Lab Assignment #6
Julia Bloom
jb4202
12/4/2018

1. Run a naive ("pooled") OLS regression on the panel data. Tell we how you expect your Xs to affect your Y and why. Apply clustered standard errors too. Interpret your results.

For my analysis, I chose to examine whether working status had any effect on respondents' opinion of whether working women make good mothers. For my dependent variable (Figure 1.1), I chose to use *FECHLD* which measures respondents' agreement with the idea that a working mother can establish just as warm and secure a relationship with her children as a mother who does not work, on a scale of 1 (Strongly agree) to 4 (Strongly disagree). For my independent variables (Figure 1.1), I chose to use *WRKSTAT* which shows respondents working status with 1 = working full time, and *SEX*, which shows respondents sex. Before running my analysis, I predicted that individuals who work full-time would agree more with the statement than individuals who do not work full-time, since people who work full-time would not want to believe that they have (or could have) "worse" relationships with their children than people who do not work.

The results of the pooled regression and variable recoding is shown in Figure 1.2. The results of the regression show that respondents who work full time score 0.26377 higher on the agreement scale than those who do not (holding all else constant), thus confirming my prediction above. Additionally, women score 0.348906 on the agreement scale than men, holding working status constant. Both coefficients are statistically significant at $p < 0.001$, though the R-squared is low at 0.05447. Running the same regression with clustered standard errors applied in Figure 1.3 shows that the coefficient for *working* (0.263767) and *female* (0.348906) remains the same, with the same significance levels and roughly the same t-values.

Figure 1.1: Variables used in analysis

257) Now I'm going to read several more statements. As I read each one, please tell me whether you strongly agree, agree, disagree, or strongly disagree with it. For example, here is the statement: A working mother can establish just as warm and secure a relationship with her children as a mother who does not work. (FECHLD)

1) Strongly agree
2) Agree
3) Disagree
4) Strongly disagree

```
> table(pd.sub$fechld)

   1    2    3    4
 802 1334  818  170
```

3) Last week were you working full-time, part-time, going to school, keeping house, or what? (WRKSTAT)

1) Working full-time
2) Working part-time
3) With a job, but not at work because of temporary illness, vacation, strike
4) Unemployed, laid off, looking for work
5) Retired
6) In school
7) Keeping house

```
> table(pd.sub$wrkstat)

   1    2    3    4    5    6    7    8
2345  466   85  185  878  131  569  148
```

44) Respondent's sex (SEX)

1) Male
2) Female

```
> table(pd.sub$sex)

   1    2
2018 2794
```

Figure 1.2: Naïve ("Pooled") Regression on variables *FECHLD, SEX*, and *WRKSTAT*

```
> pd.sub$female = ifelse(pd.sub$sex==2, 1, 0) ## if participant is female
> pd.sub$working = ifelse(pd.sub$wrkstat==1, 1, 0) ## if participant is working
> pd.sub$r.fechld = 5-pd.sub$fechld ## reverse code for how close a working mother can
be to her child
> pd.sub$year= ifelse(pd.sub$panelwave==3, 1, 0)
> lm1 <- lm(r.fechld ~ female + working,  data = pd.sub)
> summary(lm1)

Call:
lm(formula = r.fechld ~ female + working, data = pd.sub)

Residuals:
     Min       1Q   Median       3Q      Max
-2.17324 -0.82433  0.09053  0.82676  1.43943

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.56057    0.02854  89.712   <2e-16 ***
female       0.34891    0.03034  11.499   <2e-16 ***
working      0.26377    0.03008   8.768   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8275 on 3119 degrees of freedom
  (2878 observations deleted due to missingness)
Multiple R-squared:  0.05447,     Adjusted R-squared:  0.05386
F-statistic: 89.83 on 2 and 3119 DF,  p-value: < 2.2e-16
```

Figure 1.3: Applied clustered standard errors for pooled regression

```
> clusterSE <- function(fit, cluster.var, data){
+   require(plm); require(lmtest)
+
+   if (missing(data) & cluster.var %in% colnames(index(fit))){
+     cvar <- index(fit, cluster.var)
+     n <- length(unique(cvar))
+     N <- length(cvar)
+   }
+   else{
+     row.ids <- as.numeric(rownames(model.frame(fit)))
+     n <- length(unique(data[row.ids, cluster.var]))
+     N <- length(row.ids)
+   }
+
+   df <- (n/(n - 1)) * (N - 1)/fit$df.residual
+   vcov <- df*vcovHC(fit, type = "HC0", cluster = "group")
+   coeftest(fit, vcov = vcov)
+ }
> clusterSE(fit = lm1, cluster.var = "idnum", data=pd.sub)

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) 2.560568   0.028327 90.3932 < 2.2e-16 ***
female      0.348906   0.030117 11.5851 < 2.2e-16 ***
working     0.263767   0.030042  8.7801 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

2. Run a first differences regression on the same model in Question 1. Interpret your results. Do you draw a different conclusion than in Question 1? Explain.

The results of the first differences regression in Figure 2.1 show that a change in someone's working status (i.e., going from not working full-time to working full-time) leads to a 0.147374 increase on the agreement scale, holding all else in the model constant. This would indicate an agreement that a woman who works can establish just as warm and secure a relationship with her children than a woman who does not work for the same person across 3 waves of data, holding all else constant. This coefficient is statistically significant at $p < 0.01$. Additionally, the coefficient for *year* shows that there is a statistically insignificant increase of 0.018165 on the agreement scale when going from Wave 2 to Wave 3, vs. going from Wave 1 to Wave 2. Finally, there is a negative coefficient of -0.068145 for the variable *female*, though it is statistically insignificant. While it is possible for someone to change genders over time, this might indicate an error on the part of the survey taker. The R-squared for this model is also lower than the previous model, at 0.0055573. Applying clustered standard errors in Figure 2.2 show roughly the same coefficients and significance levels for all variables, though the t-values for each variable have slightly decreased.

In contrast to the model in question 1, the results from this model might suggest that the experience of working full time might make individuals more inclined to agree with the statement, rather than something inherent in the "type" of person who would work full time. In this sense, the second model might show a more casual relationship between going to work full-time, and one's opinion on whether working mothers can have a warm and secure relationship with their children.

Figure 2.1: First differences regression on variables *FECHLD*, *SEX*, and *WRKSTAT*

```
> plm1 <- plm(r.fechld ~ female + working + year,  index = c("idnum", "panelwave"),  m
odel = "fd", data = pd.sub)
> summary(plm1)
Oneway (individual) effect First-Difference Model

Call:
plm(formula = r.fechld ~ female + working + year, data = pd.sub,
    model = "fd", index = c("idnum", "panelwave"))

Unbalanced Panel: n = 1308, T = 1-3, N = 3122
Observations used in estimation: 1814

Residuals:
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-3.01816 -0.16554  0.00000  0.00445  0.14737  3.00000

Coefficients:
         Estimate Std. Error t-value Pr(>|t|)
female  -0.068145   0.195217 -0.3491 0.727076
working  0.147374   0.046459  3.1721 0.001539 **
year     0.018165   0.029533  0.6151 0.538583
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1317.8
Residual Sum of Squares: 1310.5
R-Squared:      0.0055573
Adj. R-Squared: 0.0044591
F-statistic: 5.035 on 2 and 1811 DF, p-value: 0.0065975
```

Figure 2.2: Applied clustered standard errors for first differences regression

```
> clusterSE(fit = plm1, cluster.var = "idnum", data=pd.sub)

t test of coefficients:

         Estimate Std. Error t value Pr(>|t|)
female  -0.068145   0.316547 -0.2153  0.82958
working  0.147374   0.050736  2.9047  0.00372 **
year     0.018165   0.033543  0.5415  0.58820
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```