Bölüm 8

Çoklu Bağlanım Çözümlemesi - Tahmin Sorunu

8.1 Üç Değişkenli Model

8.1.1 Gösterim ve Varsayımlar

- Önceki bölümlerde bağımlı değişken Y'nin yalnızca bir açıklayıcı değişken X tarafından etkilendiği varsayılmıştı.
- Ancak iktisat kuramı bu denli basit değildir.
- *Örnek:* Bir mala olan talep yalnızca o malın fiyatına değil; ikame ya da tamamlayıcı malların fiyatına, gelir düzeyine, nüfusa ve diğer değişkenlere de bağlı olabilir.
- *Örnek:* Tüketim harcamaları yalnızca gelir ile değil; kişinin yaşı, eğitim düzeyi, cinsiyeti, toplam serveti ve benzer değişkenler ile de ilişkili olabilir.
- Modele başka değişkenler eklemek bizi çoklu bağlanım çözümlemesine götürür.
- En basit çoklu bağlanım modeli, bir bağımlı ve iki açıklayıcı değişkenden oluşan üç değişkenli bağlanımdır:

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

- Burada Y bağımlı değişken, X_2 ve X_3 açıklayıcı değişkenler, u olasılıksal hata terimi, i gözlem no'sudur.

- β_1, modelde bulunmayan tüm değişkenlerin Y üzerindeki ortalama etkisini gösteren sabit terimdir.
- β_2 ve β_3'e de *"kısmi bağlanım katsayısı"* (partial regression coefficient) adı verilir.

Üç değişkenli modeldeki kısmi bağlanım katsayılarının anlamı şudur:

- β_2, X_3 sabit tutulurken X_2'deki bir birimlik değişmeye karşı Y'nin beklenen değeri $E(Y|X_2, X_3)$'teki değişmeyi ölçer.
- Bir başka deyişle β_2, X_3 sabitken $E(Y|X_2, X_3)$'ün X_2'ye göre eğimini verir.
- Diğer bir deyişle β_2, X_2'deki bir birimlik değişmenin Y üzerindeki X_3'ten ayrı, net etkisini gösterir.
- β_3'ün yorumu da benzer şekildedir.

Üç Değişkenli Model Varsayımları

Daha önce KDBM çerçevesinde yapılmış olan varsayımlar, k değişkenli çoklu bağlanım modeli için de geçerlidir:

1. Çoklu bağlanım modeli değiştirgelerde doğrusaldır.
2. Açıklayıcı değişkenler tekrarlı örneklemlerde değişmez.
3. Açıklayıcı değişkenlerde yeterli değişkenlik bulunur.
4. Hata teriminin ortalaması sıfırdır: $E(u_i|X_{2i}, X_{3i}, \ldots, X_{ki}) = 0$
5. Hata teriminin varyansı sabittir: var$(u_i) = \sigma^2$
6. u_i ve X'ler birbirlerinden bağımsız dağılmaktadır:

$$\text{cov}(u_i, X_{2i}) = \text{cov}(u_i, X_{3i}) = \ldots = \text{cov}(u_i, X_{ki}) = 0$$

7. *"Serisel ilinti"* (serial correlation) bulunmamaktadır:

$$\text{cov}(u_i, u_j) = 0 \ (i \neq j)$$

8. *"Model belirtim hatası"* (model specification error) yoktur.
9. X_2 ile X_3 arasında *"tam eşdoğrusallık"* (exact collinearity) bulunmamaktadır.

Eşdoğrusallık Kavramı

- X_2 ile X_3 arasında tam doğrusal ilişki olmadığı yönündeki SEK varsayımını anımsayalım.
- *"eşdoğrusal-dışılık"* (non-collinearity) varsayımına göre, aşağıdaki gibi tanımlanan iki değişken doğrusal bağımlıdır:

$$X_{2i} = aX_{3i} \qquad \text{ya da} \qquad X_{2i} - aX_{3i} = 0, \qquad a \in \text{R}$$

- Dolayısıyla, X_2 ve X_3 eğer aynı modelde yer alırlarsa tam eşdoğrusal ilişki ortaya çıkar.
- Tam eşdoğrusallık çoklu bağlanımda önemli bir konudur çünkü bu durumda açıklayıcı değişkenlerin bağımlı değişken üzerindeki tekil etkilerini bulmanın yolu yoktur.
- Tam eşdoğrusallık olması durumunda kısaca elde iki değil bir bağımsız değişken var demektir.
- *Örnek:* $4X_{2i} = X_{3i}$ olsun. Bu durumda üç değişkenli model ikili modele indirgenir:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3(4X_{2i}) + u_i \\ &= \beta_1 + (\beta_2 + 4\beta_3)X_{2i} + u_i \\ &= \beta_1 + \alpha X_{2i} + u_i \end{aligned}$$

- Diğer yandan, eğer $X_{3i} = X_{2i}^2$ ise iki değişken arasındaki ilişki doğrusal değildir. Bu durumda da eşdoğrusal-dışılık varsayımı çiğnenmiş olmaz.

8.1.2 Kısmi Bağlanım Katsayılarının Tahmini

SEK Tahmincileri

- Üç değişkenli modelin SEK tahmincilerini bulmak için önce örneklem bağlanım işlevini aşağıdaki gibi yazalım:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

- SEK yöntemi, anakütle tahmincilerini kalıntı kareleri toplamı ($\sum \hat{u_i}^2$) en küçük olacak biçimde hesaplar:

$$\min \sum \hat{u_i}^2 = \min \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2$$

- Yukarıdaki eşitliği enazlayacak en doğrudan süreç eşitliğin $\hat{\beta}$'lara göre türevini almak, bunları sıfıra eşitlemek ve daha sonra eşanlı olarak çözmektir.

Üç değişkenli model için SEK yöntemi şu tahmincileri verir:

$$\begin{aligned}
\hat{\beta}_1 &= \hat{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \\
\hat{\beta}_2 &= \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \\
\hat{\beta}_3 &= \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}
\end{aligned}$$

- $\hat{\beta}_2$ ve $\hat{\beta}_3$ tahmincileri bakışımlıdır ve paydaları aynıdır.
- Demek ki X_2 ile X_3'ün yerleri değiştirilirse $\hat{\beta}_2$ ile $\hat{\beta}_3$'ün de yeri değişir ama bu bağlanım sonuçlarını etkilemez.

Varyans ve Ölçünlü Hatalar

SEK tahmincilerinin varyansları ise aşağıdaki gibi bulunur:

$$\begin{aligned}
\text{var}(\hat{\beta}_1) &= (\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2 \bar{X}_3 \sum x_{2i} x_{3i}}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2})\sigma^2 \\
\text{var}(\hat{\beta}_2) &= \frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}\sigma^2 = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)} \\
\text{var}(\hat{\beta}_3) &= \frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}\sigma^2 = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}
\end{aligned}$$

- var($\hat{\beta}_2$) ve var($\hat{\beta}_3$) formüllerinde yer alan r_{23}, X_2 ve X_3 arasındaki örneklem ilinti katsayısı r'dir.
- Ölçünlü hatalar ise varyansların artı değerli karekökleridir:

$$\text{öh}(\hat{\beta}) = \sqrt{\text{var}(\hat{\beta})}.$$

- Varyans ve ölçünlü hata formüllerindeki σ^2'nin anakütle hata terimi u_i'nin sabit varyansı olduğunu biliyoruz.
- Bu anakütle katsayısının yansız tahmincisi şöyledir:

$$\hat{\sigma}^2 = \frac{\sum \hat{u_i}^2}{n-3}$$

- σ^2'nin bu tahmincisi ile iki değişkenli modeldeki tahmincisi $(\sum \hat{u_i}^2/n-2)$ benzerdir. Aralarındaki tek fark üç değişkenli model için serbestlik derecesinin artık $(n-3)$ olmasıdır.
- Kalıntılar bulunduktan sonra $\hat{\sigma}^2$ kolayca hesaplanabilir.
- Kalıntı kareleri toplamı ise şu eşitlik ile kolayca bulunabilir:

$$\sum \hat{u_i}^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

SEK Tahmincilerinin Özellikleri

Üç değişkenli model için SEK tahmincilerinin özellikleri iki değişkenli model ile aynıdır:

1. Üç değişkenli bağlanım doğrusu (düzlemi) $\bar{Y}$, $\bar{X}_2$, $\bar{X}_3$ ortalamalarından geçer: $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2\bar{X}_2 + \hat{\beta}_3\bar{X}_3$
2. $\hat{Y}_i$'nın ortalaması gözlenen Y_i ortalamasına eşittir: $\bar{\hat{Y}}_i = \bar{Y}_i$
3. Kalıntılar toplamı sıfıra eşittir: $\sum \hat{u}_i = n\bar{\hat{u}}_i = \bar{\hat{u}}_i = 0$
4. Kalıntılar X_{2i} ve X_{3i} ile ilişkisizdir: $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$
5. $\hat{u}_i$ kalıntıları $\hat{Y}_i$ ile de ilişkisizdir: $\sum \hat{u}_i \hat{Y}_i = 0$
6. Varyans formüllerinden görüldüğü gibi, X_2 ile X_3 arasındaki ilinti katsayısı r_{23} artarken $\hat{\beta}_2$ ve $\hat{\beta}_3$'nın varyansları yükselir.
7. Gözlem sayısı n artarken $\hat{\beta}_2$ ve $\hat{\beta}_3$'nın varyansları da azalır.
8. $\hat{\beta}_2$ ve $\hat{\beta}_3$ tahmincileri, en iyi doğrusal yansız tahminci ya da kısaca EDYT'dirler.

ÖBİ'nin Sapmalar Biçimi Gösterimi

Çok değişkenli modelde ÖBİ'nin sapmalar biçiminde gösterimi aşağıda gösterilen şekilde elde edilir:

1. Üçlü bağlanım modelini ele alalım:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

2. Bağlanım yüzeyi $\bar{Y}$, $\bar{X}_2$, $\bar{X}_3$ ortalamalarından geçtiği için:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$$

3. İkinci denklemi birinciden çıkartırsak şunu buluruz:

$$\begin{array}{rrcccc} & \hat{Y}_i = & \hat{\beta}_1 & + \quad \hat{\beta}_2 X_{2i} & + & \hat{\beta}_3 X_{3i} \\ - & \bar{Y} = & \hat{\beta}_1 & + \quad \hat{\beta}_2 \bar{X}_2 & + & \hat{\beta}_3 \bar{X}_3 \\ \hline & \hat{Y}_i - \bar{Y} = & \hat{\beta}_1 - \hat{\beta}_1 & + \hat{\beta}_2 (X_{2i} - \bar{X}_2) & + & \hat{\beta}_3 (X_{3i} - \bar{X}_3) \end{array}$$

$$\hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$$

Ençok Olabilirlik Tahmincileri

- İki değişkenli modelde olduğu gibi çoklu modeller için de bağlanım katsayılarının SEK ve EO tahmincileri aynıdır.
- Ancak üçlü modelde σ^2'nin SEK tahmincisi $\sum \hat{u_i}^2/(n-3)$ iken EO tahmincisi modelde kaç değişken olursa olsun $\sum \hat{u_i}^2/n$ olarak bulunur.
- Diğer bir deyişle SEK tahmincisi serbestlik derecesini hesaba katarken yanlı EO tahmincisi bunu dikkate almaz.
- Eğer n çok büyükse kuşkusuz EO ve SEK tahmincileri birbirlerine yaklaşırlar.

8.2 Çoklu Bağlanımda Yakışmanın İyiliği

8.2.1 Çoklu Belirleme ve İlinti Katsayıları

- İki değişkenli durum için geliştirmiş olduğumuz r^2, ikiden çok değişkenli bağlanım modellerine de genişletilebilir.
- Çoklu modelde bu istatistiğe R^2 ya da *"çoklu belirleme katsayısı"* (multiple coefficient of determination) denir.
- R^2, bağımlı değişken Y'deki değişimin $X_2, X_3, \ldots, X_n$ ile topluca açıklanabilme oranını gösterir.

Çoklu Belirleme Katsayısı

$$R^2 = \frac{\text{BKT}}{\text{TKT}} = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \cdots + \hat{\beta}_n \sum y_i x_{ni}}{\sum y_i^2}$$

- R^2 de r^2 gibi 0 ile 1 arasındadır.
- R^2 1'e ne kadar yakınsa modelin verilere yakışması da o kadar iyidir. Eğer $R^2 = 1$ ise, yakıştırılan bağlanım Y'deki değişimin tamamını açıklıyor demektir.
- İki değişken arasındaki doğrusal ilişkinin derecesini ölçen r'nin çoklu bağlanımdaki karşılığı da *"çoklu ilinti katsayısı"* (coefficient of multiple correlation) olup, R ile gösterilir:

Çoklu İlinti Katsayısı

$$R = \pm\sqrt{R^2}$$

- R değeri, bağımlı değişken Y ile tüm açıklayıcı değişkenler arasındaki ortak ilişkinin derecesini ölçer.
- Diğer taraftan uygulamada R'nin önemi azdır. Bağlanım çözümlemesi çerçevesinde asıl anlamlı büyüklük R^2'dir.
- R^2'nin önemli bir özelliği, modelde bulunan açıklayıcı değişken sayısının azalmayan bir işlevi olmasıdır.
- Diğer bir deyişle açıklayıcı değişken sayısı arttıkça R^2 hemen hemen her zaman artar, asla azalmaz.

- Bunu görebilmek için belirleme katsayısının tanımını anımsayalım:

$$R^2 = 1 - \frac{\text{KKT}}{\text{TKT}} = 1 - \frac{\sum \hat{u}_i{}^2}{\sum y_i^2}$$

- Burada TKT, X'lerin sayısından bağımsızdır. KKT ise açıklayıcı değişken sayısı arttıkça azalma eğilimine girer.
- Bu nedenle, bağımlı değişkeni aynı olan ama farklı sayıda açıklayıcı değişken içeren iki ayrı bağlanım modeline ait R^2 değerleri karşılaştırırken dikkatli olunmalıdır.

R^2 Değerlerinin Karşılaştırılması

- İki R^2 değerini karşılaştırırken modelde var olan açıklayıcı değişken sayısını da dikkate alma gereksinimi *"ayarlamalı"* (adjusted) belirleme katsayısı $\bar{R}^2$ tanımına yol açmıştır:

  Ayarlamalı Belirleme Katsayısı

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i{}^2/(n-k)}{\sum y_i^2/(n-1)} \qquad \text{ya da} \qquad \bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{s_Y^2}$$

- Burada k, sabit terimle birlikte modeldeki katsayı sayısıdır. s_Y^2 ise Y'nin örneklem varyansıdır.
- Ayarlamalı sözcüğü, giren kareler toplamının serbestlik derecesine göre ayarlanmış olduğu anlamına gelir.
- *Dikkat:* Üç değişkenli bağlanım için $\sum \hat{u}_i{}^2$ sd'sinin $(n-3)$ olduğunu anımsayınız.
- $\bar{R}^2$'nin R^2 ile ilişkisini aşağıdaki eşitlikle gösterebiliriz:

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

- Buradan da görülüyor ki $k > 1$ olduğunda $\bar{R}^2 < R^2$'dir.
- Diğer bir deyişle, X'lerin sayısı arttıkça ayarlamalı R^2 *"ayarlamasız"* (unadjusted) R^2'ye göre daha az artar.

- Ayrıca $\bar{R}^2$'nin eksi değerler de alabildiği görülmektedir. Eğer $\bar{R}^2$ eksi bulunursa uygulamada sıfır kabul edilir.
- Tüm modern ekonometri yazılımları alışıldık R^2'nin yanısıra ayarlamalı R^2 istatistiğini de verir.

İki farklı modeli ayarlamalı ya da ayarlamasız R^2 temelinde karşılaştırabilmek için iki noktaya daha dikkat edilmelidir:

1. Örneklem büyüklüğü n her iki model için aynı olmalıdır. *Dikkat:* Modele gözlem eklendiğinde ya da çıkartıldığında, hesaplanan R^2'nin de değişeceğini unutmayınız.
2. Bağımlı değişken Y de her iki model için aynı olmalıdır. *Dikkat:* R^2 değerinin, X açıklayıcı değişkenlerinin Y'deki değişimi açıklama oranını gösterdiğini anımsayınız. Eğer Y'ler farklıysa, hesaplanan R^2'ler de farklı şeylerin değişim oranını göstereceği için karşılaştırılamaz.

- Bağımlı değişkenleri aynı olmayan iki model düşünelim:

$$\begin{aligned}
\hat{Y}_i &= \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} \\
\widehat{\ln Y_i} &= \alpha_1 + \alpha_2 \ln X_{1i} + \alpha_3 \ln X_{2i}
\end{aligned}$$

- Burada R^2 değerlerini karşılaştırmak için 2 yol izlenebilir:

1. Yol
İkinci modelden tahmin edilen $\widehat{\ln Y_i}$'ların anti-logaritmaları alınır. Bulunan değerler ile Y_i arasında hesaplanan r^2 değeri birinci modeldeki R^2 ile karşılaştırılabilir.

2. Yol
Birinci modelden tahmin edilen $\hat{Y}_i$'ların logaritmaları alınır. Bulunan değerler ile $\ln Y_i$ arasında hesaplanan r^2 değeri ikinci modeldeki R^2 ile karşılaştırılabilir.

- Bağımlı değişkenleri farklı modelleri karşılaştırmak için, iki değişken arasındaki ilinti formülünün karesine dayanan şu r^2 formülü kullanılabilir:

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

- Son olarak, R^2'nin yakışmanın iyiliğini ölçmede kullanılan istatistiklerden yalnızca biri olduğu unutulmamalıdır.
- Model seçimi için başka ölçütler de bulunmaktadır:

"Akaike bilgi ölçütü" (Akaike information criterion)
"Schwarz Bayesçi ölçüt" (Schwarz Bayesian criterion)
"Hannan-Quinn ölçütü" (Hannan-Quinn criterion)

- Araştırmacının asıl ilgisi, açıklayıcı değişkenlerin bağımlı değişken ile olan mantıksal ya da kuramsal ilişkilerine ve bunların istatistiksel anlamlılıklarına yönelik olmalıdır.

8.2.2 Kısmi İlinti Katsayıları

- İki değişken arasındaki doğrudan ilişkinin bir ölçüsü olarak tanımlanan ilinti katsayısı r kavramını anımsayalım.
- Üç değişkenli model için böyle üç ayrı *"basit ilinti katsayısı"* (simple correlation coefficient) değerinden söz edilebilir:

  Basit İlinti Katsayıları

  Y ile X_2 arasındaki ilinti katsayısı: r_{12}
  Y ile X_3 arasındaki ilinti katsayısı: r_{13}
  X_2 ile X_3 arasındaki ilinti katsayısı: r_{23}

- Bunlara aynı zamanda *"sıfırıncı dereceden ilinti katsayısı"* (correlation coefficient of zero order) da denmektedir.
- Eğer bir X_3 değişkeni hem Y hem de X_2 ile ilişkiliyse, bu durumda Y ve X_2 arasındaki basit ilinti r_{12} yanıltıcıdır.
- İki değişken arasında, üçüncü bir değişkenin etkisinden bağımsız olarak bulunan *"kısmi ilinti katsayısı"* (partial correlation coefficient) ise şöyle tanımlanır:

  Kısmi İlinti Katsayıları

  X_3 sabitken Y ile X_2 arasındaki kısmi ilinti: $r_{12.3}$
  X_2 sabitken Y ile X_3 arasındaki kısmi ilinti: $r_{13.2}$
  Y sabitken X_2 ile X_3 arasındaki kısmi ilinti: $r_{23.1}$

- Bunlara *"birinci dereceden"* (first order) ilinti katsayıları denir. Buradaki derece ikincil alt imlerin sayısıdır.
- Buna göre, X_3 ve ikinci bir X_4 sabit tutulurken bulunan $r_{12.34}$ değerine de ikinci dereceden bir ilinti katsayısı denir.

- Birinci dereceden kısmi ilinti katsayılarını hesaplamak için aşağıdaki eşitlikler kullanılabilir:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$
$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$
$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}}$$

Çok değişkenli modellerde basit ilinti katsayılarını yorumlarken şu noktalara dikkat etmek gereklidir:

- $r_{12} = 0$ olsa bile, aynı anda r_{13} ya da r_{23} de sıfır olmazsa $r_{12.3} = 0$ olmaz.
- $r_{12.3}$ ile r_{12} aynı işareti taşımak zorunda değildir.
- $r_{13} = r_{23} = 0$ olması $r_{12} = 0$ anlamına gelmez.
- İkili bağlanımdaki $0 \leq r^2 \leq 1$ tanımını anımsayalım. Kısmi ilinti katsayıları kareleri için de geçerli olan bu durumdan yararlanılarak, üç sıfırıncı dereceden ilinti katsayısı arasındaki ilişki şöyle gösterilebilir:

$$0 \leq r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1$$

- Yukarıdaki eşitsizlikten de anlaşılabileceği gibi, Y ile X_2'nin ve X_2 ile de X_3'ün ilintisiz olması Y ile X_3'ün ilintisiz olacağı anlamına gelmemektedir.

8.2.3 Çoklu Bağlanım Açıklayıcı Örnek

- Çoklu bağlanıma örnek olarak 2005-2009 aylık verilerini alalım ve Türkiye için bir *"beklentilerle-genişletmeli Phillips eğrisi"* (expectations-augmented Phillips curve) modeli belirtelim:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_t$$

- Burada
  Y_t TÜFE değerini (2005 Ocak=100),
  X_{2t} işsiz sayısını (bin kişi, mevsimsel ayarlamalı),
  X_{3t} ise beklenen TÜFE değerini
  göstermektedir.
- İktisat kuramına göre β_2 eksi, β_3 ise artı değerli olmalıdır.
- Aslında kurama göre $\beta_3 = 1$ beklentisi vardır.

SEK yöntemi ile elde edilen bağlanım bulguları şöyledir:

$$\begin{array}{llllll}
\ln \hat{Y}_t = & -0{,}1879 & - \quad 0{,}0364 & \ln X_{2t} + & 1{,}1012 \quad \ln X_{3t} & \\
\text{öh} & (0{,}1072) & (0{,}0166) & & (0{,}0120) & \\
t & (-1{,}7535) & (-2{,}1960) & & (91{,}8156) & R^2 = 0{,}9963
\end{array}$$

- $\hat{\beta}_2$ ve $\hat{\beta}_3$ önsel beklentilerle uyumlu işaret taşımaktadır.
- $\hat{\beta}_1$'ya göre, X_2 ve X_3 dışındaki diğer tüm etmenler TÜFE üzerinde ortalama $e^{-0{,}1879} \approx 0{,}83$ etkiye yol açmaktadır.
- $\hat{\beta}_2$ kısmi bağlanım katsayısı ise X_3 sabit tutulduğunda işsizlikteki $\%1$'lik bir artışa karşılık TÜFE'nin de yaklaşık $\%0{,}036$ düşeceği anlamına gelir.
- Bulunan bu düşük değer, Türkiye'de enflasyon ve işsizlik arasındaki ilişkinin zayıf olduğu önsel bilgisi ile uyumludur.
- R^2 değeri, enflasyon oranındaki değişimin %99'unun bu iki açıklayıcı değişkenle açıklanabildiğini öne sürer. Bu kadar yüksek bir R^2 bağlanıma kuşkuyla yaklaşmayı gerektirir.

Model Belirtim Yanlılığı Sorunu

- Klasik doğrusal bağlanım modeli varsayımlarına göre bağlanım modeli doğru kurulmuş olmalıdır.
- Eğer çözümlemede kullanılacak bağlanım modeli yanlış kurulursa *"model belirtim yanlılığı"* (model specification bias) ortaya çıkar.
- Bu varsayımın önemini vurgulayabilmek için elimizdeki Phillips eğrisi modeli yardımcı olabilir.

(. . . devam)

- Az önce ele almış olduğumuz aşağıdaki üçlü bağlanım modelinin "doğru" model olduğunu varsayalım:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_{1t}$$

- Elimizdeki Türkiye verilerini şu iki değişkenli modele yakıştırmakta diretiyor olalım:

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + u_{2t}$$

- Y_t burada t dönemindeki TÜFE değerini, X_{2t} ise toplam işsiz sayısını göstermektedir.
- Birinci model "doğru" olduğuna göre ikinci model bir model belirtim hatası içermektedir.
- Buradaki hata, X_{3t} beklenen TÜFE değişkenini modelden dışlamış olmaktır.
- Birinci modeldeki $\hat{\beta}_2$'nın gerçek β_2'nin yansız bir tahmincisi olduğunu biliyoruz.
- Diğer yandan ikinci modeldeki $\hat{\alpha}_2$ değiştirgesi β_2'nin yansız tahmincisi değildir.
- α_2'nin aslında X_3'ün X_2'ye göre bağlanımından ortaya çıkan eğim değiştirgesi α_3 ile ilişkili olduğu gösterilebilir:

$$\alpha_2 = \beta_2 + \beta_3\alpha_3 + \text{hata terimi}$$

- Buna göre $E(\alpha_2)$ beklenen değeri β_2 değil de $\beta_2 + \beta_3\alpha_3$ olarak karşımıza çıkmaktadır.
- Sonuç olarak, ilk modeldeki β_2 değiştirgesi X_2'nin Y üzerindeki doğrudan ya da tekil etkisini ölçmektedir.
- Hatalı modeldeki α_2 değiştirgesi ise X_2'nin Y üzerindeki hem doğrudan hem de X_3 üzerinden dolaylı etkisini verir.

Hatalı modelin SEK tahmini aşağıdaki bulguları vermektedir:

$$
\begin{array}{llllll}
\ln \hat{Y}_t = & -1{,}7203 & + & 0{,}8327 & \ln X_{2t} & \\
\text{öh} & (1{,}4369) & & (0{,}1845) & & \\
t & (-1{,}1972) & & (4{,}5142) & & r^2 = 0{,}3070
\end{array}
$$

- Kuramsal beklentinin aksine α_2 burada artı değerlidir ve 0,83 gibi yüksek, gerçek dışı bir büyüklüktedir.
- Demek ki belli bir model "doğru" olarak kabul ediliyorsa bir ya da birkaç değişkeni çıkartarak modeli değiştirmek yanlı tahminlere yol açmaktadır.
- Yanlış belirtilen bir model anakütle katsayı tahminlerinin yanlı olması gibi ciddi bir soruna neden olabilmektedir.

8.3 Çokterimli Bağlanım Modelleri

- Çoklu bağlanımın bir şekli de *"çokterimli"* (polynomial) bağlanım modelleridir.
- Şimdiye kadar ele aldığımız tüm örneklerde bağlanım işlevinin değişkenlerde doğrusal olduğunu varsamıştık.
- Gerçek hayatta bu varsayımın geçerli olmadığı pek çok durum düşünülebilir.
- Örnek olarak, gelir düzeyi yükseldikçe doğurganlığın da düştüğü bilinen bir olgudur.
- Düşük gelir düzeylerinde çocuk bir tür sosyal güvence olarak düşünülebildiği için doğurganlık hızı yüksektir.
- Gelir arttıkça ortalama çocuk sayısı da azalır ancak ilişki doğrusal değildir. Belli bir gelirden sonra çocuk sayısının sıfır ya da eksi değerlere ulaşacağını beklemeyiz.
- İki değişken arasındaki doğrusal olmayan bir ilişkiyi incelemenin bir yolu çokterimli SEK modelidir.
- Genel olarak, r'inci dereceden çokterimli bağlanım modeli şöyle gösterilir:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_r X^r$$

- Buradaki tek açıklayıcı değişken olan X, farklı kuvvetlerle gösterildiği için bu model bir çoklu bağlanım modelidir.
- Çokterimli modeller β katsayılarında doğrusal oldukları için SEK yöntemi ile tahmin edilebilirler.
- Bu modelde X ve X'in kuvvetleri arasındaki ilişki güçlü olmakla birlikte doğrusal olmadığı için, KDBM'nin "çoklueşdoğrusallık yoktur" varsayımı çiğnenmemiş olur.
- Doğrusal modellerde β terimlerinin Y'nin farklı X'lere göre sabit eğimini verdiğini anımsayalım.
- Değişkenlerde doğrusal-dışı olan çokterimli modellerde ise katsayıların yorumlanması biraz daha karmaşıktır.

- Bu modellerde ele alınan ilişki eğrisel olduğu için, eğim de X'in düzeyinine göre değişir.
- Bu nedenle, X'deki bir birimlik artışın Y üzerindeki etkisini bulmak için, önce bir başlangıç X düzeyi seçilir ve buna karşılık gelen $\hat{Y}$ değeri hesaplanır.
- Daha sonra X bir birim artırılır ve $\hat{Y}$ yeniden hesaplanır.
- Aradaki fark, seçili X düzeyindeki ortalama eğimi verir.

Çokterimli Bağlanım Açıklayıcı Örnek

- Çokterimli bağlanım modeline bir örnek olarak, Türkiye'de illerdeki gelir ve doğurganlık ilişkisini *"kareli"* (quadratic) bir işlev çerçevesinde ele alalım.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

- Burada Y ortalama çocuk sayısını, X ise kişi başına düşen gayri safi yurtiçi hasılayı göstermektedir.
- Görüldüğü gibi bu modelde Y ve X değişkenleri arasındaki ilişkiyi tanımlayan iki ayrı β_1 ve β_2 bulunmaktadır.
- Kabaca, β_1 ilişkinin yönünü gösterirken β_2'nin ise eğriselliği anlattığını söyleyebiliriz.
- Önsel beklentimiz, X artarken Y'nin de azalacağı ancak bu azalmanın giderek yavaşlayacağı yönündedir. Buna göre β_1 eksi, β_2 ise artı değer almalıdır.

Modeli 2000 yılı Türkiye verilerine yakıştırdığımızda aşağıdaki bulguları elde ediyoruz:

$$\begin{array}{lllll}
\hat{Y}_i = & 5{,}9486 & -\ 0{,}0030\ X_i & +\ 4{,}978\text{e-}07\ X_i^2 & \\
\text{öh} & (0{,}3835) & (0{,}0004) & (9{,}727\text{e-}08) & R^2 = 0{,}5196 \\
t & (15{,}5094) & (-7{,}2485) & (5{,}1179) & \bar{R}^2 = 0{,}5073
\end{array}$$

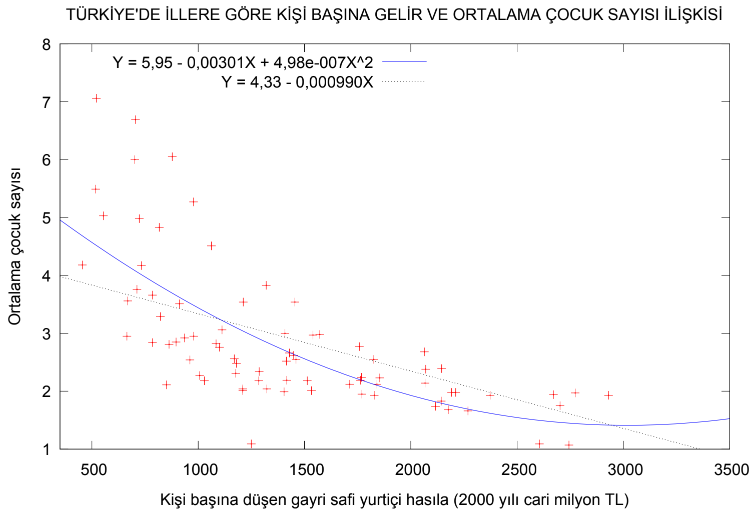
- Katsayıların işaretleri beklentilerimiz ile örtüşmektedir.
- İlişki doğrusal olsaydı, $\hat{\beta}_2$ anlamlı çıkmayacaktı. $\hat{\beta}_2$'nın anlamlı olması doğrusal-dışılığı onaylayıcı niteliktedir.
- Gelir 1000 TL olduğunda ortalama çocuk sayısı şudur:

$$\hat{Y} = 5{,}9486 - (0{,}0030 \times 1000) + (4{,}978\text{e-}07 \times 1000^2) = 3{,}44$$

- Gelir 1100 TL olduğunda çocuk sayısı ise şöyledir:

$$\hat{Y} = 5{,}9486 - (0{,}0030 \times 1100) + (4{,}978\text{e-}07 \times 1100^2) = 3{,}24$$

- Demek ki $X = 1000$ olduğunda, gelir düzeyindeki 100 TL kadar bir artış ortalama çocuk sayısını 0,2 düşürmektedir.



Uygulamaya İlişkin İki Nokta

Son olarak, çokterimli modeller kullanılırken özellikle iki noktaya dikkat etmek önemlidir:

1. Öncelikle doğrusal-dışı ilişki tanımlanmalıdır. Araştımacı, X ve Y arasındaki ilişkinin neden doğrusal olmayabileceğini sorgulamalıdır. Daha sonra, uygun bir işlev biçimi seçmek için iktisat kuramı temel alınmalıdır.

2. İkinci olarak, uygun bir çokterimli model belirtilip tahmin edildikten sonra bunun ilişkiyi iyi anlattığı ve doğrusal modelden üstün olduğu doğrulanmalıdır. Bunun için tahmin edilen bağlanım işlevinin çizdirmek ve bağlanımın verilere iyi yakışıp yakışmadığına bakılabilir. Ayrıca, anakütle bağlanım işlevinin doğrusal olduğu sıfır önsavı istatistiksel yöntemler kullanılarak sınanmalıdır.

Bu çıkarsama yöntemleri ise bir sonraki konuda ele alınacaktır.

Önümüzdeki Dersin Konusu ve Ödev

Ödev
Kitaptan *Bölüm 7* “Multiple Regression Analysis: The Problem of Estimation” okunacak.

Önümüzdeki Ders
Çoklu Bağlanım Çözümlemesi: Çıkarsama Sorunu

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011