Somebody else has done v similar to this. Using the same data, but they used event study which is probably more appropriate but we can use matching. The only problem is that we are using a cross sectional survey. We have birth cohort 1980-91, some attend before facebook was introduced, some after. But we only have data for 2014 earnings, so obviously some are at different stages of earnings than others.

Maybe we can use lasso to try to determine who those are that are in that cohort year that is affected by facebook

Treatment is defined by 4 expansion batches, as was first expanded to 100 selective unis then another batch etc. But found if we look at everyone that has facebook, so everyone younger than like 18/19 in 2004, we get 40$k effect of treatment. But when we restrict it to just the people that were in uni around the expansion, we get more sensible results, something like 1000-3000 dollars

But I think this was for any treatment in the 2 years it was expanding, so comparing cohorts in 2014. But the paper kinda overcomes this by looking at their income rank compared to the rest of their cohort, which makes sense

I am only thinking of lasso/ridge to somehow determine who the people are that go to these universities, if that is even possible im not sure

But, that is the only kinda extension thing I can think of. We could just use matching and try to put our heads together for ridge/lasso. But my thinking is there is heterogeneity in treatment, so the effect of facebook isn’t equal, so we would probably uncover that it was best for people in the first treatment batch

There is also another paper using this data but with a mental health dataset, but we don’t have access to that so unsure what direction we can go in other than matching really