1

SIRADAN EN KÜÇÜK KARELER (OLS) YÖNTEMİNİN ASİMPTOTİK ÖZELLİKLERİ

Hüseyin Taştan[1]

[1]Yıldız Teknik Üniversitesi
İktisat Bölümü

Ders Kitabı:
Introductory Econometrics: A Modern Approach (2nd ed.)
J. Wooldridge

14 Ekim 2012

2

OLS Yönteminin Asimptotik (Büyük Örneklem) Özellikleri

- Önceki derslerde SEKK (OLS) tahmin edicilerinin belirli sayıda gözlem içeren sonlu örneklem (finite sample) özelliklerini gördük.
- Bu özellikler herhangi bir örneklem büyüklüğü, n, için geçerliydi.
- Bunlar şu özelliklerdi:
  - Sapmasızlık (MLR.1-MLR4 altında)
  - BLUE özelliği (MLR.1-MLR5 altında)(Best=En Etkin)
- Varsayım MLR.6'nın işlevi: Hata terimi, u, normal dağılmıştır ve açıklayıcı değişkenlerden bağımsızdır.
- Bu varsayım kullanılarak OLS tahmin edicilerinin x'lere koşullu kesin örneklem dağılımları türetilebilir.
- Normal örnekleme dağılımı: t ve F testleri standart dağılımlara sahip olur (her gözlem boyutunda).

3

Asimptotik Özellikler

- OLS tahmin edicilerinin sonlu (finite) örneklem özelliklerinin yanında büyük örneklem özelliklerini de incelememiz gerekmektedir.
- Bu özellikler "*örneklem hacmi, n, sınırsız olarak artarken*" durumu altında incelenecektir.
- İnceleyeceğimiz özellikler: tutarlılık (consistency) ve asimptotik normallik
- OLS tahmincilerinin asimptotik özelliklerini ortaya koyarken MLR varsayımlarından bazılarını yumuşatabileceğiz.

4

Asimptotik Özellikler: Tutarlılık

Tanım

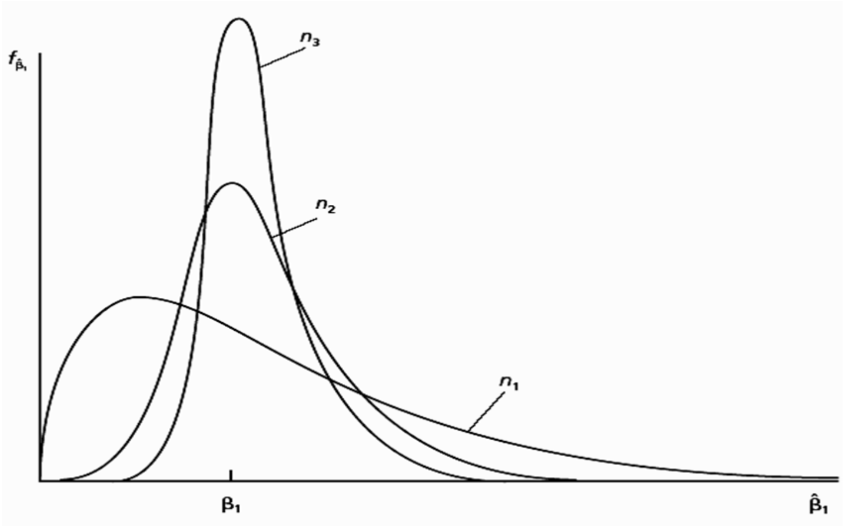
W_n bilinmeyen parametre vektörü θ'nın $\{Y_1, Y_2, \ldots, Y_n\}$ örneklemine dayanan bir tahmincisi olsun. İstediğimiz kadar küçük seçebileceğimiz herhangi bir $\epsilon > 0$ için, n sınırsız büyürken

$$P\left(|W_n - \theta| > \epsilon\right) \to 0, \quad n \to \infty$$

koşulu sağlanıyorsa, W_n, θ'nın tutarlı (consistent) bir tahmin edicisidir: plim$(W_n) = \theta$

- Örnek: Popülasyon ortalaması μ'nun tahmin edicisi: $\bar{x}$, aritmetik ortalama.
- Popülasyondan rassal örnekleme yoluyla çekilmiş (iid) örnekleme dayanılarak hesaplanan $\bar{x}$, μ'nun tutarlı bir tahmin edicisidir.

Farklı Örneklem Büyüklükleri ve Örnekleme Dağılımları: $n_1 < n_2 < n_3$



Asimptotik Özellikler: Tutarlılık

- Her bir örnek hacmi n için $\hat{\beta}_j$ bir olasılık dağılımına sahiptir. Bu dağılım, hepsi n hacimli farklı yinelenen örneklerde $\hat{\beta}_j$'nın alabileceği mümkün değerleri gösterir.
- Sapmasızlık özelliği gereği bu dağılımların ortalaması β_j'ye eşittir.
- Eğer tahmin edici $\hat{\beta}_j$ tutarlı ise, n arttıkça bu dağılımlar β_j etrafında daha dar olarak dağılır.
- $n \to \infty$ iken, $\hat{\beta}_j$'nin dağılımı tek bir nokta üzerine, β_j, düşer.
- Yani, ne kadar çok büyük örnek alabilirsek, bilinmeyen kitle parametresi β_j'yi o kadar az hata ile tahmin edebiliriz.

Asimptotik Özellikler: Tutarlılık

- Eğer n büyürken, kitle parametresi β_j'ye daha yakın bir noktaya gelmiyorsak o tahmin edici tutarsız (inconsistent) bir tahmin edicidir.
- Sapmasızlık (unbiasedness) için gereken varsayımlar tutarlılık için de yeterlidir: MLR.1-MLR.4 varsayımları altında OLS tahmin edicileri tutarlıdır.
- Ancak tutarlılık için MLR.3 varsayımından daha zayıf bir varsayım olan '**hata terimi ile açıklayıcı değişkenlerin ilişkisiz olma**' koşulu yeterlidir.

Asimptotik Özellikler: Tutarlılık

Varsayım MLR.3': Sıfır Ortalama ve Sıfır Korelasyon

$$\mathrm{E}(u) = 0$$

$$\mathrm{Cov}(x_j, u) = 0, \quad j = 1, 2, \ldots, k$$

- MLR.3 yerine MLR.3' varsayımı yapılırsa OLS tahmincilerinin tutarlı olduğu gösterilebilir.
- Bu varsayım her bir x açıklayıcı değişkeniyle hata teriminin ilişkisiz (sıfır korelasyonlu) olduğu anlamına gelir.
- MLR.3' varsayımı MLR.3 varsayımına göre daha zayıf bir varsayımdır. Sapmasızlığın sağlanması için yeterli değildir.
- MLR.3' varsayımının sağlanması MLR.3'ün sağlandığı anlamına gelmez. Ancak MLR.3 sağlanıyorsa MLR.3' otomatik olarak sağlanır.

Asimptotik Özellikler: Tutarlılık

- Basit regresyon modelinde eğim katsayısının OLS tahmin edicisinin aşağıdaki gibi yazılabildiğini biliyoruz:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)y_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

- Burada $y = \beta_0 + \beta_1 x_1 + u$ yerine yazılıp yeniden düzenlenirse:

$$\hat{\beta}_1 = \beta_1 + \frac{n^{-1}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)u_i}{n^{-1}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

- Bu ifadenin olasılık limiti (plim) alınırsa

$$\text{plim}(\hat{\beta}_1) = \beta_1 + \frac{\text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

- MLR.3' varsayımı gereği $\text{Cov}(x_1, u) = 0$ olduğundan

$$\text{plim}(\hat{\beta}_1) = \beta_1$$

Asimptotik Özellikler: Tutarlılık

- Asimptotik sapma:

$$\text{plim}(\hat{\beta}_1) - \beta_1 = \frac{\text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

- Asimptotik sapmanın işareti x_1 ile u arasındaki kovaryansın işaretine bağlıdır. $\text{Cov}(x_1, u)$'nun tahmin edilmesi, u gözlemlenemez olduğundan, mümkün değildir.
- Gerekli bir değişkenin model dışında bırakılması tutarsızlık yaratır. Örneğin gerçek popülasyon modeli aşağıdaki gibi olsun. İlk dört MLR varsayımı sağlanıyor olsun.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \nu$$

- x_2 model dışında bırakılırsa hata terimi $u = \beta_2 x_2 + \nu$ olur.

Asimptotik Özellikler: Tutarlılık

- Bu durumda olasılık limiti

$$\text{plim}(\hat{\beta}_1) = \beta_1 + \beta_2\delta_1$$

$$\delta_1 = \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

- Bu sonuca aşağıdaki ifadenin olasılık limitini alarak ulaştık:

$$\hat{\beta}_1 = \beta_1 + \beta_2\left(\frac{n^{-1}\sum(x_{i1} - \bar{x}_1)x_{i2}}{n^{-1}\sum(x_{i1} - \bar{x}_1)^2}\right)$$

$$\text{plim}\hat{\beta}_1 = \beta_1 + \beta_2 \underbrace{\frac{\text{plim } n^{-1}\sum(x_{i1} - \bar{x}_1)x_{i2}}{\text{plim } n^{-1}\sum(x_{i1} - \bar{x}_1)^2}}_{=\delta_1}$$

OLS tahmincilerinin tutarlılığı

- Bir tahmin edici tutarsız ise gözlem sayısı n arttırılarak sorun çözülemez. n sınırsız arttıkça OLS tahmin edicisi β_1'e değil $\beta_1 + \beta_2\delta_1$'e yaklaşır.
- x'lerden sadece birisinin (x_1 diyelim) u ile ilişkili olması genellikle (eğer x_1 diğer tüm x'lerle ilişkili ise) tüm OLS tahmin edicilerinin tutarsız olmasına yol açar.
- x_1 ile ilişkisiz olan x'ler varsa onların katsayıları tutarlıdır.

Tutarsızlık: Örnek

- y: ev fiyatları
- x_1: evlerin çöp arıtma tesislerine uzaklığı
- x_2: evin kalitesi (büyüklük, oda sayısı, bulunduğu semtin çekiciliği, ulaşım kolaylığı vs. gibi tüm faktörler)
- Bu modelde x_2'nin model dışında bırakıldığını düşünelim.
- $\beta_1 > 0$, $\beta_2 > 0$
- Eğer ortalamada kaliteli evler çöp arıtma tesislerinin uzağındaysa bu iki değişken (pozitif) ilişkili olur: $\delta_1 > 0$
- Bu durumda
$$\beta_1 + \beta_2\delta_1 > \beta_1$$
- β_1'in OLS tahmincisi $\hat{\beta}_1$, limitte $\beta_1 + \beta_2\delta_1$ değerine yaklaşır.

Asimptotik Normallik

- MLR.6 hata teriminin x'lere koşullu dağılımının normal olduğunu söylüyordu. Bu varsayım sonucunda y'nin x'lere koşullu dağılımı da normaldir.
- OLS tahmincilerinin sapmasızlığı için normallik varsayımına gerek yoktu. Normallik varsayımı betaşapkaların örnekleme dağılımlarının elde edilmesi ve istatistiksel çıkarsama yapılabilmesi için gerekliydi.
- Normal dağılım varsayımının sağlanmadığı durumda t ve F testlerini ve diğer istatistiksel çıkarsamaları yapamayacak mıyız?
- Eğer yeterince büyük bir örneklem varsa Merkezi Limit Teoreminden hareketle hata teriminin asimptotik normal dağılıma sahip olduğu varsayılarak testler yapılabilir.

Asimptotik Normallik: Teorem 5.2

- Gauss-Markov varsayımları altında
$$\sqrt{n}(\hat{\beta}_j - \beta_j) \sim^a \; Normal\left(0, \frac{\sigma^2}{a_j^2}\right)$$
- Asimptotik varyans:
$$\frac{\sigma^2}{a_j^2} > 0$$
$$a_j^2 = \text{plim}\left(n^{-1}\sum_{i=1}^{n}\hat{r}_{ij}^2\right)$$
- Burada $\hat{r}_{ij}^2$, x_j'nin tüm diğer x'ler üzerine regresyonundan elde edilen kalıntılardır.
- OLS tahmincileri standardize edilirse:
$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim^a \; N(0,1)$$

Asimptotik Normallik

- n arttıkça t test istatistiği standart normal dağılıma yaklaşacağından şu yazılabilir (asimptotik t istatistiği):
$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim^a \; t_{n-k-1}$$
- Teorem 5.2 geçerliyse artık MLR.6 Normallik varsayımına gerek kalmaz.
- Hata teriminin dağılımına ilişkin tek kısıt u'ların varyansının sonlu olmasıdır: $\sigma^2 > 0$. Bunu zaten her zaman varsayıyoruz.
- Teorem 5.2, sabit varyans ve sıfır koşullu ortalama varsayımlarını yapmaktadır.
- OLS tahmincilerinin standart hataları ($se(\hat{\beta}_j)$) $1/n$ oranında sıfıra yaklaşır.

Asimptotik Etkinlik

- OLS tahmincilerinin Gauss-Markov varsayımları altında BLUE olduklarını biliyoruz.
- OLS tahmincileri Gauss-Markov varsayımları altında asimptotik etkindir. Belli bir sınıf tahminciler kümesi içinde asimptotik varyansı en küçük tahminci OLS tahmincisidir.

$$Avar\left(\sqrt{n}(\hat{\beta}_j - \beta_j)\right) \leq Avar\left(\sqrt{n}(\tilde{\beta}_j - \beta_j)\right)$$

Burada $Avar$ asimptotik varyansı ifade etmektedir. $\tilde{\beta}_j$ ise OLS dışında başka bir tahmin edicidir.

Büyük Örneklem Testi: LM Test İstatistiği

- Doğrusal kısıtlamaların testinde Lagrange Çarpanı (LM - Lagrange Multiplier, score test) testi büyük örneklemlerde kullanılabilir.
- Bu test istatistiği sadece kısıtlanmış modelin tahminine dayanmaktadır. Bir yardımcı regresyon kurularak LM test istatistiği kolayca hesaplanabilmektedir.
- Örnek: Dışlama testi - kısıtlanmamış model aşağıdaki gibi olsun:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$$

- $H_0 : \beta_3 = \beta_4 = 0$, H_1: en az biri sıfırdan farklıdır.
- LM test istatistiği H_0 altında elde edilen kısıtlanmış modelin tahmininden elde edilen kalıntıların, kısıtlanmamış modeldeki tüm x'ler üzerine regresyonundan elde edilen determinasyon katsayısının R^2 gözlem sayısı n ile çarpılmasıyla elde edilir.

Büyük Örneklem Testi: LM Test İstatistiği

- $H_0 : \beta_3 = \beta_4 = 0$ altında kısıtlanmış model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- $\tilde{u}$ bu modelden elde edilen kalıntılar olsun. Kalıntıların tüm x'ler üzerine regresyonunu kuralım:

$$\tilde{u} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + hata$$

- Bu regresyondan elde edilen determinasyon katsayısı $R^2_{\tilde{u}}$ olsun. LM test istatistiği:

$$LM = nR^2_{\tilde{u}} \sim \ \chi^2_q$$

- Boş hipotez altında LM test istatistiği serbestlik derecesi $q =$ toplam kısıt sayısı olan ki kare dağılımına uyar.
- Karar: $LM > c$ ise H_0 reddedilir.

LM Test İstatistiği: Örnek

Yeni doğan bebeklerin sağlık düzeyi ve anne-babanın eğitim düzeyi: bwght.gdt

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u$$

- Bağımlı değişken: $y =$ yeni doğan bebek ağırlıkları (libre)
- Açıklayıcı değişkenler
  - x_1: annenin hamilelik süresince günde içtiği ortalama sigara sayısı
  - x_2: bebeğin doğum sırası
  - x_3: ailenin yıllık geliri
  - x_4: annenin eğitim düzeyi, yıl
  - x_5: babanın eğitim düzeyi, yıl.
- İlgilendiğimiz hipotez: $H_0 : \beta_4 = 0, \beta_5 = 0$, annenin ve babanın eğitim düzeyleri birlikte bebek ağırlıkları üzerinde bir etkiye sahip değildir.

LM Test İstatistiği: Örnek

1. adım: **Kısıtlanmış** modelin tahmini

$$\widehat{\text{bwght}} = \underset{(1.656)}{115.470} - \underset{(0.1088)}{0.5979}\,\text{cigs} + \underset{(0.6575)}{1.8323}\,\text{parity} + \underset{(0.0324)}{0.0671}\,\text{faminc}$$

$$n = 1191 \quad R^2 = 0.036 \quad F(3, 1187) = 14.953 \quad \hat{\sigma} = 19.796$$

Kısıtlanmış modelden elde edilen kalıntılar: uhat

2. adım: Kısıtlanmış modelin kalıntılarının tüm x'ler üzerine regresyonu

$$\widehat{\text{uhat}} = \underset{(3.729)}{-0.9456} + \underset{(0.110)}{0.0019}\,\text{cigs} - \underset{(0.659)}{0.0447}\,\text{parity} - \underset{(0.0366)}{0.011}\,\text{faminc}$$
$$- \underset{(0.319)}{0.370}\,\text{motheduc} + \underset{(0.283)}{0.472}\,\text{fatheduc}$$

$$n = 1191 \quad \boldsymbol{R^2 = 0.00242} \quad F(5, 1185) = 0.57491 \quad \hat{\sigma} = 19.789$$

LM Test İstatistiği: Örnek

3. adım: test istatistiğinin hesaplanması

$$LM = nR^2_{\tilde{u}} \sim \ \chi^2_q$$

$$LM = (1191)(0.00242) \sim \ \chi^2_2$$

$$LM = 2.88$$

2 serbestlik derecesinde kritik değer, $c = 5.99$. Buna göre H_0 reddedilemez. p-değeri $= 0.24$