Problem Set 1 - Economics of Education

Fern Ramoutar, Nadia Lucas, Sid Sah and George Vojta

April 21, 2021

Question 1

Part A

In the first step of their estimation, CFR-1 residualize student test scores by first removing the effect of observable characteristics. They do this by running an OLS regression with teacher fixed effects:

$$A^*_{it} = \alpha_j + \boldsymbol{\beta}\mathbf{X}_{it}$$

and then constructing residuals:

$$A_{it} = A^*_{it} - \widehat{\boldsymbol{\beta}}\mathbf{X}_{it}$$

The first difference from the previous literature is that they estimate β using within teacher variation only. Previous papers typically used within- and between-teacher variation. They argue that if teacher VA is correlated with X_{it}, then we will attribute part of the teacher effect to the covariates if we estimate β without including teacher fixed effects.

The second difference is that their VA model accounts for drift in teacher quality over time. Previous papers constructing teacher VA weighted scores in all of a teacher's classes equally. The underlying assumption of this practice is that teacher quality is fixed. CFR-1 instead estimates the autocovariance of scores across a teacher's classes non-parametrically. In practice, they regress current scores on average scores in other years, and allow the coefficients to vary across different lags.

Part B

Teacher moves within the sample period help back out a value-add estimation because if you believe that other measures of success are on average the same between both periods for classrooms on each side of the move, as the authors do (with the assumption that other test score characteristics are orthogonal to teacher moves at the high frequency level) then changes in the classroom can be associated with the difference in the value-add. Once this is in place, the movers actually give you the ability to compare individuals at these schools as well. If a teacher adds the highest units of value in one school but the lowest in another than you can say that the teachers in the latter school should be put above the former school and you'll have a way to bench mark the two schools against eachother.

The big problem comes, in fact, when you don't have movers between two schools because then it is virtually impossible to pit two teachers in those schools against eachother. How much of the difference is due to other classroom factors? The authors acknowledge that you cannot take a stance on this, and thus the analysis is done with connected sets as opposed to other larger samples of the data. Fortunately, about 30% of the teachers move per period and thus the connected set is pretty high, however many of those moves are actually just with the "other" option. This could be problematic for many reasons, teaching teachers who get run out of the profession and those who get a really cushy private school job might complicate the analysis.

Part C

There's a lot of explainers to students test scores that can change across school moves, such as social isolation that make the identifying assumption impossible. It is likely, in fact, that if we were trying to run the same placebo test we'd be dead on arrival. Basically, a students movement from one school to another would not be correlated with other student level characteristics that explain success.

Part D

Following closely with the paper and appendix we estimate teacher value add in three steps. Steps 1 and 2 give output that matches the Chetty built-in but Step 3 was off for us.

1. First regress student scores, A^*_{it} on all relevant controls and teacher fixed effects. Use this to construct the residual and add back in teacher fixed effects, we call this A_{it}. The regression of interest is

$$A^*_{it} = \alpha_j + \beta X_{it} + \epsilon_{it}$$

   If we denote the residuals as ϵ_{it} then

$$A_{it} = \epsilon_{it} + \alpha_j$$

   which we end up using as the residual in the rest of the procedure.

2. Now we estimate individual level variances of all test scores σ^2_ϵ and then estimate all the auto-covariances of A across teacher-years. We do not consider any within teacher-year covariances since in this dataset, a teacher at most teaches one class per year. The autocovariances are denoted as σ_{As} where s is the lag. Importantly, we assume stationarity in the distribution of A so the covariances are not state dependent on time, just the lag. Up until this step of the

procedure, our code matches exactly with the output of vam.ado which is the built-in from Chetty et al. The autocovariances we find are

lag	autocovariance
1	.00127258
2	.00073332
3	.0004487
4	.00020362
5	.00014041
6	.00023207

3. Finally, we construct the weights Σ_{jt} and γ_{jt} along with the vector $\overrightarrow{A_j}^{-t}$ which will give our teacher value-added estimates

$$\hat{\mu}_{jt} = \left(\Sigma_{A_{jt}}^{-1}\gamma_{jt}\right)' \overrightarrow{A_j}^{-t}$$

where

$$[\Sigma_{A_{jt}}]_{mm} = \hat{\sigma}_\theta^2 + \frac{\hat{\sigma}_\epsilon^2}{n_{ct}}$$
$$[\Sigma_{A_{jt}}]_{mn} = \hat{\sigma}_{A,|s-s'|}$$

and

$$[\gamma_{jt}]_m = \hat{\sigma}_{A,|t-s|}$$

And $\overrightarrow{A_j}^{-t}$ is the scores vector that are all the lags of t that would go into its prediction. This is the notion of drift. I know this step is doing something wrong in our code because the estimate of bias demonstrated in the next step is off.

Another indication that this isn't entirely correct is that we attempt to replicate figure 1 in the appendix but our empirical distribution of teacher value-added estimates was not cleanly normal in the same way.

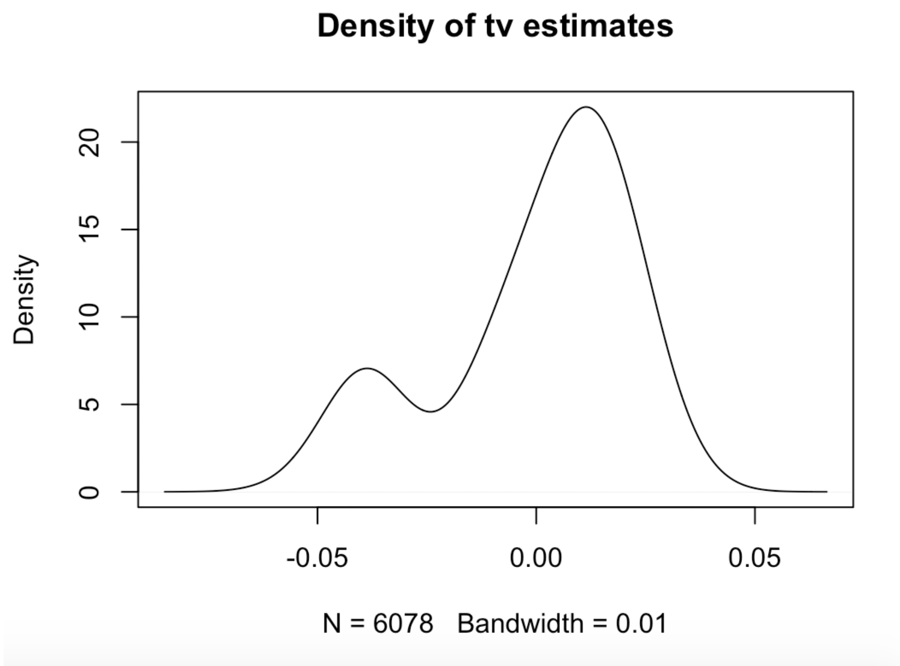

Figure 1: Empirical distribution of teacher value-added

Part E

They define forecast bias as the extent to which the VA estimates accurately predict differences in the test scores of students who are randomly assigned to teachers in a given year. Suppose λ is the effect on test score of being assigned to a teacher who is one unit higher in VA $\hat{\mu}_{jt}$. Then, the degree of forecast bias is $1 - \lambda = B(\hat{\mu}_{jt})$. We compute the bias by regressing individual level test scores on the teacher value added. By construction, this coefficient should be 1, so $1 - \lambda$ is the bias. Additionally, we run the regression where we include individual level covariates and year fixed effects for each student and should also get somewhere right around 1. I know something went wrong in our step 3 calculation so ours is 0.001087, implying that the teacher value add statistic is completely uninformative and our bias is 1 which is not true.

We think the issue is that we should have used all years in even though conceptually for this section we used only past years in $\overrightarrow{A}_j^{-t}$ but didn't have time to recode it so there isn't enough predictive power.

In the paper, bias is also calculated in the same way as above but adding back in student demographics which would reduce the predictive power of the teacher value add estimates.

Part F

If we instead do not account for drift, we compute the teacher value-added as

$$\hat{\mu}_{jt} = \bar{A}_j^{-t} \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\theta^2 + \sigma_\epsilon^2/n)/(t-1)}$$

where now $\bar{A}_j^{-t}$ is just prior years mean residual and the weights are a reliability of the VA estimate. This conceptually is a little different because we are now using less information on the stationarity of the teacher value-add time series to predict value-added. However since our estimates in part D were off, it was difficult to compare, although we attempted this in the end of the code. We did find that bias under the no drift specification was also quite high and teacher value-add was quite noisy, although the density was quite normal

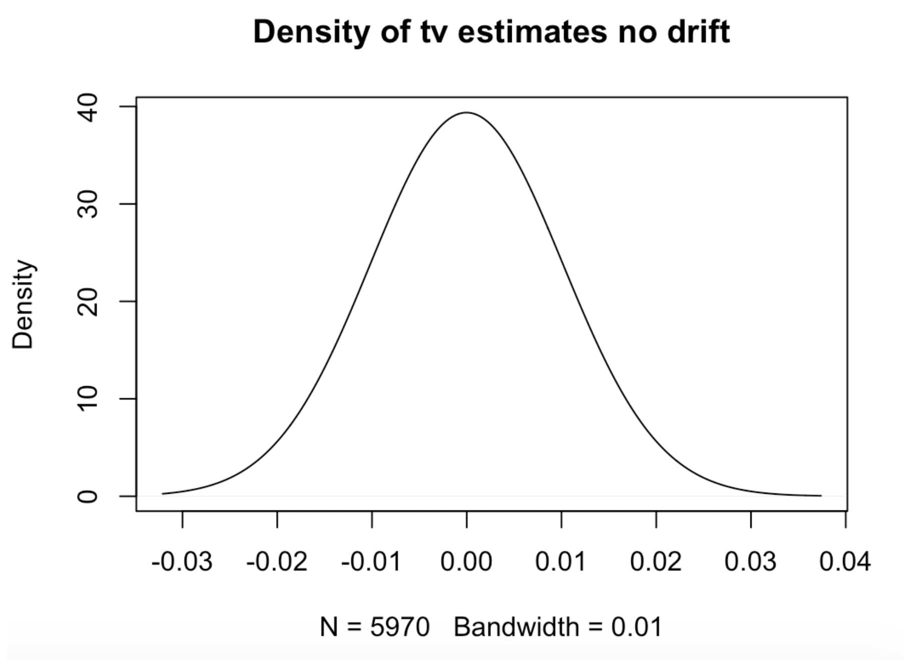

Figure 2: Teacher value-added with no drift

Question 2

Part A

Rothstein (2017) directly questions the primary underlying assumption of CFR-I's quasi-experimental design - that changes in average teacher VA across cohorts are not correlated with changes in other determinants of test scores. In particular, Rothstein (2017) expresses concern about the possibility of non-random sorting of students to teachers. In the quasi-experimental analysis, CFR-I consider both average teacher VA and average student outcomes at the school-grade-subject level before and after a teacher move, which should abstract away from concerns about non-random sorting specifically to the new teacher. However, this abstraction requires that teacher VA's and student scores are averaged across all classes in a grade. CFR-I exclude teachers and students of those teachers for whom they are unable to compute an estimated VA score, namely teachers who are only observed for one or two years. If such teachers get students who are systematically advantaged or disadvantaged compared to others, the central assumption may not hold.
CFR-I perform a placebo test of their assumption by comparing test scores as predicted by variables that were not used in the construction of the teacher VA estimates, such as parental income, before and after a teacher change, finding almost no difference. However, as these variables were excluded precisely because they are unobserved by schools and thus cannot be used in sorting students to teachers, Rothstein (2017) argues that comparing scores predicted by these variables may miss the bias introduced by sorting. When performing an analogous placebo test using scores predicted by the same variables used in the construction of the teacher VAs, such as prior test scores, Rothstein (2017) does find a significant change in predicted average scores before and after teacher move events.

Part B

In their reply to Rothstein (2017), CFR seek to explain how the finding of teacher quality affecting past test scores might arise even when their identifying assumption is correct. They argue that because prior test scores are used in constructing the estimated teacher VAs, performing a placebo test involving prior test scores effectively puts prior test scores on both sides of the regression, which can lead to a spurious positive coefficient. Specifically, any shock to a school's test scores in a given year will lead to a shock in the school's teacher's estimated VA in future years. Thus, for instance, if there is a school-wide positive shock to test scores in a given year, and a new teacher enters the school the following year, the average teacher VA will tend to rise, as the incumbent teachers will have randomly elevated estimated VAs. The rise in average teacher VA across years will come with a rise in the average previous scores of students across years, because the incoming cohort will have benefited in the previous year from the same schock that elevated the incumbent teachers' VAs. CFR confirm that this may occur in a simulation exercise in which they enforce that their identifying assumption holds.

Part C

One obvious solution would be to create an RCT with students sorted into classrooms first and then teachers being dealt into classrooms afterwards. Then we could also control the movement between teachers from school to school in this hypothetical world where we have infinite money and full experimental power and the IRB has given in to our whims and desires. However this is a boring way to do things so we offer the following proposal:

Consider looking at teachers who transfer in and out of classes due to unforseeable circumstances (e.g. car accidents, etc). Then we could use these midyear transfers, controlling of course for the shift in teachers, which we need a sort of linearity assumption on, as a way to analyze whether or not the impact of students self sorting themselves into classes is important or not. There are definitely some other quirks to work out but we think this would be an interesting direction to go.