Çoklueşdoğrusallık

Çoklueşdoğrusallığın Sonuçları

Ekonometri 2 – Konu 6

Sürüm 2,0 (Ekim 2011)

UADMK Açık Lisans Bilgisi

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

Ders Planı

1. Çoklueşdoğrusallığın Sonuçları
   - Kuramsal Sonuçlar
   - Uygulamaya İlişkin Sonuçlar
   - Açıklayıcı Örnek

Kuramsal Sonuçlar

- Çoklueşdoğrusallık tama yakın olsa bile SEK tahmincileri yansız ve enaz varyanslıdırlar.
- Diğer bir deyişle, çoklueşdoğrusallık durumunda da SEK tahmincileri EDYT'dirler.
- Çoklueşdoğrusallığın tek etkisi, ölçünlü sapması düşük tahminler yapmayı güçleştirmesidir.
- Kuramsal anlamda (1) çoklueşdoğrusallık, (2) az sayıda gözlem ve (3) yüksek varyanslı bağımsız değişkenler kavramları aynı sorunun üç farklı şekilde dile getirilmesidir.
- Goldberger gibi bazı ekonometriciler, örneklem büyüklüğü konusunu vurgulamak için çoklueşdoğrusallık terimi yerine "mikrosayıdalık" (micronumerosity) sözcüğünü yeğlerler.

Kuramsal Sonuçlar

- Çoklueşdoğrusallık temelde bir örneklem ya da örneklem bağlanımı olgusudur.
- Diğer bir deyişle, X değişkenleri anakütlede doğrusal ilişkili olmasalar bile eldeki örneklemde doğrusal ilişkili olabilirler.
- ABİ'yi tahmin etmek üzere kullanılan bir örneklemdeki X'ler yüksek bir çoklueşdoğrusallık gösterir ise bunların Y üzerindeki tekil etkilerini ayırmak zorlaşır.
- Kısaca eldeki örneklem tüm X'leri çözümlemeye katmaya yetecek kadar zengin olmayabilir.

Kuramsal Sonuçlar

- Örneklemin yeterliliği sorununa örnek olarak aşağıda verilen tüketim-gelir örneğini ele alalım:

$$\text{Tüketim} = \beta_1 + \beta_2\text{Gelir} + \beta_3\text{Servet} + u_i$$

- İktisat kuramına göre gelir ve servet, tüketim harcamalarını açıklamada önemli iki değişkendir.
- Ancak veriler derlendiğinde bu iki değişken tam olmasa bile yüksek ilişkili çıkar.
- Diğer bir deyişle, gelir ve servetin tüketim harcamaları üzerindeki etkilerini örneklemde ayırmak zor olabilir.
- Bu ayrımı yapabilmek için ise geliri az ama serveti çok olan ve geliri çok ama serveti az olan kimselerin yeterli sayıda örneklem gözlemini edinebilmek gereklidir.
- Kesit verilerinde bunu sağlamak mümkün olabilse de toplu zaman serilerinde buna erişmek neredeyse imkansızlaşır.

Uygulamaya İlişkin Sonuçlar

Tama yakın çoklueşdoğrusallık durumlarında, uygulamada şu sonuçlarla karşılaşılabilir:

- SEK tahmincileri, EDYT olmalarına karşı yüksek varyans ve kovaryanslıdırlar.
- Yüksek varyanslar nedeniyle güven aralıkları geniş olma eğilimindedir.
- Geniş güven aralıkları ise katsayı tahminlerine ilişkin sıfır önsavlarının reddedilememesine ve birçok *t* oranının istatistiksel olarak anlamlı olmamasına yol açar.
- Bir ya da daha çok katsayının anlamlı olmamasına karşın bütünün yakışma iyiliğinin ölçüsü R^2 yüksek olabilir.
- SEK tahminleri "sağlam" (robust) olmayabilirler. Diğer bir deyişle, verilerdeki küçük değişmelere duyarlı olabilirler.

Yüksek Varyans ve Kovaryans Sorunu

- Yüksek varyans ve kovaryans sorununu görebilmek için üçlü bağlanıma ait şu ilişkileri anımsayalım:

$$
\begin{aligned}
\mathrm{var}(\hat{\beta}_2) &= \frac{\sigma^2}{\sum x_{2i}^2(1 - r_{23}^2)} \\
\mathrm{var}(\hat{\beta}_3) &= \frac{\sigma^2}{\sum x_{3i}^2(1 - r_{23}^2)} \\
\mathrm{cov}(\hat{\beta}_2, \hat{\beta}_3) &= \frac{-r_{23}\sigma^2}{(1 - r_{23}^2)\sqrt{\sum x_{2i}^2 \sum x_{3i}^2}}
\end{aligned}
$$

- Buradaki r_{23} terimi X_2 ile X_3 arasındaki ilinti katsayısıdır.
- Eşdoğrusallık düzeyi yükselirken, diğer bir deyişle r_{23} 1'e yaklaşırken, iki tahmincinin varyanslarının artarak sonsuza yaklaştığına dikkat ediniz.

Yüksek Varyans ve Kovaryans Sorunu

- Çoklueşdoğrusallık altında varyans ve kovaryansların büyüme hızını görmek için "varyans şişme çarpanı" (variance inflating factor) kavramından yararlanılabilir:

$$\text{VŞÇ} = \frac{1}{(1 - r_{23}^2)}$$

- Yukarıdaki formüle göre r_{23} 1'e yaklaşırken VŞÇ değeri de sonsuza yakınsamaktadır.
- VŞÇ tanımı kullanılarak $\hat{\beta}_2$ ve $\hat{\beta}_3$'nın varyansları şöyle gösterilebilir:

$$\begin{aligned} \text{var}(\hat{\beta}_2) &= \frac{\sigma^2}{\sum x_{2i}^2}\text{VŞÇ} \\ \text{var}(\hat{\beta}_3) &= \frac{\sigma^2}{\sum x_{3i}^2}\text{VŞÇ} \end{aligned}$$

Yüksek Varyans ve Kovaryans Sorunu

- r_{23} artarken varyans ve kovaryansların büyümelerine ilişkin bir örnek olarak, şu çizelgeyi inceleyelim:

Çizelge: r_{23}'teki Artışın Etkisi

r_{23} Değeri	VŞÇ	var($\hat{\beta}_2$)	cov($\hat{\beta}_2, \hat{\beta}_3$)
0,00	1,00	$\times$ 1	0
0,50	1,33	$\times$ 1,33	$\times$ 0,67
0,70	1,96	$\times$ 1,96	$\times$ 1,37
0,80	2,78	$\times$ 2,78	$\times$ 2,22
0,90	5,76	$\times$ 5,76	$\times$ 4,73
0,95	10,26	$\times$ 10,26	$\times$ 9,74
0,97	16,92	$\times$ 16,92	$\times$ 16,41
0,99	50,25	$\times$ 50,25	$\times$ 49,75
0,995	100,00	$\times$ 100,00	$\times$ 99,50
0,999	500,00	$\times$ 500,00	$\times$ 499,50

Yüksek Varyans ve Kovaryans Sorunu

- Çizelgede görüldüğü gibi, yüksek bir ölçünlü hata anakütle katsayılarının güven aralıklarının geniş olmasına neden olmaktadır.
- Örnek olarak $r_{23} = 0{,}95$'ken β_2'nin güven aralığı da $r_{23} = 0$ durumuna oranla $\sqrt{10{,}26}$ ya da yaklaşık 3 kat büyüktür.
- Ayrıca, tahmin edilen ölçünlü hatalardaki artış t değerlerini de küçültmektedir.
- Bu yüzden anakütleye ait gerçek katsayının sıfır olduğuna ilişkin varsayımlar daha az reddedilir.
- Son olarak, katsayılar istatistiksel olarak anlamlı olmasa bile kovaryansın yüksek olmasından dolayı R^2 de yüksek, örnek olarak 0,90'ın üstünde olabilir.
- Demek ki anlamlı olmayan t değerleriyle birlikte görülen yüksek bir R^2, çoklueşdoğrusallığın belirtilerinden biridir.

Küçük Değişmelere Duyarlılık Sorunu

- Çoklueşdoğrusallık durumunda, bağlanım tahminleri ve bunların ölçünlü hataları verilerdeki küçük değişmelere yüksek duyarlılık gösterirler.
- Bunu görmek için şu iki varsayımsal veri setine bakalım:

Y	X_2	X_3
1	2	4
2	0	2
3	4	12
4	6	0
5	8	16

Y	X_2	X_3
1	2	4
2	0	2
3	4	0
4	6	12
5	8	16

- İki veri seti arasındaki tek fark X_3'ün üçüncü ve dördüncü gözlemlerinin yer değiştirmiş olmasıdır.

Küçük Değişmelere Duyarlılık Sorunu

- Birinci veri setine dayanarak şu sonuçlar bulunur:

$$
\begin{array}{llll}
\hat{Y}_i = & 1{,}1939 & +\,0{,}4463\,X_{2i} & +\,0{,}0030\,X_{3i} \\
\text{öh} & (0{,}7737) & (0{,}1848) & (0{,}0851) \\
t & (1{,}5431) & (2{,}4151) & (0{,}0358) \quad R^2 = 0{,}8101 \\
r_{23} = & 0{,}5523 & & \text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0{,}0087
\end{array}
$$

- İkinci veri seti ise aşağıdaki bağlanım bulgularını verir:

$$
\begin{array}{llll}
\hat{Y}_i = & 1{,}2108 & +\,0{,}4014\,X_{2i} & +\,0{,}0270\,X_{3i} \\
\text{öh} & (0{,}7480) & (0{,}2721) & (0{,}1252) \\
t & (1{,}6187) & (1{,}4752) & (0{,}2158) \quad R^2 = 0{,}8143 \\
r_{23} = & 0{,}8285 & & \text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0{,}0282
\end{array}
$$

- Görüldüğü gibi sonuçlar önemli farklılıklar sergilemektedir.

Açıklayıcı Örnek

- Çoklueşdoğrusallığa bir diğer örnek olarak, Türkiye'nin farklı illerinde faaliyet gösteren şehirlerarası otobüs firma sayılarını inceleyen aşağıdaki modeli ele alalım.

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

- Burada
  - Y ilde faaliyet gösteren otobüs firma sayısını (adet),
  - $X2$ ildeki toplam otomobil sayısı (bin adet),
  - $X3$ ise ildeki yetişkin nüfusu (milyon kişi)

  göstermektedir.
- Dikkat: İldeki nüfus ile otomobil sayısı arasında yüksek bir eşdoğrusallık gözleneceği açıktır.

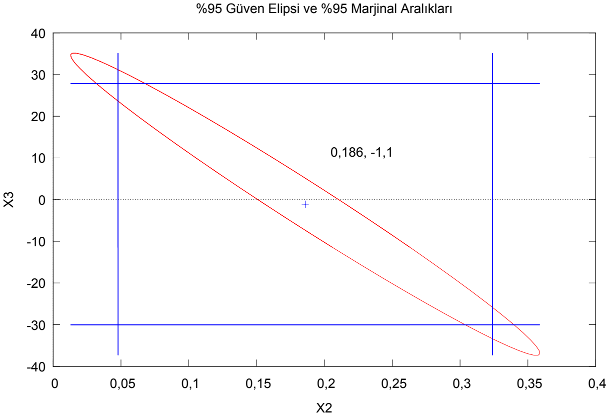
Açıklayıcı Örnek

- Otobüs firmalarının otomobil sayıları ve nüfus ile olan ilişkisinin doğrusal olduğunu varsayarsak şunu buluruz:

$$\begin{array}{llll} \hat{Y}_i = & 26{,}6672 & + 0{,}1859\, X_{2i} & - 1{,}0990\, X_{3i} \\ \text{öh} & (3{,}7763) & (0{,}0693) & (14{,}5375) \\ t & (7{,}0617) & (2{,}6808) & (-0{,}0756)\ R^2 = 0{,}7455 \end{array}$$

- Sonuçlar, otomobiller ve nüfusun birlikte firma sayılarındaki değişimin yaklaşık %75'ini açıkladığını göstermektedir.
- Diğer yandan, nüfusun eğim katsayısı istatistiksel olarak anlamlı değildir ve üstelik işareti de yanlıştır.
- Ayrıca, $\beta_2 = \beta_3 = 0$ önsavını sınamak için bir ortak güven aralığı belirlendiğinde bu önsav reddedilmez.
- Bunu görmek için bildik F sınamasına başvurulabilir.
- F sınaması yerine X_2 ile X_3'ün güven elipsinin 0 noktasını içerip içermediğine de bakılabilir.

Açıklayıcı Örnek



- Güven elipsinin doğruyu andıran şeklinin X_2 ile X_3 arasında tama yakın bir eşdoğrusallığı gösterdiğine dikkat ediniz.

Açıklayıcı Örnek

- Çözümlemeyi bir adım ileriye götürür ve X_3'ün X_2'ye göre bağlanımını hesaplarsak aşağıdaki sonuçları elde ederiz.

$$
\begin{array}{llll}
\hat{X}_{3i} = & 0{,}1620 & + \; 0{,}0047 \; X_{2i} & \\
\text{öh} & (0{,}0228) & (9{,}25e\text{-}05) & \\
t & (7{,}0913) & (50{,}7795) & r^2 = 0{,}9703
\end{array}
$$

- Buna göre X_3 ile X_2 arasında oldukça yüksek bir eşdoğrusallık bulunmaktadır.
- Ayrıca Y'nin X_2 ve X_3'e göre ayrı ayrı ikili bağlanımlarını alacak olursak, eğim katsayılarının işaretlerinin doğru ve anlamlılık düzeylerinin de yüksek olduğunu görürüz.
- Bu da gösterir ki yüksek çoklueşdoğrusallık gösteren X değişkenlerinden birini modelden çıkartmak, çoğu zaman diğer(ler)inin istatistiksel olarak anlamlı çıkmasını sağlar.

Önümüzdeki Dersin Konusu

Önümüzdeki ders

Çoklueşdoğrusallığı saptamak ve düzeltmek