Code Replication 6

Andres Perez

2022-11-14

```
suppressMessages(library(haven))
suppressMessages(library(fixest))
suppressMessages(library(AER))
suppressMessages(library(clubSandwich))

df = read_dta("E:/CalPoly SLO/Fall Quarter 2022/Advanced Econometrics
1/Working directory (datafiles too)/CodeRep6.dta")
```

a) How many observations are in the data set, and how many variables are in the dataset?

```
n = nrow(df)
k = ncol(df)
```

There are 24721 observations and 37 variables in the dataset

Column 1 of Table 3 reports averages for certain variables in the data set. What is the average of Housing vouchers in the data?

```
avgvouch = mean(df$vouch)
avgvouch

## [1] 20498
```

The average of Housing vouchers in the sample is 20498.00.

The full sample column in Table 4 reports one of the author's main analysis. Replicate the results in this column. The left-hand side variable is lrent_ut. You should be able to replicate these results very closely. Read the note in the table to see what additional controls other than the ones listed in the table need to be included in the regression. It says the regression includes individual unit and year fixed effects and unit-specific controls for: evidence of rodents, presence of washer or dryer, large cracks in the wall, and if the sewage system broke down in the last two years.

Also the standard errors need to be clustered at the MSA level, which is a lower level of aggregation than the individual unit fixed effects.

What is the coefficient estimate and standard error for log of vacancy rate? Is this variable statistically different than zero at the 1% level?

```
mod1 = feols(lrent_ut ~ lvouch + ltotpop + linc + lvacancy + evrod + drywash
+ cracks + ifsew + factor(year) + smsa | factor(control) + smsa, data = df)

## The variable 'smsa' has been removed because of collinearity (see
$collin.var).
```

```
summary(mod1)

## OLS estimation, Dep. Var.: lrent_ut
## Observations: 24,721
## Fixed-effects: factor(control): 8,388,  smsa: 135
## Standard-errors: Clustered (factor(control))
##                   Estimate Std. Error   t value   Pr(>|t|)
## lvouch            0.006527   0.012715  0.513323 6.0774e-01
## ltotpop           0.358711   0.098514  3.641216 2.7299e-04 ***
## linc              0.808238   0.127928  6.317932 2.7859e-10 ***
## lvacancy         -0.018983   0.013510 -1.405157 1.6001e-01
## evrod            -0.001021   0.008979 -0.113662 9.0951e-01
## drywash           0.035583   0.010775  3.302443 9.6247e-04 ***
## cracks            0.015992   0.009510  1.681611 9.2681e-02 .
## ifsew            -0.008442   0.018285 -0.461720 6.4429e-01
## factor(year)1999  0.002568   0.010267  0.250126 8.0250e-01
## factor(year)2001  0.079032   0.014212  5.560769 2.7684e-08 ***
## factor(year)2003  0.113606   0.014623  7.768710 8.8652e-15 ***
## ... 1 variable was removed because of collinearity (smsa)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.24335     Adj. R2: 0.639814
##                 Within R2: 0.094321

#clusterby
cs = vcovCR(mod1,df$smsa, "CR1")
round(coeftest(mod1,vcov = cs),3)

##
## z test of coefficients:
##
##                  Estimate Std. Error z value Pr(>|z|)
## lvouch              0.007      0.015   0.445    0.656
## ltotpop             0.359      0.174   2.067    0.039 *
## linc                0.808      0.209   3.867   <2e-16 ***
## lvacancy           -0.019      0.020  -0.967    0.333
## evrod              -0.001      0.007  -0.155    0.877
## drywash             0.036      0.011   3.375    0.001 ***
## cracks              0.016      0.009   1.705    0.088 .
## ifsew              -0.008      0.016  -0.512    0.608
## factor(year)1999    0.003      0.015   0.174    0.862
## factor(year)2001    0.079      0.021   3.811   <2e-16 ***
## factor(year)2003    0.114      0.020   5.751   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Copy and paste the 16 words from the abstract that describe the results from this regression

> Our results are consistent with voucher recipients renting more expensive units after receiving the subsidy

Everything in Table 4 can be replicated almost exactly, except one of the numbers is very far off from what is reported in the fixest output, which number in the table is that?

```
summary(mod1)
```

```
## OLS estimation, Dep. Var.: lrent_ut
## Observations: 24,721
## Fixed-effects: factor(control): 8,388,  smsa: 135
## Standard-errors: Clustered (factor(control))
##                   Estimate Std. Error   t value   Pr(>|t|)
## lvouch            0.006527   0.012715  0.513323 6.0774e-01
## ltotpop           0.358711   0.098514  3.641216 2.7299e-04 ***
## linc              0.808238   0.127928  6.317932 2.7859e-10 ***
## lvacancy         -0.018983   0.013510 -1.405157 1.6001e-01
## evrod            -0.001021   0.008979 -0.113662 9.0951e-01
## drywash           0.035583   0.010775  3.302443 9.6247e-04 ***
## cracks            0.015992   0.009510  1.681611 9.2681e-02 .
## ifsew            -0.008442   0.018285 -0.461720 6.4429e-01
## factor(year)1999  0.002568   0.010267  0.250126 8.0250e-01
## factor(year)2001  0.079032   0.014212  5.560769 2.7684e-08 ***
## factor(year)2003  0.113606   0.014623  7.768710 8.8652e-15 ***
## ... 1 variable was removed because of collinearity (smsa)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.24335     Adj. R2: 0.639814
##                 Within R2: 0.094321
```

There is not a single number that he reported that is wrong. We can see in our model that every number is right and if theres any discrepancies between the replication and the table then it must be because of rounding.

Conduct a joint hypothesis test that the coefficients on evidence of rodents, large cracks in the wall, and if the sewage system broke down are all jointly insignificant (equal to zero). If you are using linearHypothesis with the fixest output it will default to a chi-squared test (large sample test). What is the chi-squared statistic from this test? Do you reject the null hypothesis at the 10% level?

```
co = linearHypothesis(mod1, c("evrod=0","cracks=0","ifsew=0"),vcov = cs)
co
```

```
## Linear hypothesis test
##
## Hypothesis:
## evrod = 0
## cracks = 0
## ifsew = 0
##
## Model 1: restricted model
## Model 2: lrent_ut ~ lvouch + ltotpop + linc + lvacancy + evrod + drywash +
##     cracks + ifsew + factor(year) + smsa | factor(control) +
##     smsa
```

```
##
## Note: Coefficient covariance matrix supplied.
##
##   Df Chisq Pr(>Chisq)
## 1
## 2  3 3.267     0.3523
```

We see that the value of the chi-square statistic is 3.267 so we reject the null at the 10%

In a paragraph, summarize your conclusions from this analysis, make sure to state what the research question is and what the findings are in from your regression in practical terms. Do you think that the assumptions of the linear regression model are satisfied? Comment.

This research looked at how the low-income programs (in this case vouchers) that seek to help people pay for rent might actually be secretly increasing the cost of renting. The research found that vouchers are not the cause of rent increase, but rather that people who receive these vouchers tend to go and rent in more expensive areas as they feel like the voucher allows them to rent more expensive houses or more expensive places. The results also show that as the vouchers became available to more and more low income people the cost of rent increased, but it was noted that this is due to the same as before, people who get this vouchers tend to move to more expensive places. I think the assumptions of the linear regression model are satisfied because the data provided and the tables shown in the research journal show us the relationship between the price of rent and all the different variables provided