

Ekonometrik Modelleme

Modellemeye İlişkin Konular




Ekonometri 2 – Konu 16
Sürüm 2,0 (Ekim 2011)



UADMK Açık Lisans Bilgisi

İşbu belge, “Creative Commons Attribution-Non-Commercial ShareAlike 3.0 Unported” (CC BY-NC-SA 3.0) lisansı altında bir açık ders malzemesi olarak genel kullanıma sunulmuştur. Eserin ilk sahibinin belirtilmesi ve geçerli lisansın korunması koşulu ile özgürce kullanılabilir, çoğaltılabilir ve değiştirilebilir. Creative Commons örgütü ve “CC-BY-NC-SA” lisansı ile ilgili ayrıntılı bilgi “<http://creativecommons.org>” adresinde bulunmaktadır. Bu ekonometri ders notları setinin tamamına “<http://www.acikders.org.tr>” adresinden ulaşılabilir.

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011 

Ders Planı

- 1 Modellemeye İlişkin Konular
 - Yuvalı-Dışı Modellerin Sınanması
 - Model Seçim Ölçütleri
 - Dışadüşenler ve Eksik Gözlemler

Yuvalı-Dışı Modellerin Sınanması

- Model belirtim sınamaları bağlamında, “yuvalı” (nested) ve “yuvalı-dışı” (non-nested) model ayrımı önemlidir.
- Şu iki modeli ele alalım:

$$\text{Model A: } Y_i = \alpha_1 + \alpha_2 X_1 + u_i$$

$$\text{Model B: } Y_i = \alpha_1 + \alpha_2 X_1 + \alpha_3 X_2 + \alpha_4 X_3 + v_i$$

- Model A, B içinde yuvalıdır çünkü onun özel bir durumudur.
- Şimdi de aşağıdaki modelleri karşılaştıralım:

$$\text{Model C: } Y_i = \alpha_1 + \alpha_2 X_1 + \beta_3 X_2 + u_i$$

$$\text{Model D: } Y_i = \beta_1 + \beta_2 Z_1 + \beta_3 Z_2 + v_i$$

- Model C ve Model D yuvalı-dışıdır çünkü biri diğherinin özel bir durumu olarak türetilemez.
- Böyle modeller arasında karşılaştırma yapmak için alışıldık t ve F sınamalarından farklı bir yaklaşım gereklidir.

Yuvalı-Dışı Modellerin Sınanması

- Model C ve Model D gibi iki yuvalı-dışı model arasında seçim yapmak için kullanılacak bir yaklaşım, aşağıdaki “melez” (hybrid) modeli tahmin etmektir:

$$\text{Model E : } Y_i = \lambda_1 + \lambda_2 X_1 + \lambda_3 X_2 + \lambda_4 Z_1 + \lambda_5 Z_2 + w_i$$

- Görüldüğü gibi, yukarıdaki model diğer iki modele yuvadır.
- Bu durumda, eğer $\lambda_2 = \lambda_3 = 0$ koşulu geçerli ise Model D doğrudur. Eğer $\lambda_4 = \lambda_5 = 0$ geçerliyse Model C doğru olur.
- Her iki koşul da alışıldık F sınaması ile kolayca sınanabilir. Bu sınamaya “yuvalı-dışı F sınaması” (non-nested F test) adı verilir.

Yuvalı-Dışı Modellerin Sınanması

- Uygulaması kolay olsa da yuvalı-dışı sınamaların bazı sakıncaları da vardır.
- Öncelikle X ve Z 'lerin yüksek ilintili olma olasılığı vardır ve bu da çoklueşdoğrusallık sorununa yol açar.
- Model C'yi temel alalım ve buna Z_1 ve Z_2 'yi ekleyelim. Eğer bu değişkenler R^2 'yi anlamlı biçimde yükseltmezse, Model D'yi reddederiz. Ancak eğer Model D'yi temel alıp X_1 ve X_2 'nin katkısını anlamlı bulmazsak, bu sefer de Model C'yi reddederiz. Yani sonuç ilk modele göre değişebilmektedir.
- Son olarak, yapay olarak belirtilen F yuva modeli büyük bir olasılıkla iktisadi anlam içermeyecektir.

Model Seçim Ölçütleri

Yuvalı olsun ya da olmasın, almaşık modeller arasında seçim yapmak için bir yöntem de belli bir ölçüyü temel almaktır.

Araştırmacılar tarafından başvurulan yaygın “**model seçim ölçütleri**” (model selection criteria) şöyle sıralanabilir:

- “**R-kare Ölçütü**” (R-square Criterion, R^2)
- “**Ayarlamalı R-kare**” (Adjusted R-square, \bar{R}^2)
- “**Akaike Bilgi Ölçütü**” (Akaike Information Criterion, AIC)
- “**Bayesçi Bilgi Ölçütü**” (Bayesian Information Criterion, BIC)
- “**Hannan-Quinn Ölçütü**” (Hannan-Quinn Criterion, HQC)

Tüm bu ölçütler KKT’yi enazlamaya dayanır. Ayrıca, R^2 dışında hepsi de açıklayıcı değişken sayısında “**tutumlu**” (parsimonious) olmayı ödüllendiricidir.

AIC, BIC ve HQC özellikle zaman serileri modellerinde gecikme uzunluğunu saptamada yaygın olarak kullanılmaktadır.

R-kare Ölçütü

- Bilindiği gibi, R^2 belirleme katsayısı 0 ve 1 arası değerler alır ve aşağıdaki şekilde hesaplanır:

$$R^2 = \frac{\text{BKT}}{\text{TKT}} = 1 - \frac{\text{KKT}}{\text{TKT}}$$

- R^2 ölçütünün başlıca sakıncası, bunun bir “örneklem içi” (in sample) yakışmanın iyiliği ölçütü olmasıdır.
- Diğer bir deyişle, R^2 'si yüksek diye modelin “örneklem dışı” (out of sample) gözlemleri iyi yordayacağına güvenilemez.
- İkinci bir zayıf nokta ise iki R^2 'nin karşılaştırılabilmesi için bağımlı değişkenlerin aynı olması zorunluluğudur.
- Son olarak, modele yeni bir değişken eklendiğinde aslında yordama hata varyansları artıyor olsa da R^2 yükselir.

Ayarlamalı R-kare Ölçütü

- 1971 yılında Henry Theil tarafından geliştirilen ayarlamalı R-kare tanımını anımsayalım:

$$\bar{R}^2 = 1 - \frac{\text{KKT}/(n - k)}{\text{TKT}/(n - 1)} \quad \text{ya da} \quad = 1 - (1 - R^2) \frac{n - 1}{n - k}$$

- Bilindiği üzere burada n örneklem büyüklüğünü ve k de açıklayıcı değişken sayısını göstermektedir.
- Yukarıda görüldüğü gibi, \bar{R}^2 modele açıklayıcı değişken eklemeyi cezalandırır ve bu nedenle R^2 'den küçük çıkar.
- Modeller arası karşılaştırma açısından \bar{R}^2 daha iyidir ama karşılaştırmanın geçerli olabilmesi için burada da bağımlı değişkenlerin aynı olması zorunluluğu unutulmamalıdır.

Akaike Bilgi Ölçütü

- Akaike ölçütünü 1974 yılında Hirotugu Akaike geliştirmiştir.
- Birden çok AIC tanımı vardır. Enküçük kareler tahmininde gretl, Akaike'nin kendi tanımına dayalı şu formülü kullanır:

$$AIC = -2\ell(\hat{\theta}) + 2k$$

- Burada $\ell(\hat{\theta})$, değiştirge tahminlerinin bir işlevi olan ençok log olabilirliği göstermektedir.
- AIC ne kadar küçükse yakışma da o kadar iyidir. Modeller karşılaştırılırken AIC değeri düşük olan yeğlenir.
- $2k$ teriminin AIC değerini yükselttiğine ve böylece değişken eklemeyi (\bar{R}^2 'den daha çok) cezalandırdığına dikkat ediniz.
- AIC ölçütünün en büyük üstünlüğü hem örneklem içi hem örneklem dışı başarıyı karşılaştırmada kullanılabilmesidir.
- Hem yuvalı hem yuvalı-dışı modellerde yararlıdır.

Bayesçi Bilgi Ölçütü

- Bu ölçüt 1978 yılında Gideon Schwarz tarafından önerildiği için Schwarz ölçütü olarak da bilinir. Formülü şudur:

$$\text{BIC} = -2\ell(\hat{\theta}) + k \log n$$

- Örnekleme birlikte $\ell(\hat{\theta})$ da arttığından dolayı, modele yeni eklenen bir değişken için AIC ölçütünün verdiği ceza büyük örneklerde yetersiz kalabilmektedir.
- BIC ise, AIC formülü ile karşılaştırılınca görülebildiği gibi, modele değiştire eklemeyi daha ciddi şekilde cezalandırır.

Hannan-Quinn Ölçütü

- Tutumlu modelleri AIC'ten daha fazla ödüllendiren bir diğer ölçüt de 1979 yılında Hannan ve Quinn tarafından önerilen HQC'dir:

$$\text{HQC} = -2\ell(\hat{\theta}) + 2k \log \log n$$

- Hannan ve Quinn, yinelemeli logaritma kanununa dayanan HQC'nin almasıklarından üstün olduğunu savunmuşlardır.
- HQC kullanımı, diğer iki ölçüt gibi yaygındır. Ancak, bu üç ölçütten birinin diğerlerinden üstün olduğu tartışmalıdır.
- AIC, BIC, ve HQC hesaplamasında kullanılan formüller bilgisayar yazılımından yazılımına farklılık gösterebildiği için, asıl önemli olan nasıl yorumlanacaklarını bilmektir.
- Gretl'da her üç ölçüt için de küçük değerler daha iyidir.

Dışadüşenler

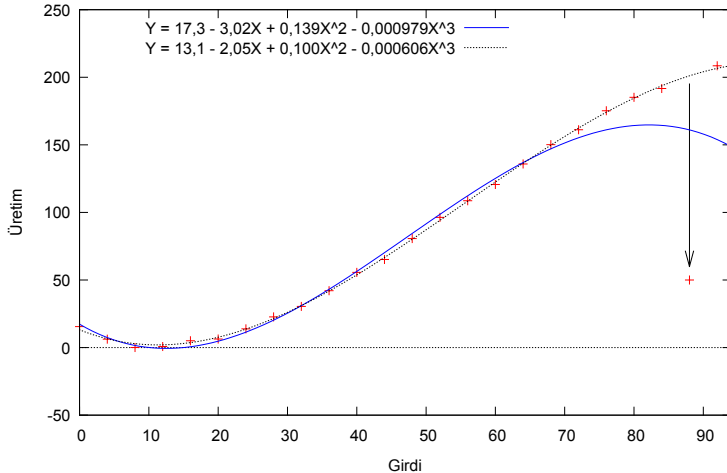
- Modelleme açısından önemli bir başlık da “**dışadüşenler**” (outliers) konusudur.
- $\hat{u}_i = (Y_i - \hat{Y}_i)$ şeklinde tanımlanan kalıntıların, bağlanım doğrusuna olan dikey uzaklığı gösterdiğini anımsayalım.
- Belli bir model bağlamında, diğer gözlemlere oranla fark edilir şekilde büyük kalıntıya sahip gözlemlere dışadüşen denir.
- Bu tür gözlemler önemlidir çünkü kaldıraç etkisi yaratarak bağlanım doğrusunu kendilerine doğru çekebilirler.
- Bağlanım doğrusunu kayda değer biçimde değiştiren böyle gözlemlere “**etkili gözlem**” (influential variable) adı verilir.
- **Dikkat:** Belli bir veri setinde birden fazla dışadüşen olabilir.

Dışadüşenler

- Dışadüşenleri saptamanın en temel yolu çizim yöntemidir çünkü bu gözlemler bağlanım doğrusundan uzaklıklarıyla dikkat çekerler.
- Biçimsel yöntemler de vardır. Gretl ve benzer ekonometri yazılımlarında, etkili gözlemleri bulmaya ve bunlara ait kaldıraç etkisini hesaplamaya yönelik işlevler de bulunur.
- Saptandıktan sonra, dışadüşenler konusunda nasıl bir yol izleneceğine karar vermek daha zor bir sorudur.
- Basitçe dışadüşenleri örneklemden çıkartmak ve geriye kalan gözlemlere odaklanmak düşünülebilir.
- Diğer taraftan, dışadüşen gözlemin sıradışı bir durumdan kaynaklandığı ve diğer gözlemler tarafından sağlanamayan bir bilgi içerebileceği de unutulmamalıdır.

Dışadüşenler

KALDIRAÇ ETKİSİ YÜKSEK BİR DIŞADÜŞENİN YOL AÇTIĞI HATALI TAHMİNLER



Eksik Gözlemler

Uygulamada kimi zaman karşılaşılan bir durum da veri setinde “**eksik gözlemler**” (missing observations) bulunmasıdır.

Bu durumun nedenleri şunlar olabilir:

- Anket verilerinde katılımcıların yanıtsız bıraktığı sorular
- Panel veri setlerinde zaman içerisinde ayrılan katılımcılar
- Güvenlik ya da özel bilgilerin korunması amacıyla gizli tutulan gözlemler
- Çeşitli ekonomik ya da siyasi nedenlerle bazı dönemlerde yapılamayan anketler ya da hesaplanmayan makro veriler

Veri setinde bir değer bile eksik olsa bağlantı hesaplanamaz. Özellikle küçük örneklerde eksik veriler veri setinin daha da küçülmesi gibi ciddi bir soruna neden olabilirler.

Eksik Gözlemler

- Farklı ailelerin tüketimlerini gelir, servet, eğitim gibi çok sayıda değişken ile açıklayan bir model düşünelim.
- Anket verilerinde ise farklı X değişkenleri için farklı birkaç aileye ait gözlemlerde eksiklik olsun.
- Diğer tüm bilgiler tamken, örnek olarak, yalnızca eğitim verisi eksik olan aile veri setinden çıkarılmak zorundadır.
- Böyle ailelerin varlığı örnekleme gözle görülür biçimde küçültebileceği gibi yanlılığa da neden olabilir.
- Örnekleme küçültmek yerine, eksik olan birkaç veri atama yolu ile tamamlanabilir.
- **“Atanan değerler”** (assigned values), eksiği olan değişkene ait örneklem ortalama ya da ortanca değeri olabilir.
- **Dikkat:** Dışadüşenler ve eksik gözlemler konusunda atılan adımlar, sonuçlar raporlanırken mutlaka açıklanmalıdır.

Önümüzdeki Dersin Konusu

Önümüzdeki ders

Nitel tepki ve doğrusal olasılık modeli