BT2101 GA5 Group 67 Submission

Lo Zhi Hao

2022-10-31

1 Fixed Effect Regression using R

Please use the jtrain dataset from Wooldridge package in R to answer this question. Please carefully read the document of data description, and refer to the cited paper in the document if necessary. For this exercise, we want to determine the effect of the job training grant on hours of job training per employee. The basic model is:

hrsempit = β0 + β1 x grantit + β2 x employit + µit (1)

```
## Setting up the environment for further studies

## install.packages("wooldridge")
## install.packages("dplyr")
## install.packages("ggplot2")

library(wooldridge)
library(dplyr)
library(knitr)
library(corrplot)
library(ggplot2)

## Downloading the dataset
data('jtrain')
summary(jtrain)
head(jtrain)
```

a. Estimate the basic model seen above in equation (1). Please interpret the meaning of β1.

```
## creating the linear model

model <- lm(hrsemp ~ grant + employ, data = jtrain)
summary(model)
```

```
##
## Call:
## lm(formula = hrsemp ~ grant + employ, data = jtrain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -41.512 -11.769  -6.894   3.219 137.339
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.56561    1.53159   8.857  < 2e-16 ***
## grant       33.71094    3.17701  10.611  < 2e-16 ***
## employ      -0.06028    0.01527  -3.948 9.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.44 on 387 degrees of freedom
##   (81 observations deleted due to missingness)
## Multiple R-squared:  0.2421, Adjusted R-squared:  0.2382
## F-statistic:  61.8 on 2 and 387 DF,  p-value: < 2.2e-16
```

The relationship is as follows:

`hrsemp` = 13.56561 + 33.71094 x `grant` + -0.06028 x `employ`

Multiple R squared is 0.2421, and Adjusted R squared is 0.2382.

For coefficient of Beta1, it is 33.71094, which suggests that firms that receive a job training grant is associated to a 33.71094 increase in the total number of training hours per individual employee compared to firms which did not receive grants holding other factors constant. This coefficient is statistically significant (p-value = 2e-16 < 0.05), which suggests that we are statistically confident that the effect of receiving grant on total number of training hours per individual employee is significantly different from 0.

b. Use log-transformed variable lhrsemp as the dependent variable and repeat a similar regression as Question (a). Please interpret the meaning of β1.

```
## creating the linear model

model2 <- lm(lhrsemp ~ grant + employ, data = jtrain)
summary(model2)
```

```
##
## Call:
## lm(formula = lhrsemp ~ grant + employ, data = jtrain)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.383 -1.346 -0.162  1.028  3.575
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.4786187  0.0914151  16.175  < 2e-16 ***
## grant        2.1065903  0.1896246  11.109  < 2e-16 ***
## employ      -0.0024020  0.0009114  -2.635  0.00874 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.339 on 387 degrees of freedom
##   (81 observations deleted due to missingness)
## Multiple R-squared:  0.2475, Adjusted R-squared:  0.2436
## F-statistic: 63.66 on 2 and 387 DF,  p-value: < 2.2e-16
```

The relationship is as follows:

`lhrsemp` = 1.4786187 + 2.1065903 x `grant` - 0.0024020 x `employ`

Multiple R squared is 0.2475, and Adjusted R squared is 0.2436.

For coefficient of Beta1, it is 2.1065903, which suggests firms that receiving a job training grant is associated to a 2.1065903 x 100 % = 210.65% increase in the total number of training hours per individual employee compared to firms which did not receive grants holding other factors constant. This coefficient is statistically significant (p-value = 2e-16 < 0.05), which suggests that we are statistically confident that the effect of receiving grant on log-transformed total number of training hours per individual employee is significantly different from 0.

c. There are many confounders that could bias the relationship between job training grants and hours of job training per employee. Suppose those confounders can be categorized into time-invariant, firm-invariant, and fully flexible (variable both in time and by firms). This division is mutually exclusive. Please identify one confounder for one category. Clearly articulate your logic for the confounding pathway and why it belongs to that particular category above.

For time invariant confounder, we identified industry for the firms as a time invariant confounder. Different firms in different industries may or may not receive the grants from local authorities based on the industry that the government wants to develop. For example, if the local government at California prefers to develop the technology industry, firms in semiconductor industry might be easier to receive grants compared to firms which are in the textile industry. For hours of employee training, some industry requires more precise and advanced machinery and technology, and thus may have to invest more hours into job training for employees. As a firm’s industry will be relatively stable for the short term (for example in a few years), it is essentially time invariant. Thus, as it may be associated to both hours of job training per employee and job training grants awarded, it is thus a time invariant confounding variable that might exist in the model.

For entity invariant confounder, we identified the state policy as a potential entity invariant confounder. For state policy, the change in policies and regulations for different manufacturing industries may be associated to different amount of hours of job training per employee. At the same time, different in state policy across different years may be associated with different amount and focus area for awarding job training grants. As every firm in the dataset is in the same state, state policy will affect all firms, regardless of the firm’s background and culture thus it is entity invariant. Therefore, as it is both associated to hours of job training per employee and job training grants awarded, it is thus an entity invariant confounding variable that might exist in the model.

For fully flexible confounder, we identified technology as a potential fully flexible confounder. For technology, the more advanced the technology needed for the production for the firm, the more hours is needed to train and educate the employees on how to use it thus it may be associated with hours of job training per employee. At the same time, firms which possess valuable technology may be more likely to receive job training grants from the state government, as it may be recognised as technology that may change and improve the lives of residents by the state government. Thus it may be associated with the job training grant as well. As each firm possess different technologies and uses different technologies for production, it differs from entity to entity, and as technologies may be advanced over time through research and development, it differs across time as well. Therefore, as it is both associated to hours of job training per employee and job training grants awarded, it is thus a fully flexible confounding variable that might exist in the model.

d. Estimate a model that adds a full set of firm dummies and year dummies into the basic model, does β1 change compared to the basic model in Question (a)? Explain why or why not.

```
## creating the linear model

model3 <- lm(hrsemp ~ grant + employ + as.factor(year) + as.factor(fcode), data = jtrain)
summary(model3)
```

```
##
## Call:
```

```
## lm(formula = hrsemp ~ grant + employ + as.factor(year) + as.factor(fcode),
##     data = jtrain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.604  -3.865   0.307   3.952  92.419
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            14.28557   10.14426   1.408 0.160297
## grant                  34.39102    2.41695  14.229  < 2e-16 ***
## employ                 -0.08119    0.05171  -1.570 0.117632
## as.factor(year)1988    -0.78872    1.88628  -0.418 0.676207
## as.factor(year)1989     4.97762    1.90377   2.615 0.009473 **
## as.factor(fcode)410440  -3.29311   12.79194  -0.257 0.797053
## as.factor(fcode)410495  19.24098   12.58121   1.529 0.127439
## as.factor(fcode)410500  -1.90978   11.66653  -0.164 0.870101
## as.factor(fcode)410509 -17.63945   17.25476  -1.022 0.307626
## as.factor(fcode)410513  -7.74328   12.73403  -0.608 0.543686
## as.factor(fcode)410517 -14.00401   12.62880  -1.109 0.268537
## as.factor(fcode)410518 -10.45248   12.05088  -0.867 0.386572
## as.factor(fcode)410521   6.54594   12.73403   0.514 0.607669
## as.factor(fcode)410523  12.26933   11.71278   1.048 0.295869
## as.factor(fcode)410529 -11.78582   12.14018  -0.971 0.332577
## as.factor(fcode)410531   0.83575   12.27613   0.068 0.945777
## as.factor(fcode)410533   7.29105   12.77008   0.571 0.568546
## as.factor(fcode)410535 -14.87000   12.85854  -1.156 0.248605
## as.factor(fcode)410536 -12.87804   12.71975  -1.012 0.312300
## as.factor(fcode)410538  10.60802   16.73271   0.634 0.526678
## as.factor(fcode)410540 -11.48722   12.07983  -0.951 0.342549
## as.factor(fcode)410544 -12.07899   12.40692  -0.974 0.331208
## as.factor(fcode)410546  -1.16542   11.66245  -0.100 0.920480
## as.factor(fcode)410547 -13.20164   12.58795  -1.049 0.295301
## as.factor(fcode)410556  -8.62771   16.41823  -0.525 0.599702
## as.factor(fcode)410560 -13.78751   12.57449  -1.096 0.273925
## as.factor(fcode)410561 -12.74319   12.92666  -0.986 0.325177
## as.factor(fcode)410562 -10.82824   12.69847  -0.853 0.394627
## as.factor(fcode)410563  -7.81794   11.78136  -0.664 0.507564
## as.factor(fcode)410564   0.22511   11.72589   0.019 0.984698
## as.factor(fcode)410565  -1.41798   11.59008  -0.122 0.902724
## as.factor(fcode)410566 -10.58184   12.76283  -0.829 0.407828
## as.factor(fcode)410567  -6.61394   12.16627  -0.544 0.587179
## as.factor(fcode)410569 -13.67926   12.54781  -1.090 0.276683
## as.factor(fcode)410571   6.55870   12.69847   0.516 0.605963
## as.factor(fcode)410577  -7.46357   11.72322  -0.637 0.524934
## as.factor(fcode)410586  32.68013   11.58302   2.821 0.005164 **
## as.factor(fcode)410591 -11.88518   12.41311  -0.957 0.339251
## as.factor(fcode)410593 -15.27594   12.97290  -1.178 0.240101
## as.factor(fcode)410596  -2.49113   12.75560  -0.195 0.845319
## as.factor(fcode)410603  19.33303   14.02195   1.379 0.169193
## as.factor(fcode)410604  -6.21007   11.58291  -0.536 0.592336
## as.factor(fcode)410606   2.70432   11.59071   0.233 0.815705
## as.factor(fcode)410609 -12.89446   12.36420  -1.043 0.298004
## as.factor(fcode)410612   1.59527   11.90246   0.134 0.893487
## as.factor(fcode)410626  -9.97840   12.47613  -0.800 0.424583
## as.factor(fcode)410635  -4.17935   11.65271  -0.359 0.720151
## as.factor(fcode)410639  -2.64167   11.62877  -0.227 0.820479
## as.factor(fcode)410640  26.15859   12.80661   2.043 0.042138 *
## as.factor(fcode)410665 -11.70554   14.09445  -0.831 0.407041
## as.factor(fcode)410680  10.79538   12.15577   0.888 0.375344
## as.factor(fcode)410686 -14.21531   12.94200  -1.098 0.273089
## as.factor(fcode)418006  -8.07600   12.01482  -0.672 0.502094
## as.factor(fcode)418008 -20.94622   12.90103  -1.624 0.105715
## as.factor(fcode)418011 -17.99731   12.15300  -1.481 0.139889
## as.factor(fcode)418013 -18.90927   12.71207  -1.488 0.138137
## as.factor(fcode)418014  72.58272   12.85629   5.646 4.44e-08 ***
## as.factor(fcode)418021 -23.58014   12.86370  -1.833 0.067976 .
## as.factor(fcode)418024  -8.39975   11.92684  -0.704 0.481915
## as.factor(fcode)418035 -18.54954   12.49128  -1.485 0.138799
## as.factor(fcode)418036  14.65283   22.99540   0.637 0.524571
## as.factor(fcode)418045  39.53593   12.74839   3.101 0.002147 **
## as.factor(fcode)418046 -16.18847   12.69807  -1.275 0.203532
## as.factor(fcode)418052  -8.32509   13.00804  -0.640 0.522759
## as.factor(fcode)418054 -21.63189   12.00600  -1.802 0.072784 .
## as.factor(fcode)418065   4.94811   15.01775   0.329 0.742064
## as.factor(fcode)418066 -11.39225   13.65088  -0.835 0.404769
## as.factor(fcode)418076  -1.26325   11.92684  -0.106 0.915733
## as.factor(fcode)418083  -3.93126   13.02953  -0.302 0.763116
## as.factor(fcode)418084  13.87903   13.50471   1.028 0.305072
```

```
## as.factor(fcode)418091  -6.59816   11.96050  -0.552 0.581670
## as.factor(fcode)418097 -16.63625   11.61133  -1.433 0.153171
## as.factor(fcode)418098  -8.68742   12.73321  -0.682 0.495701
## as.factor(fcode)418107  10.13352   12.63606   0.802 0.423339
## as.factor(fcode)418109  77.80775   12.90103   6.031 5.80e-09 ***
## as.factor(fcode)418118 -15.42989   12.78322  -1.207 0.228552
## as.factor(fcode)418124 -18.87936   12.69110  -1.488 0.138110
## as.factor(fcode)418125  -0.75144   11.61164  -0.065 0.948453
## as.factor(fcode)418126  -3.95304   12.11298  -0.326 0.744434
## as.factor(fcode)418140   5.14486   12.69807   0.405 0.685699
## as.factor(fcode)418147  -5.63242   12.23669  -0.460 0.645707
## as.factor(fcode)418163  -8.12913   11.88859  -0.684 0.494747
## as.factor(fcode)418168  11.58017   12.91901   0.896 0.370916
## as.factor(fcode)418177  -4.06979   11.61048  -0.351 0.726237
## as.factor(fcode)418213 -12.96102   13.70878  -0.945 0.345336
## as.factor(fcode)418220 -12.25118   12.62926  -0.970 0.332949
## as.factor(fcode)418225 -22.16141   12.69110  -1.746 0.081997 .
## as.factor(fcode)418229 -15.11356   12.92666  -1.169 0.243441
## as.factor(fcode)418237 -22.38548   12.52979  -1.787 0.075212 .
## as.factor(fcode)418239  -2.83073   11.81071  -0.240 0.810778
## as.factor(fcode)418243 -14.57232   12.77735  -1.140 0.255173
## as.factor(fcode)418245 -13.11095   12.41311  -1.056 0.291884
## as.factor(fcode)419198  11.90553   12.84889   0.927 0.355035
## as.factor(fcode)419201 -11.50523   12.71263  -0.905 0.366321
## as.factor(fcode)419242  -7.63265   12.07914  -0.632 0.528036
## as.factor(fcode)419268 -15.99176   13.00804  -1.229 0.220083
## as.factor(fcode)419272   6.80971   12.32722   0.552 0.581157
## as.factor(fcode)419275  16.27349   12.74030   1.277 0.202668
## as.factor(fcode)419289  -2.00893   12.82683  -0.157 0.875671
## as.factor(fcode)419297  14.98716   11.84924   1.265 0.207109
## as.factor(fcode)419298  -9.01946   11.87036  -0.760 0.448069
## as.factor(fcode)419302   4.37560   16.15393   0.271 0.786715
## as.factor(fcode)419303   0.19768   12.95409   0.015 0.987837
## as.factor(fcode)419305 -23.73961   13.03144  -1.822 0.069688 .
## as.factor(fcode)419307  10.01960   14.15369   0.708 0.479655
## as.factor(fcode)419309   3.58514   12.93883   0.277 0.781944
## as.factor(fcode)419319  36.58669   11.85656   3.086 0.002258 **
## as.factor(fcode)419328 -20.95350   12.55587  -1.669 0.096400 .
## as.factor(fcode)419335  -7.43617   11.99993  -0.620 0.536028
## as.factor(fcode)419339 -11.89315   12.15577  -0.978 0.328821
## as.factor(fcode)419343  -4.97514   11.92028  -0.417 0.676767
## as.factor(fcode)419344 -18.53251   17.43864  -1.063 0.288927
## as.factor(fcode)419351 -17.72972   12.84152  -1.381 0.168613
## as.factor(fcode)419357  59.93358   11.92684   5.025 9.57e-07 ***
## as.factor(fcode)419376 -11.08161   12.38620  -0.895 0.371818
## as.factor(fcode)419378  44.12874   12.56904   3.511 0.000529 ***
## as.factor(fcode)419379  -7.48988   12.57783  -0.595 0.552058
## as.factor(fcode)419380  -8.28919   12.10458  -0.685 0.494103
## as.factor(fcode)419381  -8.04900   12.86370  -0.626 0.532071
## as.factor(fcode)419384  -8.31770   11.90360  -0.699 0.485352
## as.factor(fcode)419388 -17.47353   12.38020  -1.411 0.159362
## as.factor(fcode)419400  57.34797   12.91139   4.442 1.34e-05 ***
## as.factor(fcode)419401   3.27106   11.81749   0.277 0.782163
## as.factor(fcode)419409  11.08450   11.84583   0.936 0.350312
## as.factor(fcode)419410  33.63342   12.60901   2.667 0.008141 **
## as.factor(fcode)419420   0.76599   12.79768   0.060 0.952320
## as.factor(fcode)419433  -8.48454   12.70506  -0.668 0.504870
## as.factor(fcode)419434  -5.50178   11.71278  -0.470 0.638960
## as.factor(fcode)419449  10.95579   12.05565   0.909 0.364344
## as.factor(fcode)419450   3.31244   12.83617   0.258 0.796576
## as.factor(fcode)419459   5.76207   11.72589   0.491 0.623575
## as.factor(fcode)419461 -13.78826   12.54279  -1.099 0.272691
## as.factor(fcode)419467  78.16565   12.94200   6.040 5.53e-09 ***
## as.factor(fcode)419472   4.56994   11.62923   0.393 0.694675
## as.factor(fcode)419473 -13.01751   12.74030  -1.022 0.307879
## as.factor(fcode)419479  -9.61108   13.77288  -0.698 0.485932
## as.factor(fcode)419482  -6.38288   12.72688  -0.502 0.616439
## as.factor(fcode)419483  -4.63523   11.58598  -0.400 0.689443
## as.factor(fcode)419486  -7.83873   11.70984  -0.669 0.503847
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.19 on 251 degrees of freedom
##   (81 observations deleted due to missingness)
## Multiple R-squared:  0.8036, Adjusted R-squared:  0.6956
## F-statistic: 7.441 on 138 and 251 DF,  p-value: < 2.2e-16
```

For coefficient of Beta1, it is 34.39102, which suggests that firms that receive a job training grant on the year itself is associated to a 34.39102 increase in the total number of training hours per individual employee. This is a slight change from the value of 33.71094 that we found out earlier in the model above. This might be because the firm-specific and year-specific confounding variables are not causing a major bias in the model as stated above. This coefficient is statistically significant (p-value = 2e-16 < 0.05), which suggests that we are statistically confident that the effect of receiving grant on total number of training hours per individual employee is significantly different from 0.

e. Estimate a model using only the entity-demeaned variables, does β1 change compared to the basic model in question (a)? Does β1 change compared to the two-way fixed effects model in Question (d)? Explain why or why not.

```
## creating the linear model

jtrain_temp <- jtrain %>% group_by(fcode) %>% mutate(grant_demeaned = grant - mean(grant)) %>% mutate(employ_deme
aned = employ - mean(employ)) %>% mutate(hrsemp_demeaned = hrsemp - mean(hrsemp))
jtrain_temp
```

```
## # A tibble: 471 × 33
## # Groups:   fcode [157]
##     year  fcode employ  sales avgsal scrap rework tothrs union grant   d89   d88
##    <int>  <dbl>  <int>  <dbl>  <dbl> <dbl>  <dbl>  <int> <int> <int> <int> <int>
##  1  1987 410032    100 4.7 e7  35000    NA     NA     12     0     0     0     0
##  2  1988 410032    131 4.3 e7  37000    NA     NA      8     0     0     0     1
##  3  1989 410032    123 4.9 e7  39000    NA     NA      8     0     0     1     0
##  4  1987 410440     12 1.56e6  10500    NA     NA     12     0     0     0     0
##  5  1988 410440     13 1.97e6  11000    NA     NA     12     0     0     0     1
##  6  1989 410440     14 2.35e6  11500    NA     NA     10     0     0     1     0
##  7  1987 410495     20 7.5 e5  17680    NA     NA     50     0     0     0     0
##  8  1988 410495     25 1.1 e5  18720    NA     NA     50     0     0     0     1
##  9  1989 410495     24 9.5 e5  19760    NA     NA     50     0     0     1     0
## 10  1987 410500    200 2.37e7  13729    NA     NA      0     0     0     0     0
## # … with 461 more rows, and 21 more variables: totrain <int>, hrsemp <dbl>,
## #   lscrap <dbl>, lemploy <dbl>, lsales <dbl>, lrework <dbl>, lhrsemp <dbl>,
## #   lscrap_1 <dbl>, grant_1 <int>, clscrap <dbl>, cgrant <int>, clemploy <dbl>,
## #   clsales <dbl>, lavgsal <dbl>, clavgsal <dbl>, cgrant_1 <int>,
## #   chrsemp <dbl>, clhrsemp <dbl>, grant_demeaned <dbl>, employ_demeaned <dbl>,
## #   hrsemp_demeaned <dbl>
```

```
model4 <- lm(hrsemp_demeaned ~ grant_demeaned + employ_demeaned, data = jtrain_temp)
summary(model4)
```

```
##
## Call:
## lm(formula = hrsemp_demeaned ~ grant_demeaned + employ_demeaned,
##     data = jtrain_temp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.000  -3.542  -0.057   4.817  96.000
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -6.644e-16  6.190e-01   0.000    1.000
## grant_demeaned   3.490e+01  1.929e+00  18.099   <2e-16 ***
## employ_demeaned -3.774e-02  4.270e-02  -0.884    0.377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.94 on 369 degrees of freedom
##   (99 observations deleted due to missingness)
## Multiple R-squared:  0.4734, Adjusted R-squared:  0.4706
## F-statistic: 165.9 on 2 and 369 DF,  p-value: < 2.2e-16
```

For coefficient of Beta1 in model4, it is 34.90, which suggests that firms that receive a job training grant on the year itself is associated to a 34.90 increase in the total number of training hours per individual employee. This is a slight change from the value of 34.39102 that we found out earlier in the model3 and 33.71094 in model1 above. This might be because that the firm-specific confounding variable is not causing a major bias in the model as stated above. This coefficient is statistically significant (p-value = 2e-16 < 0.05), which suggests that we are statistically confident that the effect of receiving grant on total number of training hours per individual employee is significantly different from 0.

f. For each confounder you identified in Question (c): explain whether or not it has been controlled for in the model in Question (d). Also explain whether or not it has been controlled for in the model in Question (e).

In question (d), the year dummy variable and firm code dummy variable are taken into consideration. Therefore, the model essentially controls for time invariant and entity invariant data. As such, as we are able to control for both across time and firms, but they each controls for time and entity respectively and does not control for fully flexible variables as they vary across time and entity. A timed fixed and entity fixed regression controls for entity invariant and time invariant confounders respectively but not for fully flexible variable .For example,A fully flexible reacts like our independent variable with two subscripts, if it is eliminated by N+T-2 regression or through entity demeaned regression, we are technically saying that we will also eliminate our primary independent variable as well. Thus, it has control over time invariant and entity invariant variables, but not for fully flexible variables which we suggest may not be able to be controlled for using panel regression technique.

In question (e), entity-demeaned hrsemp, grant and employ variables are created and being used in the model itself. Therefore, it shows that the model only controls for time invariant variables. From this, we can see that only time invariant is controlled for, while entity invariant and fully flexible variables are not included in the model.