1. Problem 11.3

EXPER

KIDS

EDUC

No. In the equation of Ln(WAGE), the right-handside endogenous variable ‘HOURS’ is correlated with error term ‘e’, which will make the least square estimator biased and inconsistent.

1. In the system, there are total 5 variables, 2 of them are endogenous. In the equation Ln(WAGE), one variable (KIDS) is omitted, 1=M-1, which makes the equation “identified”.

Identification means the parameters can be estimated consistently.

(c ) Step-by-step of 2-Stage LS

1. Identification : yes
2. Reduced form : rewrite the equation in reduced form, relating the endogenous variables to exogenous variables :

HOURS = 1+ 2\*KIDS + 3\*EDUC +4\*EXPER + v

Apply least square regression on this reduced-form to get expected value of HOURS, denoted as HAT(HOURS)= hat(1)+ hat(2)\*KIDS + hat(3)\*EDUC +hat4)\*EXPER

1. Substitute HOURS in Ln(WAGE) equation with the values of HAT(HOURS) calculated in ii).

Ln(WAGE) = 1 + 2 \*HAT(HOURS) + 3 \* EDUC + 4\*EXPER + e\* Then apply least square regression on this equation to get the consistent estimate of parameters.

1. Problem 15.8 (a)(b)(c)(d)
2. Ln(WAGEit) = 1t + 2tEXPERit +3t EXPERit2+ 4tSOUTHit + 5tUNIONit + eit

> RawTable = read.csv("nls\_panel2.csv")

> Year8788 = data.frame(RawTable$id,RawTable$year,RawTable$lwage, RawTable$exper, RawTable$exper2, RawTable$south,RawTable$union)

> colnames(Year8788) <- c("id","year","lwage","exper","exper2","south","union")

> Year87 = subset(Year8788, year==87)

> colnames(Year87) <- c("id87","year87","lwage87","exper87","exper287","south87","union87")

> summary(lm(lwage87~exper87+exper287+south87+union87, Year87))

Call:

lm(formula = lwage87 ~ exper87 + exper287 + south87 + union87,

data = Year87)

Residuals:

Min 1Q Median 3Q Max

-1.37258 -0.28968 -0.02037 0.26430 2.04774

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.934810 0.200990 4.651 3.94e-06 \*\*\*

exper87 0.127022 0.029501 4.306 1.90e-05 \*\*\*

exper287 -0.003288 0.001067 -3.083 0.002130 \*\*

south87 -0.212794 0.033804 -6.295 5.38e-10 \*\*\*

union87 0.144536 0.038227 3.781 0.000169 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4441 on 711 degrees of freedom

Multiple R-squared: 0.1494, Adjusted R-squared: 0.1446

F-statistic: 31.21 on 4 and 711 DF, p-value: < 2.2e-16

> Year88 = subset(Year8788, year==88)

> colnames(Year88) <- c("id88","year88","lwage88","exper88","exper288","south88","union88")

> summary(lm(lwage88~exper88+exper288+south88+union88, data=Year88))

Call:

lm(formula = lwage88 ~ exper88 + exper288 + south88 + union88,

data = Year88)

Residuals:

Min 1Q Median 3Q Max

-1.49763 -0.28759 -0.01068 0.27509 2.00643

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.899282 0.240720 3.736 0.000202 \*\*\*

exper88 0.126541 0.032324 3.915 9.92e-05 \*\*\*

exper288 -0.003089 0.001069 -2.891 0.003957 \*\*

south88 -0.238422 0.034386 -6.934 9.22e-12 \*\*\*

union88 0.110209 0.038738 2.845 0.004569 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4515 on 711 degrees of freedom

Multiple R-squared: 0.1391, Adjusted R-squared: 0.1342

F-statistic: 28.72 on 4 and 711 DF, p-value: < 2.2e-16

Compare 87 and 88 regression : there are little differences between parameter values of 87 and 88 regression.

We assume the parameter values are not variant across individuals

1. Pooled data model : the equation is like this

Ln(WAGEit) = 1 + 2EXPERit +3 EXPERit2+ 4SOUTHit + 5UNIONit + eit

> summary(lm(lwage~exper+exper2+south+union, data=Year8788))

Call:

lm(formula = lwage ~ exper + exper2 + south + union, data = Year8788)

Residuals:

Min 1Q Median 3Q Max

-1.48802 -0.29000 -0.01442 0.27484 2.08075

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.9481780 0.1506222 6.295 4.08e-10 \*\*\*

exper 0.1228848 0.0211092 5.821 7.20e-09 \*\*\*

exper2 -0.0030661 0.0007279 -4.212 2.69e-05 \*\*\*

south -0.2255408 0.0240803 -9.366 < 2e-16 \*\*\*

union 0.1273629 0.0271754 4.687 3.04e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4473 on 1427 degrees of freedom

Multiple R-squared: 0.1438, Adjusted R-squared: 0.1414

F-statistic: 59.93 on 4 and 1427 DF, p-value: < 2.2e-16

**Or using plm package :**

> library(plm)

> pooling=plm(lwage~exper+exper2+south+union, data = Year8788, index=c("id","year"), model="pooling")

> summary(pooling)

Oneway (individual) effect Pooling Model

Call:

plm(formula = lwage ~ exper + exper2 + south + union, data = Year8788,

model = "pooling", index = c("id", "year"))

Balanced Panel: n=716, T=2, N=1432

Residuals :

Min. 1st Qu. Median 3rd Qu. Max.

-1.4900 -0.2900 -0.0144 0.2750 2.0800

Coefficients :

Estimate Std. Error t-value Pr(>|t|)

(Intercept) 0.9481780 0.1506222 6.2951 4.080e-10 \*\*\*

exper 0.1228849 0.0211092 5.8214 7.197e-09 \*\*\*

exper2 -0.0030662 0.0007279 -4.2123 2.686e-05 \*\*\*

south -0.2255408 0.0240803 -9.3662 < 2.2e-16 \*\*\*

union 0.1273629 0.0271754 4.6867 3.042e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares: 333.49

R Adj. R-Squared: 0.14332

F-statistic: 59.9293 on 4 and 1427 DF, p-value: < 2.22e-16

**Residual Sum of Squares: 285.53**

R-Squared: 0.14383

We assume the parameter values are not variant across individuals, and not variant across years.

We assume the variance of error terms :

Var(eit) = 2 (a constant)

Cov(eit, ejs)=0, i<>j, t<>s

Cov(eit, xkit)=0 (error term has no correlation with individual, and no correlation with year)

(c)

Fixed effect model allowing individual heterogeneity:

Ln(WAGEit) = 1i + 2EXPERit +3 EXPERit2+ 4SOUTHit + 5UNIONit + eit

Compare (c) with (b) : the intercept1 is individual-invariant in (b) but individual-variant in (c). That means : in (c), we add this assumption if i=j :

Cov(eit, eis) <>0 ts)

Also Var(eit ) is not a constant among different individuals.

Compare (c) with (a) : in (a), the coefficients are variant among years rather than individuals. But in (c) all the coefficients except the intercept are fixed, and the intercept is variant among different individuals.

**(d ) We use plm with model =”within” for fixed effect**

> fixed=plm(lwage~exper+exper2+south+union, data = Year8788, index=c("id","year"), model="within")

> summary(fixed)

Oneway (individual) effect Within Model

Call:

plm(formula = lwage ~ exper + exper2 + south + union, data = Year8788,

model = "within", index = c("id", "year"))

Balanced Panel: n=716, T=2, N=1432

Residuals :

Min. 1st Qu. Median 3rd Qu. Max.

-1.25e+00 -3.78e-02 1.29e-16 3.78e-02 1.25e+00

Coefficients :

Estimate Std. Error t-value Pr(>|t|)

exper 0.0574578 0.0329942 1.7415 0.082036 .

exper2 -0.0012344 0.0011023 -1.1199 0.263146

south -0.3260524 0.1257964 -2.5919 0.009740 \*\*

union 0.0821949 0.0312071 2.6338 0.008626 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares: 23.241

**Residual Sum of Squares: 22.439**

R-Squared: 0.034495

Adj. R-Squared: 0.017151

F-statistic: 6.35957 on 4 and 712 DF, p-value: 4.9631e-05

**H0 : 11=12= 13 = … =1,716**

**H1 : 1i are not all equal**

In pooled model, SSER = 285.53

In the fixed effect model, SSEU = 22.439

J=715, NK-K =1432- 720 =712

F = ((SSER - SSEU)/J) / (SSEU/NK-K)= (285.53-22.439/715) / (285.53/712)= 11.675 **>**

F(0.95,715, 712) = 1.131

**So we reject H0 and accept that the intercept parameters are not all identical.**

**In pooled data model in (b), we assume the intercept is a fixed value among all individuals. This test approves the assumption in (b) is incorrect.**