

14.74

Ödev 2

4 Mart 2008 Çarşamba akşam saat 5'e kadar

Cevaplarınızı kısa ve öz tutun. Ödevi cevap formunda teslim edin. Ödevin iki kısmı bulunmaktadır 1) log dosyasının *gerekli* kısımlarının da dahil olduğu cevapların yazılması 2) Çalıştığınız do dosyasının *en son* hali (kapsamlı bir dosya)

1. Okullaşmanın Getirisini Ölçme: Endonezya'da INPRESS Okul İnşası Programı

Bu soru, okullaşmanın getirisini ölçmek için, derste size gösterilen, araç değişken tekniğini gösterecektir. (spesifik olarak eğitim yılının log saatlik ücretlere etkisini) Bu soru supa.dta adı verilen, Professor Duflo'nun ders izlencesinde adı geçen "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment" adlı makalesinde kullandığı Stata dosyasını kullanmaktadır. Bu veriseti 2.5Mb olduğu için, Stata'nın standart 1Mb'lık hafızası olduğundan, Stata oturumunuzu açtığınızda "set mem 20m" yazın; ayrıca "set matsize 800" de yazmalısınız. Notasyonu ve verileri anlamak için biraz zaman harcayın ve metodlar adım adım açıklanmıştır.

a. Deste tartışılan eğitim modelindeki ücret belirleme denklemini hatırlayın:
 $\ln w_i = a_i + bS_i$

Bu denklemdeki b parametresini yorumlayın. Ücretlerin log'unun okulda geçirilen yıl üzerine olmak üzere yapılan bir regresyonda EKK (en küçük kare) yönteminin kullanılması neden b'nin iyi bir tahminini vermez?

Problemin geri kalanında, b'nin yansız bir tahminini bulabilmek için birkaç araç değişken kullanmanız istenecek. Okulda geçirilen yılın log ücretler üzerindeki etkisini bulmak istiyoruz ama okulda geçirilen yıl endojen bir değişkendir. (buna "endojen regresör" diyoruz) Araç değişkenlerin mantığı ise, okullaşmada eksojen bir varyasyon yaratan entsrüman adı verilen başka değişkenler kullanmaktır. Bu soruda INPRES programı ile oluşan okullaşmadaki eksojen değişikliklerden yararlanacağız.

b. **Wald Tahmini**

- i. Önce INPRES'in **okula gidilen yıla** etkisini farkların farkı (FF) yöntemiyle tahmin edin; bu **ilk aşamadır**. Yaş düzleminde (genç-yaşlı) ve bölgesel düzlemde (çok yoğun-az yoğun) yatay varyasyon kullanın. Tahmini yorumlayın. Hangi varsayım altında, bu INPRES'in eğitim üzerine olan etkisinin yansız bir tahmindir?
- ii. Daha sonra, INPRES'in log saatlik ücretler üzerine olan etkisinin FF tahminini, i kısmındaki aynı varyasyon kaynağını kullanarak tahmin edin., bu **indirgenmiş formdur**. Tahmini yorumlayın. Hangi

varsayımlar altında bu INPRES'in ücretler üzerine olan etkisinin yansız bir tahminidir? Eldeki veriye dayanarak, bu varsayımı nasıl test ederiz?

- iii. Ücret belirleme denklemindeki b'nin Wald tahmini, indirgenmiş formdaki FF tahmininin ilk basamaktaki FF tahmininin bölümüne eşittir. Wald tahmininin yansızlığı için gerekli olan ek varsayım nedir? Eldeki veriye dayanarak, bu varsayımı nasıl test ederiz?
 - iv. Okula gidilen yılın getirisinin Wald tahminini hesaplayın ve yorumlayın.
- c. Dolaylı (İndirekt) en küçük kareler (DEKK)
B kısmında INPRES programındaki sadece bazı varyasyonları kullandık (sadece 2 yaş ve bölgesel grupları kullandık) Şimdi, tüm bölgesel varyasyonu kullanacağız. Daha önce, çok yoğun ve az yoğun ve genç-yaşlı gruplarının ortalamasını hesaplamıştınız, bu soruda ise, genç ve yaşlıların bölge-bölge ortalamasını hesaplayacaksınız.
Önce, *bölge bazında*, gençlerin ve yaşlıların ortalama okula gittikleri yılları, ortalama ücretlerini, 1971 yılındaki ortalama çocuk sayısını, ortalama program yoğunluğunu içeren yeni bir veri seti oluşturun. (İpucu: Stata kodunuzun ilk sayısı şöyle bir şey olmalı: **generate educ_yng=yeduc if young=1**. Daha sonra *collapse* komutunu kullanın)
Daha sonra, educ_dif adlı ortalama genç eğitimi eksi ortalama yaşlı eğitime eşit olan bir değişken oluşturun. Benzer şekilde, ortalama genç log ücret eksi yaşlı log ücret olarak wage_dif adlı bir değişken oluşturun.

- i. **İlk aşama:** $SY_j - SO_j = \alpha P_j + v_j$ (S okula gidilen yılı, P program yoğunluğunu gösteriyor; Y ve O genç ve yaşlı endeksleri, j ise bölgesel endeks) Denklem, tek sağ taraf değişkeni program yoğunluğuymuş gibi yazılmış ama v_j terimi hata terimini, sabit terimi ve diğer kontrol değişkenlerini (yani diğer sağ taraf değişkenlerini) de içeriyor gibi anlaşılabilir. Denklem sol tarafında sadece educ_dif olduğunu not edin. α 'yı tahmin etmek için educ_dif program yoğunluğu üzerine olmak üzere bir regresyon yapın, 1971'deki çocuk sayısını kontrol değişkeni olarak kullanın ve tahmini rapor edin.
- ii. **İndirgenmiş form (2):** $yY_j - yO_j = \gamma P_j + \epsilon_j$ (y log saatlik ücreti göstermektedir.) Denklem sol tarafında sadece wage_dif olduğunu not edin. γ 'yı wage_dif program yoğunluğu üzerine olmak üzere yapacağınız bir regresyonla tahmin edin, 1971'deki çocuk sayısını kontrol değişkeni olarak kullanın ve tahmini rapor edin.
- iii. Ücret belirleme denkleminin $y_i = a_i + bS_i$ olduğunu hatırlayın, burda y saatlik ücretin logaritmasını göstermektedir. Bölge ve yaş grubu bazında ortalama aldıktan sonra, elimizde $yY_j = aY_j + bSY_j$ ve $yO_j = aO_j + bSO_j$ oluyor. İkincisini ilkinden çıkarırsak $yY_j - yO_j =$

$aY_j - aO_j + b(SY_j - SO_j)$ elde ediyoruz. Bu denklemi, ilk aşama ve indirgenmiş form denklemleriyle birlikte $b = \gamma/\alpha$ olduğunu göstermek için kullanın.

- iv. Dolaylı en küçük kareler yöntemiyle b 'nin tahminini bu yolla hesaplayın, tahmini rapor edin ve yorumlayın. Bunu Wald tahmini ile karşılaştırdığınızda ne görüyorsunuz? Hangisini tercih edersiniz? Açıklayın.
- v. Dolaylı en küçük kareler, programı okula gidilen yılda eksojen bir varyasyon yaratmak için araç olarak kullanılmaktadır. Aslında, okula gidilen yılı, programa tabii tutulmakla öngörmekte ve daha sonrada okula gidilen yılların getirisini sadece öngörülen okula gidilen yıldaki varyasyon ile tahmin etmektedir. Şimdi, dolaylı en küçük kareler ile okula gidilen yılı tahmin etmek için ikinci bir teknik kullanın. i kısmındaki ilk aşama regresyonu yapın ve *predict* komutunu kullanarak `educ_dif` değişkeninin öngörülmuş değerlerini `edif_pred` adını verebileceğiniz yeni bir değişken olarak kaydedin. Daha sonra, yine 1971 yılındaki çocuk sayısını kontrol değişkeni olarak içermek üzere, `wage_dif` öngörülmuş `edif_pred` üzerine olmak üzere yeni bir regresyon yapın. b 'nin yeni tahminini rapor edin ve `edif_pred`'in öngörülmuş değerlerinin katsayısının iv kısmında hesapladığınızla aynı olduğunu teyit edin.

d. İki aşamalı en küçük kareler (İAEKK)

c kısmında, yaş grubundaki tüm varyasyonu kullanmadık, bunun yerine, yaş gruplarını genç ve yaşlı olarak iki grup halinde bütünleştirdik. Şimdi, bütün yaş grubu varyasyonunu kullanacağız. c kısmında veri setini değiştirmiştiniz, şimdi **orjinal veri setine geri dönün**.

Dolaylı en küçük kareler alıştırmasını her bir yaş grubu için tekrar edebileceğimizi not edin; yani her bir yaş grubu k (= doğum yılı) için, şu regresyonları yapabiliriz:

$$Sk_j - S50_j = \alpha k P_j + v_{kj} \text{ ve}$$

$$yk_j - y50_j = \gamma k P_j + \epsilon_{kj}$$

Burda, j bölgeleri göstermekte, 50 ise en yaşlı yaş grubu (doğum yılı=1950) olmaktadır.

- i. Yukardaki regresyonları yapmanız **istenmemektedir**. Ancak, bu regresyonları yapmış olsaydınız, α_k and γ_k tahminlerini nasıl yorumlardınız? İAEKK, b 'yi tahmin etmek için kullanılabilir farklı yöntemleri optimal olarak bir araya getirmektedir. (tahminleyicinin varyansını minimize ederek) Tüm araç değişken tahminleyicilerinin olduğu gibi

İAEKK endojen açıklayıcı değişkende eksojen bir varyasyon yaratmak amacıyla araç adı verilen değişkenleri kullanmaktadır. Bizim örneğimizde, okula gidilen yıllar endojen bir açıklayıcı değişkendir ve dolayısıyla biz de "INPRES programının olan katılımı" okula gidilen yıllarda eksojen bir varyasyon yaratmak amacıyla araç olarak kullanıyoruz.

Bu veri seti, 1974'te 2 yaşından 24 yaşına kadar olan bireyleri içermektedir. 1974 yılında 2 ile 12 yaş arasında olanlar (yani yob 1962 ile 1972 arasında) programa katılmış olmak için yeterince gençtirler. Dolayısıyla elimizde 11 tane araç değişken (1 program * 11 yaş grubu) bulunmaktadır.

- ii. Programa katılan her yaş grubu için kişinin doğduğu yıl o yıla 1, değilse 0 olan bir kukla değişken yaratın. (`generate d62 = (yob==62)`, vs. `d72 ye kadar`) Şimdi, her katılan yaş grubu için program yoğunluğu `carpi` yaş grubu kukla değişkenine eşit olan bir değişken yaratın. (`generate z62 = d62*prog_int`, vs. `z72 ye kadar`) Bu 11 değişken (`z62`'den `z72`'ye kadar) 11 araç değişkendir. Bu değişkenlerin iyi araç değişkenler olması için **gerekli koşulları açıkça belirtin**. Bu değişkenlerin bu iki koşulu niçin sağlayabileceğini **açıklayın**. Regresyonlarımızdaki kontrol değişkenleri olarak, 1971'deki çocuk sayısı, doğum yılı kukla değişkenleri, doğum yılı kukla değişkenlerinin 1971'deki çocuk sayısı ile etkileşimi ("etkileşim" "çarpım" anlamına gelmektedir) Doğum yılı kukla değişkenleri, her olası doğum yılı için olmak üzere, eğer kişi o yılda doğmuşsa 1, değilse 0 değerini alan kukla değişkenlerdir. Tahmin edebileceğiniz gibi, çok sayıda kontrol değişkeni vardır. (1 değişken çocuk sayısı için+ 23 tane doğum yılı kukla değişkenleri+23 tane etkileşim terimi=47 kontrol değişkeni) Neyse ki, Stata kukla değişkenleri ve halihazırda var olan bir değişkene dayalı etkileşim terimlerini kolay bir şekilde kullanmamızı sağlamaktadır. (bu örnekte doğum yılı `yob`) Bir Stata komudu içinde, `i.variable` ifadesini kullanarak `variable` adı verilen bir değişkenden türetilmiş kukla değişkenlere referans gösterebilirsiniz. `i.variable1*variable2` ifadesini de `variable`, `variable2` ve ikisi arasındaki etkileşime dayanarak oluşturulan bir kukla değişkene referans verebilirsiniz. Bir komutta `i.` ifadesini kullandığınızda komuttan önce `xi` ifadesini koymalısınız: Yani, `y'yi x`, 1971'deki çocuk sayısı, doğum yılı kukla değişkenleri ve doğum yılı-çocuk sayısı etkileşim değişkenleri üzerine olmak üzere yapacağınız bir regresyonda şu komutu kullanmalısınız
`xi: regress y x i.YOB*ch71`
- iii. `b'nin` İAEKK yöntemiyle tahminini hesaplamak için

1. Önce, araç ve kontrol değişkenlerini kullanarak, eğitim yılını öngörün. (yani, EKK yöntemiyle eğitim yılı araç değişkenler ve kontrol değişkenleri üzerine olmak üzere bir regresyon yapın ve bu regresyondan öngörülen değerleri oluşturun.)
2. Log saatlik ücretler, öngörülen eğitim yılı ve kontrol değişkenleri üzerine olmak üzere bir regresyon yapın. Öngörülen eğitim yılının katsayısının tahmini, b'nin İAEKK tahminidir. b'nin İAEKK tahminini **hesaplayın ve yorumlayın.**
3. Bu İAEKK tahminine güvenerek, INPRES programının okul kalitesini arttırdığını söyleyebilir miyiz? Neden ya da neden değil?

2. Yazı Sorusu

Dersin internet sitesinde "İkinci Ödev" başlığı altına konulan Duflo, Dupas ve Kremer'in "Peer Effects, Pupil-Teacher Ratios and Teacher Incentives: Evidence from a Randomized Evaluation in Kenya" adlı makalesini okuyun.

- Kısa ama açık olarak deneyse dizaynı tasvir edin. Önerilen farklı işlemler nedir?
- Makalenin test ettiği belirli soruları/hipotezleri özetleyin. Kendinize yön vermek için 4,5 ve 6. Grafiklere, 3,4,5,6,7 ve 8. Tablolara bakın ve rapor edilen sonuçların test ettiği hipotezleri belirleyin. Hipotezleri soru olarak yazın. Bu bölümdeki cevaplarınızı madde madde yazabilirsiniz. Örneğin, ben olsa bu soruya 3. tablodan başlardım.

- o Sınıfın büyüklüğünü ek öğretmen programı (EÖP) yoluyla azaltmak, öğretmen katılımını ve eforunu geliştirmekte midir?
- o Öğrencileri takip etmek, öğretmen katılımını/eforunu artırır mı?
- o Takip edilen okullardaki sözleşmeli öğretmenler, sözleşmesiz öğretmenlerden daha yüksek bir katılım/efor mu göstermektedirler?
- o vs.

Takip etme ile ilgili makaleyi okuduktan sonra, Glewwe, Kremer ve Moulin'in "Many Children Left Behind: Textbooks and Test Scores in Kenya" adlı yine ödev 2 başlığı altında dersin internet sitesine konan makaleyi okuyun.

- Deneysel dizaynı kısaca özetleyin.
- Kitapta doğrudan cevaplanan sorular dışında sorular cevaplamak için deneysel dizaynı nasıl değiştirebileceğinizi düşünün. Ders kitabı kullanımına yönelik deneye devam edin ama ders kitaplarıyla ilgili başka sorular düşünün ya da ders kitaplarının etkinliğiyle ilgili diğer programlar üzerinde düşünün. Örneğin: hangi türk ders kitapları etkindir? Ders kitaplarından kimler yararlanmaktadır? Ders kitapları nasıl dağıtılmalıdır? Ne tür ikincil programlar ders kitaplarının etkinliğini arttırabilir? Vs. Makalede geçen hipotezler dışında test etmek istediğiniz hipotezler üzerine net olun.
- Yeni hipotezinizi test etmek için nasıl kontrol ve deney gruplarına ihtiyaç duyarsınız? Ne tür veriler toplarsınız? Verilerinizde ne tür karşılaştırmalar yaparsınız?