Lab 2 - Fixed/Random Effects

Christine Lucille Kuryla

2024-11-16

Contents

Analysis and data description

This analysis uses panel data from NLSY97 (National Longitudinal Survey Youth 1997), a longitudinal survey conducted in the United States. More information and data can be found here: https://www.bls.gov/nls/nlsy97.htm

A fixed and random effects analysis is performed.

A total of 6 waves have the variables of interest, so we do not explore first differences in the main analysis, just fixed and random effects.

Variable Descriptions

Outcome variable: `health`

Our outcome variable of interest is self-rated health. It has been reverse coded such that in this analysis, the values represent:

- 5 - "Excellent"

- 4 - “Very good”
- 3 - “Good”
- 2 - “Fair”
- 1 - “Poor”

Main predictor: `watched_tv`

Our main predictor variable of interest is how often the participant watches TV. The interpretation ranges from 1 (No TV) to 6 (10+ hours per week). The values are described below.

- 1 - “None”
- 2 - “Less than 1 hour a week”
- 3 - “1 to 3 hours a week”
- 4 - “4 to 6 hours a week”
- 5 - “7 to 9 hours a week”
- 6 - “10 hours or more a week”

Other predictor: `felt_happy`

How often the respondent felt happy in the last month.
This variable has been recoded such that it ranges from:

- 4 - “All of the time”
- 3 - “Most of the time”
- 2 - “Some of the time”
- 1 - “None of the time”

Other predictor: `depressed`

How often the respondent was depressed in the last month.
This variable has been recoded such that it ranges from:

- 4 - “All of the time”
- 3 - “Most of the time”
- 2 - “Some of the time”
- 1 - “None of the time”

Load and explore data

```
data <- read_csv("data/panel/nlsy97_selected.csv")  %>%
  filter(year != 2008) # there were not enough observations in this wave

# Number of participants
data %>%
  pull(id) %>%
```

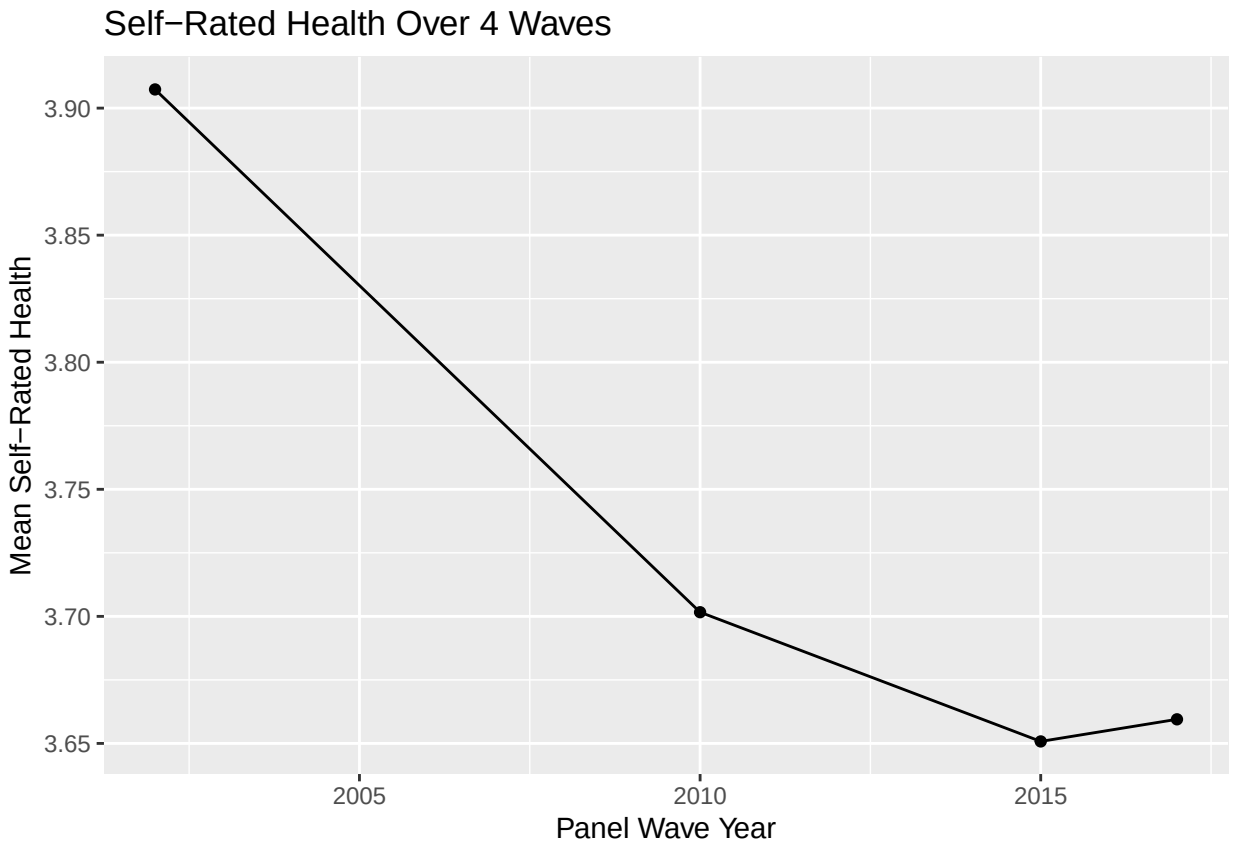
```
  unique %>%
  length()
```

```
## [1] 8522
```

```
# Number of waves
data %>%
  pull(year) %>%
  unique()
```

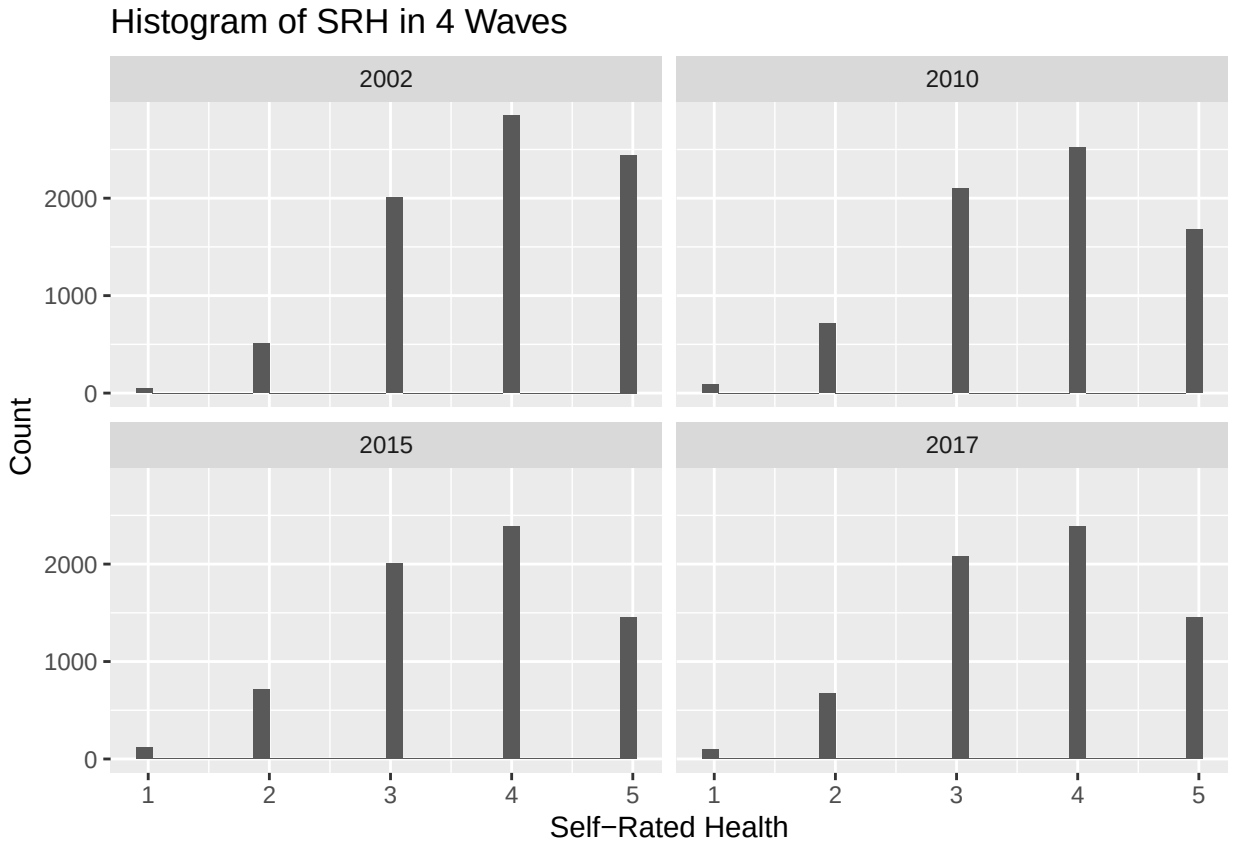
```
## [1] 2002 2010 2015 2017
```

```
# SRH over time
data %>%
  group_by(year) %>%
  summarize(mean_srh = mean(health)) %>%
  ggplot(aes(x = year, y = mean_srh)) +
  geom_line() +
  geom_point() +
  labs(title = "Self-Rated Health Over 4 Waves",
       x = "Panel Wave Year",
       y = "Mean Self-Rated Health")
```




```
# SRH in each wave
data %>%
  group_by(year) %>%
  ggplot(aes(x = health)) +
  geom_histogram() +
  facet_wrap(~ year) +
```

```
  labs(title = "Histogram of SRH in 4 Waves",
       x = "Self-Rated Health",
       y = "Count")
```



There are 8522 participants who were surveyed in 4 waves of interest: years 2002, 2010, 2015, 2017. Self-rated health overall seems to have decreased over time, although not at a constant rate.

a) OLS

1. (a) Run an OLS regression, including at least one independent variable and a time variable (as dummies). Explain how you think your independent variable relates to your dependent variable. Interpret your results. Did you find what you expected to find?

```
library(plm)

# OLS regression
summary(lm(health ~ watched_tv + as.factor(year), data))
```

```
##
## Call:
## lm(formula = health ~ watched_tv + as.factor(year), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0175 -0.6752  0.2521  0.6157  1.6203
##
## Coefficients:
```

```
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           4.09016    0.01682  243.12   <2e-16 ***
## watched_tv           -0.07271    0.00510  -14.26   <2e-16 ***
## as.factor(year)2010 -0.21526    0.01582  -13.61   <2e-16 ***
## as.factor(year)2015 -0.27425    0.01612  -17.02   <2e-16 ***
## as.factor(year)2017 -0.26960    0.01613  -16.71   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9654 on 28316 degrees of freedom
## Multiple R-squared:  0.01918,    Adjusted R-squared:  0.01904
## F-statistic: 138.4 on 4 and 28316 DF,  p-value: < 2.2e-16
```

```
eq.pool_tv <- plm(health ~ watched_tv + as.factor(year), # model formula
               index = c("id", "year"), # id & time variables
               model = "pooling",
               data = data) ## this is equivalent to above OLS ##

summary(eq.pool_tv)
```

```
## Pooling Model
##
## Call:
## plm(formula = health ~ watched_tv + as.factor(year), data = data,
##     model = "pooling", index = c("id", "year"))
##
## Unbalanced Panel: n = 8522, T = 1-4, N = 28321
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -3.01745 -0.67515  0.25215  0.61568  1.62032
##
## Coefficients:
##                       Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)          4.0901589  0.0168240 243.115 < 2.2e-16 ***
## watched_tv          -0.0727059  0.0050998 -14.257 < 2.2e-16 ***
## as.factor(year)2010 -0.2152578  0.0158203 -13.607 < 2.2e-16 ***
## as.factor(year)2015 -0.2742455  0.0161173 -17.016 < 2.2e-16 ***
## as.factor(year)2017 -0.2696017  0.0161339 -16.710 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    26905
## Residual Sum of Squares: 26389
## R-Squared:      0.019182
## Adj. R-Squared: 0.019043
## F-statistic: 138.442 on 4 and 28316 DF, p-value: < 2.22e-16
```

According to the OLS regression, *watching more TV is associated with a very slight, but statistically significant decrease in self-rated health.* Additionally, every wave has a statistically significantly lower average self-rated health score than the reference wave (2002).

Specifically,for each additional unit of time spent watching TV, the self-rated health score decreases by approximately 0.0727 points, and the 2nd, 3rd, and 4th waves have a decreased self-rated health (controling for TV watching) of -0.215, -0.274, and -0.269. respectively. All of these decreases are statistically significant to a significance level of 0.05. The R-squared was 0.019, indicating that the model accounted for 1.9% of the

variability in self-related health scores.

This is what I expected to find because watching TV can lead to sedentary behavior and other mental results that may decrease health, and thus self-rated health. I also expected the SRH to decrease in later waves because the participants are older, so their SRH should decline.

b) Fixed Effects

(b) Then run a fixed effect model version of that OLS model. Interpret your results. Did you find what you expected to find? Why? Why not?

```
eq.fe_tv <- plm(health ~ watched_tv + as.factor(year), # model formula
             index = c("id", "year"), # id & time variables
             model = "within",
             data = data) ## fixed effects ##

summary(eq.fe_tv)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = health ~ watched_tv + as.factor(year), data = data,
##     model = "within", index = c("id", "year"))
##
## Unbalanced Panel: n = 8522, T = 1-4, N = 28321
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -3.179599 -0.411371  0.032358  0.332208  2.579249
##
## Coefficients:
##                         Estimate Std. Error  t-value Pr(>|t|)
## watched_tv          -0.0117978  0.0055235  -2.1359   0.0327 *
## as.factor(year)2010 -0.2031087  0.0122910 -16.5250   <2e-16 ***
## as.factor(year)2015 -0.2588579  0.0126268 -20.5007   <2e-16 ***
## as.factor(year)2017 -0.2446321  0.0126634 -19.3180   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10616
## Residual Sum of Squares: 10327
## R-Squared:      0.027237
## Adj. R-Squared: -0.3917
## F-statistic: 138.565 on 4 and 19795 DF, p-value: < 2.22e-16
```

According to the fixed effects model, *watching more TV is still associated with a very slight, but still statistically significant decrease in self-rated health.* Additionally, every wave has still a statistically significantly lower average self-rated health score than the reference wave (2002). The *magnitude of all of the coefficients has decreased.* The adjusted R2 is now negative.

This is as expected for the reasons described above, and the decrease in magnitude of the effect is also expected, because that is usually the result when comparing a pooled to a fixed effects model. The adjusted R2 is now negative, but that sometimes what happens in fixed effects models.

c) Additional predictor

(c) Then include an additional predictor in your fixed effects model that you think might account for the initial relationship you found between your X and your Y. What effect does that new independent variable have in your new regression?

```
eq.fe_tv_depressed <- plm(health ~ watched_tv + depressed + as.factor(year), # model formula
             index = c("id", "year"), # id & time variables
             model = "within",
             data = data) ## fixed effects ##
summary(eq.fe_tv_depressed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = health ~ watched_tv + depressed + as.factor(year),
##     data = data, model = "within", index = c("id", "year"))
##
## Unbalanced Panel: n = 8522, T = 1-4, N = 28321
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -3.199011 -0.370709  0.022399  0.342587  2.721027
##
## Coefficients:
##                        Estimate Std. Error  t-value Pr(>|t|)
## watched_tv           -0.0108622  0.0055029  -1.9739  0.04841 *
## depressed            -0.1295137  0.0104685 -12.3717  < 2e-16 ***
## as.factor(year)2010 -0.2199306  0.0123193 -17.8525  < 2e-16 ***
## as.factor(year)2015 -0.2807363  0.0127022 -22.1013  < 2e-16 ***
## as.factor(year)2017 -0.2845163  0.0130205 -21.8515  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10616
## Residual Sum of Squares: 10248
## R-Squared:      0.034702
## Adj. R-Squared: -0.38109
## F-statistic: 142.316 on 5 and 19794 DF, p-value: < 2.22e-16
```

```
eq.fe_tv_happy <- plm(health ~ watched_tv + felt_happy + as.factor(year), # model formula
             index = c("id", "year"), # id & time variables
             model = "within",
             data = data) ## fixed effects ##

summary(eq.fe_tv_happy)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = health ~ watched_tv + felt_happy + as.factor(year),
##     data = data, model = "within", index = c("id", "year"))
##
## Unbalanced Panel: n = 8522, T = 1-4, N = 28321
##
```

```
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.9647059 -0.3576620  0.0084701  0.3598396  2.6141212
##
## Coefficients:
##                         Estimate Std. Error  t-value Pr(>|t|)
## watched_tv           -0.0083380  0.0054739  -1.5232   0.1277
## felt_happy            0.1774265  0.0090731  19.5553   <2e-16 ***
## as.factor(year)2010 -0.2022548  0.0121744 -16.6132   <2e-16 ***
## as.factor(year)2015 -0.2633805  0.0125090 -21.0553   <2e-16 ***
## as.factor(year)2017 -0.2719827  0.0126209 -21.5502   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10616
## Residual Sum of Squares: 10131
## R-Squared:      0.045675
## Adj. R-Squared: -0.36539
## F-statistic: 189.47 on 5 and 19794 DF, p-value: < 2.22e-16
```

Two additional variables were added to explore the relationship between watching TV and self rated health.

Depression: Adding the variable of how often the respondent has felt depressed in the last month resulted in a negative, statistically significant association (coefficient -0.1295) between frequency of feeling depressed and self rated health. This makes sense, because depression is known to affect health in a negative way. Additionally, the coefficient for watching TV only *very* slightly decreases (-0.01179 when only watched_tv is included in the model, versus -0.01086 when depression is added to the model as well), which means *feeling depressed does not explain the negative effect of watching TV on self-rated health.* Both coefficients are statistically significant, which means that both watching TV and feeling depressed are associated with lower self-rated health, even when controling for each other.

Felt Happy: Although depression proved to be unrelated to watching TV's association with self rated health, it may still be possible that watching TV could make one less happy (or vice versa), and that could be a mechanism through which watching TV could result in lower self-rated health. To explore this, the model was run with watched_tv and felt_happy. The results from the model were that felt_happy was rather strongly associated (coefficient of 0.177) with increased self rated health, which is to be expected, as happier people report better health. Interestingly, the coefficient on watched_tv decreases in magnitude and is no longer statistically significantly associated with self-rated health. This means that *watching TV was no longer statistically significantly associated with self-rated health when controlling for how often the participant felt happy.* This suggests that *the way in which watching TV has an effect on self-rated health may be through its relationship with frequency of feeling happy.*

As a quick sanity check, below I ran a fixed effects model on happiness using watched_tv as the predictor variable and explore the results. It is indeed true that watching TV is associated in lower frequency of being happy.

```
eq.fe_tv_on_happy <- plm(felt_happy ~ watched_tv + as.factor(year), # model formula
                  index = c("id", "year"), # id & time variables
                  model = "within",
                  data = data) ## fixed effects ##
summary(eq.fe_tv_on_happy)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = felt_happy ~ watched_tv + as.factor(year), data = data,
##     model = "within", index = c("id", "year"))
```

```
##
## Unbalanced Panel: n = 8522, T = 1-4, N = 28321
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -2.222033 -0.236657  0.019332  0.287717  2.293707
##
## Coefficients:
##                         Estimate Std. Error t-value  Pr(>|t|)
## watched_tv          -0.0194999  0.0042859 -4.5498 5.401e-06 ***
## as.factor(year)2010 -0.0048123  0.0095370 -0.5046  0.613850
## as.factor(year)2015  0.0254898  0.0097975  2.6016  0.009285 **
## as.factor(year)2017  0.1541518  0.0098260 15.6882 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    6338.1
## Residual Sum of Squares: 6217.5
## R-Squared:      0.01903
## Adj. R-Squared: -0.40344
## F-statistic: 96.0024 on 4 and 19795 DF, p-value: < 2.22e-16
```

d) Random Effects

(d) Then run a random effects model equivalent to your fixed effects model in step (b). Interpret the results.

```
eq.re_tv <- plm(health ~ watched_tv + as.factor(year), # model formula
             index = c("id", "year"), # id & time variables
             model = "random",
             data = data) ## random effects ##

summary(eq.re_tv)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = health ~ watched_tv + as.factor(year), data = data,
##     model = "random", index = c("id", "year"))
##
## Unbalanced Panel: n = 8522, T = 1-4, N = 28321
##
## Effects:
##                  var std.dev share
## idiosyncratic 0.5217  0.7223 0.562
## individual    0.4061  0.6372 0.438
## theta:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2501  0.4524  0.5069  0.4822  0.5069  0.5069
##
## Residuals:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.06307 -0.51673  0.06367 -0.00125  0.56580  1.96290
##
```

```
## Coefficients:
##                          Estimate Std. Error  z-value  Pr(>|z|)
## (Intercept)             4.0061048  0.0162329 246.7898 < 2.2e-16 ***
## watched_tv             -0.0380084  0.0048279  -7.8727 3.471e-15 ***
## as.factor(year)2010 -0.2093777  0.0120850 -17.3255 < 2.2e-16 ***
## as.factor(year)2015 -0.2662973  0.0123793 -21.5115 < 2.2e-16 ***
## as.factor(year)2017 -0.2559101  0.0124042 -20.6309 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    16577
## Residual Sum of Squares: 14809
## R-Squared:      0.1068
## Adj. R-Squared: 0.10668
## Chisq: 648.065 on 4 DF, p-value: < 2.22e-16
```

The coefficient for the random effects model is statistically significant, with a value of -0.038, which is between the OLS and fixed effects model, as expected due to the differences in the model assumptions. As before, it suggests that watching TV is slightly associated with a lower self-rated health. The idiosyncratic (within-individual) effect accounts for 52% of the variance and the individual effect (between individual) accounts for 40% of the variance. This indicates a large amount of between individual variability.

e) Hausman test

(e) Run a Hausman test to compare your fixed effects and your random effects models. What do you conclude?

```
library(stargazer)

phtest(eq.fe_tv, eq.re_tv) ## Hausman test comparing RE and FE ##
```

```
##
##  Hausman Test
##
## data:  health ~ watched_tv + as.factor(year)
## chisq = 97.843, df = 4, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

The Hausman test compares the FE and RE models to see if there is a significant difference in the coefficients. The very low p-value indicates that there is a difference in the coefficients, which suggests that the random effects model is inconsistent. This means that the individual effects are correlated with the regressors. Hence, the fixed effects model is preferred.

Different Models Summary

```
stargazer(eq.pool_tv, eq.fe_tv, eq.re_tv,
          title="Pooled, Fixed Effects, and Random Effects",
          align=TRUE,
          dep.var.labels=c("Equalize"),
          no.space=TRUE,
          column.labels=c("Pooled",
                          "Fixed Effects", "Random Effects"),
          dep.var.caption="",
```

```
          model.numbers=FALSE,
          type = "text", omit = "Constant")
```

```
##
## Pooled, Fixed Effects, and Random Effects
## ================================================================================================
##                                                  Equalize
##                                Pooled               Fixed Effects        Random Effects
## ------------------------------------------------------------------------------------------------
## watched_tv                   -0.073***                -0.012**             -0.038***
##                               (0.005)                  (0.006)              (0.005)
## as.factor(year)2010          -0.215***                -0.203***            -0.209***
##                               (0.016)                  (0.012)              (0.012)
## as.factor(year)2015          -0.274***                -0.259***            -0.266***
##                               (0.016)                  (0.013)              (0.012)
## as.factor(year)2017          -0.270***                -0.245***            -0.256***
##                               (0.016)                  (0.013)              (0.012)
## ------------------------------------------------------------------------------------------------
## Observations                   28,321                   28,321               28,321
## R2                             0.019                    0.027                0.107
## Adjusted R2                    0.019                    -0.392               0.107
## F Statistic          138.442*** (df = 4; 28316) 138.565*** (df = 4; 19795)   648.065***
## ================================================================================================
## Note:                                                          *p<0.1; **p<0.05; ***p<0.01
```

When modeling self-rated health as the outcome variable and watching TV as the predictor variable, all three models result in a small, statistically significant association. As expected the largest coefficient is found in the OLS, followed by RE, followed by FE. This pattern also holds for the coefficients of each wave compared to the reference wave.

Additional Analyses

FE for different combinations of predictor variables

```
eq.fe_happy_depressed <- plm(health ~ felt_happy + depressed + as.factor(year), # model formula
             index = c("id", "year"), # id & time variables
             model = "within",
             data = data) ## fixed effects ##
summary(eq.fe_happy_depressed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = health ~ felt_happy + depressed + as.factor(year),
##     data = data, model = "within", index = c("id", "year"))
##
## Unbalanced Panel: n = 8522, T = 1-4, N = 28321
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.9957883 -0.3555169  0.0038791  0.3617140  2.6878502
##
## Coefficients:
##                       Estimate Std. Error  t-value  Pr(>|t|)
```

```
## felt_happy           0.1579705  0.0095335  16.5701 < 2.2e-16 ***
## depressed           -0.0732872  0.0109418  -6.6979 2.171e-11 ***
## as.factor(year)2010 -0.2107727  0.0122242 -17.2423 < 2.2e-16 ***
## as.factor(year)2015 -0.2732668  0.0125516 -21.7714 < 2.2e-16 ***
## as.factor(year)2017 -0.2889663  0.0128311 -22.5208 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10616
## Residual Sum of Squares: 10109
## R-Squared:      0.047721
## Adj. R-Squared: -0.36246
## F-statistic: 198.385 on 5 and 19794 DF, p-value: < 2.22e-16
```

```
eq.fe_tv_happy_depressed <- plm(health ~ watched_tv + felt_happy + depressed + as.factor(year), # model
             index = c("id", "year"), # id & time variables
             model = "within",
             data = data) ## fixed effects ##
summary(eq.fe_tv_happy_depressed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = health ~ watched_tv + felt_happy + depressed +
##     as.factor(year), data = data, model = "within", index = c("id",
##     "year"))
##
## Unbalanced Panel: n = 8522, T = 1-4, N = 28321
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.9997566 -0.3552373  0.0051802  0.3593542  2.6903722
##
## Coefficients:
##                       Estimate Std. Error  t-value  Pr(>|t|)
## watched_tv          -0.0081967  0.0054679  -1.4991    0.1339
## felt_happy           0.1575486  0.0095373  16.5191 < 2.2e-16 ***
## depressed           -0.0732239  0.0109415  -6.6923 2.256e-11 ***
## as.factor(year)2010 -0.2118612  0.0122453 -17.3014 < 2.2e-16 ***
## as.factor(year)2015 -0.2752433  0.0126203 -21.8096 < 2.2e-16 ***
## as.factor(year)2017 -0.2914680  0.0129388 -22.5267 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10616
## Residual Sum of Squares: 10108
## R-Squared:      0.047829
## Adj. R-Squared: -0.36237
## F-statistic: 165.706 on 6 and 19793 DF, p-value: < 2.22e-16
```

Comparison

```
stargazer(eq.fe_tv_happy, eq.fe_tv_depressed, eq.fe_happy_depressed, #eq.fe_tv_happy_depressed,
          title="FE Results",
```

```
          align=TRUE,
          dep.var.labels=c("Equalize"),
       #   covariate.labels=c("Bills","2010"),
          no.space=TRUE,
          column.labels=c("FE TV + Happy",
                          "FE TV + Depressed",
                          "FE Depressed + Happy" #,
                         # "FE TV + Depressed + Happy"
                          ),
          dep.var.caption="",
          model.numbers=FALSE,
          type = "text", omit = "Constant")
```

```
##
## FE Results
## ================================================================================
##                                                Equalize
##                              FE TV + Happy FE TV + Depressed FE Depressed + Happy
## --------------------------------------------------------------------------------
## watched_tv                      -0.008         -0.011**
##                                 (0.005)         (0.006)
## felt_happy                     0.177***                          0.158***
##                                 (0.009)                          (0.010)
## depressed                                     -0.130***         -0.073***
##                                               (0.010)           (0.011)
## as.factor(year)2010            -0.202***      -0.220***         -0.211***
##                                 (0.012)         (0.012)           (0.012)
## as.factor(year)2015            -0.263***      -0.281***         -0.273***
##                                 (0.013)         (0.013)           (0.013)
## as.factor(year)2017            -0.272***      -0.285***         -0.289***
##                                 (0.013)         (0.013)           (0.013)
## --------------------------------------------------------------------------------
## Observations                    28,321          28,321            28,321
## R2                               0.046           0.035             0.048
## Adjusted R2                     -0.365          -0.381            -0.362
## F Statistic (df = 5; 19794)  189.470***      142.316***        198.385***
## ================================================================================
## Note:                                                *p<0.1; **p<0.05; ***p<0.01
```

Overall Summary

The goal of this analysis was to explore the relationship between watching TV and self-rated health. A normal OLS ("pooled"), fixed effects, and random effects models were run on this relationship with wave as a dummy variable. In all three scenarios, watching TV has a small, negative, statistically significant effect on self-rated health, and all three waves have a statistically significant negative association as well. This suggests that watching TV is associated with a slightly lower self-rated health, and that self-rated health decreassed in the later waves when compared to the first wave. A Hausman test was performed to explore the appropriateness of the random effects model, and found to be statistially significant, hence, the random effects model is likely bias and the fixed effects model is preferred.

When exploring the effect of frequency of depression and feelings of happiness on self-rated health, we see that depression has a negative association with self-rated health, but it is not related to the relationship of watching TV with self-rated health (that is, controlling for feeling depressed does not change the association between watching TV and SRH very much). However, when controlling for feeling happy, the effect of

watching TV and lower self-rated health becomes insignificant, suggesting that watching TV is associated with lower self rated health through its relationship with feeling happy.

These findings suggest that policies aimed at improving health, as operationalized by self-rated health, should focus on feelings of happiness in a holistic manner more than solely decreasing hours of watching TV.