HW4

1585107 이지인

2018년 11월 15일

ciga<-read.csv("ciga.csv")  
head(ciga)

## state year cpi pop packpc income tax avgprs taxs  
## 1 AL 1985 1.076 3973000 116.4863 46014968 32.5 102.18167 33.34834  
## 2 AR 1985 1.076 2327000 128.5346 26210736 37.0 101.47500 37.00000  
## 3 AZ 1985 1.076 3184000 104.5226 43956936 31.0 108.57875 36.17042  
## 4 CA 1985 1.076 26444000 100.3630 447102816 26.0 107.83734 32.10400  
## 5 CO 1985 1.076 3209000 112.9635 49466672 31.0 94.26666 31.00000  
## 6 CT 1985 1.076 3201000 109.2784 60063368 42.0 128.02499 51.48333  
## log\_Q log\_I log\_P CT T  
## 1 4.757774 2.376195 4.553502 30.20447 0.7884122  
## 2 4.856198 2.348339 4.546562 34.38662 0.0000000  
## 3 4.649403 2.551822 4.614225 28.81041 4.8052211  
## 4 4.608794 2.754509 4.607374 24.16357 5.6728627  
## 5 4.727065 2.662089 4.472877 28.81041 0.0000000  
## 6 4.693898 2.858686 4.778975 39.03346 8.8135072

#install.packages("AER")  
library(AER); library(lmtest)

## Warning: package 'AER' was built under R version 3.4.4

## Loading required package: car

## Warning: package 'car' was built under R version 3.4.4

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.4.4

## Loading required package: lmtest

## Warning: package 'lmtest' was built under R version 3.4.4

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.4.4

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 3.4.4

## Loading required package: survival

attach(ciga)

## The following object is masked from package:base:  
##   
## T

rtaxso=taxs/cpi - tax/cpi #sale tax 판매세 (도구변수1, 모형1)  
rtax=tax/cpi #ciga tax 담배세 (도구변수2, 모형2)  
perinc=income/pop/cpi #real per capita state income 소득 (외생변수-도구변수 필요없음)  
log.q=log(packpc[year==1995])-log(packpc[year==1985])  
log.p=log(avgprs[year==1995])-log(avgprs[year==1985])  
log.inc=log(perinc[year==1995])-log(perinc[year==1985])  
cigatax=rtax[year==1995]-rtax[year==1985]  
saletax=rtaxso[year==1995]-rtaxso[year==1985]  
  
#model(1)  
#도구회귀(도구변수1개): 담배수요q~가격p+소득inc | 도구변수 z1 saletax  
ivreg.fit1=ivreg(log.q~log.p+log.inc | saletax+log.inc)  
summary(ivreg.fit1)

##   
## Call:  
## ivreg(formula = log.q ~ log.p + log.inc | saletax + log.inc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.16126 -0.08153 0.02359 0.07517 0.13702   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.2085 0.1329 1.569 0.124   
## log.p -0.9380 0.2107 -4.452 5.56e-05 \*\*\*  
## log.inc 0.5260 0.3084 1.705 0.095 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09092 on 45 degrees of freedom  
## Multiple R-Squared: 0.5499, Adjusted R-squared: 0.5299   
## Wald test: 12.97 on 2 and 45 DF, p-value: 3.559e-05

coeftest(ivreg.fit1, vcov=vcovHC) #이분산성 고려

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.20855 0.14458 1.4425 0.1560991   
## log.p -0.93801 0.23175 -4.0476 0.0002011 \*\*\*  
## log.inc 0.52597 0.37225 1.4129 0.1645528   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#1st stage:   
#(내생설명변수 X) 가격log.p 를   
#(외생설명변수 W) 소득 log.inc 와 (한개의 도구변수 z1) 판매세saletax 회귀시켜  
#xhat을 구한다.  
lm.fit1.first=lm(log.p ~ log.inc + saletax)  
summary\_lm.fit1<-summary(lm.fit1.first); summary\_lm.fit1

##   
## Call:  
## lm(formula = log.p ~ log.inc + saletax)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.123986 -0.050950 -0.006608 0.049326 0.178746   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.532195 0.031249 17.031 < 2e-16 \*\*\*  
## log.inc -0.224104 0.211941 -1.057 0.296   
## saletax 0.025461 0.003737 6.813 1.93e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06334 on 45 degrees of freedom  
## Multiple R-squared: 0.5146, Adjusted R-squared: 0.4931   
## F-statistic: 23.86 on 2 and 45 DF, p-value: 8.641e-08

summary\_lm.fit1$fstatistic[1] >10

## value   
## TRUE

#model(1)의 도구변수 관련성 검정  
#검정 결과, 내생설명변수가 1개일 때, F 통계량이 10보다 크므로 내생설명변수X인 담배 가격과 연관성을 지닌다고 볼 수 있다. 도구변수 관련성 가정을 만족한다.  
#model(1)은 한 개의 내생설명변수, 한 개의 도구변수를 사용하므로 m=k exactly identified case로 J검정을 통한 도구변수 외생성을 검증할 수 없다  
  
#2nd stage:   
#(종속변수 Y) 수요 log.q를  
#(1st stage를 통해 구한 추정치) Xhat과 (외생설명변수 W) 소득 log.inc에 회귀시켜  
#TSLSE를 구한다.  
log.p.hat1=lm.fit1.first$fitted.value  
lm.fit1.second=lm(log.q ~ log.p.hat1 + log.inc)  
summary(lm.fit1.second)

##   
## Call:  
## lm(formula = log.q ~ log.p.hat1 + log.inc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.213917 -0.089565 0.008805 0.069021 0.239996   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.2085 0.1705 1.223 0.22769   
## log.p.hat1 -0.9380 0.2702 -3.471 0.00116 \*\*  
## log.inc 0.5260 0.3956 1.330 0.19037   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1166 on 45 degrees of freedom  
## Multiple R-squared: 0.2595, Adjusted R-squared: 0.2266   
## F-statistic: 7.885 on 2 and 45 DF, p-value: 0.001159

######  
#model(2)  
#도구회귀(도구변수1개): 담배수요q~가격p+소득inc | 도구변수 z2 cigatax  
ivreg.fit2=ivreg(log.q ~ log.p+log.inc| cigatax+log.inc)  
summary(ivreg.fit2)

##   
## Call:  
## ivreg(formula = log.q ~ log.p + log.inc | cigatax + log.inc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.157110 -0.075500 0.006638 0.073492 0.200945   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.4503 0.1211 3.718 0.000554 \*\*\*  
## log.p -1.3425 0.1888 -7.112 6.95e-09 \*\*\*  
## log.inc 0.4281 0.3175 1.348 0.184301   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09392 on 45 degrees of freedom  
## Multiple R-Squared: 0.5197, Adjusted R-squared: 0.4983   
## Wald test: 28.16 on 2 and 45 DF, p-value: 1.172e-08

coeftest(ivreg.fit2, vcov=vcovHC) #이분산성 고려

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.45026 0.16870 2.6690 0.01054 \*   
## log.p -1.34251 0.28042 -4.7875 1.86e-05 \*\*\*  
## log.inc 0.42815 0.32896 1.3015 0.19970   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#1st stage:   
#(내생설명변수 X) 가격log.p 를   
#(외생설명변수 W) 소득 log.inc 와 (한개의 도구변수 z2) 담배세 cigatax 회귀시켜  
#xhat을 구한다.  
lm.fit2.first=lm(log.p ~ log.inc + cigatax)  
summary\_lm.fit2<-summary(lm.fit2.first); summary\_lm.fit2

##   
## Call:  
## lm(formula = log.p ~ log.inc + cigatax)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.082015 -0.046468 0.006699 0.035803 0.128046   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.502862 0.026073 19.287 < 2e-16 \*\*\*  
## log.inc 0.029402 0.174458 0.169 0.867   
## cigatax 0.010096 0.001044 9.668 1.48e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05147 on 45 degrees of freedom  
## Multiple R-squared: 0.6796, Adjusted R-squared: 0.6653   
## F-statistic: 47.72 on 2 and 45 DF, p-value: 7.56e-12

summary\_lm.fit2$fstatistic[1] >10

## value   
## TRUE

#model(2)의 도구변수 관련성 검정  
#검정 결과, 내생설명변수가 1개일 때, F 통계량이 10보다 크므로 내생설명변수X인 담배 가격과 연관성을 지닌다고 볼 수 있다.   
#도구변수 관련성 가정을 만족한다.  
#model(2)은 한 개의 내생설명변수, 한 개의 도구변수를 사용하므로 m=k exactly identified case로 J검정을 통한 도구변수 외생성을 검증할 수 없다  
  
#2nd stage:   
#(종속변수 Y) 수요 log.q를  
#(1st stage를 통해 구한 추정치) Xhat과 (외생설명변수 W) 소득 log.inc에 회귀시켜  
#TSLSE를 구한다.  
log.p.hat2=lm.fit2.first$fitted.value  
lm.fit2.second=lm(log.q ~ log.p.hat2 + log.inc)  
summary(lm.fit2.second)

##   
## Call:  
## lm(formula = log.q ~ log.p.hat2 + log.inc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.15523 -0.05586 0.01201 0.07080 0.15522   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.4503 0.1104 4.080 0.000182 \*\*\*  
## log.p.hat2 -1.3425 0.1720 -7.805 6.64e-10 \*\*\*  
## log.inc 0.4281 0.2893 1.480 0.145924   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08559 on 45 degrees of freedom  
## Multiple R-squared: 0.6012, Adjusted R-squared: 0.5834   
## F-statistic: 33.91 on 2 and 45 DF, p-value: 1.042e-09

#####  
#model(3)  
#도구회귀(도구변수2개): 담배수요q~가격p+소득inc | 도구변수 z1 z2  
ivreg.fit3=ivreg(log.q ~ log.p + log.inc| saletax+cigatax+log.inc)  
summary(ivreg.fit3)

##   
## Call:  
## ivreg(formula = log.q ~ log.p + log.inc | saletax + cigatax +   
## log.inc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.15062 -0.07598 0.01356 0.07829 0.16785   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.3665 0.1109 3.305 0.00187 \*\*   
## log.p -1.2024 0.1712 -7.024 9.4e-09 \*\*\*  
## log.inc 0.4620 0.3081 1.500 0.14070   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09125 on 45 degrees of freedom  
## Multiple R-Squared: 0.5466, Adjusted R-squared: 0.5264   
## Wald test: 27.71 on 2 and 45 DF, p-value: 1.436e-08

coef.ivreg.fit3<-coeftest(ivreg.fit3, vcov=vcovHC); coef.ivreg.fit3

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.36654 0.14189 2.5832 0.01311 \*   
## log.p -1.20240 0.23221 -5.1782 5.072e-06 \*\*\*  
## log.inc 0.46203 0.34012 1.3584 0.18109   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#도구변수 외생성 검정  
#model(3)은 한 개의 내생설명변수, 두 개의 도구변수를 사용하므로 m>k overidentified case로 J검정을 통한 도구변수 외생성을 검증할 수 있다.  
#unknown인 u대신 uhat 도구회귀분석의 잔차를 이용  
#uhat=Y-ahat(TSLSE)-b1(TSLSE)\*X-b2(TSLSE)\*W  
uhat=log.q-coef.ivreg.fit3[1,1]-(coef.ivreg.fit3[2,1]\*log.p)-coef.ivreg.fit3[3,1]\*log.inc  
iv.exogeniety<-lm(uhat~saletax+cigatax+log.inc);   
summary\_iv.exogeniety<-summary(iv.exogeniety); summary\_iv.exogeniety

##   
## Call:  
## lm(formula = uhat ~ saletax + cigatax + log.inc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.147879 -0.067569 0.004111 0.076002 0.150489   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.002939 0.044613 0.066 0.9478   
## saletax 0.012767 0.006159 2.073 0.0441 \*  
## cigatax -0.003808 0.002118 -1.798 0.0791 .  
## log.inc -0.093406 0.297846 -0.314 0.7553   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08751 on 44 degrees of freedom  
## Multiple R-squared: 0.1008, Adjusted R-squared: 0.03948   
## F-statistic: 1.644 on 3 and 44 DF, p-value: 0.1929

iv.F<-summary\_iv.exogeniety$fstatistic[1]; iv.F

## value   
## 1.643994

#도구변수 z1과 z2가 모두 외생이라는 귀무가설하  
#J=m\*F는 자유도(m-k)인 카이제곱분포를 따른다.  
J<-2\*iv.F; J

## value   
## 3.287988

1-pchisq(J, df=2-1)

## value   
## 0.06978848

1-pchisq(J, df=2-1)<0.05

## value   
## FALSE

#p-value가 0.05보다 크므로 귀무가설을 기각할 수 없다, do not reject H0  
#도구변수 z1 판매세와 z2 담배세 모두 도구변수 외생성을 만족한다고 할 수 있다.  
  
#1st stage:   
#(내생설명변수 X) 가격log.p 를   
#(외생설명변수 W) 소득 log.inc 와 (두 개의 도구변수 z1,z2) 판매세 saletax와 담배세 cigatax 회귀시켜  
#xhat을 구한다.  
lm.fit3.first=lm(log.p ~ log.inc + saletax + cigatax)  
summary\_lm.fit3<-summary(lm.fit2.first); summary\_lm.fit3

##   
## Call:  
## lm(formula = log.p ~ log.inc + cigatax)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.082015 -0.046468 0.006699 0.035803 0.128046   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.502862 0.026073 19.287 < 2e-16 \*\*\*  
## log.inc 0.029402 0.174458 0.169 0.867   
## cigatax 0.010096 0.001044 9.668 1.48e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05147 on 45 degrees of freedom  
## Multiple R-squared: 0.6796, Adjusted R-squared: 0.6653   
## F-statistic: 47.72 on 2 and 45 DF, p-value: 7.56e-12

summary\_lm.fit3$fstatistic[1] >10 #도구변수 관련성 검정

## value   
## TRUE

#model(2)의 도구변수 관련성 검정  
#검정 결과, 내생설명변수가 1개일 때, F 통계량이 10보다 크므로 내생설명변수X인 담배 가격과 연관성을 지닌다고 볼 수 있다.   
#도구변수 관련성 가정을 만족한다.  
  
#2nd stage:   
#(종속변수 Y) 수요 log.q를  
#(1st stage를 통해 구한 추정치) Xhat과 (외생설명변수 W) 소득 log.inc에 회귀시켜  
#TSLSE를 구한다.  
log.p.hat3=lm.fit3.first$fitted.value  
lm.fit3.second=lm(log.q ~ log.p.hat3 + log.inc)  
summary(lm.fit3.second)

##   
## Call:  
## lm(formula = log.q ~ log.p.hat3 + log.inc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.15559 -0.06770 -0.02057 0.07110 0.16435   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.3665 0.1095 3.348 0.00165 \*\*   
## log.p.hat3 -1.2024 0.1690 -7.116 6.85e-09 \*\*\*  
## log.inc 0.4620 0.3041 1.519 0.13566   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09007 on 45 degrees of freedom  
## Multiple R-squared: 0.5583, Adjusted R-squared: 0.5387   
## F-statistic: 28.44 on 2 and 45 DF, p-value: 1.035e-08

