PS8_12265092

12265092

09/03/2022

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(estimatr)
```

```
## Warning: package 'estimatr' was built under R version 4.1.2
```

```
library(Rcpp)
library(readxl)
library(haven)
library(boot)
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 4.1.2

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.1.2

##
## Attaching package: 'zoo'

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

```
library(margins)
```

```
## Warning: package 'margins' was built under R version 4.1.2
```

```
library(mfx)
```

```
## Warning: package 'mfx' was built under R version 4.1.2

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 4.1.2

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

## Loading required package: betareg

## Warning: package 'betareg' was built under R version 4.1.2
```

```
library(ivreg)
```

```
## Warning: package 'ivreg' was built under R version 4.1.2
```

```
library(aod)
```

```
## Warning: package 'aod' was built under R version 4.1.2
```

```
data <- read_dta("voucher.dta")
```

```
#1a

table(data$choiceyrs)
```

```
##
##   0   1   2   3   4
## 490 211 122 111  56
```

```
# Of 990 students in sample, 56 student attended a choice school for 4 years

table(data$selectyrs)
```

```
##
##   0   1   2   3   4
## 468 116 148 150 108
```

```r
# 108 students had vouchers available for 4 years
# Of 990 students in sample, 468 students never awarded a voucher
```

```r
#1b

reg1b <- lm(formula = choiceyrs ~ selectyrs, data = data)

summary(reg1b)
```

```
##
## Call:
## lm(formula = choiceyrs ~ selectyrs, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.08725 -0.01992 -0.01992  0.21325  1.21325
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.01992    0.02461   0.809    0.419
## selectyrs    0.76683    0.01259  60.931   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.576 on 988 degrees of freedom
## Multiple R-squared:  0.7898, Adjusted R-squared:  0.7896
## F-statistic:  3713 on 1 and 988 DF,  p-value: < 2.2e-16
```

```r
#Residual standard error: 0.576 on 988 degrees of freedom
#Multiple R-squared:  0.7898,
#Adjusted R-squared:  0.7896
#F-statistic:  3713 on 1 and 988 DF,
#p-value: < 2.2e-16

#The p-value at 5% significance level is less that 0.05. So we can safely
#reject the null hypothesis that coefficient of 'selectyrs' is zero
#Also the alternate hypothesis cannot be rejected
#The t-statistic calculated says that the 'selectyrs' is statistically
#significant for estimating 'choiceyrs'

#The coefficient 0.77 shows that selectyrs and choiceyrs are positively
#correlated. This is as expected because the increase in years of voucher will
#increase the number of years student attends choice school

#Two point criteria for valid IV

#Correlation : selectyrs, choiceyrs are positively correlated
#Zero correlation with error term u
  #We can say that selectyrs to be uncorrelated with 'u' because it can be
  #understood that the voucher is offered among those who applied, the other
  #unmeasureable variables for example student background donot influnce
  #the variable selectyrs. Thus we can say selectyrs is uncorrelated with the
  #error term 'u' and is a valid IV choice for choiceyrs.
```

```r
#1c

reg1c <- lm(formula = mnce ~ choiceyrs, data = data)

summary(reg1c)
```

```
##
## Call:
## lm(formula = mnce ~ choiceyrs, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.234 -13.234   0.603  12.766  60.114
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  46.2344     0.8507  54.348  < 2e-16 ***
## choiceyrs    -1.8370     0.5255  -3.495 0.000494 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.75 on 988 degrees of freedom
## Multiple R-squared:  0.01222,    Adjusted R-squared:  0.01122
## F-statistic: 12.22 on 1 and 988 DF,  p-value: 0.0004943
```

```r
#Residual standard error: 20.75 on 988 degrees of freedom
#Multiple R-squared:  0.01222,  Adjusted R-squared:  0.01122
#F-statistic: 12.22 on 1 and 988 DF,
#p-value: 0.0004943

#Coefficient -1.837

#The p-value at 5% significance level and the t-statistic calculated says that
#the variable 'choiceyrs' is statistically significant in estimating 'mnce'
#We expect the percentile score on the math test in 1994 increase with the
#increase in number of years the student attended school in the period
#1991 to 1994
#However, we see that the coefficient is negative, which is in contrary to the
#expectations.

reg1c_1 <- lm (formula = mnce ~ choiceyrs + black + hispanic + female, data = data)

summary(reg1c_1)
```

```
##
## Call:
## lm(formula = mnce ~ choiceyrs + black + hispanic + female, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -56.122 -12.507   0.108  12.156  60.156
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.1219     1.6567  34.479  < 2e-16 ***
## choiceyrs    -0.5652     0.5307  -1.065    0.287
## black       -16.0174     1.7944  -8.926  < 2e-16 ***
## hispanic    -13.4029     2.3168  -5.785 9.73e-09 ***
## female        1.3527     1.2758   1.060    0.289
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.99 on 985 degrees of freedom
## Multiple R-squared:  0.08677,    Adjusted R-squared:  0.08307
## F-statistic:  23.4 on 4 and 985 DF,  p-value: < 2.2e-16
```

```
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept)  57.1219     1.6567  34.479  < 2e-16 ***
#choiceyrs    -0.5652     0.5307  -1.065    0.287
#black       -16.0174     1.7944  -8.926  < 2e-16 ***
#hispanic    -13.4029     2.3168  -5.785 9.73e-09 ***
#female        1.3527     1.2758   1.060    0.289

#p-value: 0.0004943

#We see that the from the p-value and t statistic calculate above, we see that
#choiceyrs variable is now not statistically significant soon as we added other
#variables black, hispanic, female to our model. Among the new variables,
#only black and hispanic are statistically significant where as female is not
#statistically significant
```

```
#1d

#Why is choiceyrs a endogenous in the equation with choiceyrs, black, hispanic, female

reg1d <- lm(formula = choiceyrs ~  black + hispanic + female, data = data)
summary(reg1d)
```

```
##
## Call:
## lm(formula = choiceyrs ~ black + hispanic + female, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4183 -1.0909 -0.1929  0.8071  3.8337
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.16633    0.09927   1.676   0.0941 .
## black        0.92456    0.10357   8.927   <2e-16 ***
## hispanic     1.15003    0.13411   8.575   <2e-16 ***
## female       0.10198    0.07648   1.333   0.1827
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1.199 on 986 degrees of freedom
## Multiple R-squared:  0.09068,    Adjusted R-squared:  0.08792
## F-statistic: 32.78 on 3 and 986 DF,  p-value: < 2.2e-16
```

```
#Coefficients
#black        0.92456    0.10357   8.927   <2e-16 ***
#hispanic     1.15003    0.13411   8.575   <2e-16 ***
#female       0.10198    0.07648   1.333   0.1827

#p-value: < 2.2e-16

#We see from the regression results that the variables black and hispanic
#are statistically significant in estimating choiceyrs and variable female is
#not. This means that the number of years a student attended choice schools
#in period 1991-1994 is dependent on race. This is like multicollinearity.
#Thus the choiceyrs is a endogenous variable in the equation with the
#variables choiceyrs, black, hispanic, female
```

```
#1e

#use selectyrs as IV for choiceyrs

ivreg1e <- ivreg(formula = mnce ~ choiceyrs + black + hispanic + female |
                     selectyrs + black + hispanic + female, data = data)

summary(ivreg1e)
```

```
##
## Call:
## ivreg(formula = mnce ~ choiceyrs + black + hispanic + female |
##     selectyrs + black + hispanic + female, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -56.0680 -12.5098  -0.0476  12.0769  59.2142
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.0680     1.6577  34.426  < 2e-16 ***
## choiceyrs    -0.2413     0.6053  -0.399    0.690
## black       -16.3169     1.8148  -8.991  < 2e-16 ***
## hispanic    -13.7754     2.3412  -5.884 5.49e-09 ***
## female        1.3197     1.2763   1.034    0.301
##
## Diagnostic tests:
##                  df1 df2 statistic p-value
## Weak instruments   1 985  3281.836  <2e-16 ***
## Wu-Hausman         1 984     1.241   0.265
## Sargan             0  NA        NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.99 on 985 degrees of freedom
```

```
## Multiple R-Squared: 0.08643, Adjusted R-squared: 0.08272
## Wald test: 23.15 on 4 and 985 DF,  p-value: < 2.2e-16
```

```
#Coefficients:
#            Estimate Std. Error t value Pr(>|t|)
#(Intercept)  57.0680     1.6577  34.426  < 2e-16 ***
#choiceyrs    -0.2413     0.6053  -0.399    0.690
#black       -16.3169     1.8148  -8.991  < 2e-16 ***
#hispanic    -13.7754     2.3412  -5.884 5.49e-09 ***
#female        1.3197     1.2763   1.034    0.301

#p-value: < 2.2e-16

#selectyrs as IV for 'choiceyrs' doesnt produce a positive effect on 'mnce'

#The variable black and hispanic are statistically significant at the
#5% level as seen from the t value calculated -8.991(Pr 2e-16) and -5.884
#(Pr 5.49e-09). The variable female is not statistically significant
#as seen from the t-statistic calculated 1.034 (Pr 0.301) at 5% significance
#level. Thus we can say that mnce is more dependent on the race (black or
#hispanic) of the student.
```

```
#1f

reg1f <- lm(formula = mnce ~ choiceyrs + black + hispanic +
              female + mnce90, data = data)

summary(reg1f)
```

```
##
## Call:
## lm(formula = mnce ~ choiceyrs + black + hispanic + female + mnce90,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.921 -11.669   0.773  10.686  50.838
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.1529     3.6204   6.119 2.73e-09 ***
## choiceyrs     0.4106     0.7359   0.558  0.57726
## black        -8.3052     2.5461  -3.262  0.00123 **
## hispanic     -4.1050     3.3624  -1.221  0.22303
## female       -0.8829     1.7760  -0.497  0.61945
## mnce90        0.6204     0.0484  12.817  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.03 on 322 degrees of freedom
##   (662 observations deleted due to missingness)
## Multiple R-squared:  0.4237, Adjusted R-squared:  0.4147
## F-statistic: 47.34 on 5 and 322 DF,  p-value: < 2.2e-16
```

```
#Coefficient of choiceyrs = 0.4106
```

```
ivreg1f <- ivreg(mnce ~ choiceyrs + black + hispanic + female + mnce90|
                     selectyrs + black + hispanic +
                          female + mnce90, data = data)

summary(ivreg1f)
```

```
##
## Call:
## ivreg(formula = mnce ~ choiceyrs + black + hispanic + female +
##     mnce90 | selectyrs + black + hispanic + female + mnce90,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.171 -11.160   1.067  10.883  50.802
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.53886    3.64548   5.908 8.79e-09 ***
## choiceyrs    1.79938    0.86019   2.092 0.037236 *
## black       -9.06711    2.57142  -3.526 0.000483 ***
## hispanic    -5.00373    3.39279  -1.475 0.141240
## female      -1.02048    1.78630  -0.571 0.568205
## mnce90       0.62881    0.04874  12.901  < 2e-16 ***
##
## Diagnostic tests:
##                  df1 df2 statistic p-value
## Weak instruments   1 322    916.06 < 2e-16 ***
## Wu-Hausman         1 321     10.43 0.00137 **
## Sargan             0  NA        NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.12 on 322 degrees of freedom
## Multiple R-Squared: 0.4173,  Adjusted R-squared: 0.4082
## Wald test: 47.64 on 5 and 322 DF,  p-value: < 2.2e-16
```

```
#Coefficient of choiceyrs = 1.79938

#The coefficient of choiceyrs in IV is 1.79938 which is greater than the
#coefficient of choiceyrs in OLS in 0.4106. The variable is statistically
#significant in IV


#The p-value 0.037 with t-statistic 2.092 for choiceyrs shows that the
#statistically significance. This is a large effect that each year in a choice
#school is worth 1.7993 times in crease in the math percentile score.
```

```r
#1g

summary(data$mnce90)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    1.00   30.50   43.00   43.93   57.25   99.00     662
```

```r
nobs(ivreg1e)
```

```
## [1] 990
```

```r
nobs(ivreg1f)
```

```
## [1] 328
```

```r
#The stark difference in the number of observations in both the iv regressions
# 990 and 328 can be acoounted for low precision in the coefficients
#estimated in 1f which can be a reason for why the analysis 1f to be unconvincing
```

```r
#1h

ivreg1h <- ivreg(formula = mnce ~ choiceyrs1 + choiceyrs2 + choiceyrs3 +
                   choiceyrs4 + black + hispanic + female + mnce90|
                   selectyrs1 + selectyrs2 + selectyrs3 + selectyrs4 +
                   black + hispanic + female + mnce90, data = data)

summary(ivreg1h)
```

```
##
## Call:
## ivreg(formula = mnce ~ choiceyrs1 + choiceyrs2 + choiceyrs3 +
##     choiceyrs4 + black + hispanic + female + mnce90 | selectyrs1 +
##     selectyrs2 + selectyrs3 + selectyrs4 + black + hispanic +
##     female + mnce90, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -58.4487 -11.2629   0.9051  10.4675  54.5467
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.91132    3.76450   5.555 5.86e-08 ***
## choiceyrs1  -2.15904    5.16369  -0.418  0.67614
## choiceyrs2   1.49306    4.62755   0.323  0.74717
## choiceyrs3   1.08156    7.14421   0.151  0.87976
## choiceyrs4  13.92571    6.35406   2.192  0.02913 *
## black       -8.34689    2.66087  -3.137  0.00187 **
## hispanic    -3.58953    3.61733  -0.992  0.32180
## female      -1.03110    1.81870  -0.567  0.57115
## mnce90       0.63574    0.05028  12.645  < 2e-16 ***
```

```
##
## Diagnostic tests:
##                                 df1 df2 statistic p-value
## Weak instruments (choiceyrs1)   4 319    74.480  <2e-16 ***
## Weak instruments (choiceyrs2)   4 319   101.952  <2e-16 ***
## Weak instruments (choiceyrs3)   4 319    37.555  <2e-16 ***
## Weak instruments (choiceyrs4)   4 319    62.387  <2e-16 ***
## Wu-Hausman                      4 315     3.126  0.0153 *
## Sargan                          0  NA        NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.34 on 319 degrees of freedom
## Multiple R-Squared: 0.4067,  Adjusted R-squared: 0.3918
## Wald test: 29.22 on 8 and 319 DF,  p-value: < 2.2e-16
```

```
#From the t-statistic and p value observed above, we can say that the variables
#choiceyrs4 and black are statistically significant.
#From the coefficeints
#The coefficient of variable black -3.137 indicates the negative correlation
#between black race and the math percentile score
#The coefficient of variable choiceyrs4 13.925 indicates that 4 additional
#years at the school of choice increases the math percentile score in 1994
#mnce by 13.925 times
#This makes sense as the coefficient found here is in line
#with the coefficients in parts d,e and f
```

```
#Question 2

#Corn - percapita consumption of corn in bushels at country level
#price  price per buschel of corn
#income - percapita country income
#rainfall - inches of rainfall during last corn-growing season


#from the given equations we can say that the exogenous variables are
#rainfall and income and the endogenous variables are corn and price as they are
#dependent on the rainfall and income variables.

#Equation 1 is the demand equation as the equation demonstrates the consumption
#of corn in form of price of the corn and the income of country. Thus this is
#the demand function

#Equation 2 is supply function. Supply or availability of corn can be said from
#the rainfall in the last season , usually increase in rainfall leading to
#increase in supply of corn and also depends on the price. Thus 2 is supply eqn.
```

```
#Question 3

#a

#In the basic model where only poorhealth is estimated on smokes yes or no
#the tax can be a valid instrument if we observe that we observe some correlation
```

```
#between smoker and other health factors or if we see some correlation with
#the error term u. Criteria for the same can be evaluated if below is true
# Cov(tax, u ) = 0. tax is not correlated with error term u
# Cov(tax, smoker) != 0. tax is correlated with smoker


#b
#We need to run the following regressions in order to obtain instrumental
#variables estimator

#smoker = alpha0 + alpha1*tax + u
#smoker^ = alpha0^ + alpha1^*tax

#and

#poorhealth = beta0 + beta1*smoker + u
#poorhealth^ = beta0^ + beta1^*smoker


#c

#Tax as a valid instrument, ommitting does not create ommitted variable bias while
#estimating poorhealth. However, we can say that it can result in
#multicollinearity, depends on the strenght of tax as a valid instrument on
#regression for variable smoker.
#As tax is a valid instrument, that means the correlation between tax and the
#error term u is zero. And the error term u accounts for the ommitted variable
# bias in the regression. So ommitting tax will not create ommitted variable
#bias in regression.These two statemtnts says that variable tax is related to
#poorhealth via smoker. Thus we may not get supportive coefficient to prove
#statistical significance of tax in OLS but can prove multicollinearity with
#a less precision.
```

```
#Question 4

votedata <- read_dta('vote2.dta')
```

```
#a

#Difference the equation across two years

#Regression

reg4a <- lm_robust(cvote ~ clinexp + clchexp + cincshr,
                   data = votedata)

summary(reg4a)
```

```
##
## Call:
## lm_robust(formula = cvote ~ clinexp + clchexp + cincshr, data = votedata)
##
## Standard error type:  HC2
```

```
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)  -2.5559     0.5873 -4.3519 2.456e-05  -3.7162  -1.3957 153
## clinexp      -1.2915     1.3076 -0.9877 3.249e-01  -3.8749   1.2918 153
## clchexp      -0.5985     0.5859 -1.0215 3.086e-01  -1.7561   0.5591 153
## cincshr       0.1559     0.0543  2.8707 4.676e-03   0.0486   0.2631 153
##
## Multiple R-squared:  0.2437 ,    Adjusted R-squared:  0.2288
## F-statistic: 17.27 on 3 and 153 DF,  p-value: 1.037e-09
```

```
#p value 1.037e-09

#From the values we observed in the regression summary table, at 5% significance
#level we can say that cincshr is statistically significant whereas the
#other two variables clinexp and clchexp are not statistically
#significant in estimating cvote
```

```
#4b

wald.test(Sigma = vcov(reg4a), b = coef(reg4a), Terms = 2:3)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 3.3, df = 2, P(> X2) = 0.19
```

```
#The p value of 0.19 is more than 0.05, so we cannot reject the null hypothesis
#THus we can say that the coefficients of the variables delta_ln(inexp) and
#delta_ln(chexp) are zero and these variables are statistically insignificant
#Thus we can say that the ln change in incumbent expenditures and the ln change
#in challenger expenditures do not have joint role in estimating the incumbent
#vote share
```

```
#4c

reg4c <- lm_robust(cvote ~  cincshr, data = votedata)

summary(reg4c)
```

```
##
## Call:
## lm_robust(formula = cvote ~ cincshr, data = votedata)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)  -2.6811     0.6218  -4.312 2.871e-05  -3.9095  -1.4528 155
## cincshr       0.2176     0.0374   5.818 3.311e-08   0.1437   0.2915 155
##
## Multiple R-squared:  0.2287 ,    Adjusted R-squared:  0.2237
## F-statistic: 33.85 on 1 and 155 DF,  p-value: 3.311e-08
```

```
#The t statistic 5.818 and p 2.817e-05 prove that the variable delta_incshr is
#statistically significant. The coeffcient estimate is 0.2176 which means that
#the incumbent's share of vote rises by 2.176 percentage points for an increase
#of incumbents share of spending by 10 percentage points
```

```
#vote_rptchall <- vote %>% filter(rptchall == 1)
#dim(vote_rptchall)

table(votedata$rptchall)
```

```
##
##   0   1
## 150  36
```

```
rptchall1 <- votedata %>% filter(rptchall == 1)

rptchall1
```

```
## # A tibble: 36 x 26
##    state district democ vote90 vote88 inexp90 chexp90 inexp88 chexp88 prtystr
##    <chr>    <dbl> <dbl>  <dbl>  <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 AZ           2     1     66     73  112373    1445   99607    3065      55
##  2 CA           3     1     60     71  734005    4628  638688    7695      49
##  3 CA          11     1     66     71  620782   97030  269510   95575      58
##  4 CA          23     1     62     64  201404  360389  140486  100956      56
##  5 CA          24     1     69     72  287505    1830  191334   15449      65
##  6 CA          28     1     73     76  113669    3799  114523      NA      73
##  7 CA          32     1     62     67  462503    7579  457410   20608      49
##  8 CA          42     0     59     64  398963   28350  494487   11889      65
##  9 GA           1     1     71     67  399035   19130  337048   40461      39
## 10 GA           5     1     76     78  108118    7755  101540    6047      68
## # ... with 26 more rows, and 16 more variables: rptchall <dbl>, tenure <dbl>,
## #   lawyer <dbl>, linexp90 <dbl>, lchexp90 <dbl>, linexp88 <dbl>,
## #   lchexp88 <dbl>, incshr90 <dbl>, incshr88 <dbl>, cvote <dbl>, clinexp <dbl>,
## #   clchexp <dbl>, cincshr <dbl>, win88 <dbl>, win90 <dbl>, cwin <dbl>
```

```
#rptchall1$incshr9088 <- rptchall1$incshr90 - rptchall1$incshr88

reg4d <- lm_robust(cvote ~  cincshr, data = rptchall1)
summary(reg4d)
```

```
##
## Call:
## lm_robust(formula = cvote ~ cincshr, data = rptchall1)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
## (Intercept) -2.24982    0.95077  -2.366   0.0244 -4.18893  -0.3107 31
## cincshr      0.09239    0.06185   1.494   0.1453 -0.03375   0.2185 31
```

```
##
## Multiple R-squared:  0.03694 ,   Adjusted R-squared:  0.005875
## F-statistic: 2.232 on 1 and 31 DF,  p-value: 0.1453
```

```
#The t statistic of 1.494 with p-value 0.1453 greater than 0.05 says that the
#variable difference in increase in share is not statistically significant
#in estimating the votes.
#By using the pairs that have repeated challengers,
#we are eliminating any unmeasurable quality effects of the incumbent. THe
#assumption that holds true here is that the quality characteristics of the
#incumbent will not change over the period 1988 - 1990. THis is
#because we are differencing and the challenger is same in both times making the
#approach preferable.
```