First paper - Draft

Jorge Cimentada

July 1, 2017

Contents

1 Introduction

2 Literature Review

Educational inequality and its long term impacts is a topic that has been prominent in the social science literature for the past 30 years. Are we destined to belong to the same social status as our parents? Is meritocracy over, if it ever existed? These questions have motivated, in conservative terms, much of the research on social mobility and social inequalities, for a good part of the 20th century and 21st century. When James Colem(1966) released his famous Coleman report it raised the topic to starlight by suggesting that family SES and student performance are tightly linked. Fast-forward today, we've understood the relationship much better, although far from being clear. The work of James Heckman, for example, has clearly showed that cognitive and non-cognitive inequalities are present before entering school (cite heckman). Moreover, these inequalities can be narrowed significantly if the children coming from unfavourable conditions are exposed to high quality learning environments. One particular finding coming from (cite heckman skill begets skill) is that the cognitive

level of a child in time t is a direct function of the experiencies from time $t - 1$. Although this sounds logical, it is often not acknowledged properly.

The implications of this cummulative model suggests that it is much less cost-effective to invest in children in time t that in $t - 1$. That is why the whole debate about narrowing inequality between polarized families should start in early education – it is the cheaper than investing in later education and it brings about the largest returns. Despite all of these findings and investments to reduce the gaps, we still find that the relationship between parental education or socio-economic status, is present in virtually all empirical studies of social mobility and inequalities (cite breen and goldthorpe, cite waldfogel what children need, cite bradbury 2015, cite chetty for recent evidence). What we don't know is how and why this relationship is formed from such an early age. Moreover, many have suggested different explanations and the topic has been very contested (duncan and brooks-gun 1997).

Social mobility has deepend our knowledge on the odds and relationship between intergenerational transmission of SES. The field of economics of education has also helped to understand that education is by far the investment that yields the best outcomes not only for the child and his/her family, but to the economy as a whole, as it boosts economic activity, it helps the labour market to improve job conditions and maintain a rapid economic growth (cite hanushk and woesmann study). We've become so intertwined with the subject that we've virtually spent most of our time studying the mechanisms through which this inequality comes about. Naturally, we want to do that to be able to reverse it and help disadvantaged children reach their fullest potential. Despite our good intentions, we've concentrated little on the magnitud of the gaps, specially in terms of cognitive abilities. In addition, since most research has concentrated on the American case, we haven't been able to learn which countries have the biggest or smallests gaps, and whether they've been narrowing or closing over time.

Virtually in all countries, be it developed or developing, there is inequality of opportunity. But there is considerable variation in the magnitud of this effect. For example, the Scandinavian countries, particularly Denmark and Sweden, prove to be very mobile countries (Esping-Andersen and Wagner, 2012; Breen and Jonsson, 2007; Shavit and Blossfeld, 1993).

Taking the Danish example, we learned a great deal about how to improve social mobility by first learning that Denmark is by far one of the most mobile countries, boosting the chances of upward mobility for less advantaged children (Bjorklund and Jantti, 2009; Jaeger and Holm, 2007). But the important finding came when we discovered the main reason behind their social elevator: its' educational system. For example, research by Esping and Waldfogel and Gosta's phd student on Denmark, show that the Danish early education system is particularly efficient at giving free high quality education, regardless of socio-economic status. Moreover, it tracks students into different curricula very late compared to other Europen countries Europe (age 16). The importance of this finding, is that we should step back and first study whether the effect is there, how big is it, and then proceed to find the causes behind it.

Recent research on educational inequality has found that achievement gap in cognitive abilities between High-SES and Low-SES kids has been widening over the years. Literature on the topic has mainly concentrated on studying the American case (Reardon, 2011) but other international evidence is emerging with a very similar landscape. The United States is usually the case of study because it is the only country where cognitive testing is present as early as 1940. This trend has allowed researchers to study achievement gaps in a lengthy period of time. In fact, using this information, (Reardon, 2011) is the first to investigate the evolution of the cognitive gap and the results are very surprising. Not only has the cognitive gap between the 90th income percentile and the 10th income percentile grown over time, but it has grown faster and to be wider than the highly contested white-black gap (cite study of white-black gap -waldfogel study). Surprisingly, the gaps have reversed and now a days we find that the income achievement gap is nearly twice as large as the black-white achievement gap. The widening of the achievement gap has been happening in parallel to the growth of income inequality,

although very suggestive, it is hard to link both things causally. However, a recent report by the Economic Policy Institute finds that the black-white income gap has been growing since 1979. If income inequality would be a strong a direct function of the achievement gap, we should expect for the white-black achievement gap to be widening rather than narrowing, as it is now (Magnuson and Waldfogel 2008) [1].

Reardon (2011) finds that the increase has occured predominantely from the 1970's until the 2000's. In fact, the hard numbers suggest that the gap widened by 40-50%. The author also estimates the rate of change using data from the 1940's and finds an even higher increase of 75%. Given that the studies before 1970 are less reliablle in terms of comparability and sampling design, the author computes all results for both before/after 1970. To provide a definitive answer to the size of the gap, Reardon (2011) concludes that the U.S gap between the 90th and 10th income percentile is at about 1.25 standard deviations in the year 2010. Using longitudinal data, Bradbury et al. (2015) find similar results. Their empirical analysis suggests that for American 14 year olds, the gap is above 1 standard deviation but lower than 1.25. Surprisingly, Duncan and Magnuson (2011) find similarly to the previous studies and confirm a gap of 1.25/1.50 standard deviations. [2]

In a follow-up study, Reardon and Portilla (2015) unexpectedly uncovered a new finding: the reversal of the trend. This follow-up study concentrated solely on the Kindergarten gap in which they took a snapshot of the achievement gap in Kindergarten in 1998, 2006 and 2010 and found that the 90th/10th income gap in readiness closed modestly. Furthermore, using data from fall and spring in the same kindergarden year, it seems like the gap in narrowed at a rate of 0.01 and 0.008 standard deviations per year for mathematics and literacy between 1998 and 2010. In contrast to (Reardon, 2011), we find that in a 30-year span the gap was systematicaly increasing at a rate of 0.02, something reasonably close to the previous estimates. Their results not only hold for the income achievement gap, but they also recorded a particular decline in the white-hispanic gap.

The American situation is a very distinctive one relative to Europe. The reasons the authors find a reversal in the trend could be numerous and should be studied very close. For example, the American educational system lacks any formal curricular differentiation *la* European style. It is true nonetheless that Reardon (2011) did not cover from the period 2000 onwards, which is specifically the trend they studied. In contrast, Reardon (2011) suggests that the reversal is likely due to the high increase of preschool enrollment. They suggest similarly that in this same period (1998 - 2010) the income achievement gap in early schooling enrollment decreased substantially. This explanation still needs further confirmation.

Motivated by these recent results, some authors have taken this analysis to an international context in order to discover between-country trends. The unique work of Bradbury et al. (2015) employs a unique comparative analysis of Australia, United Kingdom, United States and Canada. Their research design is very distinctive in that they use longitudinal data from children as early as age 2 and study the evolution of the achievement gap up until age 14 [3]. The core finding behind the book is that the American achievement gap is much wider than in any other comparison country, specially Australia and Canada. They find that once the achievement gap is present in early school entry, it doesn't seem to narrow or neither widen very much over the life course. In fact, they calculate that the schooling environment once they enter school can only explain about 30%-40% of the high school SES gap. This suggests that once the achievement gap widens before entering school, it carries a social-scar effect [4]. Once exception is the United Kingdom, which they find closed the gap in early

---

[1]The black-white achievement gap narrowed significantly for cohorts between 1950 and 1970.

[2]One important drawback of these studies is that they don't present the uncertainty of this estimates. Not necessarily to gauge their statistical significance, but to simply asses how much we can trust their accuracy. It could be that the gap is at 1.25 standard deviations, varying up to 1.75 and down to 0.80. Right now we don't know the upper/lower bound of this estimate, making it difficult to compare when new evidence arises.

[3]To the best of my knowledge this is not only the first study that uses panel data to study achievement gaps, but to also do it between countries

[4]However, schooling could be preventing the gap to widen even more, and very rigorous RCT's show that high quality

primary years. This can likely be due to the comprehensive schooling and also the public support by the welfare state in dimensions like health.

One limitation of their study is that they concentrate on countries which have a very particular educational structure, namely the fact that there is little formal stratification in terms of curricula. These four countries have no major jump in school selection, quite the opposite to the average European country. A thorough review by Van de Werfhorst and Mijs (2010) sheds some light on the subject. First and foremost, they gather substantive evidence showing that countries which have a highly tracked curriculum tend to have high levels of inequality, measured in terms of achievement gaps. Hanushek et al. (2006), use the Progress in International Reading Literacy Study (PIRLS), the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) surveys to gauge whether highly tracked countries do indeed increase inequality after students pass the age at first selection of tracking. The results suggest that early tracking increases educational inequality. While less clear, there is also a tendency for early tracking to reduce mean performance. Moreover, Micklewright and Schnepf (2007) using similar data but a different empirical strategy, finds that countries which have a high level of tracking, are distinctively unequal in the difference between the top 95th and bottom 5th performers. In fact, the difference in test scores between these two groups is about 10 times higher then the average annual gain of a year of schooling.

Another limitation of their study is that their analysis is based on four surveys that have significant differences in terms of questions, sampling and populations and cannot be easily compared. Their findings are very reliable but should be taken as suggestive at best. For this reason we should also pay particular attention to studies such as Chmielewski and Reardon (2016) and (Chmielewski, 2016) which have attempted not only to compare gaps between countries, but to evaluate whether there is a general increase in educational inequality in many countries using much more comparable studies. These studies tackle a completely different question from the above, namely to use study cross-sectional differences between many countries, instead of over-time analysis of student gaps. However, they do provide support for the overall finding that the achievement gap is certainly not narrowing over time.

Chmielewski and Reardon (2016), again using PIRLS, TIMSS and PISA, assess whether there are patterns of cross-national variation in the achievement gap. In other words, does the achievement gap differ between countries? Their work suggests that there is considerable variation in the achievement gap between top and bottom earning families across many developed countries. In comparison to the literature on achievement gaps, they find that the U.S has a gap of 1.20 SD in 2001 which increase to around 1.30 in the year 2006 while Germany has a decreasing gap from around 1.25 to 1 standard deviation in the same year-span. However, these numbers vary a lot and carry a great deal of overlapping uncertainty.

They go even further and link this achievement gap to several country-level indicators related to income inequality, school differentiation, central exams, among other things. The correlations are very suggestive as explanatory mechanisms but they are very cautios in drawing causality, specially because they are all reasonably correlated. However, one interesting question is whether these country gaps have evolved over time. With their data, they only have 3 countries which are present in all waves but they also have very few waves because their question of interest (income categories) was only asked in three waves. That is why their study is more about between country gaps rather than the magnitude and evolution of the gaps.

Chmielewski (2016), building on the work of Chmielewski and Reardon (2016) and Reardon and Portilla (2015) pooled together all the previously mentioned data, together with over 10 more studies ranging from the year 1964 until 2015 in order to discover differences between and across countries. With over 50 years of data, and over 100 countries, she finds that there seems to be a general pattern of increasing achievement gap. However, once she disentangles the relationship by country, she finds

---

schooling can indeed help ease the gap, in some instances even close it.

a sizable amount of heterogeneity, with some countries seeing the achievement gap narrowing, others no change at all, while others record a steady increase. This is revealing because it doesn't really pay off to look at general average once each country has their own distinctive gap and evolution. Another important finding is that the achievement gap is clearly no universal and should be studied in context.

One clear limitation of their study (as well as Reardon (2011)) is that the adjust for the age of each child in all studies. Although for their modeling purposes this is the right thing to do [5], they are masking age-specific achievement gaps by controlling for age, such as Reardon (2011) did. We clearly see in Reardon and Portilla (2015) that paying attention to age-specific gaps varies reasonably even in short periods of time.

In fact, the evolution of high/low SES gaps for preschool children might be much less marked than the same gap for high school children. The explanation, although very debated, has been gaining much support in recent years. In countries with high levels of curricular differentiation, the transition from early schooling into the tracking system has been found to increase inequality of learning (Hanushek et al., 2006). Moreover, the vast sociological literature on educational transitions systematically finds that tracking tends to foster between-track inequality rather than erode their differences by tackling their specific needs (Van de Werfhorst and Mijs, 2010). Based on this, we cannot simply assume that the achievement gap has been neither constant across cohorts (because there have been tracking reforms in many countries, introducing as well as elimination tracking structures) nor the same between ages, because tracking/no tracking might exarcebate the achievement gap.

Talk about how the achievement gap has been widening in Malaysia and South Korea (or Japan?) Is efficiency related to opportunity here? Is efficiency what explains the inequality of opportunity? Note how many of these studies haven't really concentrated on who is getting better or worse: top or bottom? Talk more about how countries with high social mobility has been linked to smallest achievement gaps Talk about Durpiez and Dumay and how there's not relationship between inequality income - inequality achievement in contrast to reardon and anna who find some relationship.

3   Research questions

With this being said, this paper introduces one novelty in the literature which is to evaluate the evolution of the high/low SES achievement gap in the past 15 years for all PISA participant countries. This is different from previous work because it concentrates solely on 15 year old children, and it attempts to capture the evolution of the achievement gap for many countries. The advantages of this study are twofold.

First, we concentrate on the evolution of the gap for only 15 year olds, which will serve as a comparison to the single year-country snapshot of Chmielewski and Reardon (2016) and the evolution of the Kindergarten gap in Reardon and Portilla (2015). As we've seen before, there are reasons to think that specific age-groups have seen changes in the achievement gap. Moreover, in almost all countries with a tracked curriculum children are either at or in the process of tracking by the age 15, meaning that we will be able to link whether tracked countries are the most variable in their evolution of achievement gaps.

Second, we will be able, for the first time, to study the evolution of the achievement gap for several countries other than the United States. This will help asses the mangitude of the inequality as well as the rate of growth/decline. Thirdly, this paper not only documents the size and changes of the SES gap, but studies with particular attention the source of the changes. That is, we study whether bottom performers are falling behind, if top performers are gaining advantages, or if both

---

[5]The differences in achievement could simply be due to changes in cognitive abilities across the lifetime. However, as we've noted before, Bradbury et al. (2015) find that the achievement gap is very stable across the life time

phenomenas are simultaneously at play.

More formally, We're interested in studying a) The size of the achievement in a comparative perspective and its relationship to overall achievement levels; b) Which countries are experiencing changes in the achievement gap; c) the rate at which each country-gap is widening/narrowing and; d) establishing whether the gaps are widening/narrowing because particular groups are getting ahead/behind.

We develop each question separately for more detail.

a) We want to test the notion that better performing countries have a smaller degree of inequality than other countries. (Van de Werfhorst and Mijs, 2010) emphasize that there is empirical evidence that suggests this. This pattern is not so obvious, however. For example, countries with high leves of tracking could maximize student performance, specially the high SES students, raising their overall performance and thus raising the national performance score. But if the bottom performers are not gaining at the same rate, then the achievement gap will inevitable grow resulting in a high performing countries with a widening achievement gap.

b) The seminal work of Reardon (2011) suggests that achievement gaps change, and they do so much quicker that we though. After recording a SES gap increase of about 40% in only 30 years, (Reardon and Portilla, 2015) stress that they also found a significant decrease in only 15 years of data. This shows that it is important to study the changes in the achievement gap over time. This will bring a useful comparison and pinpoint which countries are doing the best to reduce the gap.

c) We want to compare the percentage change at which the gap widened/narrowed from the first to the last year available. This will give us a general idea of the overall change over time, and will allow us to compare our estimates to the actual literature [6]

d) The widening/narrowing of the achievement gap has a source, which has been often studied to be related to everything from educational spending, income inequality, time allocation to students and preschool enrollment. The literature has concentrated very narrowly on whether the gap is increasing because the top performers are getting ahead, because both are distancing or because the bottom is falling behind. We shall pay particular attention to identifying the rate at which the top/bottom groups are evolving over time. This type of analysis is particularly useful in jotting down the mechanisms through which the gap is evolving. If the bottom performers are stagnated whereas the top is gaining ground, the efficiency-trade off dilemma of tracking might be much more credible than other explanations.

4   Methods

4.1   Data

To investigate the above mentioned questions I will be using the Programme for International Student Assessment (PISA). PISA is a survey carried out every three years that aims to evaluate education systems by testing the skills and knowledge of 15-year-old students. Currrently, PISA has six waves starting which started in 2000 up until 2015, where recently, over half a million students were tested in mathematics, literacy and science in over 70 countries, both developed and developing ones.

PISA collects data through a two-stage stratified sampling design. With the help of official governments, PISA randomly chooses 150 schools in each country, where they then randomly pick thirty 15 year olds to undertake the two hour tests. The sample size for PISA 2000 is 127,388, 276,165

---

[6]Although no study has performed this age-specific achievement gap for comparable tests over such a long time. Our results will serve as comparison for other studies that use age-specific groups, such as 4th graders.

for PISA 2003, 398,750 for PISA 2006, 515,958 for PISA 2009, 480,810 for PISA 2012, and 519,334 for PISA 2015. Together with the subject tests, PISA collects personal information from students, their families and their school environment (including teacher surveys), that serves as relevent background information that can be matched to the students performance. With the recent inclusion of PISA 2015, these six waves make up a time-series analysis of 15 years, enough to visualize changes in the structure of an educational system. None of the literature cited so far has used the last PISA wave, which was recently released in December 2016.

To identify a family's socio-economic status PISA collects several variables that measure different dimensions. Classically, they ask student's their parent's educational level. Scholars have considered this to be a reliable recall given that we expect fifteen year olds to know their parent's level of education (Reardon, 2011)(cite more here). This question has been asked in every wave and holds a somewhat similar coding across time, although the first two waves have a different coding. In spite of this, another limitation is the fact that parent's education is measured with the ISCED classification, something that has changed over time. For example, until PISA 2009, the preferred framework was ISCED 1997, whereas the next wave switched to the newly developed ISCED 2011 classification. Both these classification schemes have equivalent look-up tables, but this requires a detailed inspection of the codings. Aside from all these limitations, this is very reliable and important variable that measures an important factor of socio-economic status.

Another social background variable PISA collects is the International Socio-Economic Index of Occupational Status (ISEI). This variable captures the economic status of the family, the closest it can without asking for income information. This index variable was developed by Ganzeboom (cite) and it attempts to measure occupational status using a continuous measure. This indicator has been found to ourperform classical occupational classifications such as the Erikson-Goldthorpe-Portocarero schema (cite). It has been scaled for comparability between waves and some authors have used it for inequality studies finding expected results to be consistent with social origin literature (Chmielewski, 2016). PISA also includes a plethora of indicators on family wealth, home educational resources, the number of books in the home, among many other material resources in the household.

Yet one of the most relevant variables for our study is a composite SES index created by the PISA team. The index of economic, social and cultural status (ESCS) was created on the basis of the following variables: the International Socio-Economic Index of Occupational Status (ISEI), the highest level of education of the students parents, the PISA index of family wealth (which measured the material wealth of the family), the PISA index of home educational resources; and the PISA index of possessions related to "classical" culture in the family home (mainly about books in the household) (OECD, 2002). This variable, aside from capturing all relevant dimensions of SES, such as education, occupation, and material resources, takes care of transforming the variables into comparable metrics across waves. To the best of our knowledge, this is the first paper that uses the newly-released ESCS index (OECD, 2016) which was rescaled so that all ESCS indexes are suitable for over-time analysis [7] In other words, the ESCS index does not need any transformation or coding updates because it is ready for comparison over time.

Aside from SES, the other most relevant variables are test scores for Mathematics, Literacy and Science. PISA does not provide a single test result for each respondent. Instead, it provides a *series* of 'plausible values' that the child could actually score. As explained in the PISA manual (OECD, 2012), these are imputed values that resemble individual test scores and have approximately the same distribution as the latent trait being measured (the true distribution of the possible scores a student can achieve). Suppose we have μ_i, the average student test score in Mathematics for student i. Instead of estimating μ_i alone, plausible values estimate a distribution of possible μ's for student i, together with the likelihood of each μ_i based on the respondents answers on the test. This is defined as the posterior distributions of μ's for student i. The reason why we use this procedure is because estimating a single estimate μ_i is plagued with measurement error, among other types of bias (see Wu, 2005).

---

[7]These rescaled indices can be found in the PISA website under *Rescaled Indices for Trend Analyses.*

The number of plausible values for PISA waves are usually five (although ten for PISA 2015) random draws from this distribution. In practice each student has 5 scores for each test, that resembles this distribution. They are continuous, ranging from 0 to 500, with a mean of 250.

4.2   Data analysis

The aim of this paper is to identify and disaggregate country trends in the achievement gap for several countries. To represent the SES gap, most of the literature on achievement gaps has concentrated on indicators such as parental education, parental occupational status, income achievement gaps and actual SES achievement gaps (Fryer and Levitt, 2004; Hanushek et al., 2006; Saw, 2016; Bradbury et al., 2015; Byun and Kim, 2010). The actual calculation of the achievement gap varies substantially and different strategies have been implemented. For example, Micklewright and Schnepf (2007) calculates the difference in achievement by crudely subtracting the gap between the 95th and 5th percentile of the mathematics distribution. Although in principle you should be able to capture some type of SES effect like this, theoretically, it should be much more accurate to difference out the mean score of, for example, parental education or some other SES proxy. Saw (2016), for instance, used parental education as a proxy of SES, whereas Byun and Kim (2010) use a similar SES index as ours, but created by them.

Reardon and Portilla (2015); Chmielewski and Reardon (2016); Chmielewski (2016) used a different method developed by Reardon (2011) which we partially adopt in this paper. SES achievement gaps are measured as the difference in standardized achievement between the 90th and 10th percentiles of the chosen SES variable. The rule of thumb to choose the 90th, 50th and 10th percentile is arbitrary, others have used, for example, the 5th, 50th and 95th (Micklewright and Schnepf, 2007) while the literature has been following the standard set by Reardon (2011) to use 90th/10th percentiles.

For each country in each study, SES disparities in achievement are measured as the gap in standardized achievement between the 90th and 10th percentiles of each countrys distribution of each SES variable, following Reardons (2011) method for income achievement gaps.

The original strategy of Reardon (2011) is as follows: first, achievement is standardized (see below for a mathematical explanation of the standardization). We then use it to calculate the mean achievement (and standard error) for each category of the SES variable of interest (parent's education, income categories, etc..). "Category means are plotted at their percentile ranks and cubic models are fit through the points using weighted least squares. Finally, achievement at each countrys 90th and 10th SES percentiles is interpolated from the model" (Chmielewski, 2016). The result is a SES gap from an ordinal variable of interest.

As mentioned before, PISA does not provide a single achievement indicator. Instead, we take the median of all plausible values for each student [8], resulting in one single score.

To standardize the test score we fit a linear model

$$Y_i = \alpha + \beta_1 * AGE_i + \epsilon_i, \epsilon_i\ N(0, \sigma^2) \tag{1}$$

for each wave, where Y_i is the median student test score and AGE_i is their age measured in months (because all children should be 15 year olds) weighted by the student sample weights.

We then adjust $\hat{Y}_i$ by

---

[8]Since each plausible value is a random draw from a theoretical latent normal distribution of possible student achievement scores, the median should be precise in getting a central measure of the latent distribution.

$$\hat{Y}_i = \frac{\hat{\epsilon}_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}} \quad (2)$$

Where $\hat{\epsilon}_i$ is the residual for student i and the denominator is the root mean square error of the model.

This new standardized variable has a mean of zero. Standardizing the median test score solves the problem of comparability of gaps measured with different tests, and across waves. However, if the variance of academic achievement changes over time, then standardizing the overall score at each country-wave pair actually makes the variable more biased. That is, by standardizing we're forcing the standard deviation of test scores to be zero across all waves. But if the true deviations of the median academic achievement changes over time, then the estimated trend in the SES gaps will be underestimated, or vice versa.
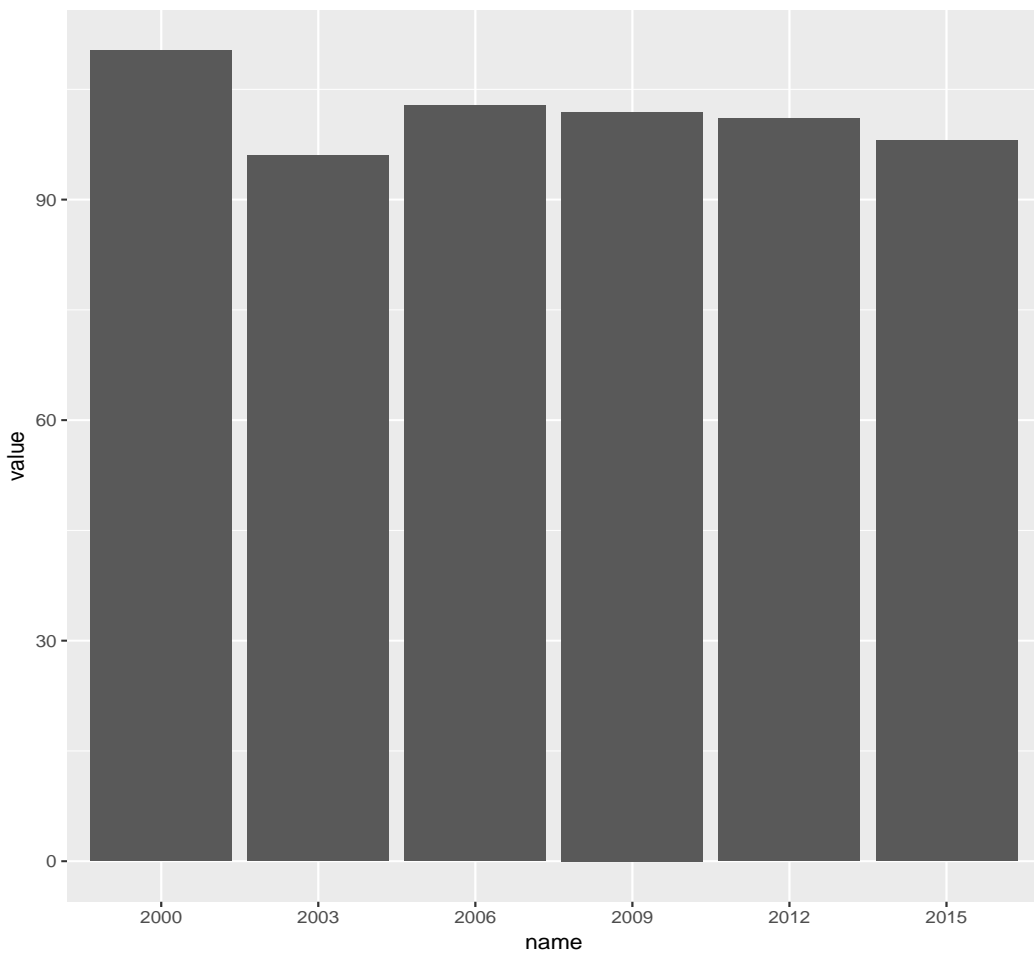
Inspection of the data suggests that it's something we shouldn't be deeply concerned with. The standard deviation of each wave seems to be following a very similar pattern with a not so drastic exception of the year 2000.

Another concern is that if the gaps at different waves are measured with tests that have different amounts of measurement error, then the amount of bias will not be the same in each measure of the gap. This can be very misleading and suggest erroneous intepretations regarding trends in the of the gaps over time (Reardon, 2011). PISA has tried to make sure the tests are comparable across waves (PISA 2012 technical report), but we still have to adjust for this imprecision.

In order to correct gap estimates for measurement error, I adjust student test scores by

$$\hat{Y}_i = \hat{Y}_i * \frac{1}{\sqrt{r}} \tag{3}$$

where r is the reliability score of the wave. [9]. Each PISA survey provides a reliability indicator which we use accordingly. This yields estimates of the true gaps, and eliminates any bias in the trend that may arise from differential reliability of the tests.

After constructing the adjusted test score measurement, we estimate the SES dummy by calculating the weighted 90th and 10th quantiles for the SES composite index, and then generating a dummy of 1 for those above (including) the 90th percentile and 0 for those below (including) the 10th percentile.

We then fit

$$Y_i = \alpha_{j[i]} + \beta_{j[i]} * SES_i + \epsilon_i, \text{ for } i = 1, 2, ..., n \text{ within country } j \tag{4}$$

where SES_i is whether the student is in the top/bottom SES dummy and we allow both α_i and β_i to vary by country j. We implement this separately for each wave and weight by the wave-specific student sample weights.

Finally, we calculate the SES achievement gap for each country by extracting the fitted α and β for each country j and calculating the difference of the predicted score of High and Low SES.

$$\begin{aligned} \text{High SES}_j &= \alpha_j + \beta_j \\ \text{Low SES}_j &= \alpha_j \\ \text{SES gap}_j &= \text{High SES}_j - \text{Low SES}_j \end{aligned} \tag{5}$$

We also calculate the standard error of this difference and use it to create uncertainty intervals in the estimation of time trends. We fit a multilevel rather than a linear model because by pooling the information together, we weight countries appropriately to their sample size. Given that including the SES dummy reduces the sample size considerably, we want to be able to estimate each country-difference as accurately as possible. For further exploration, we also compute the 90th/50th and 50th/10th SES gaps following the same method outlined above.

---

[9]Other proceedures multiply each country by their own reliability measure for each year-subject pair (Chmielewski, 2016). Unfortunately, PISA 2000 did not provide any reliability measure separately for each country and PISA 2015 has to yet release their own. At the moment of writing this paper, they were unavailable. For this reason we implement the analysis following the original work of Reardon (2011)
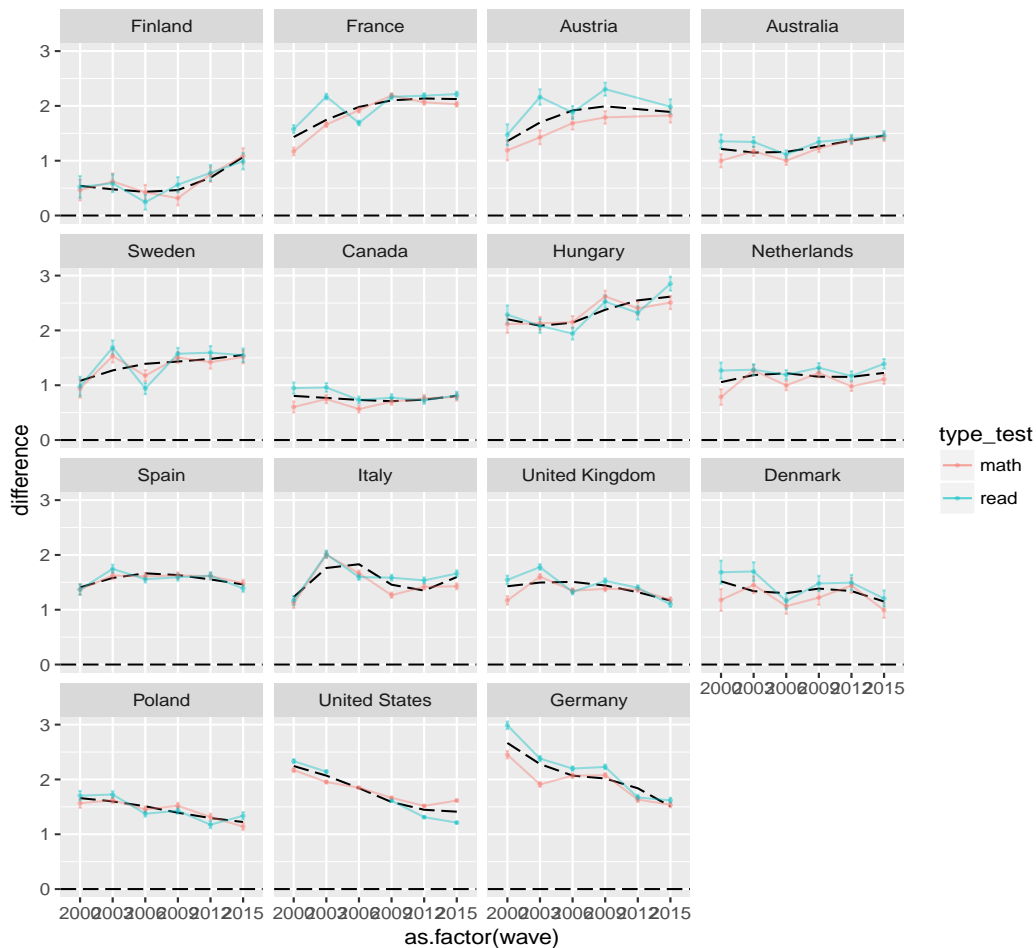
5   Analysis

5.1   Descriptives
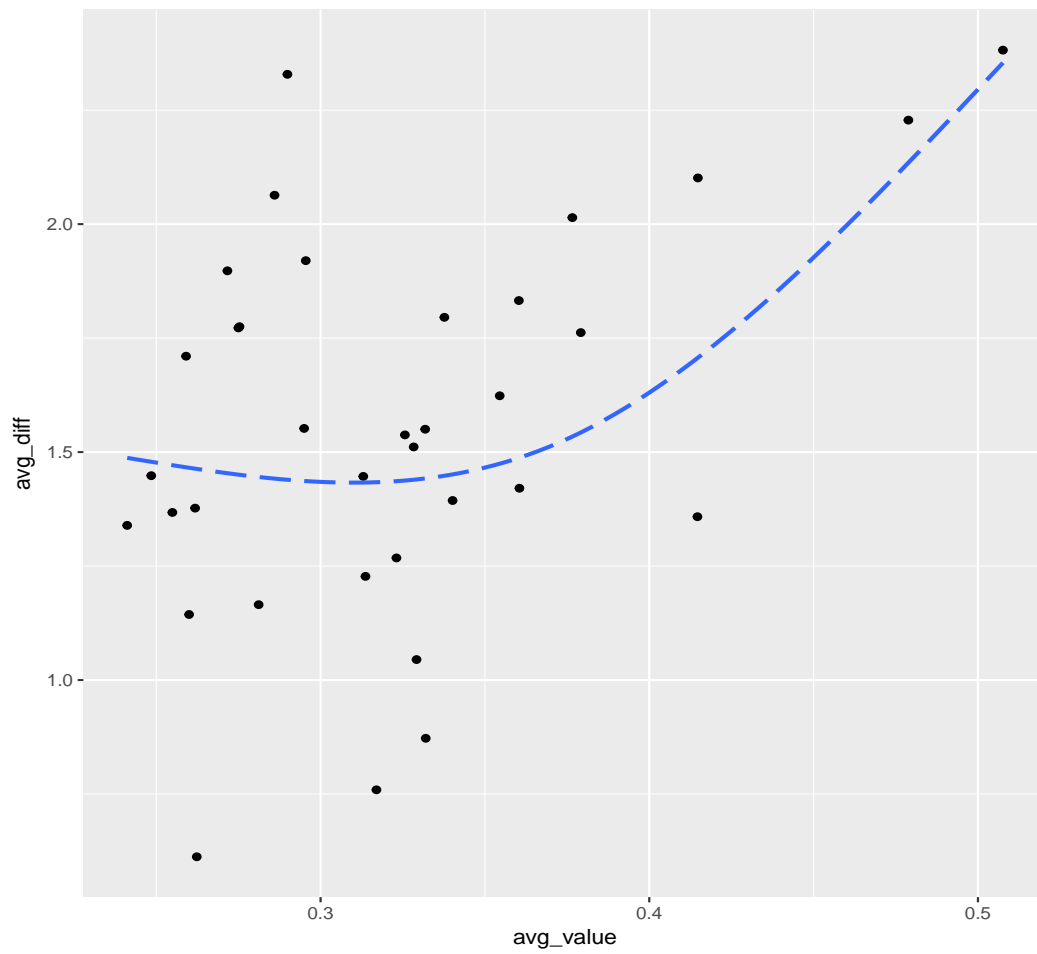
5.2   Results

Calculate the difference between 90/10 and then the ratio of this with the annual gain of one year of schooling (Micklewright and Schnepf (2007) did it).

We start by look at the achievement for some countries. We plot both mathematics and reading gaps for each country and also a cubic trend spline for both test scores pooled. Some countries have increased their achievement very strongly. For example, France, Austria and surprisingly Sweden. In mathematics, France increase the gap by roughly 0.9. This pattern happens similarly for literacy with an increase of 0.6. For such a short period of time, the magnitud of the increase is reasonably big.

Given that no one has estimated the evolution of the gap we can't cross-check how other empirical estimations put France at. However, we can make some educated inferences. For example, Micklewright and Schnepf (2007) also finds that France is a low dispersion country, specifically in the year 2000.

However, we don't know what happens over time. But Ballarino and Bernardi (2016) did study the relationship of social origin effects and found that they increased very much in France in the last decades.

Other countries have reasonable increases such as Austria and Finland. Aside from these countries, some other countries have no change at all. Specially, Canada, Netherlands and Spain. Canada excels here not only because the gap has been stable over time, but because it has the smallest gap of all countries presented here. It's nearly 0.5 standard deviations in 2000 and it increased only by 0.2 in 2015. We can't really say Italy has had a significant since it has a sort of wiggly pattern that looks like it cancels itself out.

Furthermore, we do have some countries which show a decrease in the SES achievement gap. United Kingdom, Israel, Poland, United States and Germany show some steep downward slopes. Most notably are the U.S and German cases. These two countries show a very high level of dispersion in the year 2000 with SES gaps of over 2 standard deviations. However, both countries reduced the gaps by -0.6 and -0.9 respectively. Their distinctively big gaps in 2000 also show up in the work of

Micklewright and Schnepf (2007).

Analyzing this graph we might naively conclude that these trends are not very steep and they should not be very important to consider. However, remember that the Y axis is measured in standard deviations. Small changes are actually very huge in practical terms. Take the case of Sweden. The slope doesn't look that steep but in reality it increased the gap from 1 standard deviation in 2000 to around 1.5 standard deviation. The actual difference is of about 0.6. Now that we know that, the trends of Poland, United States, France and Germany are particularly shocking.

Something reassuring is that mathematics and reading follow basically the same trend across all countries. This means that the result is not an artifact of chance. However, it's important to disentangle where is this gap originating from. Is this because the top are improving why the bottom decreases? Or is it that the bottom is catching up? Up next we plot the same graph but show the divergent patterns between high/low SES origins.

The top line belongs to the high SES group whereas the bottom line belongs to the low SES group. The middle line is a summary line showing whether one of the groups is growing/shrinking faster than the other. We find plenty of variation in these patterns. For example, in the United States the top seems to be equalizing much stronger than than the bottom is catching up. The U.K seems to be following the same path as the United States as well. On the other hand, in Poland the gap seems to be closing because the bottom SES group is catching up much faster. The Netherlands shows a similar pattern as the U.S, U.K patterns but upon closer inspection the explanation is very different. We see that the slope is negative (like in the previous two countries) but that's because the low SES group is decreasing at a greater rate than the high SES is decreasing.

Finally, we pay attention to the flat summary lines. These lines should be interpreted with caution because it doesn't mean that the gap is not increasing. Denmark, Belgium, Netherlands, Norway and Spain show a flat line because the gap is growing very little, if at all. On the other hand, Sweden, France, Finland, New Zealand, Austria, among other countries, show a flat line because both groups are distancing themselves at a very similar rate. there's a group of distinctive groups where both low/high SES rates are very great resulting in the summary line being flat.

Ths source of this gap can be deconstructed into two broader gaps. We want to know if these gaps have increased because the 90th SES group has distanced itself from the 50th SES group or because the middle has been also distancing itself from 10th SES group. Up next we plot all the combinations of the 90th/50th/10th gaps and pay attention to the magnitude of each gap. For example, we're interested in searching which gap is bigger between countries, the 90th/50th or the 50th/10th? Moreover, are some of these gaps contributing more to the 90th/10th gap we saw earlier. Because we visualize three plots for every country, we only plot a selected number of countries that represent the overall patterns from the previous plots [10].

---

[10]United States and Germany show a decrease, France shows a steep increase and Denmark shows no significant change.
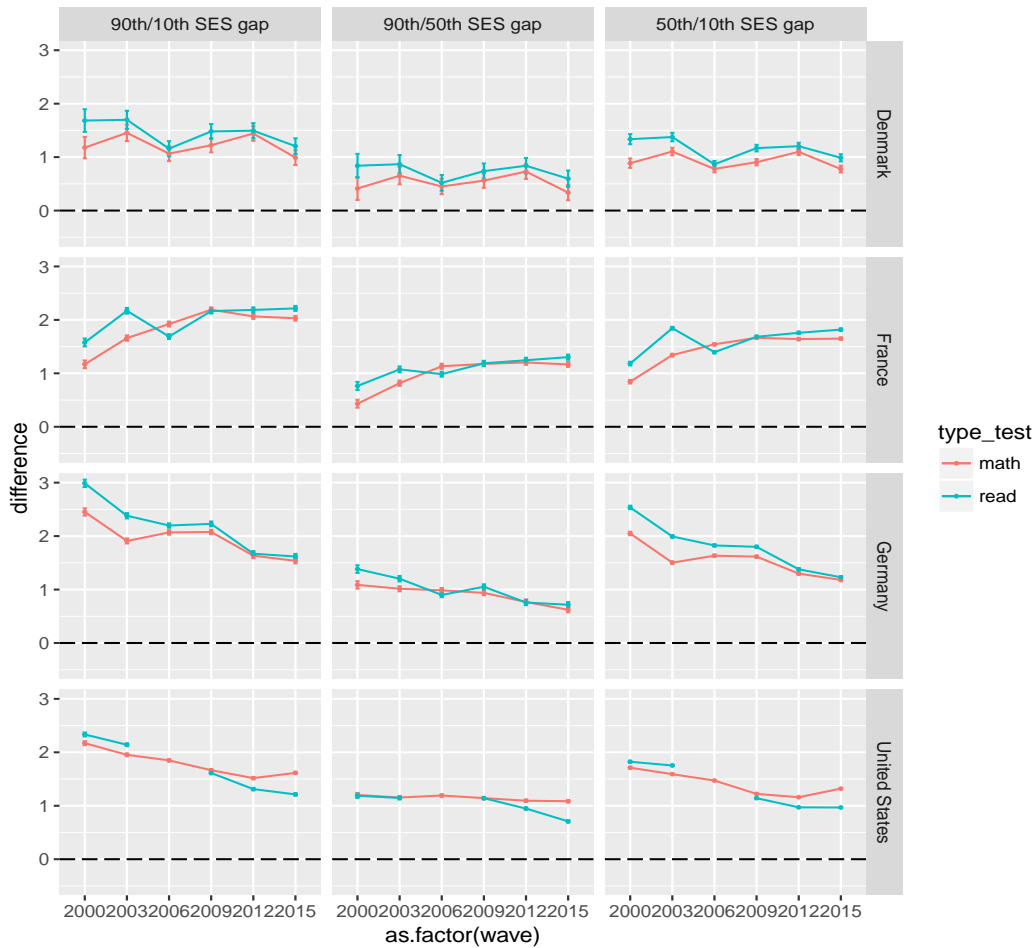
We start off with Denmark. Generally speaking, we find that the 50th/10th gap is contributing more than the 90th/50th gap as the gap is bigger for both tests. Furthermore, we find, just as in the 90th/10th SES gap, that there isn't a significant decrease, with the exception of reading in the 50th/10th SES gap which seemed to decrease from nearly 1.5 to 1 standard deviation. France shows a similar pattern as Denmark in that the 50th/10th SES gap is the main contributer and that the three gaps have the similar trend.

Germany and U.S, on the other hand, show a very different landscape. For example, Germany, shows that the 90th/10th gap is primarily driven by the 50th/10th SES gap and that the slopes of the 90th/50th gap is much weaker than the other two gaps. This means that Top/Middle SES groups have shrunk much less than the other groups. The big catching up is actually coming from the Middle/Bottom SES groups. However, we don't know if this is because the bottom is catching up or that the middle is coming down. However, based on the gap plots from above, we can infer that it is because the bottom is catching up rapidly because against the 90th SES group, it is increasing steeply every year. The U.S shows a different pattern from all other countries. Both the 90th/50th/10th gaps seem to contribute rather similarly to the overall 90th/10th gap (although 50th/10th is marginally higher) but more interestingly, the 90th/50th gap has stagnated in the last 15 years. The top/middle groups have a relatively big gap and it hasn't changed at all. On the other hand, the steep decrease we saw for the U.S is coming exclusively from the 50th/10th gap shrinking. The reason, nonetheless, might be different from the German case, where it's probable that the bottom is catching up. Here, as in the 90th/10th gap, the middle might be 'dumbing' down closer to the low SES group.

But, similar as before, we don't know which group is causing the decrease/increase of the gap in the 90th/50th and 50th/10th gap. For that we plot the evolution of these gaps below.
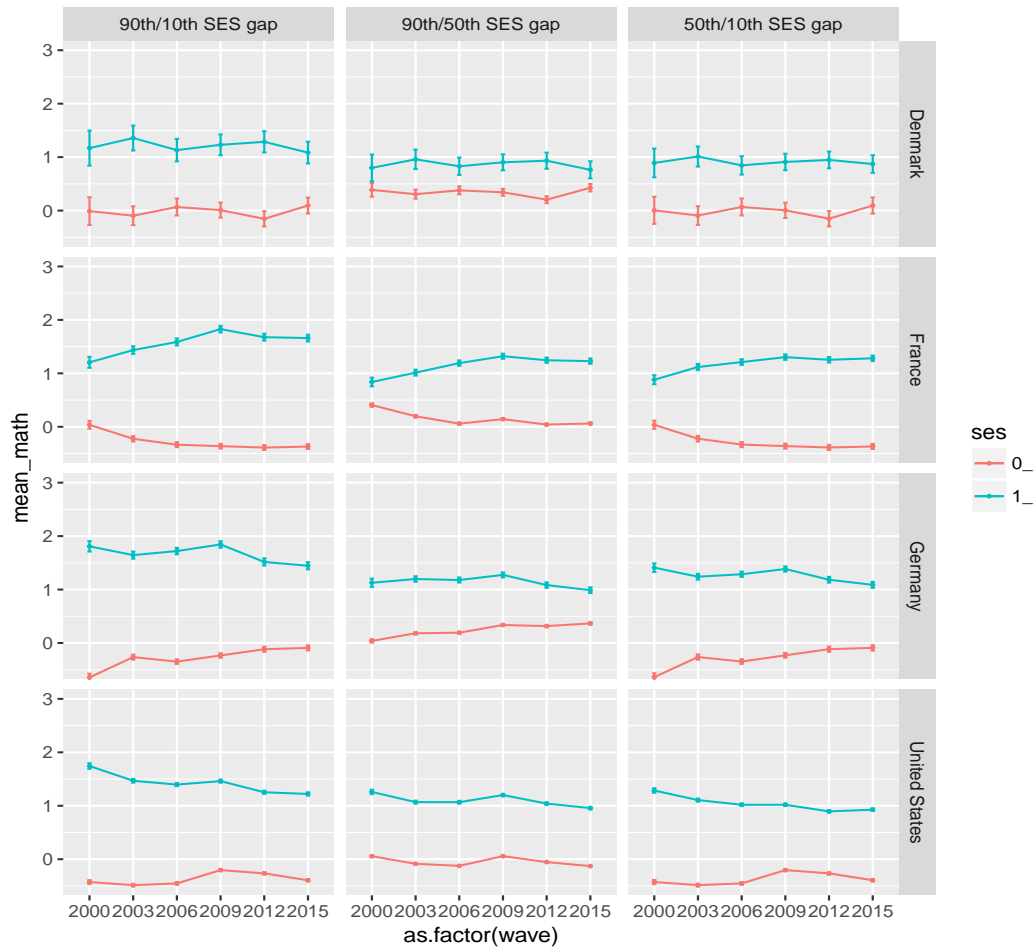
With Denmark there's no average change across all years. However, for France, we see that the gap is growing much faster for the 50th/10th SES gap. For Germany we see that the gap is closing in all three groups although much more strongly for the 50th/10th SES gap. It looks like the 50th/10th gap is mainly driving the changes in the gaps for most of the countries that actually see a change over time. For the U.S, there's very little change in the 90th/50th SES gap, with most of the change coming from the 50th/10th gap.

Given that we have the data for the average performance for each country and the achievement gap, we can also investigate whether there's an efficiency/equality tradeoff and if it stands over time. One particular topic that has been researched in the past is whether improving the overall cognitive development of a country compromises the distribution of the countries overall performance. For example, in order to improve the average score in, let's say, mathematics, tracking can improve the average score by making children in the higher trackers perform better, raising the overall performance. This, however, can come at a price. If tracking actually benefits the top but doesn't increase the performance of the bottom tracks then we see a net increase of inequality accompanied with an increase in the overall performance.

For all country-years pooled we see a very strong correlation but contrary to what we would expect. We find that countries with high performance have also a very low level of inequality. In fact, even if we separate it by year.

In fact, we see a strong correspondence. In countries like Germany and France we see quite a functional form between both inequality and performance. This pattern is also seen in in the UK, Austria, Australia and New Zealand. Poland and Italy are quite interesting cases because it shows how inequality has been decreasing that much whereas average performance has been improving notably. On another hand, only Finland, Canada and Netherlands (three very equal countries) have their

average performance in standard deviations above their level of inequality.
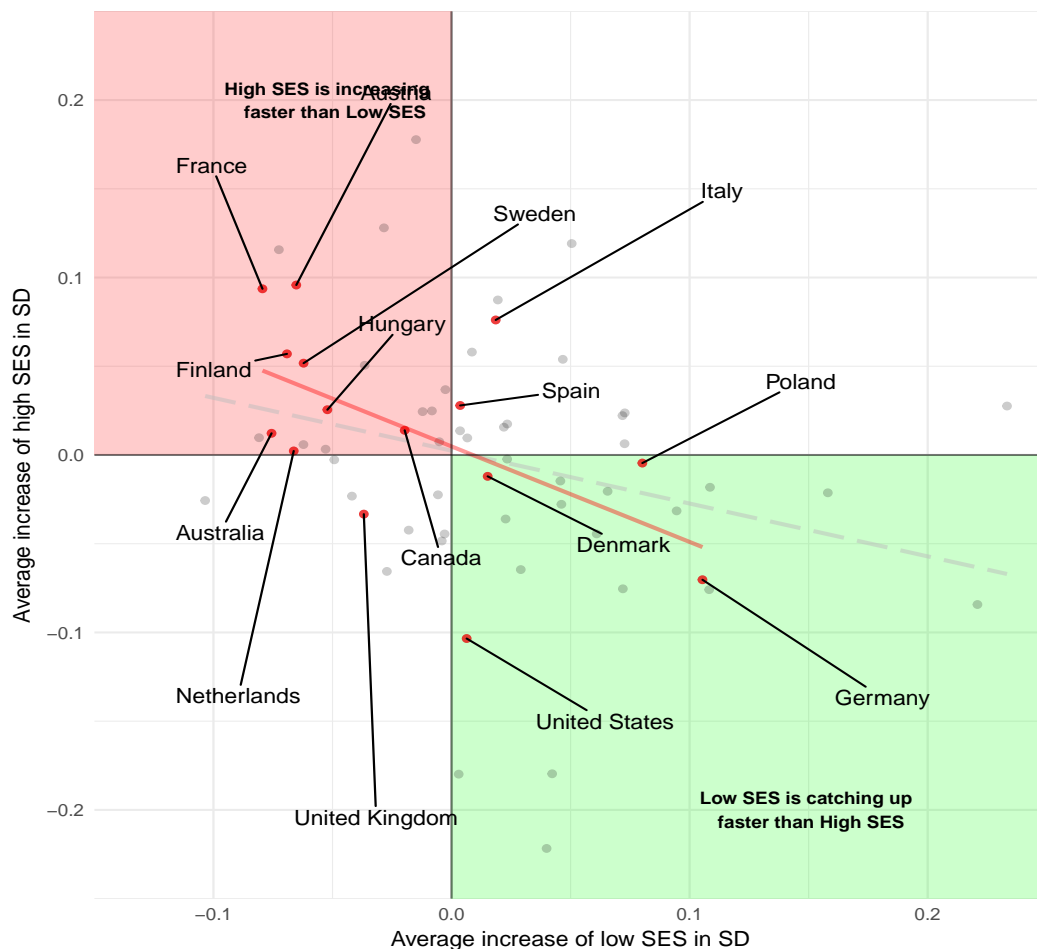
In a more convenient way, we can see that there is a lot of variance in the rate at which top/bottom groups are changing. We see that there are more countries where the inequality is increasing (red panel) but there's also a fair share where Low SES is catching up faster, such as United States and Germany. We see a negative relationship where the faster the low SES catch up the slower the High SES decrease. Quite an interesting pattern because we see very few countries where low SES are decreasing (so going even further down) and High SES is also decresing (bottom left panel). On the other hand, we do see more ocuntries where the both groups are increasing (top right panel). We plot all remaining countries in the back, together with a trendline and we see that the relation holds for all PISA countries.

Figures showing "90/10", "90/50", and "50/10" panels of Average % difference by Country.
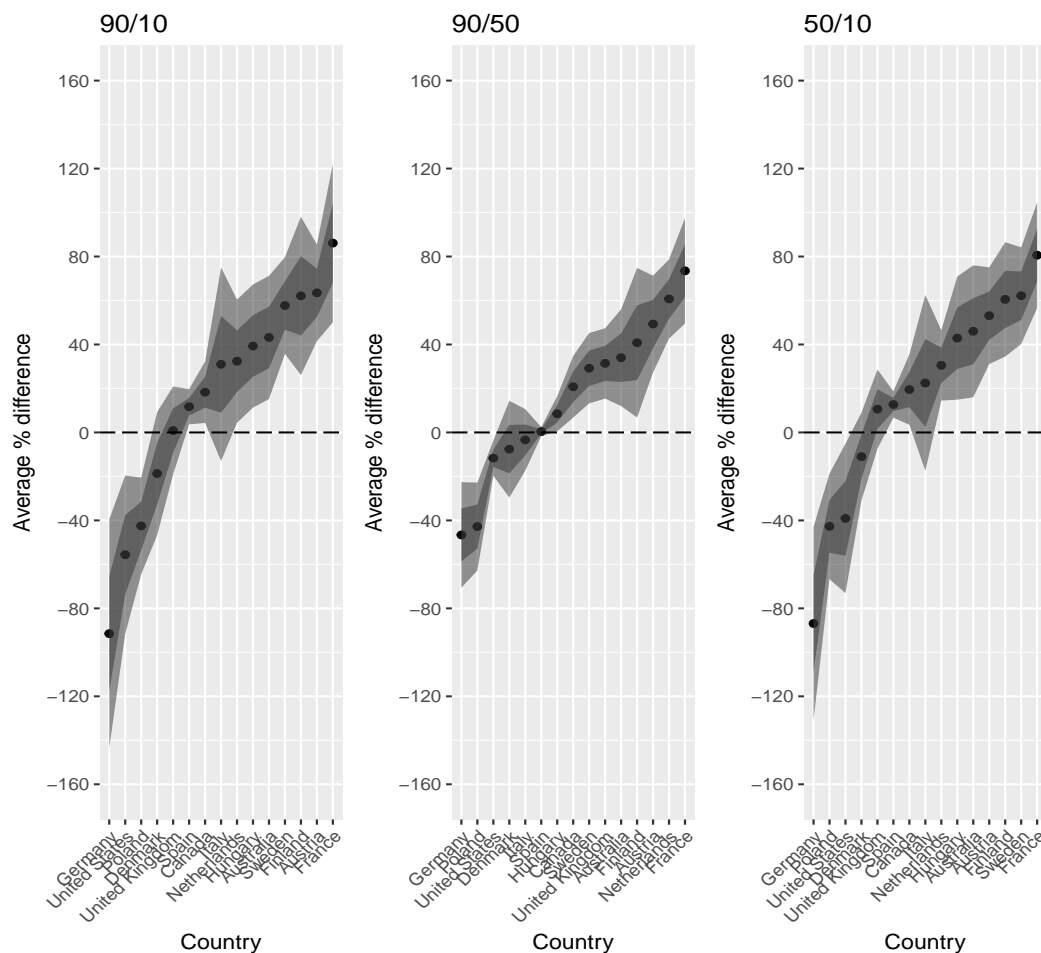
References

Bradbury, B., Corak, M., Waldfogel, J., and Washbrook, E. (2015). *Too many children left behind: The US achievement gap in comparative perspective.* Russell Sage Foundation.

Breen, R. and Jonsson, J. O. (2007). Explaining change in social fluidity: Educational equalization and educational expansion in twentieth-century sweden 1. *American Journal of Sociology*, 112(6):1775–1810.

Byun, S.-y. and Kim, K.-k. (2010). Educational inequality in south korea: The widening socioeconomic gap in student achievement. In *Globalization, changing demographics, and educational challenges in East Asia*, pages 155–182. Emerald Group Publishing Limited.

Chmielewski, A. K. (2016). Changes in socioeconomic achievement gaps in international comparison, 1964-2012. In *2016 APPAM International Conference*. Appam.

Chmielewski, A. K. and Reardon, S. F. (2016). Patterns of cross-national variation in the association between income and academic achievement. *AERA Open*, 2(3):2332858416649593.

Duncan, G. J. and Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. *Whither opportunity*, pages 47–70.

Esping-Andersen, G. and Wagner, S. (2012). Asymmetries in the opportunity structure. intergenerational mobility trends in europe. *Research in Social Stratification and Mobility*, 30(4):473 – 487. Consequences of Economic Inequality.

Fryer, J, R. G. and Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2):447–464.

Hanushek, E. A. et al. (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. *The Economic Journal*, 116(510):C63–C76.

Micklewright, J. and Schnepf, S. (2007). *Inequality and Poverty Re-Examined.* Oxford Univ. Press.

OECD (2002). *Education At A Glance: OECD Indicators 2002.* OECD Publishing.

OECD (2012). *PISA 2009 Technical Report.* OECD Publishing.

OECD (2016). *PISA 2015 Results (Volume I).* OECD Publishing.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, pages 91–116.

Reardon, S. F. and Portilla, X. A. (2015). Recent trends in socioeconomic and racial school readiness gaps at kindergarten entry. *Stanford, CA: Center for Education Policy Analysis. Retrieved December*, 20:2015.

Saw, G. K. (2016). Patterns and trends in achievement gaps in malaysian secondary schools (1999–2011): gender, ethnicity, and socioeconomic status. *Educational Research for Policy and Practice*, 15(1):41–54.

Shavit, Y. and Blossfeld, H.-P. (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries. Social Inequality Series.* ERIC.

Van de Werfhorst, H. G. and Mijs, J. J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual review of sociology*, 36:407–428.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3):114–128.