Çoklu Bağlanım – Tahmin Sorunu

Çoklu Bağlanımda Yakışmanın İyiliği

TÜBA
TÜRKİYE BİLİMLER AKADEMİSİ

UADMK
ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Ekonometri 1 – Konu 22

Sürüm 2,0 (Ekim 2011)

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011 BY-NC-SA

Ders Planı

ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Çoklu Belirleme Katsayısı

- 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 $X_2, X_3, \dots, X_n$ 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.

Çoklu İlinti Katsayısı

- 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.

Ayarlamalı Belirleme Katsayısı

- 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.

R^2 Değerlerinin Karşılaştırılması

- $\bar{R}^2$'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.

R^2 Değerlerinin Karşılaştırılması

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.

R^2 Değerlerinin Karşılaştırılması

- 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.

(. . . devam)

R^2 Değerlerinin Karşılaştırılması

- 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.

Basit 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.

Kısmi İlinti Katsayıları

- 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.

Kısmi İlinti Katsayıları

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

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

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

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

Kısmi İlinti Katsayıları

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.

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.

Çoklu Bağlanım Açıklayıcı Örnek

SEK yöntemi ile elde edilen bağlanım bulguları şöyledir:

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

- $\hat{\beta}_2$ 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)

Model Belirtim Yanlılığı Sorunu

- 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.

Model Belirtim Yanlılığı Sorunu

- 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.

Model Belirtim Yanlılığı Sorunu

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.

Önümüzdeki Dersin Konusu

Önümüzdeki ders

Çokterimli bağlanım modelleri