Çoklueşdoğrusallık

Çoklueşdoğrusallığı Saptamak ve Düzeltmek

TÜBA
TÜRKİYE BİLİMLER AKADEMİSİ

UADMK
ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Ekonometri 2 – Konu 7

Sürüm 2,0 (Ekim 2011)

UADMK
ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

UADMK Açık Lisans Bilgisi

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

Ders Planı

1. Çoklueşdoğrusallığı Saptamak ve Düzeltmek
   - Var Olup Olmadığını Anlamak
   - Çoklueşdoğrusallığı Düzeltici Önlemler

Var Olup Olmadığını Anlamak

Bir bağlanımda çoklueşdoğrusallığın varlığını anlama konusu ile ilgili olarak şu noktalara dikkat edilmelidir:

- Çoklueşdoğrusallık nitelik değil nicelik sorunudur. Anlamlı bir ayrım çoklueşdoğrusallığın çeşitli dereceleri arasında yapılmalıdır.
- Çoklueşdoğrusallık örneklemin bir özelliği olduğu için çoklueşdoğrusallığa ilişkin bir sınama yapılamaz. Ancak derecesi ölçülebilir.
- Çoklueşdoğrusallığın var olup olmadığını anlamak ve eğer varsa derecesini ölçmek için tek bir yöntem yoktur. Bunun yerine izlenebilecek birkaç gevşek kural vardır.

Var Olup Olmadığını Anlamak

Çoklueşdoğrusallığın var olup olmadığını anlamak için kural olarak yararlanılabilecek bazı belirtiler şunlardır:

1. Yüksek R^2'ye karşı anlamlı olmayan t oranları
2. Değişken çiftleri arasında yüksek ilinti
3. Yüksek dereceli kısmi ilintilerin yüksek olması
4. Yardımcı bağlanımlarda görülen güçlü ilişkiler
5. Düşük özdeğerler ya da yüksek koşul endeksi değeri
6. Yüksek varyans şişme çarpanları

Var Olup Olmadığını Anlamak

Kural 1: Yüksek R^2'ye karşı anlamlı olmayan t oranları

Kısmi eğim katsayıları tekil olarak sıfırdan farklı değilken R^2 değerinin yüksek (örneğin 0,8 ve üzeri) bulunması.

- Bu klasik belirtinin kötü yanı aşırı güçlü olmasıdır.
- Diğer bir deyişle, bu tanı ancak X'lerin Y üzerindeki tüm etkileri birbirinden ayırt edilemeyecek noktadaysa çoklueşdoğrusallığı zararlı sayar.
- Öyleyse bu durum çoklueşdoğrusallığın varlığı için yeterli ama gerekli değildir.

Var Olup Olmadığını Anlamak

Kural 2: Değişken çiftleri arasında yüksek ilinti

İki açıklayıcı değişken arasındaki ilinti katsayısının 0,8 gibi yüksek bir değer olması.

- Bu ölçütteki sorun ise yalnızca sıfırıncı dereceden ilintilere bakmanın tek başına yeterli olmamasıdır.
- İkiden fazla açıklayıcı değişken olması durumunda, basit ilintiler tekil olarak düşük (örneğin 0,5 ve altı) olsa bile çoklueşdoğrusallık ciddi derecede yüksek olabilir.

Var Olup Olmadığını Anlamak

Kural 3: Yüksek dereceli kısmi ilintiler

Sıfırıncı dereceden ilintilere güven sorunu nedeniyle bakılan yüksek dereceli kısmi ilinti katsayılarının yüksek çıkması.

- Örnek olarak Y'nin X_2, X_3, X_4'e göre bağlanımında yüksek bir $R^2_{1.234}$ ama düşük $r^2_{12.34}$, $r^2_{13.24}$, $r^2_{14.23}$ değerleri bulmak.
- Böyle bir durum; X_2, X_3 ve X_4'ün kendi aralarında yüksek ilintili olduğu ve dolayısıyla bunlardan en az birinin gereksiz olduğu izlenimini verir.
- Çoklueşdoğrusallık bir ya da daha çok değişkenin diğer değişkenlerin tam ya da tama yakın bir doğrusal bileşimi demek olduğu için, çok karmaşık şekillerde oluşabilir.
- Dolayısıyla kısmi ilintileri incelemek yararlıdır ama bu da yanılmaz bir gösterge değildir.

Var Olup Olmadığını Anlamak

Kural 4: Yardımcı bağlanımlarda görülen güçlü ilişkiler

Hangi X'in diğer X'ler ile ilişkili olduğunu bulmak amacıyla her bir X_i değişkeninin diğerlerine göre bağlanımını tahmin etmek ve buna karşılık gelen R_i^2 değerini hesaplamak.

- Bu bağlanımlara "yardımcı" (auxiliary) bağlanım denir.
- Örnek olarak, $X_{2i} = a_1 + a_3 X_{3i} + a_4 X_{4i} + \cdots + a_k X_{ki} + u_i$ bağlanımından $R^2_{X_2}$ elde edilir.
- Daha sonra (k-2) ve (n-k+1) sd ile F dağılımına uyan şu istatistik hesaplanır:

$$F_i = \frac{R^2_{x_i.x_2x_3\ldots x_k}/(k-2)}{(1-R^2_{x_i.x_2x_3\ldots x_k})/(n-k+1)}$$

- Bulunan F_i eğer kritik değeri aşıyorsa, X_{2i}'nin diğer X'lerle çoklueşdoğrusal olduğu önsavı reddedilmez.

Var Olup Olmadığını Anlamak

- Yardımcı bağlanım yönteminde eğer hesaplanan bir F_i anlamlıysa, ilgili X_i'nin çıkartılıp çıkartılmayacağına ayrıca karar vermek gereklidir.
- Çok sayıda karmaşık doğrusal ilişki varsa karşılıklı ilişkileri saptamak güç olacağından, bu yöntem pek yararlı olmaz.
- Bütün R_i^2'leri tek tek sınamaya almaşık olarak "Klein'ın başparmak kuralı" (Klein's rule of thumb) da uygulanabilir.
- Bu kurala göre bir yardımcı bağlanımdan elde edilen R^2 bütünün R^2'sinden büyükse, çoklueşdoğrusallık dikkate alınmaya değecek kadar yüksek demektir.
- Diğer kurallar gibi bu kural da dikkatli kullanılmalıdır.

Var Olup Olmadığını Anlamak

Kural 5: Düşük özdeğerler ya da yüksek koşul endeksi değeri

Doğrusala yakın bağımlılıkların bir işareti olarak bir değişkene ait "özdeğer" (eigen value) büyüklüğünün düşük olması.

- Ekonometri yazılımları ile kolayca bulunabilen özdeğerler kullanılarak "koşul sayısı" (condition number) k ve "koşul endeksi" (condition index) KE değerleri şöyle hesaplanır:

$$k = \frac{\text{En Yüksek Özdeğer}}{\text{En Düşük Özdeğer}}, \qquad KE = \sqrt{k}$$

- Çoklueşdoğrusallık, k eğer 100 ile 1000 arasındaysa orta ya da güçlü derecededir. Eğer 1000'i aşıyorsa da ciddidir.
- Almaşık olarak, çoklueşdoğrusallık eğer KE 10 ile 30 arasındaysa orta ya da güçlüdür. 30'u aşıyorsa da ciddidir.
- Bu gevşek kural da diğerleri gibi dikkatli kullanılmalıdır.

Var Olup Olmadığını Anlamak

Kural 6: Yüksek varyans şişme çarpanları

X_i'nin diğer değişkenlerle ilişkisi artarken "varyans şişme çarpanı" (variance inflation factor) ya da kısaca "VŞÇ" (VIF) değerinin de artmasının bir ölçüt olarak kullanılması.

- k değişkenli modeldeki bir kısmi bağlanım katsayısının varyansı, VŞÇ cinsinden şu şekilde gösterilebilir:

$$\text{var}(\hat{\beta}_i) = \frac{\sigma^2}{\sum x_i^2}\left(\frac{1}{1-R_i^2}\right) = \frac{\sigma^2}{\sum x_i^2}\text{VŞÇ}_i$$

- $\hat{\beta}_i$ ve R_i^2 değerleri burada X_i'nin kısmi bağlanım ve belirleme katsayılarıdır. VŞÇ_i ise varyans şişme çarpanıdır.
- Bir başparmak kuralı olarak, bir değişkenin VŞÇ değeri 10'dan büyükse çoklueşdoğrusallığı da yüksektir denebilir.

Hoşgörü ve Varyans Şişme Çarpanı

- Bazı ekonometriciler VŞÇ yerine almaşık olarak "hoşgörü" (tolerance), kısaca "HOŞ" (TOL) değerini kullanırlar:

Hoşgörü

$$\text{HOŞ}_i = \frac{1}{\text{VŞÇ}_i} = (1 - R_i^2)$$

- Buna göre X_i diğer değişkenlerle tam ilişkiliyse $\text{HOŞ}_i = 0$, ilişkisizse de $\text{HOŞ}_i = 1$ olur.
- $\text{var}(\hat{\beta}_i)$ tanımından, yüksek bir HOŞ_i değerinin düşük bir σ^2 ya da yüksek bir $\sum x_i^2$ ile dengelenebildiği görülmektedir.
- Dolayısıyla küçük bir HOŞ (ya da büyük bir VŞÇ) yüksek ölçünlü hatalar bulmak için ne yeterli ne de gereklidir.

Çoklueşdoğrusallığı Düzeltici Önlemler

Çoklueşdoğrusallığın nasıl giderileceğine ilişkin kesin kurallar yoktur. Uygulanabilecek gevşek kurallardan bazıları şunlardır:

1. Önsel bilgilere başvurmak
2. Havuzlamalı verilerden yararlanmak
3. Bazı değişkenleri bırakmak
4. Verileri dönüştürmek
5. Ek ya da yeni veriler derlemek
6. Diğer iyileştirici önlemler

Önsel Bilgilere Başvurmak

Yöntem 1: Önsel bilgilere başvurmak

Çoklueşdoğrusallık sorununu gidermek için, modele önsel bilgilere dayalı sınırlamalar getirilebilir.

- Aşağıdaki modeli ele alalım:

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

- Burada Y_i tüketimi, X_{2i} geliri, X_{3i} de serveti göstermektedir. Gelir ile servet yüksek derecede eşdoğrusaldır.
- $\beta_3 = 0{,}1\beta_2$ olduğunu "önsel" (a priori) olarak bildiğimizi varsayalım. Bundan yararlanarak şunu elde edebiliriz:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + 0{,}1\beta_2 X_{3i} + u_i \\ &= \beta_1 + \beta_2 X_{4i} + u_i \end{aligned}$$

- Burada $X_{4i} = X_{2i} + 0{,}1 X_{3i}$'dir.
- $\hat{\beta}_2$ bir kez bulunduktan sonra $\hat{\beta}_3$ da β_2 ile β_3 arasında var olduğu düşünülen ilişkiden kolayca bulunabilir.

Önsel Bilgilere Başvurmak

- Önsel bilgiden yararlanabilmek için katsayılar arasındaki ilişkiye ait böyle bir bilginin öncelikle var olması gereklidir.
- Önsel bir bilgi daha önceki görgül çalışmalardan ya da modelin gerisinde yatan kuramdan gelebilir.
- Örnek olarak, Cobb-Douglas türü üretim işlevine dayanan bir modelde ölçeğe göre sabit getiri olması bekleniyorsa, $\beta_1 + \beta_2 = 1$ sınırlaması geçerli olur.
- Diğer yandan, modele sınırlama getirmek konusunda dikkatli olunmalıdır.
- Öncelikli amacımızın kuramın ileri sürdüğü önsel bilgileri modele zorla sokmak değil, bu beklentilerin kendisini sınamak olduğunu unutmamalıyız.

Havuzlamalı Verilerden Yararlanmak

Yöntem 2: Havuzlamalı verilerden yararlanmak

Dışsal ya da önsel bilginin bir biçimi de "havuzlamalı veriler" (pooled data) kullanmak, diğer bir deyişle yatay kesit ve zaman serisi verilerini bir araya getirmektir.

- Aşağıdaki bağlanımı ele alalım:

$$\ln Y_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln I_t + u_t$$

- Burada Y satış sayısını, P ortalama fiyatı, I geliri ve t ise zamanı göstermektedir.
- Zaman serisi verilerinde fiyat ve gelir değişkenleri yüksek bir eşdoğrusallık gösterme eğilimindedir.
- Diğer yandan, zaman içerisinde tek bir noktada derlenen kesit verilerinde fiyat çok değişikliğe uğramadığı için bu sorunla fazla karşılaşılmaz.

(...devam)

Havuzlamalı Verilerden Yararlanmak

- Yatay kesit verileri kullanılarak β_3'ün güvenilir bir tahmini bulunduktan sonra, zaman serisi bağlanımı şöyle yazılır:

$$Y_t^* = \beta_1 + \beta_2 \ln P_t + u_t$$

- Burada $Y^* = \ln Y - \beta_3 \ln I$ dönüştürmesi kullanılmıştır.
- Gelir etkisinden arındırmalı Y değerleri kullanılarak, artık β_2 tahmin edilebilir.
- Yatay kesit ve zaman serisi verilerini bir araya getirmenin bazı yorum sorunları doğurabileceği unutulmamalıdır.
- Örnek olarak, burada kesit verileriyle bulunan esnekliğin zaman serisiyle bulunan değere eşit olduğu örtük olarak varsayılmaktadır.

Bazı Değişkenleri Bırakmak

Yöntem 3: Bazı değişkenleri bırakmak

Ciddi bir çoklueşdoğrusallıkla karşılaşınca izlenebilecek bir diğer yol da değişkenlerden bir ya da birkaçını bırakmaktır.

- Diğer yandan, modelden değişken çıkartmak bir model "belirtim yanlılığı" (specification bias) ya da "belirtim hatası" (specification error) sorununa yol açabilir.
- Örnek olarak, doğru model aşağıdaki gibi olsun:

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

- Yanlışlıkla aşağıdaki modeli yakıştırmış olalım:

$$Y_i = b_1 + b_{12} X_{2i} + \hat{u}_i$$

- Bu durumda şöyle bir yanlılık ortaya çıkar:

$$E(b_{12}) = \beta_2 + \beta_3 b_{32}$$

- b_{32} burada X_3'ün X_2'ye göre bağlanımındaki eğimdir.

Bazı Değişkenleri Bırakmak

- Örnekte gösterilen b_{12}, β_2'nin "yanlı" (biased) tahmincisidir.
- Diğer bir deyişle b_{12} katsayısı, $\beta_3 b_{32}$ çarpımının işaretine bağlı olarak β_2'yi düşük ya da yüksek tahmin eder.
- Bu noktada, tama yakın çoklueşdoğrusallık varken bile SEK tahmincilerinin EDYT olduğunu anımsayalım.
- Çoklueşdoğrusallık modeldeki anakütle katsayılarının keskin olarak tahmin edilmesini engellemektedir.
- Bir değişkeni çıkartmak ise yanlılığa yol açarak anakütle katsayılarının gerçek değeri konusunda bizi yanıltabilir.
- Demek ki bazı durumlarda ilaç hastalıktan daha kötü olabilmektedir.

Verileri Dönüştürmek

Yöntem 4: Verileri dönüştürmek

Çoklueşdoğrusallık, verileri dönüştürerek de yok edilebilir.

- Uygulamada sıkça kullanılan veri dönüştürme yollarından biri, "oran dönüşümü" (ratio transformation) yöntemidir.
- Aşağıdaki modeli ele alalım:

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

- Burada Y_i tüketim, X_{2i} milli gelir ve X_{3i} de toplam nüfustur.
- Toplam gelirin nüfus ile eşdoğrusallık göstermesi sorunu, modelin kişi başına olarak belirtilmesiyle çözülebilir:

$$\frac{Y_i}{X_{3i}} = \beta_1 \left(\frac{1}{X_{3i}}\right) + \beta_2 \left(\frac{X_{2i}}{X_{3i}}\right) + \beta_3 + \left(\frac{u_i}{X_{3i}}\right)$$

- Buradaki sorunsa ilk bağlanımdaki u_i terimi sabit varyansla dağılıyor olsa bile dönüştürmeli bağlanımındaki u_i/X_{3i}'nin "farklıserpilimsellik" (heteroscedasticity) göstermesidir.

Verileri Dönüştürmek

- Bir diğer dönüştürme yöntemi olarak şu modeli ele alalım:

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

- Buradaki gelir (X_{2t}) ve servetin (X_{3t}) eşdoğrusallıklarının bir nedeni, bunların zaman içinde birlikte değişmeleridir.
- Zamanın ilk noktası t isteğe bağlı olduğu için şu yazılabilir:

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1}$$

- Yukarıdaki ikinci denklemi birinciden çıkartırsak, modeli "birinci fark" (first difference) biçiminde yazmış oluruz:

$$Y_t - Y_{t-1} = \beta_2 (X_{2t} - X_{2,t-1}) + \beta_3 (X_{3t} - X_{3,t-1}) + v_t$$

- Bu işlem eşdoğrusallık sorununu azaltır çünkü X_2 ile X_3'ün farklarının eşdoğrusal olması için önsel bir neden yoktur.
- Ancak birinci fark dönüşümü gözlemlerin sıralı olmadığı yatay kesit verileri için uygun değildir.
- Ayrıca, fark alma nedeniyle baştaki gözlem yitirildiği için serbestlik derecesi de bir azalır.

Yeni Veriler Derlemek

Yöntem 5: Yeni veriler derlemek

Çoklueşdoğrusallık bir örneklem özelliği olduğuna göre, daha büyük ya da aynı değişkenlerin yer aldığı farklı bir örneklemde daha az ciddi olabilir.

- Üç değişkenli model için varyans formülünü anımsayalım:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

- Görüldüğü gibi, örneklem büyürken $\sum x_{2i}^2$ de büyümekte ve buna koşut olarak azalan $\text{var}(\hat{\beta}_2)$ değeri β_2'nin daha kesin tahmin edilmesini sağlamaktadır.
- Ancak, iktisadi çalışmalarda ek veriler bulabilmek ya da "daha iyi" veriler derleyebilmek her zaman kolay değildir.

Diğer Düzeltici Önlemler

Yöntem 6: Diğer düzeltici önlemler

Çoklueşdoğrusallığı gidermeye yönelik başka dönüştürme ve tahmin yöntemleri de bulunmaktadır.

- Örnek olarak, açıklayıcı değişkenlerin çeşitli üstlerle girdiği "çokterimli" (polynomial) modellerde, çoklueşdoğrusallığı azaltmanın bir yolu X'leri sapmalar biçiminde kullanmaktır.
- Bunların dışında, çoklueşdoğrusallık sorununu çözmede
  - "etmen çözümlemesi" (factor analysis),
  - "baş bileşenler" (principal components),
  - "sırt bağlanımı" (ridge regression)

  gibi yöntemler de sıkça kullanılır.
- Bunlar daha ileri düzeydeki bir tartışmanın konusudur.

Önümüzdeki Dersin Konusu

Önümüzdeki ders

Farklıserpilimselliğin niteliği