Research Module papers

Winter 2019

These notes will span the collection of papers focused on human capital, inter-generational mobility and other papers confined to the Research Module. Each paper will include a **Quote, Argument, Question, Reflection** (QAQR) report, a **dense summary** and **notes** on the respective paper. The QAQR report follows the format defined in the research module github page. The dense summary contains what I consider key parts of the paper. Notes contain substantive remarks on the key parts of the paper. In blue are lightning talk papers.

Contents

1 Life Cycle Earnings, Education Premiums and Internal Rates of Return[1]

1.1 QAQR

The paper argues the profitability of additional schooling, given that the rate of returns from schooling are higher than the market interest rate. The results show that after considering taxes and entitlements, the internal rate of return is 10increase in lifetime earnings and steeper age-earnings profiles. Furthermore, in comparison, the cross-sectional Mincer regressions understate the returns to schooling; coming from two assumptions: (1) students generate zero earnings while in school (2) exogenous post schooling employment. These assumptions definitely do not hold in reality today, but the model serves as a baseline to make comparisons. To Mincer's defense, these assumptions held in 1974 for his data.

For details on the paper, see the notes. The focus will be on the Mincer regression and the mechanisms explaining its difference with the paper's results. I am not familiar with the Mincer regression and for completion, thought I would address this directly. Holding the strong assumptions above, the Mincer regression can estimate the internal rates of return. why?

Figure 6 shows a graph comparing OLS estimates of the cross-sectional returns to the results of the paper. First, the Mincer estimations show a visible constant increase over time. Taking a first difference will remove any time-constant fixed effects; thus also, time-constant unobserved heterogeneity. Nevertheless, this is not done in the paper and they explicitly state violation of the stationarity assumption. A possible explanation for not first differencing comes from not being able to estimate demographic variables or any other time-constant variables; these variables would be removed in the demeaning step of a fixed effects estimation. In the paper, they do attempt to control for calendar time effects in the Mincer estimations; to account for changes in skill prices (what a firm would pay for a skill today could be different tomorrow), which they believe is causing the upward trend.

As a result, the estimates decreased even further. The paper offers two possible explanations for the overall downward bias: (1) students do generate earnings while in school (2) schooling affects employment. First, due to discounting, earnings received early on in life carry the most weight. To remove the earnings capacity while in school drives estimates downward. The paper runs an estimation assigning zero earnings while in school and the results become comparable to the Mincer regression estimates. Second, schooling affects employment; in particular for those in the lower income percentiles. Chow tests are not done to compare the estimates; probably because the differences are salient.

Question: The paper used Norwegian panel data spanning the lifetime of individuals. They instrument for schooling using a school reform near the mid-point of the last century. The schooling reform is exogenous to people's ability. How would this result hold when school accessiblity is limited? Conditional on limited accessibility, how would the returns to education look over one's lifecycle? Also, it would be interesting to compare outcomes of individuals from impoverished and affluent backgrounds, given they went to the same college. These ideas arise from the setting in which this data holds. Norway is not stricken with inequality. Education is also free in Norway. When these two former statements do not hold, how do the returns look? Would it still be the case to invest in education?

In relation to the lectures and the literature I am familiar with, Norwegian data is well sought. It is powerful in information and use because of the expansive panel of years and sample size. Because of it, questions on schooling can be addressed like in Black (2005); but, it is still Norwegian data. Although highly administrative data, the returns could be different for a different cohort in a different country in a different economic environment. Let these papers serve as motivation for subsequent research ventures; but only applicable to where the data was confined. (Will continue working on reflection).

1.2 Dense Summary

- **Question**: Tackles three fundamental questions in economics:

    1. What do the education premiums look like over the life cycle?
    2. What is the impact on schooling on earnings over time?

3. How does the internal rate of return (IRR) compare with the opportunity cost of funds? And does considering progressive taxes lessen the incentive to invest in education?

- **Data**: Only Males. 1943-1963 birth cohorts. Norwegian Panel Data 1967-2010 merged with 1960 census data. Public domain. Linking child to parent possible. Schooling measured until age 40. three measures: (1) Earnings over the life-cycle (discretized) (2) IRR (3) Lifetime Education Premium (mean lifetime earnings)

- **Methods**: OLS, IV

- **Identification Strategies**

  1. compulsory schooling reform as an instrument for education (it is exogenous, thus extremely useful BUT it just increases **required** schooling from 7 to 9 years)
  2. Controlling of course for ability test scores to account for selection bias ('more able' students will get more schooling).
  3. Within-twin-pair estimation (to control for environment/genetics)

- **Conclusion**: Additional Schooling gives higher lifetime earnings and steeper age-earnings profile. IRR around 10% after considering taxes and earnings related to entitlements.

1.3 Notes

- **Mincer Regression**: Assumes *exogenous* variation in schooling, quadratic potential experience in explaining log earnings. Mincer also assumes college students do not have any earnings and post-schooling employment is exogenous. (both relaxed here to compare estimates and make bounds).

- **Education Premium**: Interpreting the Mincer regression as the price to pay for a laborer given a certain education level, we can interpret the schooling's coefficient as the *education premium*. With respect to the paper, this is done at each age. They consider the education premium's mean and present value. The IRR is rate ρ which satisfies the typical PV problem:

$$\sum_{a=17}^{62} \frac{\beta_a}{(1+\rho)^{a-16}} = 0$$

  They compare this rate to the real-interest rate (proxy for opportunity cost). Thus, trying to answer the question: Is it more profitable to invest in education or invest in the market at the inflation-adjusted rate? Furthermore, they use the **ordinary annuity** expression to compute the lifetime education premium. The **average real interest rate of 2.3%** was used for 1967-2010 (should another rate be considered?)

- **First Identification Strategy - IV**: Since schooling is endogenous, we seek to extract only the exogenous portion from the first stage regression. They instrument using a boolean variable for compulsory schooling. Estimation results from equation (7) show a strong rst stage with an estimated coefficient on the instrument of 0.213. This means that exposure to the compulsory schooling reform increased years of schooling by about one-fth of a year. The F-statistic for the instrument is around 93, implying that weak instrument bias is not a concern for our analysis.

- **Second Identification Strategy - Controlling for ability**: The second method controls for ability to mitigate selection bias. **The exam**: The arithmetic test mirrors the test in the Wechsler Adult Intelligence Scale (WAIS); the word test is similar to the vocabulary test in WAIS; and the gures test is comparable to the Raven Progressive Matrix test. They use **test score indicators**. Problem: only those that tried for the military took the exam (for 1950 or later).

- **Third Identification Strategy - Within-twin-pair estimation**: To control for environment and genetics - used as a fixed effect.

- **Unbalanced Panel**: When the number of time periods are *not* the same for all individuals. They wish to study those born in 1943-1963 during the period 1967-2010 for ages 17 to 62. Problem:

    1. The person born in 1963 at most will be 47, thus we don't know what happens between 48-62. A difference exists if the cohort was born after 1949. A robustness check (using a balanced panel) is done to reassure the results.

- **Table 2 and Table 3**: The main results show the following:

    1. IRR robust to the specification. Between 10%-14%
    2. Returns to education negative for 17-24 year olds (since they are giving up earnings compared to those going straight to work). Returns increase (still negative) when considering those in the workforce having to pay taxes on their wages.
    3. For ages 25-44, positive, but again for the IV estimate, really noisy (standard errors 15 times larger than that of the twins estimation). This is typically a sign of weak instruments, but the F-stat, t-stat and the fact that the instrument is exogenous helps.
    4. For ages 45-62, IV returns no difference and again high standard errors.
    5. All estimates range between 2% and 9% of the mean.
    6. **Table 3** shows that the IRR results are robust after considering taxes and pension income considering all ages. Looking at Figure 5, we can see that taxes decreases the negative effect of going to college since those who did not go to college and chose to work, must pay taxes on their earnings. The pension entitlements were estimated based on earnings information.
    7. As part of the sensitivity analysis, they had to make sure that cohort trends in earnings were unrelated to the reform. In other words, that nothing else made this earnings 'jump'. This is proven by creating a specification where they use interaction terms for years and municipality characteristics (plus the fixed effects) to explain the timing of the reform (e.g. poor municipalities would be the first to implement the reform. No evidence of such a relationship was found.
    8. Controlling for pre-reform trends, the IRR stays sizeable.
    9. In considering different ways in treating the missing values, whether by imputing zeros or past earnings, results are robust and considering a balanced panel also rendered consistent results.
    10. Mincer Regression estimates are lower because of the restriction of no earnings while in school and exogenous post schooling employment accounts also for the downward bias.

2 Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital[2]

2.1 QAQR

The paper dives directly into addressing a key dichotomy in the literature when explaining the variation between parents' levels of education and children's level of education. The core argument arises from whether this relationship stems from pure selection theory or from causal implications of education. The paper states:

"Distinguishing between these scenarios is important from a policy perspective. One of the key roles of publicly provided education in our society is to increase equality of opportunity, and many policies have been implemented to further that goal in recent years. Proponents of this type of education policy often cite as one of the benefits the spillover effects on to later generations; having more educated citizens will have longer run effects by improving the outcomes of their children. While many argue that this is the case, there is little causal evidence to suggest it is true; the research to date has been limited in its ability to distinguish between selection and causation"

They use the same reform as the previous paper to instrument for parental education; to capture variation in children educational attainment given an exogenous change to parent education attainment. In light of the results, the highly-correlated relationship between the educational attainments of children and parents is mostly a pure-selection story. For that reason, I want to discuss the estimates dervived from IV and OLS. Typically, IV estimates are larger than those of OLS. The instrument can be upward biased from unobserved ability; but here, education reform and the ability to control for both age and municipality effects leads to greater confidence that the instrument is not correlated with unobserved ability. Having also administrative data reduces the downward bias from measurement error. In the literature, the high IV estimates are rationalized by heterogenous returns to education with particularly high returns for the group of people who is impacted by the instrument being used. Education is free in Norway, so it is possible that those in the lower tail of the educational attainment distribution are impacted no higher than those on average.

Mark presented a paper about the intergenerational transmission of educational attainment in Germany; also using a compulsory schooling reform to instrument for schooling. Surprisingly, two different countries with two different histories with different size reforms capture the effect between mother and son educational attainment. They both argue similar mechanisms; but here are a few: (1) Children tend to spend more time with their mother. Naturally, extra schooling for the mothers increases their own human capital, thus their optimal human capital choice of children. (2) To see if there exists a relationship between mother's and father's education (testing the hypothesis that higher educated women marry highly educated men), they ran a regression using the reform as an instrument for mother's education and nothing is conclusive. This implies that the greater education indued by the reform did not have any mojor effect ont he type of father chosen by the mother. (3) To test whether women with more education have less children and ivest more in each of the,, testing the quality/quantity tradeoff, they see whether increasing the mother's education has an effect ont he number of children they have. There is a lack of evidence on this (except for mother's on the lower end of the educational attainment distribution).

2.2   Dense Summary

- **Question**: What goes into the composition of the variation when trying to explain the relationship between the education levels parents attain compared to their children's education levels? Is it mostly family characteristics and inherited ability (selection theory explanation) or education spillovers (being educated has implications on parenting and leads to your children having more educational outcomes)?

- **Significance**: If it is the case that education spillovers exist, education policy seems to be the recourse. Thus, this paper proposes to provide evidence on the *causation* argument and relieve some questions on the intergenerational transmission of education.

- **Data**: Norwegian data and children data from 2000 aged 20-35. Census data from 1960, 1970, 1980. Educational attainment from Statistics Norway. Cohorts of parents born between 1947 and 1958. Analysis mostly focused on parents with less than 9 years of schooling.

- **Methods**: OLS, IV

- **Identification Strategies**:

  1. They instrument for *parental* education using the compulsory schooling reform (1960s Norway) since it is exogenous to parental ability.
  2. They make two assumptions:
     (a) That individuals who get 9 years of education after the reform would have gotten less than or equal to 9 before the reform.
     (b) Those who got less than 9 years of education before the reform, get 9 years of education after the reform.

- **Conclusion**: There is little evidence of a causal relationship between the father's education and the children's education. A mother and son comparison generates a weakly significant result, but no causal relationship between mother's and daughter's educations.

2.3   Notes

- They checked for significant determinants with respect to the timing of the reform (as did Paper 1) such as average earnings of the people, taxable income, size of the municipality, unemployment rates, urbanization, labor force composition, and education levels and found no relationship.

- Their estimation uses parental education, children age, sex, and fixed effects for parents' age and parents' municipality. The fixed effect for parents' age is their to control for secular (persisting nearly indefinitely) changes in educational attainment over time that may be completely unrelated to the reform. The municipality effects control for variation in the timing of the reform across municipalities that may not have been exogenous to parents' educational choice like them seeking higher education because their municipality required highly skilled workers. As long as we fix these characteristics, make sure they are not correlated with the timing of the reform nor the schooling of the children of this generation.

- **Small things to be aware of**

  1. They assume in cases of one-parent households that the child lived with the mom. This comes up because the mother's parent place of residence can be found in the 1960 census. Thus, they make the 'safe' assumption that the child lived their too (this is of course relative to the epoch).

  2. **Selective Migration**: There could be cases where the parents anticipated a change in another municipality and moved there. This affects the estimates, but not the consistency - thus with large enough sample size, the estimate is fine.

- Looking at the distribution of parents and their education levels plus minus 2 years from the reform shows that the *cumulative* distribution of people getting ≤ 9 years of education is constant. Thus, the reform itself does not have spillover effects.

- Looking at family characteristics, there is no evidence of compositional change.

- **Results**

  1. OLS - positive and significant results (5% level) for discrete and Boolean education DV.

  2. 2SLS - insignificant, large s.e.'s, but this is due to the focus on all parents and not those in the bottom of the distribution where the reform has 'more bite'. The instrument is weak and the t-stat, although showing variation exists between the reform and parental education, is mostly minor at the higher education levels.

  3. Focusing on the sample of parents who received less than 10 years of education, 2SLS shows mother's education has an impact over on child's education level and son. Standard errors decreased after subsetting, this the estimate is more precise.

  4. Overall, results indicate that the positive correlation between parent and child and their education levels mostly comes from other positively correlated factors with education like ability, family background, income or other factors. The causal effect is weak.

  5. In the literature, high IV estimates are rationalized by heterogeneous returns to education (a particular group whose behavior is impacted by the instrument i.e. child goes to school longer).

  6. To control for the fact that educated women marry educated men, they control for father's education and thus, both parents, in explaining child's educational attainment. They also only include cases where at least 1 parents attained less than 10 years of education. IV estimates still weak.

- **Robustness/Specification Checks**: (1) Municipality-specific time trends, county cohort fixed effects (2) In case they were inaccurate in selecting the reform year, they removed data for that year, and the years preceding it and following it. (3) Sample selection from 20-35 to 25-35 years old. The initial reason for 20-35 comes from the fact that less educated parents have less educated children. Thus, since the reform hits the lower distribution, the children were expected to be finished by 20 years of age. To be sure they are finished, they consider 25-35 years. Still, significant, positive result for mother/son, but nothing else. (4) Subsetting to the highly educated rendered nothing. (5) Conditional on human capital, high able women are less likely to have children. Thus, when the reform hits, some woman will experience an increase in human capital. These women will not have children and the group of women left hit by the reform having children have **less ability** than those mothers not hit by the reform. To check whether the reform increases one's probability to have children, they run a linear probability model with the DV equal to 1 if parent has child and 0 if not. Not significant; but, could also use probit/logit. (6) Sibling fixed effects. This is supposed to capture some unobserved ability. In the first stage regression, if unobserved ability is positively correlated with parental education, then of course the OLS estimates will decrease with including the covariate. First, they consider the sample of parents who have brothers or sisters and seeing how they impact the same-sex child. The results, as hypothesized, reduce the estimates in the OLS regression; implying the OLS estimates are upward biased by unobserved ability.

- **Exploring the mechanisms for explaining why mother's education affects child's education**: (1) Children tend to spend more time with their mother. Naturally, extra schooling for the mothers increases their own human capital, thus their optimal human capital *choice* of children. (2) To see if there exists a relationship between mother's education and father's education (testing the hypothesis that higher educated women marry highly educated men), they ran a regression using the reform as an instrument for mother's education and again, nothing. This implies that the greater education induced by the reform did not have any major effect on the type of father chosen by the mother. (8) To test whether women with more education have less children and invest more in each of them, testing the quality/quantity trade off, they see whether increasing mother's education has an effect on the number of children they have. There is a lack of evidence on this (except for mother's on the lower end of the educational attainment distribution

3   Estimating ATE/LATE of Education when Compulsory Schooling Laws Really Matter
Summary by Florian Schoner[5]

3.1   QAQR

To be written.

3.2   Dense Summary

- Analyzes the impact of the increase of the minimum school-leaving age from 14 to 15 in the UK and Northern Ireland on subsequent earnings in the labor market.

- **Specialty 1)**

  – Prior to the reform, 60% of pupils left school at 14 and this share decreased to 10% after the reform. Therefore, half the population of 14-year-olds are compliers in the sense that they take up the treatment (an additional year of schooling) because they were exposed to the instrument (the minimum school-leaving age is 15).

- In contrast to previous studies, the group of compliers therefore is a majority of the whole population, hence the LATE should be closer to the ATE (In fact, the author argues that LATE $\approx$ ATE).

- **Specialty 2)**

  - Achieves relatively clean identification by using a (fuzzy) RDD

- Previous studies based on IV-estimates of the causal effect of education suffered from the caveat that identification was driven by a very particular group.
  For instance, proximity to college has been used to estimate the LATE in the US. However, that dummy was equal to 1 for only 10% of the population, of which only a fraction *actually* went to college. The literature was concerned whether these particular groups of compliers may exhibit larger effects than the population at large because those individuals are more credit constrained, have a greater distaste for school, or a greater immediate need to work.

- This belief is driven by the fact that IV estimates usually exceed those from OLS-specifications in the literature. This is surprising since we would usually assume that, due to omitted variable bias, OLS *overestimates* the causal effect of schooling. Therefore one potential explanation was that would-be dropouts must have higher gains from additional year of schooling, thus leading to LATE > ATE

- **Findings**

  - Finds increase in earnings of around **15 - 17**% due to the additional year of schooling.

  - In a corrigendum published by Oreopoulos/the AEA in 2008, revised estimates are around **7 - 13**%, and some of them were rendered statistically insignificant

  - Further caveat: Used the wrong SEs for a RDD with a discrete running variable. Using the correct ones renders even more of the estimated effects statistically insignificant.

- **Method** Uses RDD for identification.

  - First Stage: Regress average age of leaving full time education on the year aged 14 (from 1947 on, the reform was in place) and polynomials of it and a dummy for 1947 or later. (T on Z)

$$AGELEAVE_i = \gamma_0 + \tau_{FS} \cdot \mathrm{I}(MINAGE_i = 15) + \sum_{j=1}^{4} \gamma_j \cdot YEAR14_i^j \tag{1}$$

  - Second Stage: Regress $\ln(EARN)$ on polynomials of year aged 14, and instrument the indicator: and indicator whether year aged 14 was greater or equal than 14. (Y on Z)

$$\ln(EARN_i) = \beta_0 + \tau_{SS} \cdot \mathrm{I}(MINAGE_i = 15) + \sum_{j=1}^{4} \beta_j \cdot YEAR14_i^j \tag{2}$$

  - then divide $\widehat{\tau}_{SS}/\widehat{\tau}_{FS}$

- **Conclusion**

  - Since he finds similar LATEs in the UK, US, and Canada and the latter two were suspected to be higher than the respective ATEs, he concludes that the ATE of education is indeed high and that would-be dropouts do not gain more from additional education than the reference population.

4 How does your kindergarten classroom affect your earnings? Evidence from STAR[3]

4.1 QAQR

The point of classroom interventions and test scores is to see how they participate in the labor market. Does having higher test scores or being exposed early on to a special classroom structure indicative of better life outcomes in young adulthood or even later in life? The paper attests to those outcomes between high school graduation and a few years after college graduation. To say that these factors persist into further future outcomes is something to test; where we can consider personality traits or rather, non-cognitive skills because these are factors that follow you for a large portion of your life. The paper uses highly administrative data to make the argument that earlier childhood interventions are not entirely necessary when trying to make a difference in one's life. The STAR experiment intervenes when students are between the ages of 5-8 years old and from the results, are already generating significant differences in means in the young adult-hood outcomes than those who did not receive the treatment. The Perry Pre-school intervention occurred in pre-school and the life-outcomes are now marketed across the U.S. Point is, these interventions are causal effects for the data for which this took place. As always, these should motivate further research in early-childhood interventions; but, not conclusive.

Chetty alludes to this by quanitfying the observable classroom characteristics. He does this for two reasons: (1) To measure whether such class effects matter (do interventions matter) (2) In cases that they matter, what are the impacts one can expect from them. (1) is tested doing an ANOVA for class effects. (2) is measured by a correlated random-effects model; where the change in an outcome is measured given a standard deviation change in class effects.

Another argument that had to be made in response to the results is the concept of "Fading-Out". Why is it that the small class intervention and class quality impact test scores today; but, not my 8th grade test scores? Also, we see that the small class interventions and class quality have significant impacts in young adulthood outcomes. Thus, there is "re-emergence" of this effect. The rationalization that the paper takes on these two concepts is non-cognitive skills as seen in the literature (e.g. Heckman 2000, Heckman, Stixrud, and Urzua 2006, Lindqvist and Vestman 2011). An increase in class quality (measured by peer scores) by one percent, increases ones non-cognitive score by .15%. Intuitively, a better kindergarten classroom might simultaneously increase performance on end of-year tests and improve untested non-cognitive skills. A 1% increase in your grade 4 non-cognitive score translates to $106 increase in average earnings. Throughout the entire paper, they do not regress class quality on small class treatment probably given that class quality is just demeaned-jackknifed test scores; so the result could be similar to projecting test scores on small class treatment. If all this is the case, then it could be that the non-cognitive skills generated from being in a small class that are causing the earnings increase in the future. Non-cognitive skills in 4th grade are indicative of wages and math+reading test scores in 8th grade; something KG test scores could not connect. Because non-cognitive skills appear to be correlated with earnings through channels that are not picked up by subsequent standardized tests, they could explain fade-out and re-emergence. These non-cognitive skills may not be well measured by standardized tests, leading to very rapid fade-out immediately after KG. There is imprecision in the estimates and probability of attrition bias, hence more work could be done in analyzing the value of non-cognitive skills.

I will continue to read more on the literature of early childhood school interventions, starting with some references made in the paper. A cool idea that I had in mind is the the IB intervention. IB stands for International Baccalaureate program. It is a boxed up curriculum qualified schools can make to students. The program is rigorous and designed for select students; yet, in particular, the structure of the classes, exams and requirements are completely different. To compare apples to apples, we can look at students who are in the IB curriculum and compare them to students who take a comparable amount of Advanced Placement classes (AP). Students taking AP classes are also indicative of some level intelligence comparable to an IB student; but, the structures they are in are different. In particular, one is aligned with U.S. college standards (designed by the college board) where you take a course and you take an exam with respect to the course. IB is an international curriculum designed to broaden the horizon of students. They deliver

this by requiring more hours of community service from the students, more hands-on involvement in the sciences, internally assessed exams as well as externally assessed exams for each subject, requirement of learning a language and notoriously known for a final paper of at least 4000 words on a particular research question. The program is completely exempt from the typical high-school requirements of 24 credits to graduate and are allowed to study topics which are not allowed in the typical U.S. curriculum.

4.2 Dense Summary

- **Questions**:

  1. What are the long-run outcomes from Project STAR?
  2. What are the long-term impacts of early childhood education? But also rather, do the environments that raise test scores also translate to better adult outcomes (i.e. earnings, college attendance, marriage, home ownership, savings, etc)?
  3. Although the effects of classroom attributes on test scores do not persist through education, why do they persist to adulthood?

- **Data**: Linking Project STAR data (student/teacher achievement ratio) and U.S. Tax data from 1996-2008 (1040 form, W-2, and 1098-T, the latter two being available from 1990), they were able to follow 95% of the STAR participants into adulthood.

- **Methods**: OLS, ANOVA

- **Identification Strategy**: Using school-by-entry-grade fixed effects, peer/teacher characteristics and the **Experimental Design** of STAR.

- **Conclusion**:

  1. Kindergarten test scores are highly correlated with (1) earnings at age 27 (2) college attendance (3) home ownership (4) retirement savings.
  2. *Experimental Impacts*:
     - Students in the small classes are significantly more likely to attend college and exhibit improvements in other outcomes.
     - Class size in general does not have a significant effect on earnings at age 27, but it is imprecisely estimated.
     - Having experienced teachers translates to higher earnings
     - Classroom effects (higher quality versus lower quality defined by student test scores) have higher earnings, college attendance rates, and other outcomes.
     - Classroom effects lessen on test scores in later grades (by grade 8, class quality in kindergarten holds no weight), but gains in non-cognitive measures persist.

4.3 Notes

- **Brief intro about STAR**: The Student/Teacher Achievement Ratio (STAR hereforth) experiment randomly assigned 11,571 students to different size classrooms (and different types/quality of classrooms) for grades K-3 for 79 schools in 1985-1989.

  - **Assignment**: Students were assigned to small classes (15 students on average), and large classes (22 students on average). Teachers were also randomly assigned (they may differ in quality, defined by years of experience) and by definition, so were your peers.
  - **Objective**: To see the causal effects of such assignment on test scores.
  - **Critique**:

– 1. The schools were predominantly of lower socio-economic status compared to the rest of Tennessee and the rest of the U.S.

  2. Since kindergarten was not required, almost half the sample entered in 1st grade; not really an issue, but something to keep in mind. The paper uses school-by-entry-grade fixed effects to control for unobserved characteristics that may have risen from entering the school in different years.

  3. No record of the *location* of the classrooms, only classroom attributes. Details on the students' peers and teachers are also limited.

– In cases where deviations occurred (e.g. moving from a small classroom to a large classroom), typical protocol was adopted (this sort of non-random sorting is treated by just granting treatment based on the **initial** random assignment, meaning there was an **intent to treat**.)

– Score distributions of the students from large classes were made, then the scores from the students from the smaller classes were placed in the appropriate percentiles created by the large class distribution so see how treatment class fared in comparison.

- **Variables**: **Earnings** is in 2009 dollars, adjusting for inflation using the CPI. Earnings capped at 100k since less than 1% of the sample made over that much (done to increase precision) and also the earnings outcome was an **average** between 2005-2007 (also done to increase precision of the estimate). 1098-T forms indicate college enrollment. All universities, colleges and vocational institutions receiving federal aid must submit one per student (one student can have multiple forms if they switch schools for example). They have payments made by the student (which is interesting because you can translate that to years of college completion based on the cost of the university). This was not done/considered.

- To measure **college quality**, they used an earnings-based approach, where they group students based on university and see what there average earnings are. This correlates .75 with the U.S. news ranking system, thus a decent measure.

- **Married** is an indicator for whether a person was *ever* married by age 27 and indicated if the taxes were filed *jointly*. **401k savings** is reported on the W-2. **Home ownership** can be found from the 1040 form, but mostly from the 1098 (evidence of mortgage payments). Naturally, this excludes people that own a home and make zero mortgage payments; but, the likelihood of owning a home before 27 is very small. **Cross-state mobility** indicated whether a STAR student ever lived *outside* Tennessee between age 19-27. **Neighborhood quality** is based on the percentage of college graduates from that zip-code in 2000. **Parents** are defined as those who claim a STAR student as a dependent from 1996-2008 (the likelihood of being a dependent to a non-biological guardian is small). By 1996, STAR students were all 16-17 years old, so still going to high school - 86% were able to be linked. **Parental Income** average of 1996-1998 adjusted gross income (for all observations, income zeroed if no taxes were filed).

- **Other Variables**: Age, gender, race, whether they receive free or reduced lunch, teacher experience, teacher highest degree, test scores from the *entry grade*, student outcomes, parent characteristics, classroom characteristics

- **Validity of Experimental Design**: **Do characteristics vary across small vs large class types and do they vary across classrooms within schools?** Causal inferences can be made due to the randomization of students into classrooms and no differences in attrition across classrooms. To *verify*, a **balance check** is done using *pre-determined variables*. For instance, before the experiment, we knew the student's age, race, whether they qualified for free lunch or not, and gender. For an experiment to be random, these variables should not impact the decision of classroom assignment. Furthermore, they used parental income, 401k savings, home ownership, marital status, and mother's age at child's birth as another balance check. The balance check is just to establish *randomization*, that allocation of the students to small or large classes or any other attribute did not arise from

having a certain background (e.g. poor people were allocated to small classes) - that it is indeed **random**. This can be done by regressing an indicator for whether a student was in a small class on the set of variables above plus age, gender, whether they received free or reduced lunch, and race. No difference implies characteristics across both groups are no different. *Although* these new predetermined variables are AFTER the experiment, the researchers believe these variables are unlikely to be impacted by the random assignment. **Table 2** shows the randomization tests and concludes the experiment was *indeed* random. Meaning, pre-determined variables showed no difference *across* classrooms. The last test used the predicted values of Column 1, Table 2 and regressed those values on kindergarten classroom indicators (including school FE). Results showed clustering of earnings (created using pre-determined variables) is non-existent by classroom indicators. Clustering here just means the values do not *shift* or cluster in space conditional on the treatment. Then, to see if those 95% of people linked to the tax data were matched non-randomly (e.g. rich people match more than poor people), they run a regression of being matched on whether you were assigned to a small class with and without a full-set of controls - no differences. Finally, do see if there is some variation across groups with respect to the death rate, they regressed the death rate on whether you were assigned to a small class or not - no differences.

- **How to Assess a Relationship**: When trying to make implications about the relationship between variables (in this case, test score percentiles and adult outcomes), descriptive statistics is a great benchmark assessment. Showing correlations and raw differences (without controls), strengthens your argument and also shows you are not tossing in controls for a significant result - that, the *significant difference* was there from the beginning. Figure Ia plots the correlations and illustrates a summary index (some composite measure for home ownership, 401k savings, if ever married, ever lived outside Tennessee and neighborhood quality).

- **First Results: Regressing adult outcomes on scores**

  - **Dependent Variables**: Earnings, College attendance in 2000, College attendance by 2007, college quality, and the summary index
  - **Independent Variables**: Entry-grade percentile score, 8th grade percentile score, parental income percentile, class FE, race, gender, age, free lunch status, mother's age at child birth, indicator for whether they have a 401k, home ownership and a quartic in parent's household income interacted with an indicator for whether they were married between 1996 and 2008.
  - See Table 4 for complete results. Considering the first 5 columns, wage earnings increase given a unit change in the kindergartener's (KG) percentile score. When controlling for the 8th grade test score, this new score absorbs the entire effect of the kindergarten test score - still significant. Even more interesting is that parent's income percentile is more indicative of a child's earnings than test scores.
  - For college and the summary index, a unit change in KG's score translates to higher probabilities in attending college but also, better school quality. Summarized by the summary index of other adult outcomes, a greater KG score implies a better young adulthood.

- **Second Results: Regressing adult outcomes on class size**

  -

$$y_{icnw} = \alpha_{nw} + \beta SMALL_{cnw} + X_{icnw}\delta + \varepsilon_{icnw}$$

  - We seek to explain earnings y of student i in classroom c of school n, for wave w (defined as when they entered the school).
  - α_{nw} are the school-by-entry-grade FE
  - X_{icnw} is the vector of characteristics (parents' and child's).
  - As we saw earlier, X_{icnw} is balanced across classrooms, so it should not impact the class size estimate, β.

– Standard errors are done at the school level, but also done at the classroom level for robustness.

– result should be interpreted as the effects of attending a class that is 33% smaller for 2.14 years (2.27 -.13 = 2.14 which is the difference in time spent in small classes for the small class group and large class group).

– attending a smaller class **increased KG test scores** as founded by Krueger (1999)

– Splitting college attendance (i.e. one indicator for college at 20 years old and another indicator for college before 27) is necessary to estimate effects of class size on college attendance properly. The effect is smaller in the latter group and they explain the mechanism behind as students tending to take up more classes early on than in the third or fourth year (i.e. part time vs full time). Attending a small class **increased the likelihood of attending college**, yet this was a weak result.

– Overall, students in small classes **have a higher probability of attending a better quality college**. Quality is defined as the mean earnings of the attendees. Yet, *conditional on attending college*, students from the small classrooms attend **lower quality schools** (based on their measure) than those in the larger classrooms.

– With respect to earnings, the regression estimate is **negative** (-$124). They blame the imprecision of the estimate, but using some cleverness: The confidence interval's upperbound is $535. We saw earlier that a unit increase in percentile from KG test scores meant an increase of $90 in earnings. We also saw that those in a small classroom can expect to score 5 percentile points higher than those otherwise. Thus, being in a small classroom means $5 \times 90 = \$450$ increase in earnings. **Nevertheless**, the conclusion reached says class size is not powerful enough to detect earnings increases of some magnitude as of age 27.

– The effect of class size is present in the summary index, yet the individual factors in the index are insignificant. Side note: Having a 401k is indicative of having a good job; however, being assigned to a small classroom in KG shows no variation in that.

– Considering **subgroups of the sample** (e.g. males, females, whites, blacks, rich or poor (proxied by free lunch status)), they show consistent positive effects when assigned to a small class. Most results are insignificant, but typically those with greater estimates for test scores indicates large estimates across the other adult outcomes. Yet, the STAR experiment is not powerful enough to detect heterogeneity in the impacts of class size on adult outcomes with precision.

• **Third Results: Teacher and Peer Effects**

–

$$y_{icnw} = \alpha_{nw} + \beta_1 SMALL_{cnw} + \beta_2 z_{cnw} + X_{icnw}\delta + \varepsilon_{icnw}$$

– z_{cnw} is a vector of teacher and peer characteristics.

– On average, for kindergarteners, **having a more experienced teacher** translates to $1104 more in earnings. The effect is significantly reduced for grades 1-3. A complete explanation cannot be given on the variation for lack of data. To be clear, **Experience was not randomly assigned**, thus the result is more along the lines of: Placing a child in KG classroom with a more experienced teacher will yield better outcomes. It *does not* mean we should increase the experience of teachers to yield better outcomes. The difference in earnings could come from intrinsic characteristics of experienced teachers. Put more concretely, they have a graph showing a linear relationship between teacher experience and earnings. Papers found that teacher performance plateaus after a few years. Thus, it must be other factors driving this linearity, in the particular in the later years. They therefore conclude that early childhood teaching has a causal impact on long term outcomes but that they cannot isolate the characteristics of teachers responsible for this effect.

13

- There is odd evidence of large classes with experienced teachers showing significantly positive results.
- **Peer Effects** definitely impact the study environment and can make for conductive settings. Proxying for peer abilities using fraction black, fraction female, fraction eligible for free and reduced lunch, and age; they find nothing conclusive from the data with respect to peer effects on earnings (but predictive of test scores). Estimates are imprecise (i.e. small estimates with large standard errors). The lack of power to measure these observables comes from the design itself; randomization across classrooms means no variation across classrooms for these pre-determined variables.

- **Fourth Results: Impacts of Unobservable Classroom Characteristics**

  - There are certain aspects of the classroom that were not assessed and should be at least quantified. Teachers could have been more enthusiastic, students could have been disruptive — all confounding the effect of class quality on test scores and earnings. A **random effects model** is used to test whether such unobservables have an impact in the long run. The random effects model assumes the parameters are **random variables** and for it to work, the individual specific effects must be uncorrelated with independent variables (which was tested when we looked at the validity of the experiment). Such models are useful when trying to control for unobserved heterogeneity between multiple groups. Heterogeneity can be removed from the data after first differencing. Let the following section explain.

  - **A Theoretical Model of Class Effects (without peer effects)**
    * $s_{icn} = d_n + \sum_k \mu_k^S Z_{cn}^k + a_{icn}$
    * $y_{icn} = \delta_n + \sum_k \mu_k^Y Z_{cn}^k + \rho a_{icn} + \nu_{icn}$
    * Let d_n and δ_n be equal to school fixed effects to account for variation across schools. Let Z_{cn}^k be classroom characteristics and μ_k be their corresponding effects. Let ρ be the effect for intrinsic ability, a_{icn} (orthogonal to ν_{icn}). This term controls for the correlation between intrinsic ability and earnings ability. Thus, let ν_{icn} represent the component of intrinsic earnings ability uncorrelated with academic ability. Looking at these two equations, something to consider is the ratio: $\frac{\mu_k^Y}{\mu_k^S}$. There are cases where teachers might've taught to the test, increasing test scores, but not necessarily increasing earnings.
    * We can rewrite the equations above as:
    * $s_{icn} = d_n + z_{cn} + a_{icn}$
    * $y_{icn} = \delta_n + \beta z_{cn} + z_{cn}^Y + \rho a_{icn} + \nu_{icn}$
    * where $z_c n = \sum_k \mu_k^S Z_{cn}^k$ and z_{cn}^Y denotes the effects on earnings othorgonal to z_{cn}. **The reason for doing this** is to put it in a nice form to prepare it for the **Within-Transformation**. It is important to interpret β as the correlation between the class effects on scores and the class effects on earnings.
    * By definition of experimental design, the earnings ability or intrinsic ability of a student is orthogonal to the classroom characteristics. $\nu_{icn}, a_{icn} \perp z_{cn}$. Because of this condition, the equations can be estimated using OLS.
    * **To Analyze the Unobservable Attributes**:
      1. *ANOVA*: They perform an analysis of variance over classrooms; just to see if the outcome varies across classes by more than just chance. **Limitations** are: (1) doesn't answer the question of correlation β (2) This can only be done for KGs, since the later grades have less power.
      2. *Covariance between effects on scores and earnings*: This is estimated using the **Within-Transformation**. Since we cannot observe the class effects on scores, z_{cn}, they proxy it using peer-test scores, which were taken at the end of the year (done later). To continue the model, suppose:

14

3. $\frac{1}{I}\sum_{i=1}^{I} s_{icn} = s_{cn} = d_n + z_{cn} + \frac{1}{I}\sum_{i=1}^{I} a_{icn}$ The average was simply taken and of course, the fixed effect and the linear combination of dummies, z_{cn}, are constants so unchanged. We wish to subtract this from the **mean scores of the schools**. This is called the **First-difference**.

4.

$$\Delta s_{cn} = s_{cn} - s_n = z_{cn} + \frac{1}{I}\sum_{j=1}^{I} a_{jcn} - \frac{1}{IC}\sum_{c=1}^{C}\sum_{j=1}^{I} a_{jcn}$$

5. Notice the consistency — As $I \longrightarrow \infty$, $\Delta s_{cn} \longrightarrow z_{cn} = \sum_k \mu_k^Y Z_{cn}^k$. Thus, in finite samples, it's simply a noisy measure.

6. Substituting that in, we can then estimate $y_{icn} = \alpha_n + b^M \Delta s_{cn} + \varepsilon_{icn}$

7. This is estimate is naturally *upward-biased* since high ability students raise average scores and also earnings himself. To control or eliminat this problem, we can use a **Jackknife** measure, removing the self-observation and then, computing the mean of the sample. You then take the average of those means. $\Delta s_{cn}^{-i} = s_{cn}^{-i} - s_n^{-i}$

8. Thus, $y_{icn} = \alpha_n + b^{LM}\Delta s_{cn}^{-i} + \varepsilon_{icn}$, where Δs_{cn}^{-i} is **class quality**, defined by *peer-scores*, which estimates **the correlation between random classroom effects on scores and earnings**.

9. The fixed effects approach just makes dummies for the effects.

10. the Random effects approach makes it a normally distributed random variable.

* **Actual Attempt at measuring unobservables**
  · **Analysis of Variance** - $y_{icn} = \alpha_n + \gamma_{cn} + X_{icn}\delta + \varepsilon_{icn}$. γ is the class effect and y is any outcome. We test the null by an **F-test** for whether $\gamma_{cn} = 0$ for all c, n. To capture the magnitude, they compute the variance of γ_{cn} by running a random-effects specification (assuming γ_{cn} is normally distributed). Random effects is apropos because the covariance of class effects and the demographic variables X_{icn} are independent.

5 Returns to education: The causal effects of education on earnings, health, and smoking[4]

5.1 QAQR

To be written.

5.2 Dense Summary

References

[1] Manudeep Bhuller, Magne Mogstad, and Kjell G Salvanes. "Life-cycle earnings, education premiums, and internal rates of return". In: *Journal of Labor Economics* 35.4 (2017), pp. 993–1030.

[2] Sandra E Black, Paul J Devereux, and Kjell G Salvanes. "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital". In: *American economic review* 95.1 (2005), pp. 437–449.

[3] Raj Chetty et al. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR". In: *The Quarterly journal of economics* 126.4 (2011), pp. 1593–1660.

[4] James J Heckman, John Eric Humphries, and Gregory Veramendi. "Returns to education: The causal effects of education on earnings, health, and smoking". In: *Journal of Political Economy* 126.S1 (2018), S197–S246.

[5] Philip Oreopoulos. "Estimating average and local average treatment effects of education when compulsory schooling laws really matter". In: *American Economic Review* 96.1 (2006), pp. 152–175.