bileşenin çıkarılması istenir.
- TÜFE ve ÜFE gibi önemli iktisadi zaman serileri genellikle *"mevsimsel ayarlamalı"* (seasonally adjusted) yayınlanır.
- *"Mevsimsellikten arındırma"* (deseasonalization) işleminin çeşitli yolları vardır ve bunlardan birisi de kukla değişkenler yöntemidir.
- Konuya ilişkin olarak, Türkiye'de inşaat kesimi için üç aylık üretim ve toplam maliyet endeksleri verilerini ele alalım.

Çizelge: İnşaat Kesiminde Üretim ve Maliyet (2005=100)

Dönem	Üretim	Maliyet	Dönem	Üretim	Maliyet
2005Ç1	78,07	98,44	2008Ç1	100,01	138,79
2005Ç2	105,47	98,65	2008Ç2	128,49	153,78
2005Ç3	114,56	100,97	2008Ç3	128,62	142,16
2005Ç4	101,90	101,93	2008Ç4	105,08	136,62
2006Ç1	90,34	105,63	2009Ç1	81,24	135,44
2006Ç2	126,56	118,86	2009Ç2	101,53	136,62
2006Ç3	136,43	119,73	2009Ç3	106,09	137,37
2006Ç4	120,15	119,72	2009Ç4	97,61	137,47
2007Ç1	101,83	124,53	2010Ç1	88,76	142,26
2007Ç2	135,27	125,75	2010Ç2	121,29	142,77
2007Ç3	142,52	125,89	2010Ç3	128,60	145,52
2007Ç4	120,08	126,56	2010Ç4	115,21	147,81

- İnşaat üretim faaliyetlerinde mevsimsel bir etki olup olmadığını görmek için aşağıdaki modeli inceleyelim:

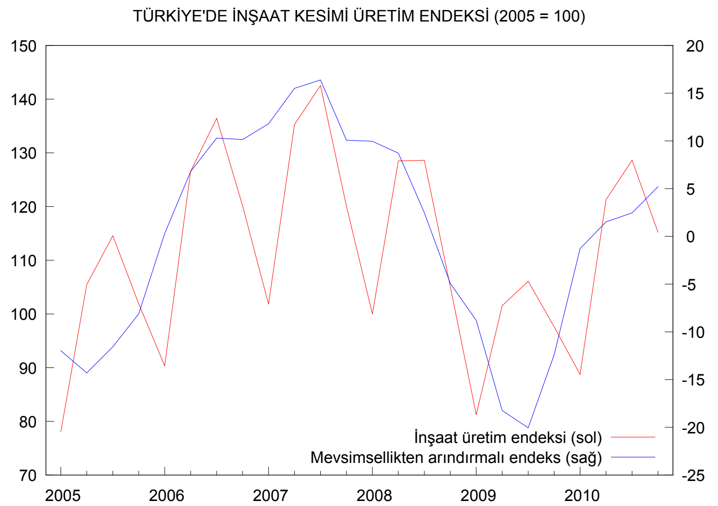
$$\hat{Y}_t = \alpha_1 + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_i$$

- $D_{2t} = 1$, ikinci üç ay ise; $D_{2t} = 0$, eğer değilse. $D_{3t} = 1$, üçüncü üç ay ise; $D_{3t} = 0$, eğer değilse. $D_{4t} = 1$, dördüncü üç ay ise; $D_{4t} = 0$, eğer değilse.

- Burada mevsim değişkeni dört ulamdan oluştuğu için üç farklı kukla değişken kullanılmıştır.
- Nicel bir değişken kullanmayıp, Y_t'nin yalnızca sabit terime göre bağlanımını hesapladığımıza dikkat ediniz.
- Tahmin sonuçları aşağıdaki gibidir:

$$\begin{array}{lllll}
\hat{Y}_t = & 90{,}0422 & +\ 29{,}7242\ D_{2t} & +\ 36{,}0968\ D_{3t} & +\ 19{,}9633\ D_{4t} \\
\text{öh} & (4{,}7960) & (6{,}7826) & (6{,}7826) & (6{,}7826) \\
t & (18{,}7744) & (4{,}3824) & (5{,}3220) & (2{,}9433) \\
R^2 = & 0{,}6183 & & &
\end{array}$$

- İnşaat üretim endeksinin taban dönem olan kış döneminde ortalama 90 düzeyinde olduğunu görüyoruz.
- Endeks bahar döneminde 30 puan yükselmekte, yazın bu yükselişini sürdürmekte, ve güz döneminde gerilemektedir.
- Yukarıdaki bağlanım tahminine ait kalıntılar, inşaat üretim endeksinin mevsimsellikten arındırmalı bir serisini verir.
- Bu kalıntılardan daha sonra serinin *"eğilim bileşeni"* (trend component), *"çevrimsel bileşen"* (cyclical component) ve *"rastsal bileşen"* (random component) unsurları bulunabilir.
- *Dikkat:* Sözü edilen bu mevsimsellikten arındırma işlemi her zaman serisi için uygun değildir.



- Şimdi, önceki yılın aynı dönemine ilişkin toplam maliyeti (X_{t-4}) bir nicel değişken olarak modele ekleyelim.Tahmin sonuçları şöyledir:

$$\begin{array}{lllllll}
\hat{Y}_t & = 145{,}4342 & +\ 32{,}9016\ D_{2t} & +\ 38{,}0662\ D_{3t} & +\ 20{,}9031\ D_{4t} & -\ 0{,}4396\ X_{t-4} \\
\text{öh} & (15{,}7171) & (5{,}6225) & (5{,}5994) & (5{,}5901) & (0{,}1262) \\
t & (9{,}2533) & (5{,}8517) & (6{,}7983) & (3{,}7393) & (-3{,}4831) \\
R^2 = & 0{,}8020
\end{array}$$

- İkinci, üçüncü, ve dördüncü çeyreklerdeki üretimin ilk çeyrekten yüksek olduğunu bu modelde de görüyoruz.
- Ayrıca, mevsimsel etkiler gözönüne alındığında, maliyet endeksindeki 1 puanlık artışın üretim endeksinde yaklaşık 0,44 puanlık bir azalmaya yol açtığı anlaşılmaktadır.
- Bu noktada ilginç bir soru X_t'nin de Y_t gibi bir mevsimsel örünüm sergileyip sergilemediği sorusudur.
- Az önce ele almış olduğumuz modelin bir özelliği, bağımlı değişken Y_t'yi mevsimsellikten arındırırken aynı zamanda X_t'yi de mevsimsellikten arındırmasıdır.
- Bunu görmek için ilk bağlanımı tahmin edelim ve kalıntıları saklayalım. Bu, mevsimsellikten arındırmalı $Y_{2,t}$ olsun.
- Şimdi de aynı modeli bu sefer de maliyet bağımlı değişken olacak şekilde tahmin edip kalıntıları saklayalım. Bu da mevsimsellikten arındırmalı $X_{2,t}$ olsun.
- $Y_{2,t}$ ve $X_{2,t-4}$ bağlanıma birlikte sokulursa, $X_{2,t-4}$'ün eğim katsayısının önceki beş değişkenli bağlanımdaki X_{t-4} ile aynı olduğu görülür. Yani bir taşla iki kuş vurmuş oluyoruz.

10.3.2 Yarı-Logaritmasal İşlevler

- Eğitim deneyimi (yıl) ve cinsiyete (1 = erkek) göre öğretim görevlisi işe başlama ücretlerini (yıllık, bin dolar) gösteren şu varsayımsal verileri ele alalım:

Ücret	Deneyim	Cinsiyet
23,0	1	1
19,5	1	0
24,0	2	1
21,0	2	0
25,0	3	1
22,0	3	0
26,5	4	1
23,1	4	0
25,0	5	0
28,0	5	1
29,5	6	1
26,0	6	0
27,5	7	0
31,5	7	1
29,0	8	0

- Verileri şu log-doğ modeline yakıştırmak istiyor olalım:

$$\ln Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + u_i$$

- Y_i başlama ücreti, X_i ise eğitim deneyimidir.
- $D_i = 1$, eğer erkekse; $D_i = 0$, kadınsa.
- β_2 katsayısı burada X_i'deki bir birimlik değişmeye karşılık Y_i'deki göreli değişmeyi göstermektedir.
- Göreli değişme 100 ile çarpılır ise yüzde değişme olur.
- Ancak yukarıdaki açıklama, değişkenin yalnızca sürekli bir değişken olması durumunda geçerlidir.
- Kukla değişkenin ortalama Y_i'deki göreli etkisini bulmak için, tahmin edilen $\hat{\beta}_3$ katsayısının e tabanına göre ters logaritmasının alınması ve bundan 1 çıkartılması gerekir.
- Örnekteki modeli tahmin edersek şunu buluruz:

$$\begin{array}{llll} \widehat{\ln Y_i} = & 2{,}9298 & +0{,}0546\,X_i & +0{,}1341\,D_i \\ t & (481{,}524) & (48{,}3356) & (27{,}2250) \\ & & & R^2 = 0{,}9958 \end{array}$$

- Buna göre cinsiyet farkı dikkate alındığında ortalama işe başlama ücreti, deneyim yılı başına % 5,46 artmaktadır.
- Ancak, D_i'nin katsayısına bakarak ücretlerin erkekler için yüzde 13,41 daha fazla olduğunu söylemek doğru olmaz.

- 0,1341'in ters logaritması alınır ve bundan 1 çıkartılırsa 0,1435 bulunur. Demek ki erkek öğretim görevlisi ücretleri kadınlara göre yüzde 14,35 daha yüksektir.

10.3.3 İleri Çalışma Konuları

Rastsal Değiştirge Modelleri

- Ele almış olduğumuz modellerde β anakütle katsayılarının bilinmeyen ama sabit büyüklükler olduğunu anımsayalım.
- Kukla değişkenlere ilişkin ileri konulardan biri de *"rastsal değiştirge"* (random parameter) modelleridir.
- Yazında çeşitli biçimlerde karşımıza çıkan bu modeller, β değiştirgelerinin de rastsal olduğunu varsayar.

Değiştirilen Bağlanım Modelleri

- İki bağlanımın hem sabit terim farkı hem de eğim farkı kullanılarak karşılaştırıldığı kukla değişken modellerinde, kırılma noktasının bilindiği örtük olarak varsayılır.
- Diğer yandan, kırılma noktasının örneğin 1994'te mi ya da başka bir dönemde mi olduğu çoğu zaman bilinemez.
- Dolayısıyla, bir diğer ileri çalışma konusu da *"değiştirilen bağlanım"* (switching regression) modelleridir.
- Bu modeller, kırılma noktasının da rastsal olmasına izin vererek bağlanımın *"yinelemeli"* (iterative) olarak tahmin edilmesini sağlarlar.

Dengesizlik Modelleri

- Pazarın dengeye gelmediği, arzın talebe eşit olmadığı durumlar için özel tahmin yöntemleri gerekir.
- Örnek olarak bir malın talebi, fiyat ve çeşitli değişkenlerin bir işlevi olarak modellenirken, aynı malın arzı da yine fiyat ve diğer değişkenlerin bir işlevi olarak modellenebilir.
- Arzda yer alan değişkenler taleptekilerden farklı olursa, gerçekte alınıp satılan mal miktarı arzın talebe eşitlendiği noktada olmayabilir ve bu da dengesizliğe yol açar.

- İşte böyle durumları kukla değişkenler yardımıyla ele alan modellere de *"dengesizlik"* (disequilibrium) modelleri denir.

Farklıserpilimsellik ve Özilintinin Etkisi

- Türkiye'de 1994 sonrası tüketimde yapısal bir değişiklik olup olmadığını inceleyen örneği anımsayalım:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_i$$

- Kukla değişken kullanılan böyle bir modelde örtük olarak *"aynıserpilimsellik"* (homoscedasticity), diğer bir deyişle var(u_{1i}) = var$(u_{2i}) = \ldots = \sigma^2$ varsayımı söz konusudur.
- Eğer bu varsayım sağlanamıyorsa tutarsız sonuçlar elde edilmesi olasıdır.
- Öyleyse, kukla değişkenli modellerde *"farklıserpilimsellik"* (heteroscedasticity) sorununun olmadığı doğrulanmalıdır. (*Not:* Bunun için Chow yerine Wald sınaması yapılabilir.)
- Bu tür modellerde özilinti olmadığı varsayımı da önemlidir. Bu konu daha sonra ele alınacaktır.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 9* “Dummy Variable Regression Models” okunacak.

Önümüzdeki Ders
Doğrusal Bağlanım Modeline Dizey Yaklaşımı

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011