

A graphical exploration of fixed effects panel models

Matthew Alampay Davis

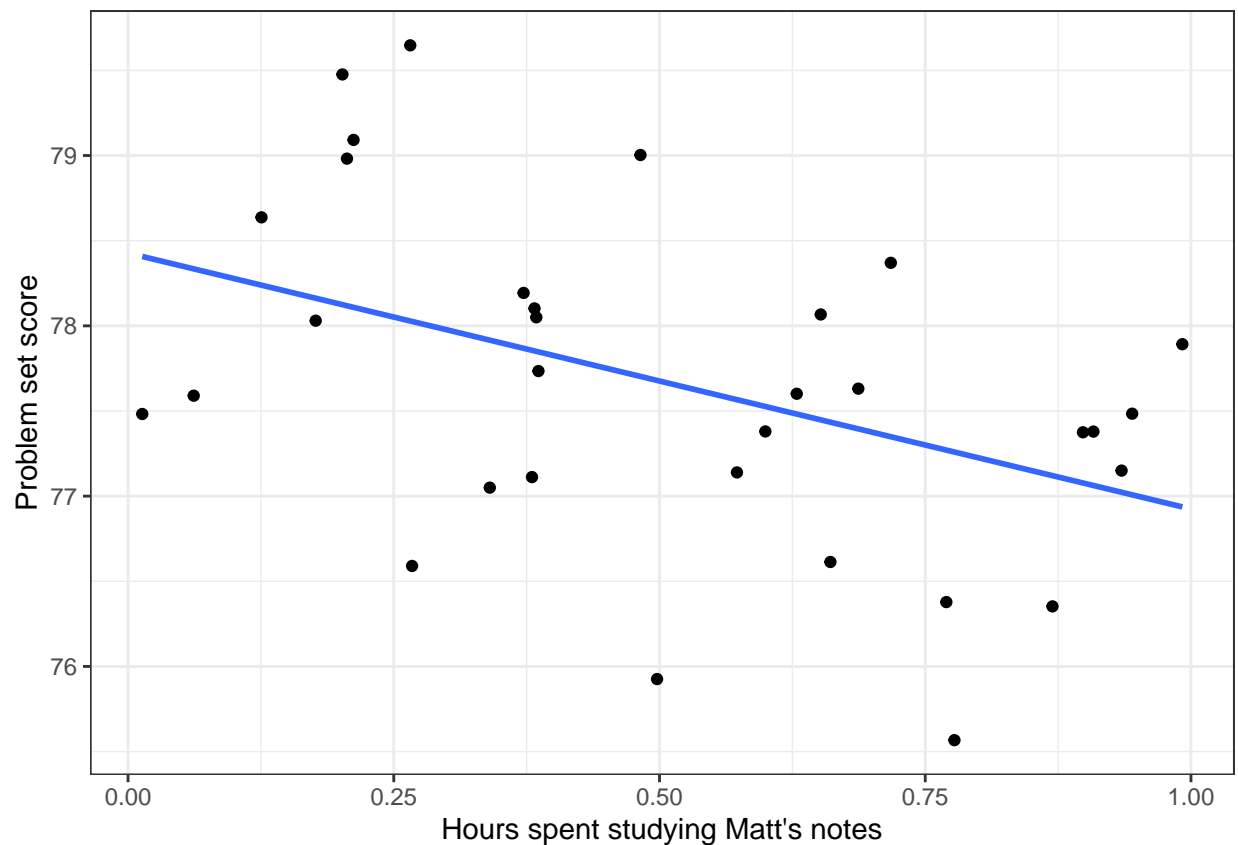
June 27, 2023

This note is just to demonstrate what fixed effects panel models do and how they differ from pooled OLS estimation. If you find it just adds to your confusion, then feel free to disregard but I think students in the past have found it helpful. The commands used here are also not important, the discussion should be pretty self-contained and follow directly from the various graphs.

Suppose I have 32 datapoints. This (fake) dataset has two variables: $Notes_i$ is the number of hours a student indexed by i spent studying my recitation notes and $Score_i$ is the score they received in a problem set:

```
ggplot(test, aes(x = study, y = score)) +  
  theme_bw() +  
  geom_smooth(method = 'lm', se = F) +  
  ylab('Problem set score') + xlab("Hours spent studying Matt's notes") +  
  geom_point()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



This plot seems to suggest that studying my notes actually makes students perform worse. Indeed, when we run a basic regression, we get a negative effect significant at the 5% level:

```
feols(score ~ study, test) %>% etable
```

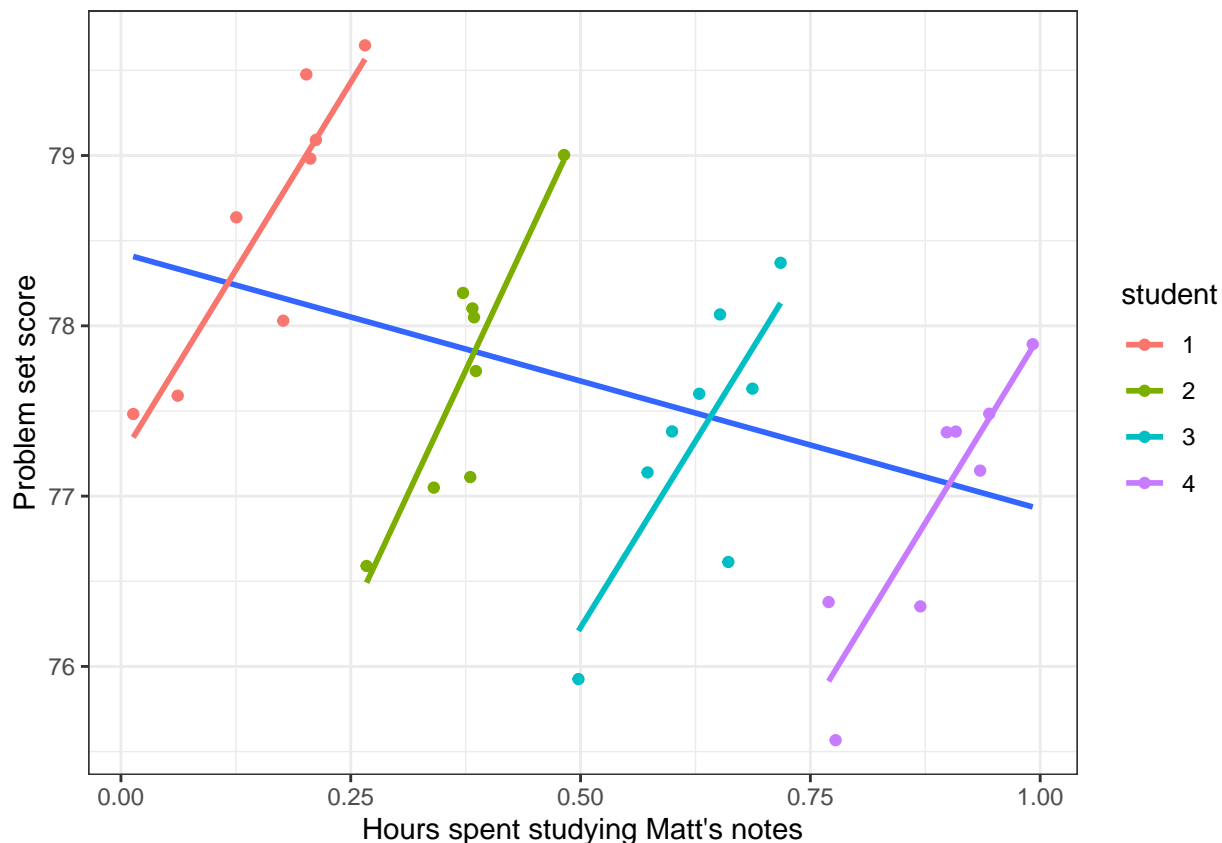
```
##
## Dependent Var.:      score
##
## Constant      78.43*** (0.3283)
## study         -1.503* (0.5623)
## -----
## S.E. type      IID
## Observations   32
## R2             0.19231
## Adj. R2        0.16539
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This regression suggests that for every additional hour a student spends studying my notes, their expected problem set score falls by 1.5 points. Obviously, this doesn't make sense given that I'm such a good TA. How did this happen?

It turns out that our data isn't cross-sectional after all. The 32 observations actually correspond to just four students and their performances on the eight problem sets. Thus we can say the dataset has a panel structure where we have $T = 8$ "periods" corresponding to the eight problem sets and $N = 4$ students as different "entities" or "units". If we color the points by student, we get the following:

```
ggplot() +
  theme_bw() +
  geom_smooth(data = test, aes(x = study, y = score), method = 'lm', se = F) +
  ylab('Problem set score') + xlab("Hours spent studying Matt's notes") +
  geom_point(data = test, aes(x = study, y = score, color = student)) +
  geom_smooth(data = test, aes(x = study, y = score, color = student), se = F, method = 'lm')

## 'geom_smooth()' using formula = 'y ~ x'
## 'geom_smooth()' using formula = 'y ~ x'
```



Now the data is starting to make some sense: each student individually is seeing positive performance gains from reading my notes—this is captured by the positive student-specific slopes—but if we regarded all the points as independently drawn the way we do with pooled OLS, we’d infer a negative effect (as seen in the blue line of best fit) because we’d be actually be running a pooled OLS regression of data with a panel structure. This gives misleading estimates if there are student-specific factors that are unaccounted for that are causing omitted variable bias. Fixed effects panel models account for these even without having exhaustive data on those potentially relevant variables. Formally, these individual-specific omitted factors are contained in the α_i term here:

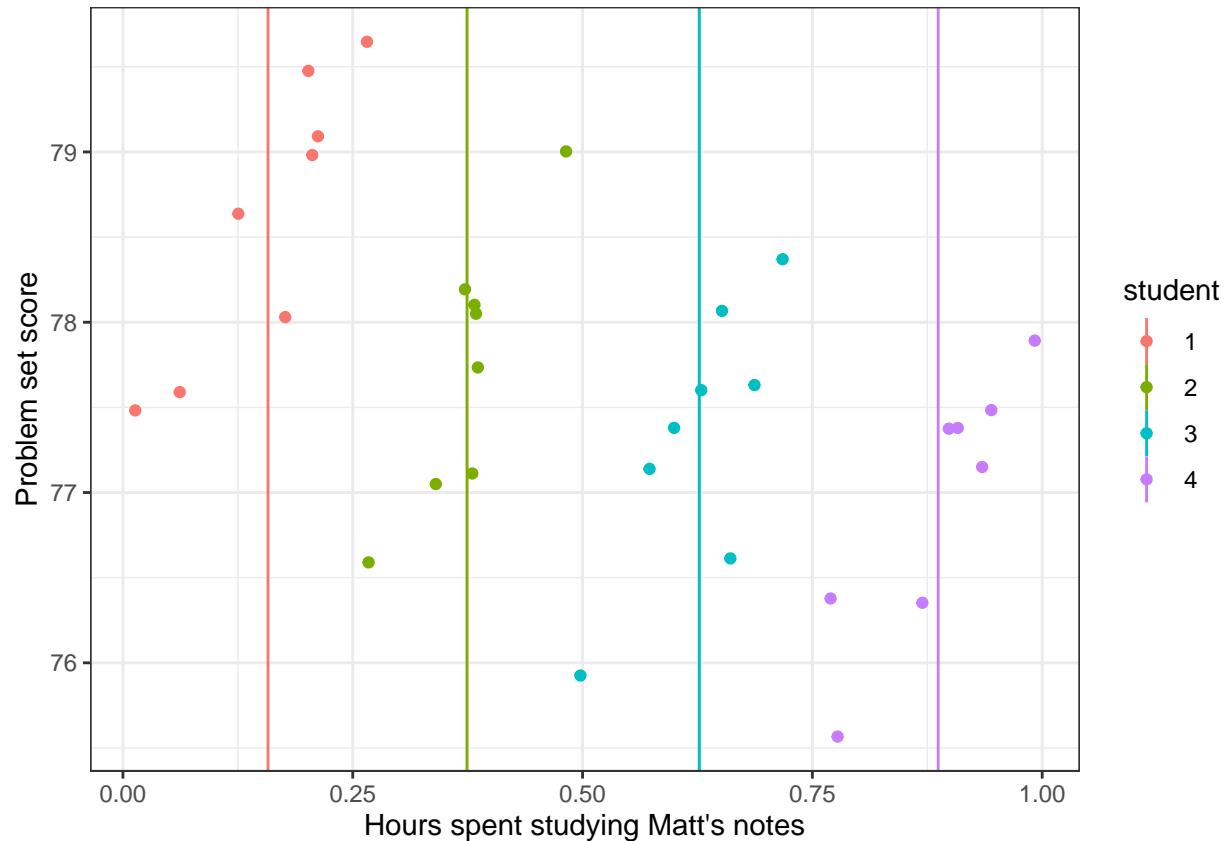
$$Score_{it} = \alpha_i + \beta Notes_{it} + u_{it}$$

For example, α_4 might absorb the negative effect of Student 4 having a particularly annoying roommate who prevents Student 4 from focusing on his work. Even if different students experience similar benefits to studying my excellent notes (meaning that reading my notes have the same per-hour positive effect β on their problem set score), systematic differences in student-level characteristics could give each linear relationship a different intercept α_i for each student i .

Fixed effects models allow us to account for these unobservable characteristics simply by entity/unit or time demeaning. The lecture and textbook guides us through the algebraic justification for this, but basically the individual effects are differenced away in the demeaning process so that the identifying variation comes from deviations from entity means: this is why it’s often called the “within” estimator. We can visualize what demeaning variables accomplishes graphically here for intuition.

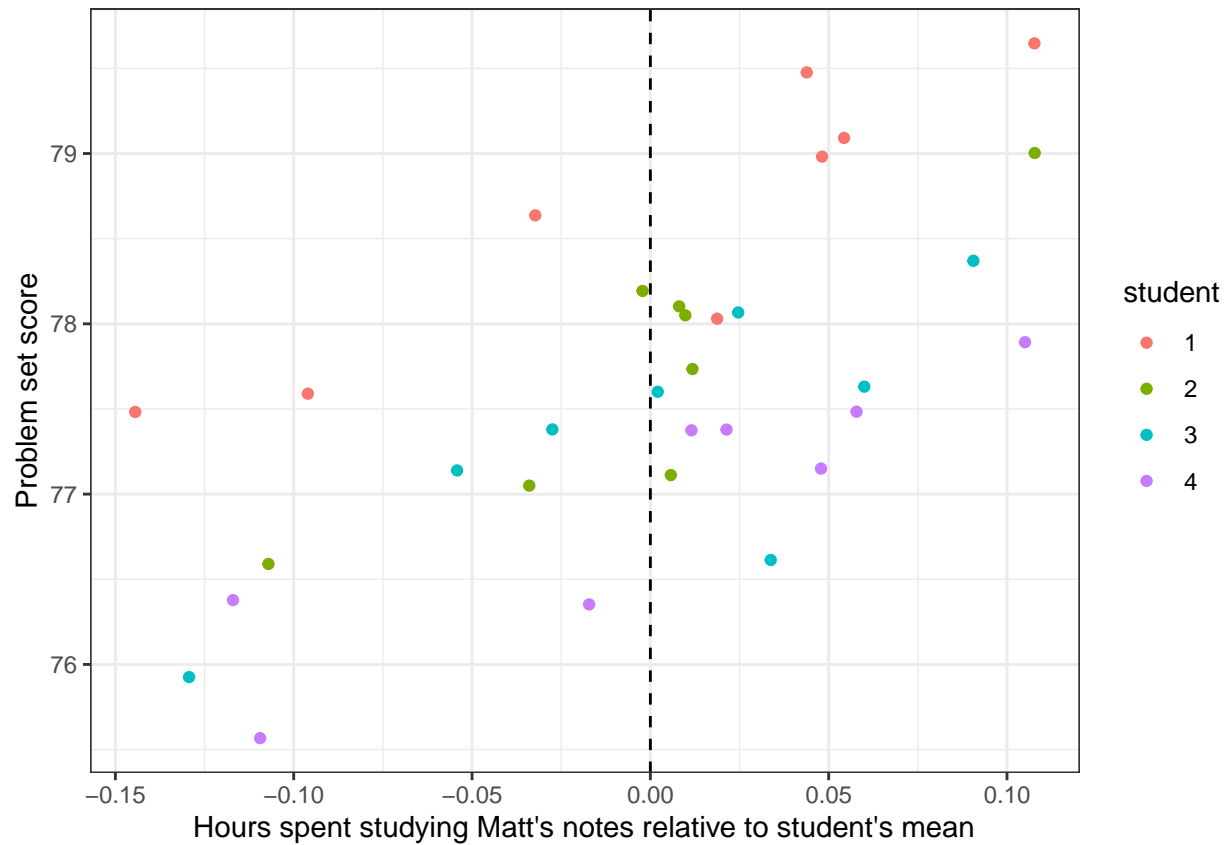
First, let us mark the mean time spent studying my *Notes* for each student:

```
ggplot(test, aes(x = study, y = score, color = student)) +
  theme_bw() +
  ylab('Problem set score') + xlab("Hours spent studying Matt's notes") +
  geom_point() + geom_vline(aes(xintercept = mean.study, color = student))
```



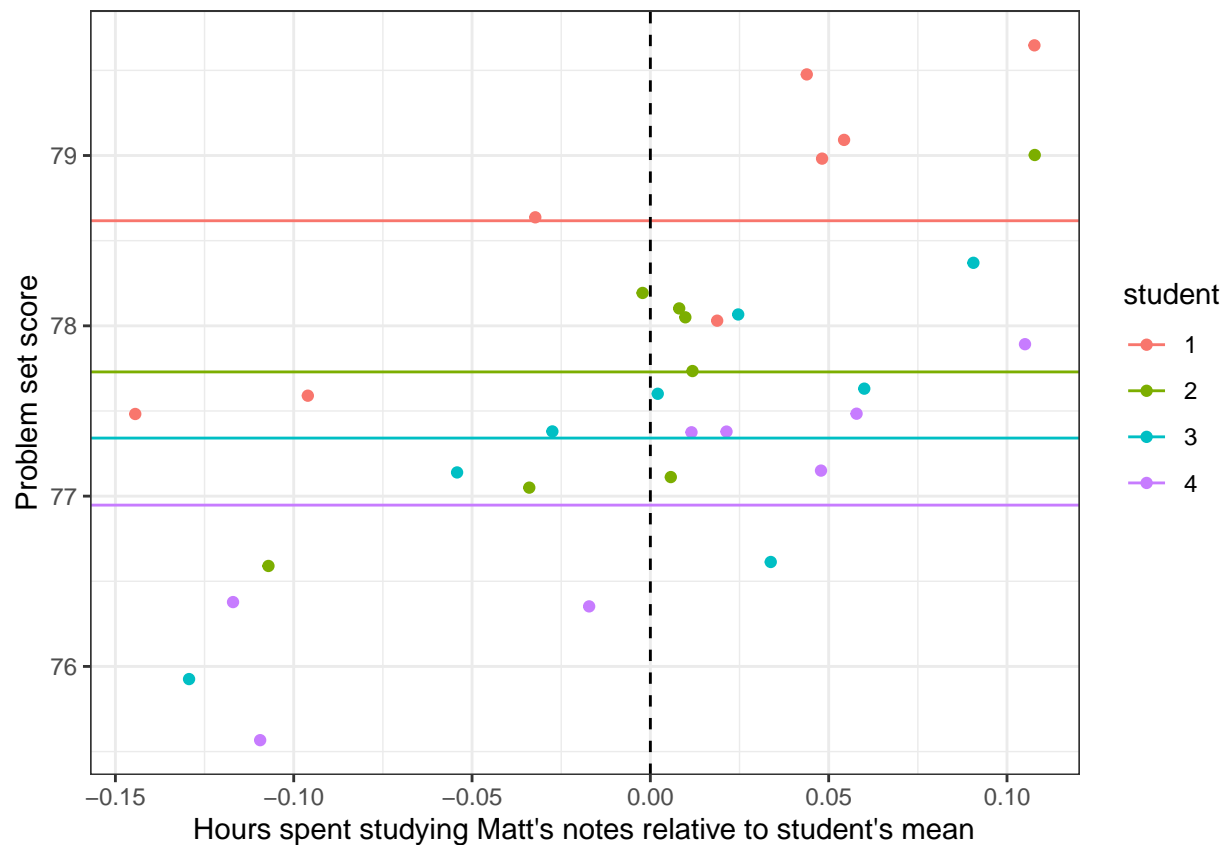
Then, for each observation, subtract their mean hours from their student-specific mean hours so that each student's hours are mean zero (the vertical line in black below):

```
ggplot(test, aes(x = study-mean.study, y = score, color = student)) +
  theme_bw() +
  ylab('Problem set score') + xlab("Hours spent studying Matt's notes relative to student's mean") +
  geom_point() +
  geom_vline(xintercept = 0, color = 'black', lty = 2)
```



Now for the dependent variable. Let's mark the mean problem set score for each student:

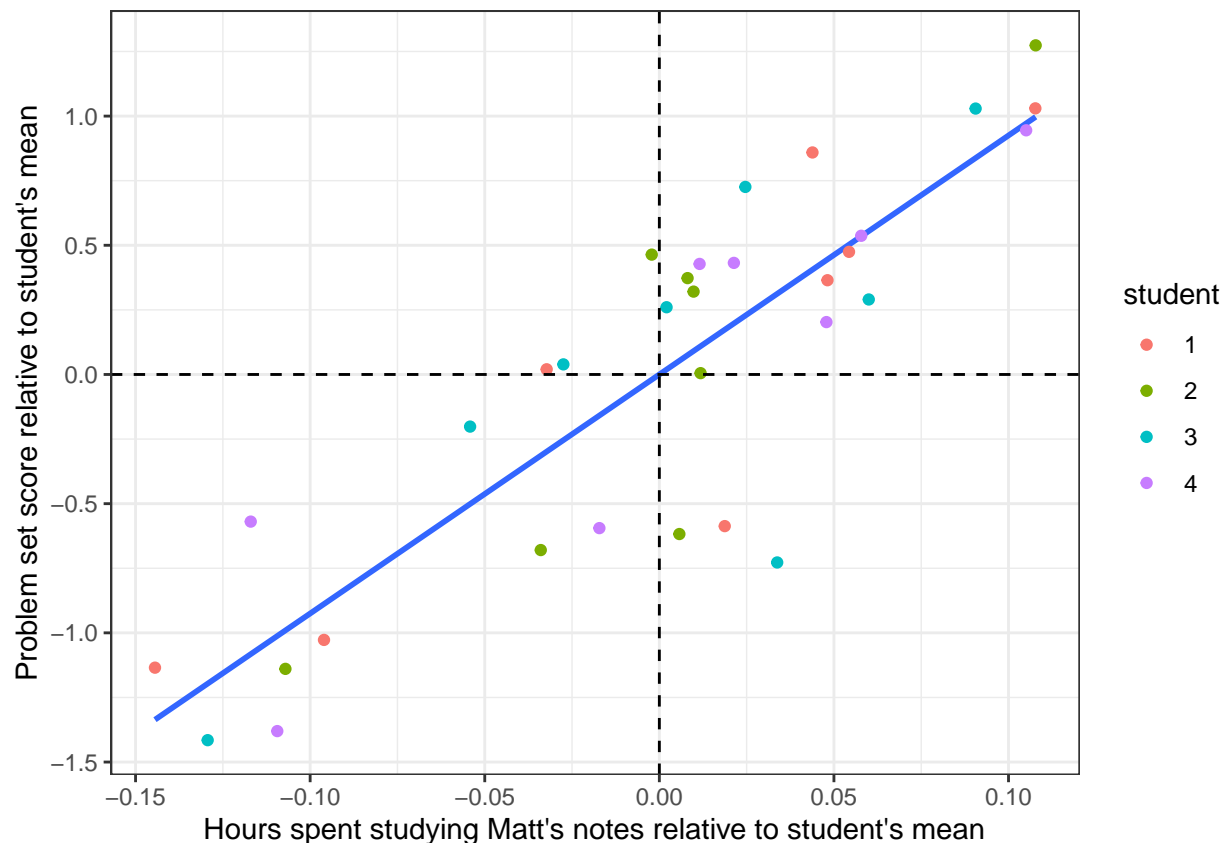
```
ggplot(test, aes(x = study-mean.study, y = score, color = student)) +
  theme_bw() +
  ylab('Problem set score') + xlab("Hours spent studying Matt's notes relative to student's mean") +
  geom_point() + geom_hline(aes(yintercept = mean.score, color = student)) +
  geom_vline(xintercept = 0, color = 'black', lty = 2)
```



Then as before, for each observation, subtract their student-specific mean score so the vertical axis is now relative to their student-specific means. Now we have a horizontal line whose y-intercept is zero corresponding to the student's mean deviation from their mean score (which by definition is 0)

```
ggplot(test, aes(x = study-mean.study, y = score-mean.score)) +
  theme_bw() +
  ylab("Problem set score relative to student's mean") + xlab("Hours spent studying Matt's notes relative to student's mean") +
  geom_smooth(method = 'lm', se = F) +
  geom_point(aes(color = student)) +
  geom_vline(xintercept = 0, color = 'black', lty = 2) +
  geom_hline(yintercept = 0, color = 'black', lty = 2)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



Aha. Now when we run the OLS regression on this demeaned data, we can get an estimate of the effect of studying my notes that makes sense:

```
test %<>% mutate(score.demeaned = score - mean.score,
                  study.demeaned = study - mean.study)
feols(score.demeaned ~ study.demeaned, data = test, cluster = 'student') %>% etable
```

```
##
## Dependent Var.:      score.demeaned
##
## Constant            -8.88e-16 (1.73e-15)
## study.demeaned      9.250*** (0.5066)
## -----
## S.E.: Clustered      by: student
## Observations         32
## R2                   0.74110
## Adj. R2              0.73247
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This gives us the same result as using the original data with unit (here, student) fixed effects:

```
fe1 <- feols(score ~ study | student, data = test, cluster = 'student')
etable(fe1)
```

```
##                                fe1
```

```
## Dependent Var.:          score
##
## study          9.250*** (0.5066)
## Fixed-Effects:  -----
## student                          Yes
## -----
## S.E.: Clustered      by: student
## Observations          32
## R2                    0.84749
## Within R2             0.74110
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

These estimates suggest that for every hour spent studying my notes, a student's expected problem set score increases by 9.25 points and that this estimate is very significant.

We can also extract the estimated values of the student fixed effects:

```
fixef(fe1)
```

```
## $student
##      1      2      3      4
## 77.15713 74.26691 71.54054 68.74349
##
## attr("class")
## [1] "fixest.fixef" "list"
## attr("exponential")
## [1] FALSE
```

The fixed effects suggest that Student 1 is the highest performing student with a mean expected problem set score of 77.2. Students 2 to 4 have decreasing mean scores, corresponding to systematic differences in their implied intercepts. This is also what led to our initial negative relationship: the students with the lowest average scores spent the most time studying. But this does not mean that studying more decreased their scores. Instead, the fixed effects model reveals that once we account for differences in average time spent studying my notes, additional time spent studying my notes is associated with better scores for all students.

Note that all the above analysis describes entity fixed effects. We could perform a similar exercise to capture time fixed effects μ_t if we suspect the presence of time-specific (here, problem-set specific) characteristics of each observation. For example, a particularly difficult problem set would cause trouble for all students relative to the average problem set. If a regression with these problem-set fixed effects still produces a positive coefficient on study, this would still be consistent with the idea that my notes help increase students' scores. And of course we could include both entity (student) and time (pset) fixed effects together to account for both sources of systematic difference.

Hopefully this was helpful but if not, feel free to disregard. The main point is that if anyone does poorly on the exam, even if it might seem like it's my fault it's really not. Best of luck!