Problem Set 5

eitc <- read.csv("eitc.csv", header = TRUE)  
attach(eitc)

#### 1.

eitc$posttreat <- (eitc$year >= 1994)\*1  
eitc$anykids <- (eitc$children != 0)\*1  
head(eitc, 5)

## state year urate children nonwhite finc earn age ed work unearn  
## 1 11 1991 7.6 0 1 18714.394 18714.3940 26 10 1 0.000000  
## 2 12 1991 7.2 1 0 4838.568 471.3656 22 9 1 4.367203  
## 3 13 1991 6.4 2 0 8178.194 0.0000 33 11 0 8.178194  
## 4 14 1991 9.1 0 1 9369.570 0.0000 43 11 0 9.369571  
## 5 15 1991 8.6 3 1 14706.608 14706.6080 23 7 1 0.000000  
## posttreat anykids  
## 1 0 0  
## 2 0 1  
## 3 0 1  
## 4 0 0  
## 5 0 1

#### 2.

mean(eitc[which(eitc$posttreat == 0 & eitc$anykids == 0),'work'])

## [1] 0.5754597

mean(eitc[which(eitc$posttreat == 0 & eitc$anykids == 1),'work'])

## [1] 0.4459619

mean(eitc[which(eitc$posttreat == 1 & eitc$anykids == 0),'work'])

## [1] 0.5733862

mean(eitc[which(eitc$posttreat == 1 & eitc$anykids == 1),'work'])

## [1] 0.4907615

**i)**

Mean of posttreat = 0 & anykids = 0 is 0.5754597

**ii)**

Mean of posttreat = 0 & anykids = 1 is 0.4459619

**iii)**

Mean of posttreat = 1 & anykids = 0 is 0.5733862

**iv)**

Mean of posttreat = 1 & anykids = 1 is 0.4907615

#### 3.

mean(eitc[which(eitc$posttreat == 1 & eitc$anykids == 1),'work']) - mean(eitc[which(eitc$posttreat == 0 & eitc$anykids == 1),'work'])

## [1] 0.04479962

Effect = 0.04479962 increase in employment for women with children

#### 4.

eitc$interact = eitc$posttreat\*eitc$anykids  
reg = lm(work ~ anykids + posttreat + interact, data = eitc)  
summary(reg)

##   
## Call:  
## lm(formula = work ~ anykids + posttreat + interact, data = eitc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.5755 -0.4908 0.4245 0.5092 0.5540   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.575460 0.008845 65.060 < 2e-16 \*\*\*  
## anykids -0.129498 0.011676 -11.091 < 2e-16 \*\*\*  
## posttreat -0.002074 0.012931 -0.160 0.87261   
## interact 0.046873 0.017158 2.732 0.00631 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4967 on 13742 degrees of freedom  
## Multiple R-squared: 0.0126, Adjusted R-squared: 0.01238   
## F-statistic: 58.45 on 3 and 13742 DF, p-value: < 2.2e-16

DID estimate = 0.046873

Standard Error = 0.017158

#### 5.

**i)**

where

**ii)**

is independent of , so

**iii)**

and are independent, so

**iv)**

OLS estimator is unbiased because

**v)**

, so

OLS estimator is consistent

#### 6.

**i)**

**ii)**

Consistent estimates of indicates that and

, so while there is no correlation with the error term, so

**iii)**

No

**iv)**

Instrument relevance and instrument exogeneity.

**v)**

By using instrumental variables that are uncorrelated with the error term, it is possible to estimate the model parameter . These instrumental variables are correlated to the endogenous variables (which are correlated to the error term) but by itself has no correlation with the error term.