Ekonometrik Modelleme

Model Belirtimi ve Tanısal Sınamaları

Yrd. Doç. Dr. A. Talha YALTA
Ekonometri 2 Ders Notları
Sürüm 2,0 (Ekim 2011)

TÜBA
TÜRKİYE BİLİMLER AKADEMİSİ

UADMK
ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Açık Lisans Bilgisi

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

ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Ders Planı

Ders Planı

Model Belirtim Konusu

KDBM'nin 9. varsayımı, kullanılan modelin "doğru" belirtilmiş olduğudur. Bu varsayım altında şu ana kadar katsayı tahmini ve buna ilişkin sınamalar üzerine odaklanılmıştı.

Ancak, eğer model doğru belirtilmediyse "model belirtim hatası" (model specification error) ya da "model belirtim yanlılığı" (model specification bias) sorunu ortaya çıkar.

Bu bölümde şu sorulara yanıt arayacağız:

1. Uygulamada karşılaşılan belirtim hataları nelerdir?
2. Bu hatalar hangi sonuçları doğurur?
3. Belirtim hataları nasıl saptanabilir?
4. Düzeltmek için ne gibi önlemler alınabilir?
5. Almaşık modeller arasında nasıl seçim yapılır?

Model Belirtim Konusu

Model belirtimi konusu, uzmanlar arasında zaman zaman bakış ayrılıkları da olabilen geniş bir alandır. Ancak yaygın görüşe göre çözümlemede kullanılacak model şu özellikleri taşımalıdır.

1. Onanırlık: Model çıkarımlarının kabul edilebilir olması.
2. Kuram ile uyumluluk: Modelin iktisat düşüncesi açısından anlamlı olması.
3. Açıklayıcı değişken dıştürelliği: Bağlayanların hata terimi ile ilintisiz olması.
4. Değiştirge değişmezliği: Değiştirge tahminlerinin farklı örneklemlerde değişmemesi.
5. Veriler ile bağdaşma: Kalıntıların tümüyle rastsallık, diğer bir deyişle "beyaz gürültü" (white noise) özelliği göstermesi.
6. Kapsayıcılık: Modelin açıklama gücü bakımından almaşık modeller içinde en iyisi olması.

Belirtim Hatası Türleri

Bir modelin yukarıda sözü edilen özellikleri kaybetmesine yol açabilecek dört önemli hata türü şunlardır:

- "Atlanan değişken hatası" (omitted variable error)
- "İlgisiz değişken hatası" (irrelevant variable error)
- "Yanlış işlev biçimi" (wrong functional form)
- "Ölçüm hataları yanlılığı" (measurement errors bias)

Şimdi bu sorunları ve neden oldukları olumsuz sonuçları kısaca ele alalım.

Modeli Eksik Belirtme

- Atlanan değişken hatasını açıklamak için, aşağıdaki üç değişkenli modelin "doğru" olduğunu varsayalım:

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

- Bunun yerine ise aşağıdaki "eksik belirtimli model" (under specified model) kullanılsın:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + v_i$$

- X_3'ün X_2'ye göre ikili bağlanımındaki eğim katsayısı b_{32} olsun. Bu durumda şu eşitliğin geçerli olduğu gösterilebilir:

$$E(\hat{\alpha}_2) = \beta_2 + \beta_3 b_{32}$$

- Eşitlik gösteriyor ki α_2, β_2'nin yanlı bir tahmincisidir.
- Örnek olarak, X_3'ün Y üzerindeki etkisi (β_3) ile X_3'ün X_2 üzerindeki etkisi (b_{32}) aynı anda artı değerli ise, $\hat{\alpha}_2$ yukarı doğru yanlı olacak ve gerçek β_2'den hep yüksek çıkacaktır.

Modeli Eksik Belirtme

- Şimdi de $\hat{\alpha}_2$'nın ve $\hat{\beta}_2$'nın varyanslarını karşılaştıralım:

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

- $\hat{\beta}_2$, her ne kadar yansız olsa da, daha büyük varyanslıdır. X_2 ve X_3 arasındaki eşdoğrusallığın göstergesi olan ilinti katsayısının karesi arttıkça, aradaki fark da artmaktadır.
- Anlaşılıyor ki yanlılık ve varyans arasında bir "ödünleşim" (trade off) bulunmaktadır.
- Bu durumda yüksek eşdoğrusallık altında X_3'ü modelden çıkartıp, yanlı olsa da, $\hat{\beta}_2$ yerine $\hat{\alpha}_2$ kullanmak yeğlenebilir.
- Diğer yandan, iktisat kuramına dayanarak oluşturulan bir modelden değişken çıkartmanın zorunlu kalmadıkça asla önerilmediği unutulmamalıdır.

Modeli Eksik Belirtme

Özetle, modelde bulunması gereken X_3 değişkenini atlamak şu sonuçları doğurmaktadır:

1. Hatalı modeldeki sabit terim mutlaka yanlıdır ve tutarsızdır. Diğer bir deyişle, örneklem büyüdükçe yanlılık yok olmaz.
2. Hatalı modeldeki diğer $\alpha_2, \alpha_3, \ldots$ katsayıları da yanlıdır.
3. Eğer X_3 ile atlanılmayan bir değişken arasındaki eğim sıfır ise (örneğimizdeki $b_{32} = 0$ durumu), o zaman katsayı yanlı olmaz. Ancak uygulamada bu durum neredeyse hiç yoktur.
4. Yanlı katsayı tahminlerinden dolayı alışıldık güven aralıkları ve önsav sınama sonuçları yanıltıcı olabilir.
5. Hatalı modelde varyanslar genellikle daha küçüktür. Ancak yanlılık sorunu olduğu için, hatalı modelin yeğlenebilmesi eşdoğrusallığın çok yüksek olduğu durumlar ile sınırlıdır.

Modeli Aşırı Belirtme

- İlgisiz değişken hatasınının sonuçlarını gösterebilmek için, şimdi de doğru modelin şu olduğunu varsayalım:

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

- Araştırmacı ise aşağıdaki "aşırı belirtimli" (over specified) modeli kullanmakta diretiyor olsun:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_i$$

İlgisiz değişken eklemenin katsayılar üzerindeki etkisi şöyledir:

1. Hatalı modeldeki tüm katsayı tahminleri yansız ve tutarlıdır.
2. Bu nedenle alışıldık güven aralıkları ve önsav sınamaları geçerlidir.
3. Diğer taraftan katsayılar etkin değildir. Diğer bir deyişle, varyanslar doğru modeldekilerden daha büyüktür.

Modeli Aşırı Belirtme

- Aşırı belirtimli modeldeki $\hat{\alpha}_2$ tahmininin etkin olmadığını, varyansları karşılaştırarak görebiliriz:

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

- Doğru modelde X_3 olmadığı için paydada $(1 - r_{23}^2)$ teriminin yer almadığına dikkat ediniz.
- Hatalı modelde ise $\hat{\alpha}_2$ varyansı görece yüksek çıkacaktır.
- Aradaki fark X_3 ile X_2 arasındaki ilinti katsayısının karesi ile doğru orantılıdır.
- Demek ki ilgisiz bir değişken eklemek tahmin sonuçlarının kesinliğini azaltmak gibi ciddi bir sonuca yol açabilmektedir.
- Ayrıca, bilimde en az karmaşık açıklama yeğlendiği için, Model belirtiminde "tutumluluk ilkesi" (parsimony principle) her zaman özen gösterilmesi gereken önemli bir konudur.

Ölçüm Hataları

- Ölçüm hataları bir model belirtim hatası değildir. Ancak doğurabileceği sonuçlar ekonometrik modellemede ölçüm hatalarını da dikkate almayı gerekli kılar.
- Şimdiye kadar olan varsayımımızın aksine, çözümlemede kullandığımız veriler
  "kaydedici hatası" (clerical error),
  "atanan değerler" (assigned values),
  "yuvarlama" (rounding),
  "içdeğerleme" (interpolation),
  "dışdeğerleme" (extrapolation)
  gibi nedenlerden dolayı kesin doğru olmayabilir.
- İkincil kaynaklar tarafından yayınlanan verilerde yer alan hataları bilmek oldukça güçtür. Çoğu çalışma böyle verilere dayandığı için, uygulamada bu hata ile sıkça karşılaşılır.

Bağımlı Değişkendeki Ölçüm Hataları

- Bağımlı değişkendeki ve açıklayıcı değişkenlerdeki ölçüm hatalarının etkileri farklıdır. Öncelikle şu modele bakalım:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Burada Y Friedman tarafında öne sürülen "kalıcı tüketim" (permanent consumption) harcamasını, X ise cari geliri göstermektedir.
- Gerçekte kavramsal bir araç olan kalıcı tüketim doğrudan ölçülemediği için, elimizde, gözlenebilen tüketime dayalı şu değişken vardır:

$$Y_i^* = Y_i + v_i$$

- Yukarıdaki v, Y^*'daki ölçüm hatalarını gösteren rastlantısal bir terimdir.

Bağımlı Değişkendeki Ölçüm Hataları

- Y yerine Y^* kullanıldığında tahmin edilen model şu olur:

$$Y_i^* = \beta_1 + \beta_2 X_i + \epsilon_i$$

- Görülüyor ki yukarıdaki bağlanımda katsayılar aynı ve doğru şekilde tahmin edilebilmektedir.
- Diğer taraftan, $\epsilon_i = u_i - v_i$ biçimindeki bileşik hata teriminin varyansı daha yüksektir:

$$\mathrm{var}(u_i - v_i) = \mathrm{var}(u_i) + \mathrm{var}(v_i) + 2\mathrm{cov}(u_i, v_i)$$

- Öyleyse bağımlı değişkendeki ölçüm hataları katsayı nokta tahminlerini etkilememekte ancak güven aralıklarının geniş olmasına yol açarak etkinliği azaltmaktadır.

Açıklayıcı Değişkenlerdeki Ölçüm Hataları

- Açıklayıcı değişkende yer alan ölçüm hatalarına yönelik olarak, şimdi de şu modeli ele alalım:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Bu modelde Y cari tüketim, X ise "kalıcı gelir" (permanent income) olarak tanımlanmıştır.
- Kalıcı gelir de doğrudan ölçülemediği için, uygulamada gözlenebilen gelire dayalı bir değişken tanımı kullanılır:

$$X_i^* = X_i + w_i$$

- Burada w_i, X_i^*'deki ölçüm hatasını göstermektedir.

Açıklayıcı Değişkendeki Ölçüm Hataları

- X yerine X^* kullanılması aşağıdaki modele yol açar:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 (X_i + w_i) + u_i \\ Y_i &= \beta_1^* + \beta_2^* X_i + z_i \end{aligned}$$

- Buradaki bileşik hata terimi $z_i = u_i + \beta_2 w_i$ biçimindedir.
- İçerdiği β_2 teriminden dolayı, z_i, KDBM'nin hata terimi ve açıklayıcı değişkenlerin ilişkisiz olduğu varsayımını çiğner.
- Öyleyse açıklayıcı değişkenlerdeki ölçüm hataları ciddi bir sorundur, çünkü yukarıdaki durumda SEK tahminleri hem yanlı hem de tutarsızdır.
- Bu yanlılık sorununu gidermek zordur. Başvurulabilecek bir yol "araç değişkenler" (instrumental variables) yöntemidir.
- Eğer ölçüm hataları küçükse, ki bunu bilebilmek güçtür, uygulamada sorunu gözardı etmek zorunda kalınabilir.
- En doğru yol hatasız, doğru ölçülmüş verilerle çalışmaktır.

Ders Planı

Belirtim Hatalarının Sınanması

Görgül çalışmada kullanılan modelin “doğru” olduğu kesinlikle bilinemez. Bu nedenle, önce kurama dayanılır ve bir konunun özünü yakaladığı düşünülen model belirtilip tahmin edilir.

Daha sonra eldeki model çeşitli sınamalar ve almaşık modeller ile karşılaştırılarak değerlendirilir ve yeterliliğine karar verilir.

Uygulamada modelleme sorunlarını saptamada kullanılabilecek geleneksel yöntemlerden bazıları şunlardır:

- Kalıntıların incelenmesi
- Katsayı anlamlılık sınamaları
- Ramsey RESET sınaması
- Lagrange Çarpanı sınaması

Belirtim Hatalarının Sınanması

- Konuya ilişkin olarak toplam üretim maliyeti örneğini ele alalım. "Doğru" model aşağıdaki küplü işlev olsun:

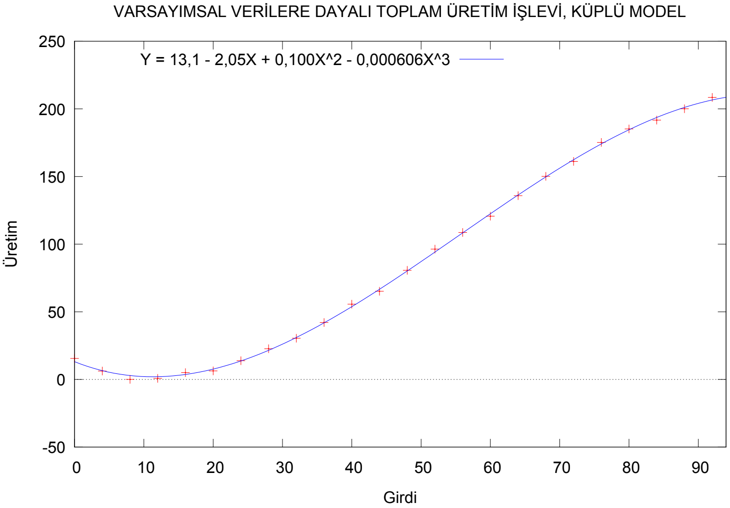
$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i$$

- Yukarıdaki model "doğru" olduğuna göre, aşağıda verilen doğrusal ve kareli modelleri kullanmak belirtim hatasına yol açacaktır:

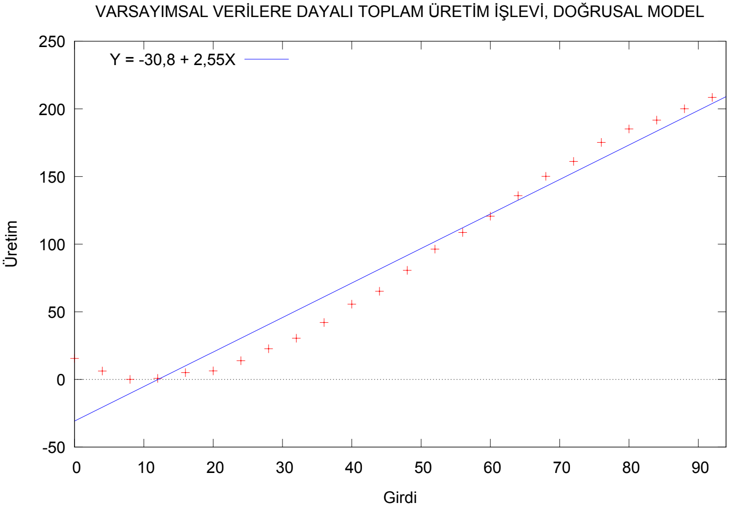
$$Y_i = \alpha_1 + \alpha_2 X_i + v_i$$
$$Y_i = \lambda_1 + \lambda_2 X_i + \lambda_3 X_i^2 + w_i$$

- Varsayımsal veriler kullanarak hatalı belirtimin yol açtığı yakıştırma sorunlarını sınayalım.

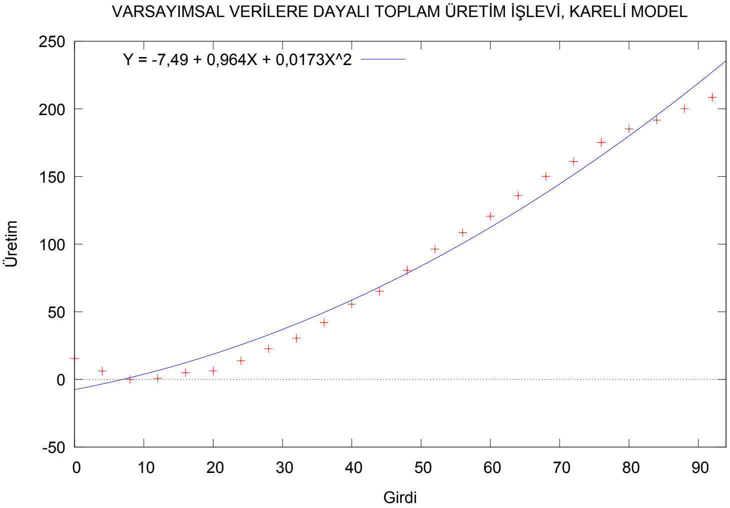
Belirtim Hatalarının Sınanması



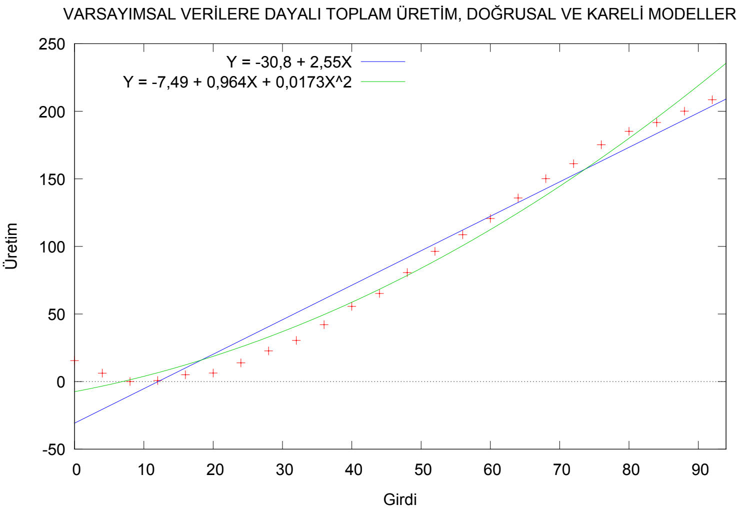
Belirtim Hatalarının Sınanması



Belirtim Hatalarının Sınanması

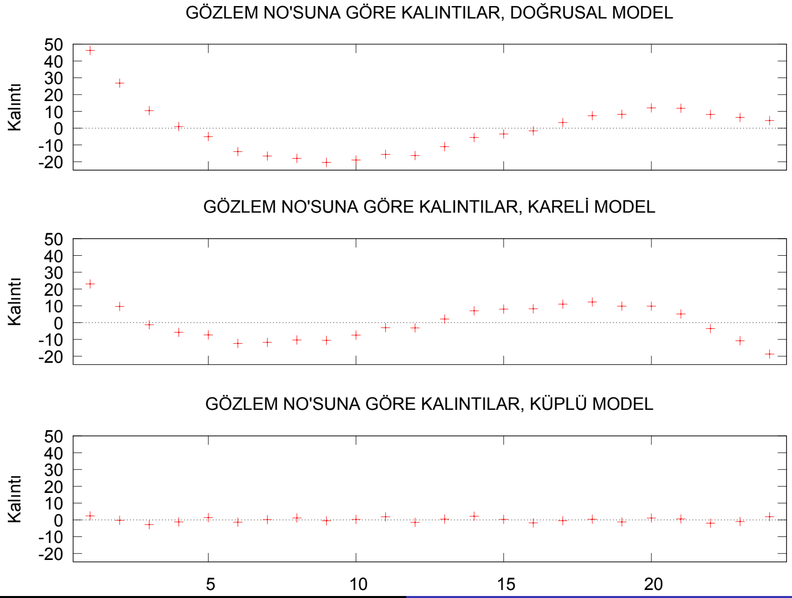


Belirtim Hatalarının Sınanması



Kalıntıların İncelenmesi

- Bağlanım kalıntıları, özellikle de yatay kesitsel verilerde, model belirtim hatalarını saptamak için yararlı bir görsel tanı aracıdır.
- Önemli bir değişkenin atlanması ya da yanlış işlev biçimi seçimi gibi sorunlar olduğunda kalıntılar da dikkat çekici örüntüler sergiler.
- Bir sonraki sayfada görüleceği üzere, hatalı yakıştırılan doğrusal ve kareli modellere ait kalıntı çizitleri çevrimsel salınımlar göstermektedir.
- Kareli modelde kalıntılar doğrusal bağlanıma göre belirgin biçimde azalmaktadır. Doğru yakıştırılan küplü modelde ise kalıntılar iyice azalmakta ve dalga görüntüsü de ortadan kaybolmaktadır.


GÖZLEM NO'SUNA GÖRE KALINTILAR, DOĞRUSAL MODEL
Kalıntı
50
40
30
20
10
0
-10
-20
GÖZLEM NO'SUNA GÖRE KALINTILAR, KARELİ MODEL
Kalıntı
50
40
30
20
10
0
-10
-20
GÖZLEM NO'SUNA GÖRE KALINTILAR, KÜPLÜ MODEL
Kalıntı
50
40
30
20
10
0
-10
-20
5
10
15
20

Katsayı Anlamlılık Sınamaları

- Modelde yer alan ilgisiz bir değişkenin yanlı tahminlere yol açmayıp, yalnızca katsayı varyanslarının büyümesi gibi daha az ciddi bir soruna yol açtığını anımsayalım.
- Bu nedenle aşırı belirtimin sınanması ve düzeltilmesi eksik belirtim sorunu yanında görece daha kolaydır.
- Aşağıdaki modeli ele alalım:

$$Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 X_i^2 + \gamma_4 X_i^3 + \gamma_5 X_i^4 + z_i$$

- Bu modelde X^4 değişkenin gerçekten anlamlı bir katkıda bulunup bulunmadığını saptamak için alışıldık t ve F sınamalarından yararlanılabilir.

Katsayı Anlamlılık Sınamaları

- Örnek olarak, küplü model şu sonuçları vermektedir:

$$\begin{array}{lccccccc} \hat{Y}_i = & 13{,}1307 & - & 2{,}0503\ X_i & + & 0{,}1009\ X_i^2 & - & 0{,}0006\ X_i^3 \\ \text{öh} & (1{,}0705) & & (0{,}1030) & & (0{,}0026) & & (1{,}88\text{e}{-}05) \\ p & (9{,}21\text{e}{-}11) & & (1{,}17\text{e}{-}14) & & (3{,}38\text{e}{-}20) & & (1{,}01\text{e}{-}18) \end{array}$$

- İlgisiz değişken içeren modele ait tahminler ise şöyledir:

$$\begin{array}{lccccccccc} \hat{Y}_i = & 14{,}2905 & - & 2{,}3622\ X_i & + & 0{,}1169\ X_i^2 & - & 0{,}0009\ X_i^3 & + & 1{,}49\text{e}{-}06\ X_i^4 \\ \text{öh} & (1{,}1463) & & (0{,}1805) & & (0{,}0082) & & (0{,}0001) & & (7{,}32\text{e}{-}07) \\ p & (1{,}36\text{e}{-}10) & & (5{,}93\text{e}{-}11) & & (1{,}44\text{e}{-}11) & & (3{,}27\text{e}{-}06) & & (0{,}0558) \end{array}$$

- Görüldüğü gibi, aşırı belirtimli modelde yer alan γ_5 tahmini büyüklük olarak sıfıra çok yakındır ve $\alpha = 0.05$ düzeyinde anlamlı da değildir.
- Bu noktada, X^4'ün yanında X^3 değişkeninin de ilgisiz olup olmadığını anlamak istersek $H_0 : \gamma_4 = \gamma_5 = 0$ sınırlamasını F sınaması ile sınayabiliriz.
- Bu doğrultuda hesaplanan $F = 604{,}4$ sınama istatistiğinin p değeri $6{,}327 \times 10^{-18}$ olduğu için, sıfır önsavı reddedilir.

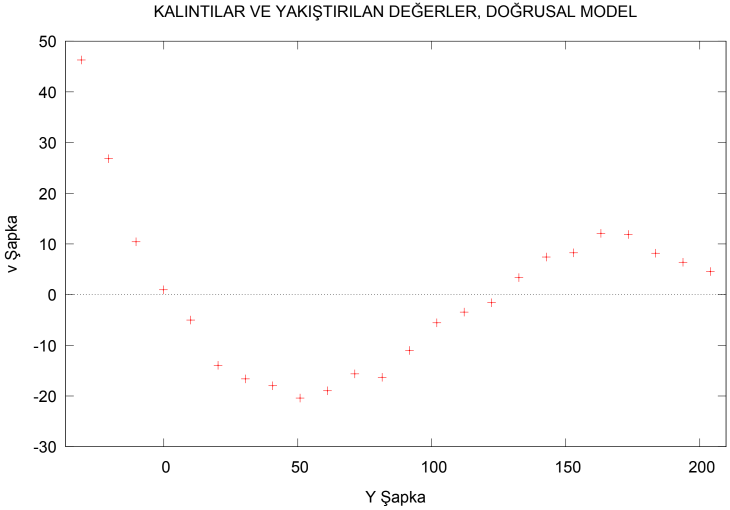
Ramsey RESET Sınaması

- Modelleme hatalarına ilişkin olarak J.B. Ramsey "Bağlanım Denklemi Belirtim Hatası Sınaması" (Regression Equation Specification Error Test), kısaca RESET adını verdiği genel bir sınama önermiştir.
- Bu sınama yaklaşımını açıklamak için $\sum \hat{u}_i$ ve $\sum \hat{u}_i \hat{Y}_i$'nın zorunlu olarak sıfır olduğunu anımsayalım ve toplam üretim işlevi örneğimizdeki doğrusal modele geri dönelim:

$$Y_i = \alpha_1 + \alpha_2 X_i + v_i$$

- Yukarıdaki hatalı modele ait $\hat{v}_i$ kalıntılarını alıp yakıştırılan $\hat{Y}_i$'lere karşı çizersek, düzenli bir örüntü ortaya çıkar.
- Bu durum ise yakıştırılan değerler ilk bağlanımda açıklayıcı değişken olarak dikkate alınırlarsa R^2'nin yükseleceği anlamına gelir.

Ramsey RESET Sınaması



Ramsey RESET Sınamasının Adımları

Ramsey RESET sınamasının adımları şöyledir:

1. Sınanacak model tahmin edilir ve yakıştırılan değerler kaydedilir.
2. Önceki çizitte de görülebildiği gibi, $\hat{v}$ ve $\hat{Y}$ arasındaki ilişki doğrusal-dışı olabilmektedir. Bu nedenle $\hat{Y}_i$'ların kareleri ve gerekli olduğu düşünülüyorsa küpleri ilk modele açıklayıcı değişkenler olarak katılır ve bağlanım yeniden hesaplanır.
3. Yeni modele eklenen değişkenlerin R^2'yi anlamlı biçimde artırıp artırmadığı bilindik F sınaması ile sınanır:

$$F = \frac{(R^2_{yeni} - R^2_{eski})/m}{(1 - R^2_{yeni})/(n-k)}$$

4. Hesaplanan F sınama istatistiği anlamlı ise, belirtim hatası olmadığını öne süren sıfır önsavı reddedilir.

Lagrange Çarpanı Sınaması

- "Lagrange çarpanı" (Lagrange multiplier) ya da kısaca "LÇ" (LM), RESET sınamasına benzeyen almaşık bir yöntemdir.
- Adından, kısıtlamalı bir eniyileme sorusundaki Lagrange çarpanları yöneyine dayandığı anlaşılan LÇ, uygulamada seyrek olarak bu yolla hesaplanır.
- Sınamada bağımlı değişken olarak tahmin edilen hatalar kullanılır ve bunların X'ler ve X^2, X^3 gibi değişkenlere göre bağlanımı tahmin edilir.
- Hata teriminin sıfır ortalamalı ve özilintisiz beyaz gürültü olduğu varsayımı nedeniyle, açıklayıcı değişkenler anlamlı olmamalıdır.
- Dikkat: LÇ kavuşmazsal bir sınamadır. Diğer bir deyişle, sonucuna ancak büyük örneklemlerde güvenilebilir.
- Dikkat: Kullanılan ek değişken sayısına özen gösterilmeli ve aşırı belirtimli bir modeli sınamaktan kaçınılmalıdır.

Lagrange Çarpanı Sınamasının Adımları

SEK'in aynı zamanda EO tahmincisi olduğunun varsayılabildiği durumlar için, LÇ sınamasının adımları şöyledir:

1. Model tahmin edilir ve $\hat{v}_i$ kalıntıları kaydedilir.
2. Model eğer hatalı ise, eldeki kalıntıların doğru modelde yer alması beklenen terimler ile ilişkili olması gereklidir. Buna göre, örnek olarak, şu yardımcı bağlanım hesaplanabilir:

$$\hat{v}_i = \theta_1 + \theta_2 X_i + \theta_3 X_i^2 + \theta_4 X_i^3 + \epsilon_i$$

3. Yukarıdaki bağlanıma ait gözlem sayısı ve R^2 çarpımının kavuşmazsal olarak dışlanacak değişken sayısı kadar sd ile χ^2 dağılımına uyduğu gösterilmiştir. Doğrusal-dışılık sınandığı için X kalır, X^2 ve X^3 ise dışlanır.
4. nR^2 çarpımı hesaplanır. Bu sınama istatistiği anlamlı ise, sınırlı bağlanımın doğru olduğu sıfır önsavı reddedilir ve baştaki modelde belirtim hatası olduğu sonucuna varılır.

Ders Planı

ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

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 diğerinin 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 kullanılabilecek 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:

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

- 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 başarımı 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:

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

- Örneklemle birlikte $\ell(\hat{\theta})$ da arttığından dolayı, modele yeni eklenen bir değişken için AIC ölçütünün verdiği ceza büyük örneklemlerde yetersiz kalabilmektedir.
- BIC ise, AIC formülü ile karşılaştırılınca görülebildiği gibi, modele değiştirge 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 almaşıklarından 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.

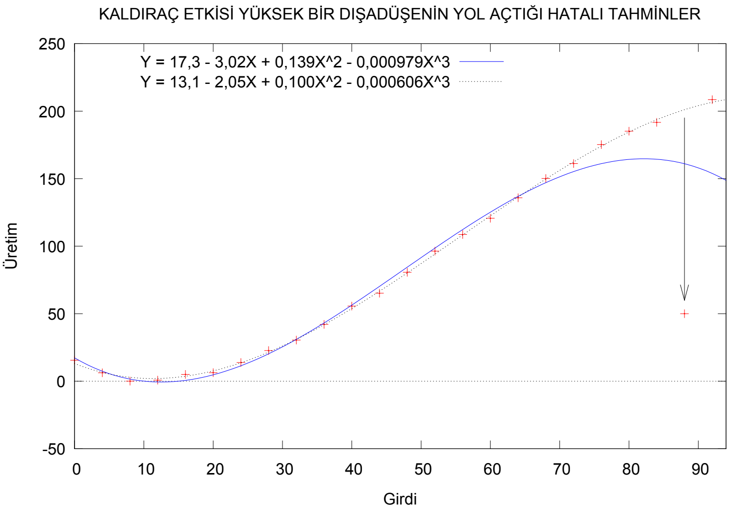
Dışadüşenler

- 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



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 bağlanım hesaplanamaz. Özellikle küçük örneklemlerde 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 örneklemi gözle görülür biçimde küçültebileceği gibi yanlılığa da neden olabilir.
- Örneklemi 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 ve Ödev

Ödev

Kitaptan Bölüm 13 “Econometric Modeling: Model Specification and Diagnostic Testing” okunacak.

Önümüzdeki Ders

Nicel Tepki Bağlanım Modelleri