İki Değişkenli Bağlanım Modeli

SEK Yönteminin Güvenilirliği

Ekonometri 1 – Konu 11

Sürüm 2,0 (Ekim 2011)

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011

Ders Planı

1. SEK Yönteminin Güvenilirliği
   - SEK Tahmincilerinin Ölçünlü Hataları
   - Belirleme Katsayısı r^2
   - Monte Carlo Yöntemi
2. Sayısal Bir Örnek

SEK Tahmincilerinin Ölçünlü Hataları

- Sıradan en küçük kareler tahmincilerinin örneklem verilerinin birer işlevi olduğunu anımsayalım:

$$\hat{\beta}_2 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^2 - (\sum X_i)^2} = \bar{Y} - \hat{\beta}_2\bar{X}$$

- Veriler örneklemden örnekleme değişeceği için tahminler de buna bağlı olarak değişecektir.
- Öyleyse $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmincilerinin güvenilirliği için bir ölçüte gereksinim vardır.

SEK Tahmincilerinin Ölçünlü Hataları

- İstatistikte rastsal bir değişkenin doğruluk derecesi "ölçünlü hata" (standard error), kısaca "öh" (se) ile ölçülür:

Ölçünlü Hata

Ölçünlü hata, bir tahminciye ait örneklem dağılımının kendi ortalamasından ortalama olarak ne kadar saptığını gösterir. Örneklem dağılımı varyansının artı değerli kare köküdür.

SEK Tahmincilerinin Ölçünlü Hataları

- Başta sözü edilmiş olan Gaussçu varsayımlar geçerli iken SEK tahmincilerinin ölçünlü hataları aşağıdaki gibidir:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

$$\text{öh}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2}\sigma^2$$

$$\text{öh}(\hat{\beta}_1) = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}}\sigma$$

- Burada
  var değişirlik ya da varyansı,
  öh ölçünlü hatayı,
  σ^2 ise bağlanımın sabit varyansını
  göstermektedir.

SEK Tahmincilerinin Ölçünlü Hataları

- u_i'nin sabit varyansını veren σ^2 şöyle tahmin edilir:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}$$

- Buradaki $\hat{\sigma}^2$, bilinmeyen σ^2'nin SEK tahmincisidir.
- $\sum \hat{u}_i^2$ terimine "kalıntı kareleri toplamı" (residual sum of squares), kısaca "KKT" (RSS) denir ve şöyle bulunur:
$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2$$
- $n-2$ değeri ise iki değişkenli çözümleme için geçerli serbestlik derecesidir.

Serbestlik Derecesi Kavramı

Serbestlik Derecesi

"Serbestlik derecesi" (degree of freedom), örneklemdeki toplam gözlem sayısı (n) eksi bunlar üzerine konulmuş olan bağımsız ve doğrusal sınırlama sayısıdır.

- Örnek olarak, KKT'nin hesaplanabilmesi için önce $\hat{\beta}_1$ ve $\hat{\beta}_2$ değerlerinin bulunmuş olması gereklidir:

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

- Dolayısıyla bu iki tahminci KKT üzerine iki sınırlama getirir.
- Bu durumda, KKT'yi ve dolayısıyla da ölçünlü hatayı doğru hesaplayabilmek için aslında elde n değil $n - 2$ sayıda bağımsız gözlem vardır.

SEK Tahmincilerinin Ölçünlü Hatalarının Özellikleri

SEK tahmincileri $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın varyans formüllerini anımsayalım:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \qquad \text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2}\sigma^2$$

$\hat{\beta}_1$ ile $\hat{\beta}_2$ tahmincilerinin varyanslarının ve dolayısıyla bunların ölçünlü hatalarının şu özellikleri önemlidir:

1. Örneklem büyüklüğü n arttıkça $\sum x_i^2$ toplamındaki terim sayısı da artar. Böylece n büyüdükçe $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın doğruluk dereceleri de artar.
2. $\hat{\beta}_1$ ve $\hat{\beta}_2$, verili bir örneklemde birbirleri ile ilişkili olabilirler. Bu bağımlılık aralarındaki kovaryans ile ölçülür:
$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\bar{X}\,\text{var}(\hat{\beta}_2)$$
3. Eğer $\bar{X}$ artı değerli ise kovaryans da eksi değerli olur. Bu durumda eğer β_2 katsayısı olduğundan büyük tahmin edilir ise β_1 de olduğundan küçük tahmin edilmiş olur.

Belirleme Katsayısı r^2

- Eldeki gözlemler çoğunlukla bağlanım doğrusu üzerinde yer almazlar.
- Artı ya da eksi işaretli $\hat{u}_i$ hataları ile karşılaşıldığına göre örneklem bağlanım doğrusunun eldeki verilerle ne ölçüde örtüştüğünü gösteren bir ölçüte gereksinim vardır:

Belirleme Katsayısı

"Belirleme katsayısı" (coefficient of determination) ya da r^2 (çoklu bağlanımda R^2), örneklem bağlanım işlevinin verilere ne kadar iyi yakıştığını gösteren özet bir ölçüttür.

Belirleme Katsayısının Hesaplanması

Belirleme katsayısını hesaplamak için, $y_i = \hat{y}_i + \hat{u}_i$ eşitliğinin iki yanının karesi alınır ve örneklem boyunca toplanır:

$$
\begin{aligned}
\sum y_i^2 &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i\hat{u}_i \\
&= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \\
&= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2 \\
\text{TKT} &= \text{BKT} + \text{KKT}
\end{aligned}
$$

Burada
- TKT “Toplam Kareleri Toplamı” (Total Sum of Squares),
- BKT “Bağlanım Kareleri Toplamı” (Regression Sum of Squares),
- KKT “Kalıntı Kareleri Toplamı” (Residual Sum of Squares)

anlamına gelmektedir.

- Yukarıdaki $\sum \hat{y}_i\hat{u}_i$ teriminin SEK bağlanım doğrusunun 3. özelliğinden dolayı sıfıra eşit olduğuna dikkat ediniz.

Belirleme Katsayısının Hesaplanması

$$\begin{array}{ccccc} \sum y_i^2 & = & \hat{\beta}_2^2 \sum x_i^2 & + & \sum \hat{u}_i{}^2 \\ \text{TKT} & = & \text{BKT} & + & \text{KKT} \end{array}$$

Yukarıdaki eşitliğin her iki yanını TKT'ye bölelim:

$$1 = \frac{\text{BKT}}{\text{TKT}} + \frac{\text{KKT}}{\text{TKT}}$$

Buna göre r^2 aşağıdaki gibi tanımlanır:

Belirleme Katsayısı

$$r^2 = \frac{\sum \hat{y}_i{}^2}{\sum y_i^2} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\text{BKT}}{\text{TKT}} = 1 - \frac{\text{KKT}}{\text{TKT}}$$

Belirleme Katsayısının Özellikleri

r^2'nin iki temel özelliğinden söz edilebilir:

1. r^2 eksi değer almayan bir büyüklüktür.
2. Sınırları $0 \leq r^2 \leq 1$'dir.

Buna göre:

- Eğer $r^2 = 1$ olursa bu kusursuz bir yakışma demektir. Bu durumda rastsal hata yoktur ve tüm gözlemler bire bir bağlanım doğrusu üzerinde yer almaktadır.
- Sıfıra eşit bir r^2 ise bağımlı değişkenle açıklayıcı değişken arasında hiçbir ilişkinin olmadığı ($\hat{\beta}_2 = 0$) anlamına gelir.

İlinti Katsayısı

r^2 ile yakın ilişkili ama kavramsal olarak çok uzak bir büyüklük "ilinti katsayısı" (coefficient of correlation), kısaca r'dir:

İlinti Katsayısı

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

- r değeri, bağımlı ve açıklayıcı değişkenler arasındaki doğrusal bağımlılığın bir ölçüsüdür.
- -1 ve $+1$ arasında yer alır: $-1 \leq r \leq 1$.
- Bakışımlıdır: $r_{XY} = r_{YX}$.
- Sıfır noktasından ve ölçekten bağımsızdır.
- Herhangi bir neden-sonuç ilişkisi içermez.
- İki değişken arasında sıfır ilinti ($r = 0$) mutlaka bağımsızlık göstermez çünkü r yalnızca doğrusal ilişkiyi ölçer.

Monte Carlo Yöntemi

- KDBM varsayımları altında SEK tahmincilerinin EDYT (En iyi Doğrusal Yansız Tahminci) olmalarını sağlayan bazı arzulanan özellikler taşıdıklarını anımsayalım.
- EDYT özelliklerinin geçerliliği, bir “benzetim” (simulation) yöntemi olan Monte Carlo deneyleri ile doğrulanabilir.
- Bu yöntem, anakütle katsayılarını tahmin eden süreçlerin istatistiksel özelliklerini incelemede sıkça kullanılmaktadır.
- Monte Carlo aynı zamanda istatistiksel çıkarsamanın temeli sayılan “tekrarlı örnekleme” (repeated sampling) kavramının anlaşılması için de yararlı bir araçtır.

Monte Carlo Yönteminin Adımları

Bir Monte Carlo deneyi aşağıdaki gibi yapılır:

1. Anakütle katsayıları seçilir. Örnek: $\beta_1 = 20$ ve $\beta_2 = 0{,}6$.
2. Bir örneklem büyüklüğü seçilir. Örnek: $n = 25$.
3. Her gözlem için bir X değeri belirlenir.
4. Bir rastsal sayı oluşturucu kullanılarak u_i kalıntıları üretilir.
5. β_1, β_2, X_i'ler ve u_i'ler kullanılarak Y_i değerleri bulunur.
6. Bu şekilde üretilen Y_i değerleri X_i'ler ile bağlanıma sokulur ve $\hat{\beta}_1$ ve $\hat{\beta}_2$ SEK tahmincileri hesaplanır.
7. İşlem tekrarlanır (örneğin 1000 kez) ve rastsallıktan dolayı her seferde değişen tahminlerin ortalamaları ($\bar{\hat{\beta}}_1, \bar{\hat{\beta}}_2$) alınır.
8. Eğer $\bar{\hat{\beta}}_1$ ve $\bar{\hat{\beta}}_2$ değerleri β_1 ve β_2'ye aşağı yukarı eşit ise, deney SEK tahmincilerinin yansızlığını, diğer bir deyişle $E(\hat{\beta}_1) = \beta_1$ ve $E(\hat{\beta}_2) = \beta_2$ olduğunu saptamış sayılır.

Sayısal Bir Örnek

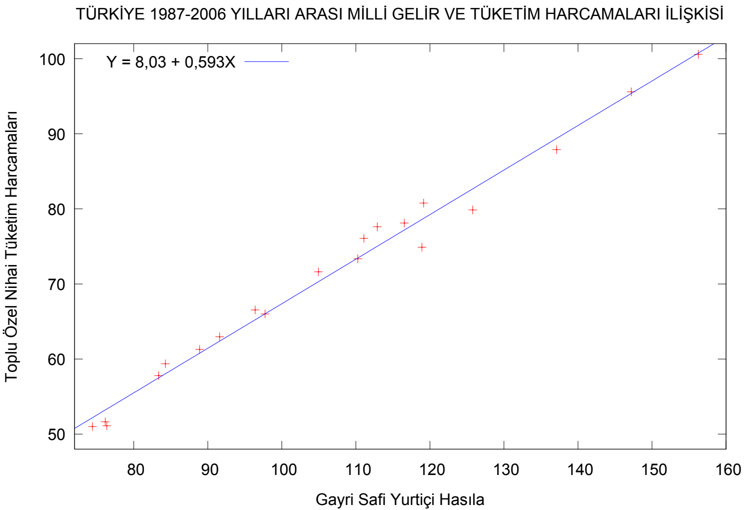
- Ele almış olduğumuz bazı kavramları sayısal bir örnek yardımı ile gözden geçirelim. Türkiye'de 1987–2006 arası toplam tüketim harcamaları ve GSYH verileri şöyledir:

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

- Toplu özel nihai tüketim harcamalarını (Y), gayri safi yurtiçi hasıla (X) ile ilişkilendirmek istiyor olalım.

Sayısal Bir Örnek



SEK Bağlanımı gretl Çıktısı

gretl: model 1

Dosya Düzenle Sınamalar Kaydet Çizitler Çözümleme LaTeX

```
Model 1: SEK (OLS), kullanılan gözlemler: 1987-2006 (T = 20)
Bağımlı değişken: tuketim

             katsayı     ölç. hata    t-oranı    p-değeri
  -----------------------------------------------------
  const       8,03438    1,85509      4,331     0,0004    ***
  gsyih       0,593351   0,0170331   34,84      5,68e-18  ***

Bağımlı değişken ort    71,20574   Bağımlı değişken ö.s.   14,07370
Kalıntı kareleri top    55,00622   Bağlanım ö.h.            1,748114
R-kare                  0,985384   Ayarlamalı R-kare        0,984572
F(1, 18)                1213,489   P-değeri(F)              5,68e-18
Log-olabilirlik        -38,49591   Akaike ölçütü           80,99182
Schwarz ölçütü          82,98329   Hannan-Quinn            81,38058
ro                      0,613231   Durbin-Watson            0,748726
Not: ö.s. ve ö.h. ölçünlü sapma ve ölçünlü hatayı göstermektedir.
```

SEK Bağlanım Çıktısının Yorumlanması

- Gretl çıktısına göre marjinal tüketim eğilimi (MTE) 0,59'dur.
- Buna göre gelir 1 lira arttığında tüketimin de 59 kuruş artması beklenmektedir.
- Sabit terim, toplam gelir sıfır olduğunda toplam tüketimin yaklaşık 8 milyon lira olacağını göstermektedir.
- Sıfır gelirin gözlem aralığı dışında kalan ve gerçek hayatta olanaksız bir değer olmasından dolayı, sabit terimin böylesi bir mekanik yorumu iktisadi anlam içermemektedir.
- Gretl $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{u}_i$ için ölçünlü hataları sırasıyla 1,85509 ve 0,0170331 ve 1,748114 olarak hesaplamıştır.
- Yukarıdaki değerlerin karesi alınarak $\text{var}(\hat{\beta}_1) = 3{,}44136$ ve $\text{var}(\hat{\beta}_2) = 0{,}000290126$ ve $\hat{\sigma}^2 = 3{,}05590$ varyansları da kolayca bulunabilir.
- $r^2 = 0{,}985$ değeri ise bağlanım modelinin verilere gerçekçi kabul edilemeyecek kadar iyi yakıştığını göstermektedir.

Önümüzdeki Dersin Konusu

Önümüzdeki ders

Normallik varsayımı ve ilişkin dağılımlar