

Do peers matter? Only if you need them (and meet them)

Andreas Bjerre-Nielsen* & Mikkel Høst Gandil†

September 2018

Abstract

Educational policies are often motivated by a desire to increase equality of opportunity and social mobility. However, the empirical findings on what actually works are often highly contested. This paper investigates one tool available to policy makers: the mechanism whereby students are allocated to schools. Using administrative changes in school attendance boundaries, we investigate the causal effect of changes in the expected socioeconomic characteristics of cohorts on student outcomes. We find weak overall effects. However, we show that this null result is driven by the combination of low compliance and effect heterogeneity. We find a strong, positive effect of stronger intended peers for disadvantaged children who are likely to enroll in the intended school. We document a tightly estimated zero effect for children with strong socioeconomic backgrounds. This non-linearity in effects suggests that associational redistribution may increase overall efficiency by improving outcomes for disadvantaged children at little cost to other children.

*University of Copenhagen, andreas.bjerre-nielsen@econ.ku.dk

†University of Copenhagen and the Economic Council of the Labour Movement, mga@econ.ku.dk

1 Introduction

Much economic research has documented large discrepancies in outcomes of individuals dependent on place of residence in childhood or school attendance. Chetty et al. (2014) document large discrepancies both in inequality and mobility across the US as well as within much more narrowly defined regions. In addition, Graham (2018) (among many others) documents strong correlations between outcomes and socioeconomic composition of neighborhoods. Thus, to borrow a phrase from Graham (2018), “place matters”. In this paper, we look at one possible explanation for such correlations, the importance of schools.

Schools play two fundamental roles in the childhood human capital formation: firstly, they provide educational inputs such as teachers, textbooks and other educational resources. Secondly, they provide a context wherein children interact. If child outcomes are affected by other children, often referred to as peer effects, then the student composition in schools may be an important part of the explanation for the variation in performance between schools. However, educational inputs and peer composition may interact in highly non-linear ways, making the educational production function notoriously difficult to identify. It is, in other words, difficult to identify whether exposure to better peers generate better outcomes or merely correlate with unobserved school inputs. Even so, if these features continue to correlate, it may matter little for policy purposes. By manipulating one characteristic, policymakers effectively manipulate other characteristics as well.

A more pressing concern is unobserved household characteristics. If unobserved parental investments correlate with peer composition we may attribute variance to schools while neglecting the role of parents. Thus, any assessment of the effectiveness of schools must handle household sorting across neighborhoods and schools.

To identify the importance of schools, one should ideally randomly allocate children to schools. This paper investigates a quasi-experiment approximating such an experiment in Danish primary schools. The Danish sys-

tem allocates children to primary schools through geographically defined attendance boundaries. Parents can therefore implicitly choose the school for their child by moving within the attendance boundary. However, these boundaries change over time, thereby allocating children to different public schools. If these changes are unexpected, then households are unable to sort in the short term. Families who thought their children would attend the same school can therefore now act as control groups for each other.

We measure the change in school as a change in expected socioeconomic composition. The effect we seek is, therefore, best thought of as an intention to treat effect (ITT). We find little overall effect of a change in intended student composition on test scores. However, we document a social gradient. We find precisely estimated zero effects of exposure to stronger peers for children of high socioeconomic status. The effect is positive for children in families with low socioeconomic status.

In Denmark, households have a degree of choice and can choose to opt out of their assigned school, a feature of the Danish system documented in Bjerre-Nielsen and Gandil (2018). This is important, as a change in the mechanical student composition may have little effect if the *a priori* likelihood of attending the designated public school is low. We, therefore, expect the magnitude of ITT-effects to be a function of the likelihood to be exposed. In the spirit of Gruber and Mullainathan (2006) we construct a prediction model for the *baseline* likelihood to comply with the school assignment. To fit this model, we use data on children living within unchanged attendance boundaries. We show that the prediction model performs extremely well out-of-sample. With this model, we predict baseline compliance for the children exposed to the exogenous changes in attendance boundaries. We then interact this prediction with the mechanical change in characteristics and document overall small and insignificant effect. However, when we interact socioeconomic status with the baseline compliance, we find that the effect for disadvantaged children is almost solely driven by children, who are likely to comply in the first place. For these children, we find large and positive effects.

Our method is not informative about the exact mechanism, but the results indicate that resourceful households are able to compensate for any adverse changes in access to specific schools. These parents could choose to compensate by moving their children out of the public sector. Though we do find evidence of this, it cannot explain the precisely estimated zero result, as the *intended* peer composition and the *actual* peer composition of these children correlate.

As we estimate positive effects for disadvantaged children and no effects for children from strong socioeconomic background, there is scope for associational redistribution. A policy whereby one mixes children from different backgrounds is likely to increase efficiency in primary education. In other words, outcomes for vulnerable children can be improved at little cost to other children.

We would like to get closer to the mechanism of these effects but are limited by our research design. Ideally, one would use instrumental variable techniques (IV) to rescale the intention-to-treat effect and thereby obtain estimates of the parameters of the production function. We argue that this is not feasible for two reasons. Firstly, the educational production function is complex and likely have multiple parameters of interest. We however only have one valid “instrument”, the attendance boundary change. Thus, we find the required exclusion restrictions implausible. Secondly, we argue that the assumption of monotonicity of treatment with respect to the instrument is not satisfied. An important way that socioeconomic strong households may compensate for an adverse shock is through opting out of the assigned schools. If households exposed to a *negative* composition change end up opting for private school, the child may end up with a *positive* change. We observe strong indications of this effect in our data and therefore limit ourselves to the ITT. Importantly, given that the government uses attendance boundaries to allocate children to schools, the ITT is the effect of interest for policymakers.

The issue with compliance is twofold. Firstly, it reduces the possibilities to obtain knowledge about the educational production function. This

paper documents that taking compliance seriously limits the knowledge of peer effects we can obtain from such natural experiments. Secondly, it is a constraint for the opportunities to effectuate policies to improve outcomes in primary school by manipulating student compositions.

The paper proceeds as follows: In the next section, we briefly introduce the peer effects literature along with the methodological challenges. We then proceed to present the Danish primary school allocation system in section 3 and our econometric model in section 4. In section 5 we present the data and section 6 describes our enrollment prediction model. We present the results in section 7 and section 8 concludes.

2 Peer effects and relevant literature

How education works is one of the most studied fields in economics. The literature on estimating the educational production function is enormous and impossible to cover in this short paper. We refer to Hanushek et al. (2016) for a general introduction. We will in this section focus on the research on peer effects, in other words, how students affect each other.

Theoretically, Epple and Romano (1998) and Durlauf (1996b) have shown the importance of peer effect in explaining linkages between inequality and intergenerational mobility: If socioeconomic strong children have a positive effect on other children, well-off families have an incentive to "hoard" strong peers to the detriment of less well-off families. The authors thus underscore the importance of the nature of peer effects for the desirability of a policy seeking to affect socioeconomic compositions of schools, a policy which Durlauf (1996a) refers to as *associational redistribution*. For associational redistribution to increase overall performance and thereby the efficiency of educational production, Hoxby and Weingarth (2005) argue that one needs complementarities between own and peer ability. In the absence of these complementarities, one merely reallocates outcomes and the policy therefore solely serves a goal of redistribution. For the policy to be desirable, the policymaker must in this case be inequality-averse. These studies all moti-

vate a careful analysis of heterogeneity in peer effects for determining the scope of associational redistribution to increase aggregate welfare.

The empirical research on peer effects has a long history dating back to at least Coleman et al. (1966). The study of peer effects is made difficult by the amount of knowledge required about assignments of peers, shared inputs and group heterogeneity, brought into focus by Manski (1993) and Angrist (2014) among others. Using network analysis concepts, these studies show that partitions of agents which can be described by a block adjacency matrix can deliver little in terms of causal identification of peer effects, often referred to as the reflection problem. A block matrix is the exact way to describe classroom interactions, where children only interact with other children in the same classroom. However, Blume et al. (2015) and Bramoullé et al. (2009) have recently shown that these worries in many cases are overstated.¹ However, even in the absence of the reflection problem, the identification of peer effects entail other challenges such as endogenous network formation and unobserved covariates.

Peer effects are often thought of as an externality and are therefore not tradeable in traditional markets, see Sacerdote (2011). The channels through which children may affect each other can be numerous. Children with high socioeconomic index (SES) may affect the learning of low-SES children by directly interacting in the classroom. However, it may also be that high SES parents are able to demand better teachers for their own child, thereby incurring a positive externality on the low SES-children. In some cases, this will not be thought of as a peer effect. In the present context, however, this is a mechanism whereby attendance boundary changes may affect the outcomes of children. In this paper, we are therefore working with an expansive definition of peer effects.

At least since Manski (1993), much of the literature has been focused on

¹Blume et al. (2015) show that some of these worries may be overcome if the network has non-transitive triads. In other words, if agent A and B are friends, then if there exists another agent C who is a friend of A but not of B then identification of a (structural) model is possible.

linear-in-means models, where the influence of peers can be described as a linear function of average peer characteristics. These models are often given a structural interpretation, for example as a function of a game, see Blume et al. (2015) for an example. Carrell et al. (2013) and Sacerdote (2011), however, show that these types of models may not be a good representation of the true social interactions as they do not account for homophily within groups and selective interactions. Using a randomized experiment, they show that when students are in very stratified groups, they interact less with students different from their own type. Similar findings have been documented by Hoxby and Weingarth (2005) and Imberman et al. (2012). An implication of these findings is that policy aimed at increasing efficiency by allocating students to affect mean characteristics may have negligible or even negative effects in practice. Accordingly, it warrants that empirical work focus on effect heterogeneity. Hoxby and Weingarth (2005) use instrumental variable estimation (IV) to get at effect heterogeneity. They use a mechanical peer composition as an instrument for true peer composition and exploit other moments than the mean to investigate the heterogeneity.

From the short outline of the literature above it is evident that any empirical approach to estimating peer effects must take endogeneity, group assignment and heterogeneity very seriously. Before presenting our econometric approach we briefly present the Danish primary school system in the next section.

3 Primary school allocation in Denmark

In what follows, we briefly describe the institutional context and our source of variation.² Danish public primary schools are run by the municipalities. The usual assignment mechanism is school attendance boundaries (SAB), which associate children with schools based on their residential location. Every child has a right to be admitted to the public school to which they are

²This section mirrors the corresponding section in Bjerre-Nielsen and Gandil (2018).

associated. These boundaries change over time due to administrative decisions, which are usually taken in the spring before schools start in August.

Once enrolled, the child is not directly affected by changes to attendance boundaries. Hence, the boundaries are therefore only important for households at the time of enrollment. The parents have the option of enrolling the child in another public school if there is sufficient capacity. Defining sufficient capacity is a decentralized and somewhat opaque process. Furthermore, parents can choose private schools, which are publicly subsidized and thus fairly cheap. The private schools are, however, often heavily oversubscribed. In Bjerre-Nielsen and Gandil (2018) we investigate the choice of these two options in detail and find that especially high-SES households avoid allocations to schools deemed undesirable through both outside options. In the short term, however, the public option completely dominates the private option as a means of avoidance.³

We now move on to present our econometric approach and the limits to what we can infer in the present setting.

4 Econometric model

A large literature shows that choice of residence is affected by the local socioeconomic makeup, see Baum-Snow and Lutz (2011) for an example. Because schools are local goods, the socioeconomic makeup of the school will often resemble the area in which the school is located. This is especially the case when school attendance boundaries (SAB) delineate who has a right to attend which school. The implication is that similar households may locate in the same area. This poses obstacles for identifying a causal relationship between peer composition and performance. If a low-SES child in a rich area performs better than a child with similar SES in a poor area, we cannot readily conclude that the peer group explains this variation. We simply do not

³This is important, as it diminishes issues of non-random attrition due to the lack of test scores for students in private school. See the last part of section 7 for sensitivity checks.

know why the two children live in different places in the first place. In other words, the non-random residential location decision makes any estimated effects of school peers susceptible to omitted variable bias.

In this paper, we will rely on natural experiments induced by administrative changes in attendance boundaries. In order for such an approach to yield unbiased results, we assume that household sorting is fully controlled for if we know the school to which the households *sought* to associate themselves before the administrative change. Assuming that households do not anticipate the boundary changes, we can control for residential sorting by including “original SAB”-year-fixed effects in our regressions.

To characterize the relocation, we need a way to summarize the variation in the peer characteristics induced by the boundary changes. We, therefore, construct a “mechanical” school characteristic in the vein of Hoxby and Weingarth (2005). We observe all children, their characteristics and addresses at age 5. Using this data, we construct our measure of mechanical school characteristics by averaging over the set of children living within the SAB to which their addresses will be associated *two years later* when the children are supposed to be enrolled in primary school. We focus on the average socioeconomic index and call this measure the *mechanical school-SES* (MSS henceforth). We describe the construction of the index in section 3. For the MSS to be perfectly correlated with the actual characteristics, we would need stable SABs and restrict households from moving, delaying school start or choosing other options than the local school. Thus, the MSS will naturally differ from the actual measure to which the child is exposed. It is, therefore, best thought of as an instrument for actual peer composition.⁴

In the language of the LATE framework, we estimate a reduced form regression of this instrument, the mechanical schools SES, directly on the outcome of interest. We perform regressions of the following kind:

⁴However, we later argue that it is not entirely clear for which characteristic the MSS is an instrument. While we use IV parlance, we do not actually use IV.

$$Y_{iss't} = \alpha X_i + \beta MSS_{s'(t-k)} + \mu_{s(t-k)} + \varepsilon_{iass't} \quad (1)$$

where $Y_{iss't}$ is the outcome of child i living at an address which belongs to district s at time $t-k$ but to district s' at time t . This outcome is assumed to be a function of child's own characteristics, X_i , which does not vary over time. We are interested in β , which is the parameter on MSS . As mentioned, we control for residential sorting by including a SAB-year fixed effect. The only variation in MSS therefore comes from administrative boundary changes, where children within the original attendance boundary are allocated to different schools and therefore have different realizations of MSS .

If we were to interpret equation (1) as a structural model it would leave little room for associational redistribution. This is because the model does not allow for complementarity between own and peer characteristics. We, therefore, estimate variations of the model with heterogeneity in own socioeconomic status. It should therefore also be clear, that we do not interpret equation (1) as reflecting a structural relationship between peer characteristics and own outcomes.

From reduced form to IV? The reduced form model in (1) amounts to the aggregate effect of the change of SAB characteristics on outcomes. We do not know how much of this effect comes from actual peer characteristics and how much comes from other factors related to the intended changes in peer composition. The natural next step would therefore be to estimate the first stage and combine this with the reduced form estimate to obtain an IV-estimate, as is done by Hoxby and Weingarth (2005). However, this approach is fraught with peril.

From the literature we have good reason to expect effect heterogeneity. In this case, an IV-regression is best interpreted as a local average treatment effect (LATE). For such an estimate to be unbiased, we need three assumptions; exogeneity of the instrument, an exclusion restriction and monotonic-

ity in the instrument. In the present context of estimating peer effect, we have faith in the exogeneity of attendance boundary changes. We, however, doubt that the other two restrictions are fulfilled.

Firstly, if we run a first stage, where MSS is used as a predictor of actual cohort composition, and scale our reduced form estimate with this first stage, we assume that it is actual cohort-SES which explains the variation in performance induced by the boundary change. However, household-SES correlates with other socioeconomic factors such as ethnicity. As we only have one instrument, we cannot separate out factors correlated with SES. We therefore only view our SES-index as a proxy for the set of factors which might affect child outcomes. Further, we do not know whether the link between SES and performance is due to actual *interaction* or whether variation in peer composition induces different teaching strategies or other educational inputs. If this is a function of actual peer composition this is not a problem, but if it is a function of *expected* composition, then the exclusion restriction is not valid.

Even if the exclusion restriction is valid, the issue of monotonicity remains. This assumption normally receives relatively little attention but is crucial in our framework. If the residential address of a household becomes associated with a school with sufficiently low expected SES, the household may opt for an outside-option such as private school or the possibility of enrolling in other public schools with sufficient capacity. If this outside option has a higher SES than the original district, monotonicity breaks down; variation in the instrument may induce treatment of the opposite sign from what the instrument would suggest. This would imply that some groups would have negative weight in the computation of the average local treatment effect, which then ceases to be meaningful. This issue is a function of the behavioral responses of households documented in Bjerre-Nielsen and Gandil (2018). In appendix A we show direct evidence that the assumption of monotonicity is indeed invalid in our setting. In other words, even if the exclusion restriction holds, the presence of heterogeneity in treatment combined with defiance prevents us from interpreting an IV estimate as a local average treatment effect.

These two issues are sufficiently severe for us to abstain from estimating IV regressions. We therefore focus exclusively on the reduced form regression and estimate Intention to Treat effects (hereafter ITT). Importantly, from the policymakers perspective, this may not be an issue as the “instrument” is policy relevant. Regardless of the household responses, we can estimate the effect of redrawn attendance boundaries. Thus, in our setup, the reduced form regression is informative about the effect of the policy tool, which municipalities have at hand.

4.1 Identifying affected children

Having shown that there are fundamental problems with an IV approach, we need alternative strategies to understand the mechanisms for a given reduced form effect. The implication of the discussion above is that not all types of households are affected the same way by the instrument and that the behavior of the households may be non-monotonic in the MSS. If an effect goes through the actual exposure to peers, we would expect larger effects for children who comply with the assignment to primary school, compared to children who do not attend the intended school. Using the actual enrollment compliance, however, will constitute a problem of being a ‘bad control’ as the likelihood of compliance depends on the treatment, i.e. changes in the mechanical school-SES. Because the choice to comply is not random, conditioning on enrolling would introduce selection issues, which are not present in the baseline setup. To circumvent such issues, we take an approach inspired by Gruber and Mullainathan (2006). We seek to identify the effect on those children who are *likely* to comply by the assignment mechanism in the absence of a change to the attendance boundary, rather than those who *actually* comply.

To do this, we divide our sample into two subsamples. The children in the first sample experience changes in their school association through attendance boundary changes. This sample constitutes our main dataset and we call this the *causal analysis sample*. The other sample, the auxiliary sam-

ple, consists of children who experience no changes in the boundaries. We use this dataset to construct a prediction model for enrolling in designated school given school characteristics at age five and family background. Note, that we only claim to have exogenous variation in our causal analysis sample. Hence, we do not consider our prediction model to be causal. We now move to describe in general terms how we construct the prediction model.

4.2 Predictive modelling

We construct our model based on household and neighborhood variables and expect many of these variables to interact in highly non-linear ways. We could model this with interactions of the variables in a linear regression framework. However, a central worry would be that we construct a model that predicts well on the estimated data but with poor out-of-sample performance. To avoid this overfitting problem, we apply a *Random Forest* algorithm (RF), introduced by Breiman (2001). This algorithm relies on fitting classification trees which handle non-linear data well, see Wager and Athey (forthcoming). However, instead of fitting a single tree, RF constructs multiple trees on different subsamples of the variables and observations. The method thereby reduces the risks of overfitting. Using the multiple decision trees, RF forms a prediction based on the average classification of the estimated trees. We can interpret this average as a probability.

A drawback of RF is that it is less transparent than a linear model. Evaluating the impact of a variable on the prediction may, therefore, require some form of simulation. In order to validate the model, we use out-of-sample prediction. We construct our model on eighty percent of the data in the auxiliary dataset and use the remaining twenty percent for model evaluation. Importantly, this entire process is kept separate from our actual analysis using changes in attendance boundaries.⁵

⁵We use the implementation of RF in ‘scikit-learn’ using the Classification and Regression Tree algorithm (Breiman, 1984; Pedregosa et al., 2011). We estimate the Random Forest with the hyperparameters recommended by Breiman (2001). Most importantly this means

It is important to compare the prediction model to the alternative strategy of simply including covariates directly in the regression of test scores on the intended peer characteristics. This would implicitly control for compliance but would not be informative about the mechanisms. We would therefore not know whether any sensitivity to controls is caused by actual effect heterogeneity or differing levels of compliance. However, our prediction is essentially a non-linear function of socioeconomic variables, as we use these variables to construct the prediction. Thus, the use of this prediction together with controls implicitly exploit the same variation twice. We return to this issue in section 6.2.

5 Data and measurement

This section describes the choices made in structuring our data and closely mirrors Bjerre-Nielsen and Gandil (2018) as the identifying variation used in these two paper is the same.

We base our analysis on Danish registry data covering the years 2008-2015. From the registries, we obtain information on households such as child gender, ethnicity, number of adults in the household, parent income and education and, importantly, geographical location. This data is of very high quality and cover the universe of Danish children and their parents. From the CPR-vej-registry we obtain information on school attendance boundaries (SABs). The municipalities report these boundaries as sections of roads. Submission is voluntary, and for the municipalities own use in their administrative IT systems. Statistics Denmark do not verify the data accuracy. We clean the SAB-data and merge it unto the register data using the variables *kom* and *opgikom*.

We sample all 5-year old children who are observed two years later, en-

using at most a number of variables for each tree equal to the square root of the total number of variables. We estimate our models using 1000 trees to ensure minimal overfitting. The mean of the votes cast by the thousand trees is interpreted as the predicted compliance rate.

rolled in primary school. Most children should be enrolled at age 7, even if parents have chosen to postpone school by a year. For outcomes, we use test scores from low-stake tests taken in *public* primary schools, which we obtain from the registries. Tests are taken almost every year, alternating between Danish and Math in the early years. We focus primarily on tests in Danish language, taken in second grade – the earliest possible test. Within a subject, students are scored along three different dimensions. We take the mean of these three dimensions and rank within cohort. Our measure of performance, therefore, follows a uniform distribution on the unit interval. As private schools do not take the tests, the sample suffers from non-random attrition, A point to which we will return in section 7.3.

To summarize across the multidimensional socioeconomic space we construct a socioeconomic index as the first component from a principal component analysis. We rank this measure across cohorts to achieve a uniform index. The socioeconomic index is increasing in income, employment and high cycle education as expected. See the appendix of Bjerre-Nielsen and Gandil (2018) for further details. We refer to this index as *household-SES*. We use an average of the SES as a proxy for the socioeconomic student composition in public schools and for calculating the mechanical school-SES - the central measure of interest in the analysis as described in section 4. After calculating school measures, we exclude non-Western children from the sample. We elaborate on the reasons for this restriction in section 6.2.

5.1 Descriptive statistics

Table 1 presents descriptive statistics for the causal sample. The second row is the mechanical school socioeconomic index (MSS), demeaned at the “attendance boundary”-year level. This is our identifying variation. Even in the causal sample, we see that the mean is zero. Appendix Figure B.1 shows that the distribution is fairly symmetric. Household-SES is slightly higher than 0.5, which means that the households living within unstable boundaries are somewhat higher on the socioeconomic spectrum than the

	Mean	Std.	Median	N
MSS	0.550	0.114	0.552	15,648
MSS, demeaned	0.000	0.035	0.000	15,648
Household SES	0.590	0.284	0.625	15,648
Female	0.488	0.500	0.000	15,648

Table 1: Descriptive statistics for causal sample

This table describes the data excluding children of non-Western descent, as they are taken out of the regression analysis. In the second row we demean the MSS with the SAB-year fixed effect.

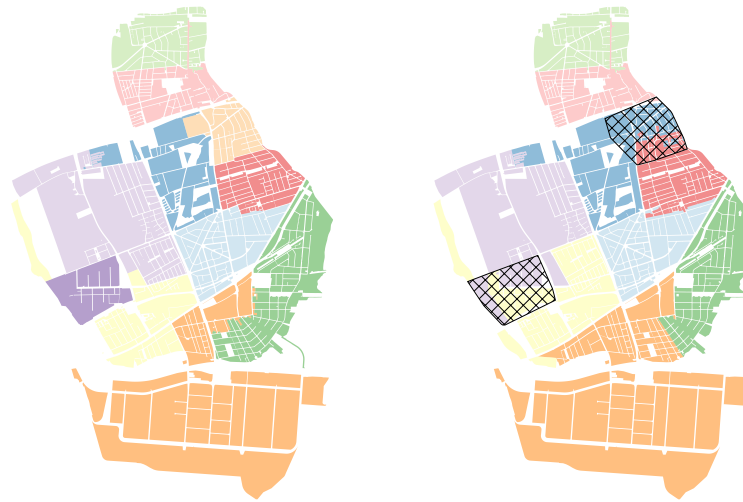
average household with children in primary school.

5.2 Empirical example

To clarify the kind of variation we exploit, we briefly illustrate a case present in our data. Figure 1 shows the school districts of the municipality of Hvidovre in the Copenhagen metropolitan region in 2011 and 2012. Between these years, two schools close down. Thus, households living within the attendance boundaries of now the closed schools unexpectedly become associated with new schools. If these households have unobservable preferences for schools, which may correlate with outcomes, we can expect these to be shared among all the households within the same original boundary. The households can therefore act as control groups for each other. By including a fixed effect for those households, which have children of the same age and thought they would send their kids to the same school, we reduce the risk of omitted variable bias caused by residential sorting.

The mechanical SES is essentially calculated as the population distribution in the map in Figure 1a with the attendance boundaries of the map in 1b superimposed.⁶ Observe that besides the two school closings, other boundaries also change. With our measure of mechanical SES, we pick up this variation as well.

⁶In reality, the base map would be 2010, as we measure the MSS using the boundaries two years later, rather than one year later as the example in Figure 1.



(a) School districts, 2011

(b) School districts, 2012

Figure 1: District changes in the municipality of Hvidovre

The figures depicts the school districts in the municipality of Hvidovre in the autumn of 2011 and 2012. The hatched areas in 2012 show the convex hull of two closed schools. In order to enroll students who would live in districts of the now-closed schools, a range of other changes was made. See section ?? for a description of the district data. Some areas differ from the official documentation. These areas are mostly not populated but some measurement error occurs. The map is constructed by merging addresses on to official geo-data. In the analysis we use addresses directly to bypass mismeasurement of geographical entities.

Variations in the mechanical SES thus occur for two reasons. Firstly, some children become associated with a new school. Secondly, there are children who are continuously associated with the same schools, but where other students become associated or disassociated with that school, thus altering the expected socioeconomic composition of students. All children within changed boundaries, therefore, experience “treatment”, regardless of whether the actual school assignment changes.

6 A model of enrollment

We proceed with developing a model for predicting enrollment of households into the intended school. As mentioned in section 4, we estimate the model on the auxiliary sample without exogenous variation and use the model to infer compliance in our main causal sample, where there is variation in the attendance boundaries. An illustration of the sample-splitting process is shown in figure 2.

The model forms a probability of attending the school to which the address of a child is associated based on two kinds of variables. Firstly, we record household variables such as family SES, income, type of residence, gender and ethnicity. Secondly, we use average cohort characteristics of the children within the same attendance boundary at age 5.

6.1 Evaluating the prediction model

We begin by inspecting which variables are important for the predictions in the model. Figure 3 displays the importance of the variables used. Household income rank and SES are the most important variables for determining predictions in the model. Furthermore, most of the cohort averages are more important than the remaining household level variables.

For all observations in the test sample, we predict the probability that the child enrolls in the school to which they were associated at age 5. We achieve an accuracy of 83.8 pct. in the out-of-sample prediction, which we regard as high.⁷

To visualize the fit, we bin household into SES-percentiles and take the average of the probabilities. We compare this to the mean of actual compliance for the test sample. Figure 4a shows the result of this exercise. The

⁷We decompose model accuracy by error type. Our model has a precision score of 77.8 pct. and is measured as the ratio of those who actually enroll in the intended school over those who are predicted to enroll. The rate is around 5 pct. higher than the mean rate of enrollment in the test data set of 73.5 pct. Our model has a recall accuracy of 90.7 pct. which is measured as the share predicted to enroll among those who actually enroll.

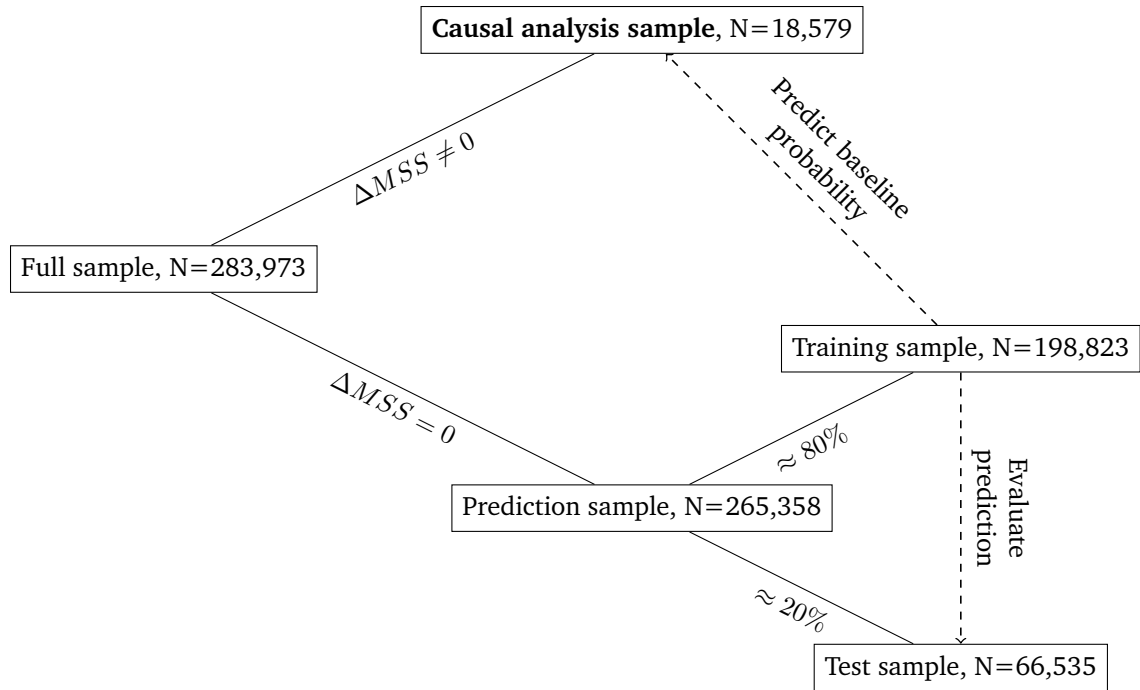


Figure 2: Definition of samples

This figure documents the sample definitions. We divide the full sample into two. The first sample, the causal analysis sample, contains those children who experience a change in the expected mechanical SES from age 5 to age 7 which imply a change in attendance boundaries. The children in the second sample, the prediction sample, experience no change in the expected mechanical SES. We further subdivide the prediction sample into two, a training sample and a test sample. We train our prediction model for assignment compliance on the training sample and evaluate the fit on the test sample. After evaluation we then use the models to construct a new variable in our causal analysis sample; a baseline probability of compliance.

black line is the actual mean of compliance in the test sample. For low SES-households compliance is in general low. We speculate that this feature is most likely due to urbanization and therefore the geographic density of schools. Approaching the middle of the SES-distribution, compliance rises and thereafter falls slightly. The functional relationship between household-SES and compliance is therefore not monotonic. Nevertheless, our prediction model is able to fully capture this, as evidenced by the blue line in Figure 4a. The apparent discontinuities are to be expected as SES is an index

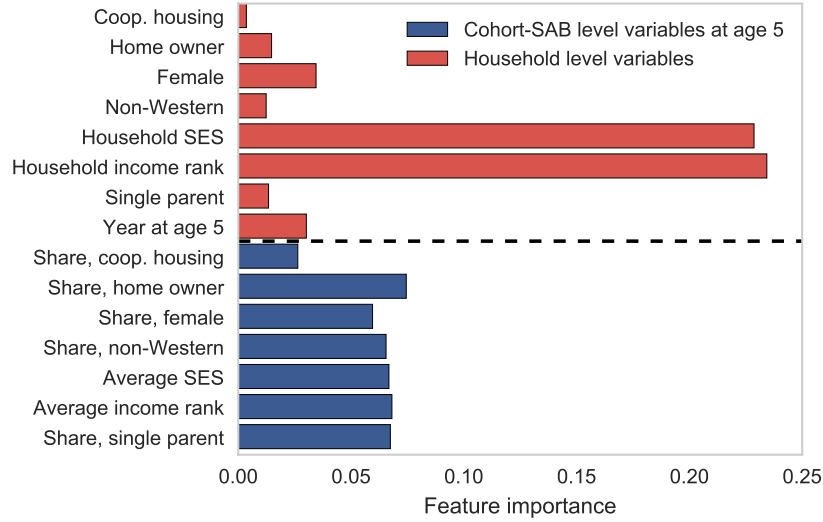
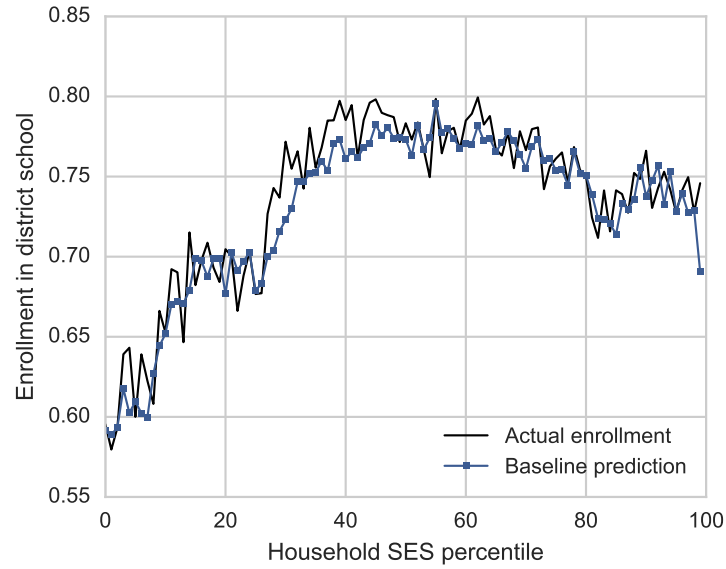


Figure 3: Feature importance in fitted prediction model

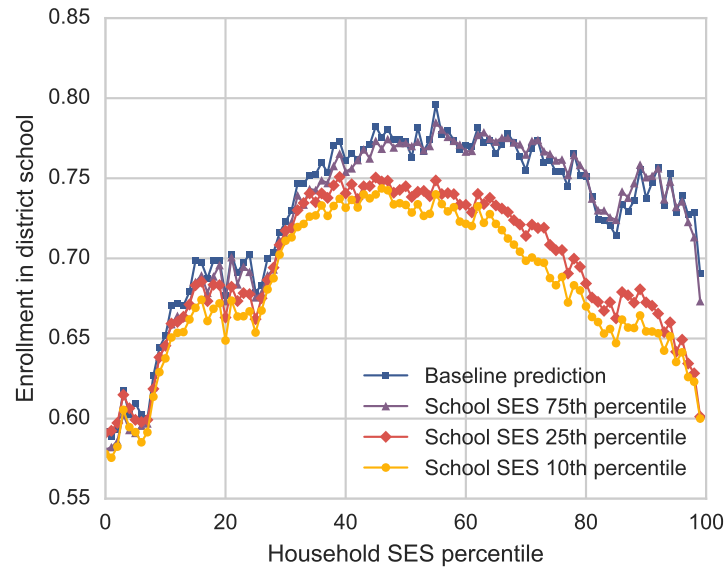
The figure presents the normalized Gini impurity measures for the features in the fitted Random Forest model. Intuitively a higher measure implies that predictions would suffer more from excluding the variable from the model. Only characteristics at age 5 are used as covariates. The measures sum to 1.

constructed from multiple variables. Most of these underlying variables are however not part of the variables used for training the model. Nevertheless, the algorithm is able to learn this process and accurately predict propensities to comply. This provides evidence that our model captures central elements of the decision process for households about whether to comply with the allocation mechanism.

To get a sense of how the model works, we can synthetically manipulate the cohort characteristics in our test sample and observe the changes in prediction. We do this in Figure 4b, where we repeat the baseline prediction from Figure 4a in blue. We first raise every household's expected school-SES to the 75th percentile of the school-SES distribution. We see almost no change in the predictions, as evidenced by the purple line. However, when we lower the school-SES to the 25th percentile we see strong heterogeneous effects (red line). While lower-SES households barely change behavior, higher SES households opt out of the assigned school. From around the



(a) Out of sample prediction



(b) Synthetic shocks to school SES

Figure 4: Performance of prediction model

The figures present the results from the fitted Random Forest model. Only characteristics at age 5 are used as covariates. Figure 4a compares out-of-sample predictions to actual behavior. For each SES-percentile in the test sample we calculate the mean of a dummy for compliance (black line) and mean predicted compliance (blue line). Figure 4b evaluates the model by changing the expected cohort SES while keeping other school and household characteristics constant.

40th household percentile, the drop in compliance is increasing in household-SES. If we decrease school SES further, the magnitudes only become larger.

These results are in line with our findings documented in Bjerre-Nielsen and Gandil (2018). However, this is *solely* a predictive model for a stable environment. We do not exploit any exogenous variation and cannot claim that these effects are causal. Nevertheless, the out-of-sample predictions imbue us with confidence that the model provides a good approximation of baseline compliance in the absence of exogenous changes in attendance boundaries.

6.2 Predicting baseline propensities in causal sample

We now take our prediction model to the causal analysis sample and construct a predicted baseline probability for each observation in the data. The density of predicted baseline probability in the causal sample is displayed in Figure 5. We group the predictions into three roughly equally sized groups: below 60 percent, between 60 and 80 percent and above 80 percent. These are marked by the dashed vertical lines in Figure 5.

In our causal analysis sample, the “prediction errors” are of interest in and of themselves. They elucidate how the enrollment decision, i.e. compliance, is a fundamental hindrance for estimating peer effects. We plot the misclassification errors as a function to the variation in mechanical school-SES in Figure 6. The classification errors are divided into false positives (falsely predicted to be in the associated school) and false negatives (falsely predicted to opt out of the associated school). Reassuringly, errors are minimized when there is approximately zero change in MSS. However, as the changes in MSS become larger, so does the error. The false positive rate rises as the shock become more negative. In other words, when there is a negative shock to the average expected SES, parents tend to defy the school assignment. Conversely, the false negative rate rises as the shocks become more positive. This implies that faced with a positive shock, parents tend to opt in at a larger rate than the baseline prediction would suggest. This closely

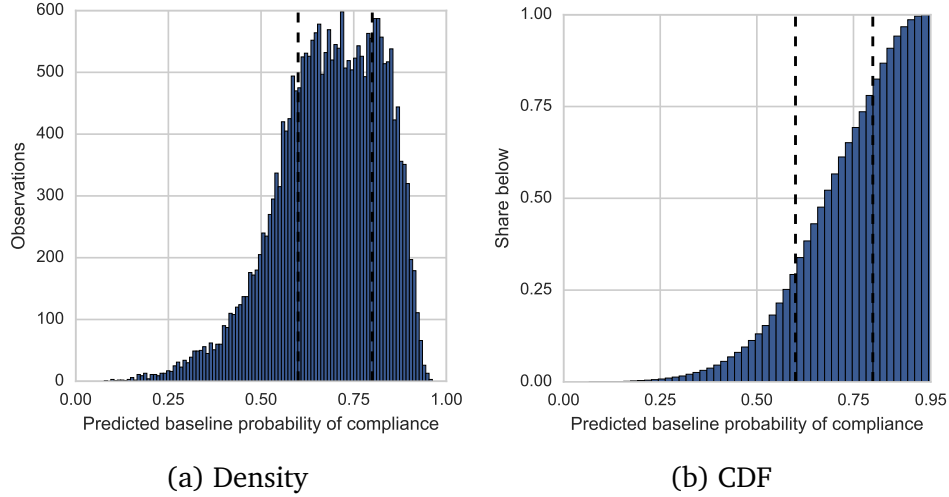


Figure 5: Distribution of baseline predictions in causal sample

The two figures display the density and cumulative distribution function of predicted baseline probabilities. The probabilities are out-of-sample predictions based on the Random Forest model presented in the text. Only baseline characteristics at age five are used for the model. The vertical dashed lines represent our grouping of the predictions into three groups; below 0.6, between 0.6 and 0.8 and above 0.8. These groups correspond roughly to a third in each group as one third of the sample is below 0.63, two thirds are below 0.77.

mirrors our findings in Bjerre-Nielsen and Gandil (2018). The changes in boundaries play no role in the baseline prediction as we only use neighborhood and school variables at age 5, which is prior to variations in the boundaries. The predictions are therefore the counterfactual, the predict compliance in the case where the attendance boundaries had not changed.

The results of Figure 6 show the perils of estimating the peer effect by IV. The magnitude of the instrument is correlated with compliance rates. The non-compliance explains why we observe non-monotonicity in the instrument. In other words, when the instrument is “large”, the actual treatment may go in the opposite direction. We document this in Appendix A. Once again, and in light of these observations, only estimating the reduced-form seems like the most sensible option.

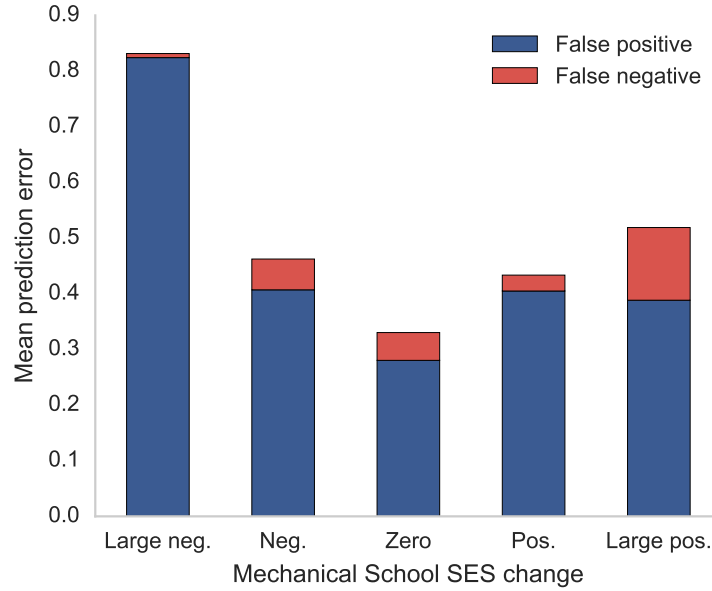


Figure 6: Prediction errors in causal sample

The role of controls and identification of likely compliance As the attendance boundaries are not stable in the causal sample, the calculated prediction is counterfactual. In other words, we are predicting the counterfactual compliance *in the absence of changes* in attendance boundaries. The nature of the prediction model implies that the probability is essentially an index of covariates. This has implications for which variables to include as controls in our regressions. In the limit, where we include covariates in a fully saturated model, the baseline prediction should have no explanatory power as it is a function of these same variables. If we excluded control variables in the main regressions and these same variables are used to form the predicted compliance, it amounts to an assumption that the variables affect test scores primarily through exposure and not through direct effects. This is a non-trivial assumption and we can not test whether it is true. In other words, we do not have sufficient variation to partial out the two mechanisms.

A poignant example of this issue is the question of including or excluding children of specific ethnicities from the sample. A child being non-western is important for predicting compliance in our model. At the same time, average

Danish test scores are significantly lower for non-Western children in the non-causal sample.⁸ We found that our initial estimations were very sensitive to the inclusion of non-Western children in the sample. However, we do not claim to know whether being of non-Western descent primarily affects children through compliance (and thus exposure to peers), inherent ability or parental inputs and therefore exclude this group from the regressions. One could use this argument for other socioeconomic characteristics as well.

Due to the out-of-sample performance of our compliance predictions, we believe that the model carries weight and that the covariates may work through the propensity to comply. However, while we think we capture effects stemming from exposure, the example with non-Western children illustrates that the essential uncertainty about mechanisms remains. The results from using the predicted compliance should therefore be regarded as qualified guesses at the underlying mechanisms, rather than concrete proof that predicted exposure is a driver of effects.

7 Results

We begin by regressing test scores on SES-quartile and mechanical changes in school-SES (i.e. MSS). The estimates are found in Table 2. We see from column 1 in Table 2 that the average effect of MSS is positive but insignificant. This implies little overall effect once we control for residential sorting. This changes when we interact the mechanical school-SES with household-SES quartile. In column 2 the reference group is the lowest quartile. There is a large and significant positive effect of MSS on language test scores, as seen by the parameter on MSS. If the mechanical school-SES rises by a standard deviation (≈ 0.1), the expected test score rises by 2 percentiles. In other words, low SES children gain from potential exposure to children from more

⁸A simple linear regression of test scores on a non-Western-dummy gives a coefficient of -0.18 with a t-value of -25.56. Baring in mind that the test scores are bounded between 0 and 1, this effect is quite large.

	<i>Danish, grade 2</i>		<i>Math, grade 3</i>	
	(1)	(2)	(3)	(4)
MSS	0.117 (0.1000)	0.226* (0.104)	0.0534 (0.112)	0.127 (0.133)
MSS x SES Q2		-0.132 (0.0813)		-0.0885 (0.0784)
MSS x SES Q3		-0.130 (0.0789)		-0.0778 (0.0768)
MSS x SES Q4		-0.160* (0.0807)		-0.107 (0.0863)
N	11669	11669	10298	10298

Standard errors in parentheses

[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Effect of mechanical SES changes

This table present different specifications of the model presented in equation (1). The dependent variables are test scores for Danish language in second grade and mathematics in third grade. All models include fixed effects for year-district at age five and a full set of controls including household socioeconomic ventile and gender. We cluster standard errors at the school associated with SAB at age 5. In other words, the clusters consist of all children who originally sought to sort into the same school.

advantaged backgrounds. The interaction terms are all negative and rising in magnitude. This implies that the effect becomes smaller as household-SES rises. The point estimates show that for a standard deviation in MSS, the expected test score of high-SES children rises by 0.6 percentiles. This effect represents less than a third of the effect for the low-SES children.

When we change the dependent variable to math test scores from third grade, we observe the same patterns, though the effects are in general insignificant. As the test is taken a year later, this is not surprising. We would expect the mechanical school-SES to be less important for the test variable when households have a longer time to adjust to the change. In an IV setup, this smaller reduced form effect would be counteracted by a smaller first stage, which would adjust the IV estimate upwards. As explained, we do not deem this approach suitable due to the lack of a valid exclusion restriction and non-monotonicity of the instrument.

7.1 Predicted compliance

We now proceed to include our compliance measure into the regressions. We fully interact the mechanical change with household-SES quartile. We include a fully interacted set of household-SES-ventile-dummies and baseline-prediction-ventiles as well as interactions to control for nonlinearity, which might confound our findings. Thus, the controls are more extensive than in the previous section. We continue to include the SAB-year fixed effects and cluster standard errors at SAB-level. To ease interpretation, we plot the interaction terms in Figure 7a. The parameters are all insignificant, though we stress that the number of controls greatly reduces statistical power. The effect is however still decreasing in SES as before. For the highest SES quartile, the interaction term is a precisely estimated zero.

Next, we perform the same regression, but instead of interacting the mechanical school-SES with household-SES, we interact the former with dummies for the three prediction groups described in section 6. As can be seen from Figure 7b, the effect of MSS is increasing in the probability of compliance, though all estimates are insignificant.

In Figure 8 we interact MSS with our three groups of predicted compliance and household SES quartile. We maintain SES-SAB-year fixed effects and interactions between household SES and baseline predictions. For the three highest quartiles, we find small insignificant effects. For the highest quartile, we identify a precisely estimated zero regardless of prediction group.

For the lowest quartile, however, we see that the cross between likely compliance and household-SES is important. For the subset of children with a high probability of compliance and a low socioeconomic background, we see a relatively large significant effect. For the lowest prediction group in the lowest SES-quartile, the estimates are close to zero and insignificant. In other words, we find weaker effects for those households, who we a priori do not expect to enroll their child in the assigned school. The differences between predicting groups within SES-quartile are however not statistically

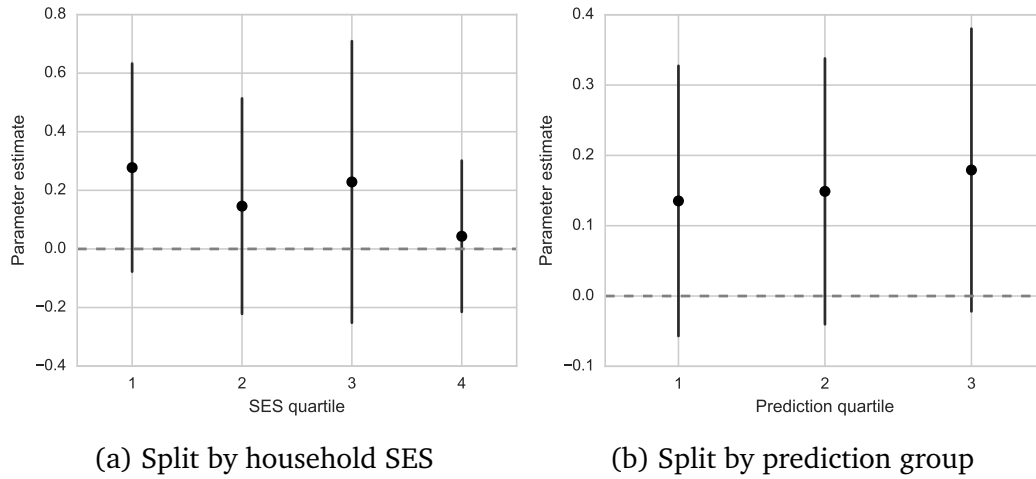


Figure 7: Heterogeneity in effects of MSS on test scores

The parameters are estimate in a model with three way interactions between prediction group dummies, SES quartile dummies and MSS as a continuous measure. Two-way interactions between prediction vintiles and SES vintiles and interactions between non-Western and gender dummies are included as controls. The dependent variable is test scores in Danish language taken in second grade.

significant.

We once again stress that the baseline predicted compliance is a function of observed covariates and that the non-linearities are essential for identifying the heterogeneity stemming from compliance separate from other controls. However, with this in mind, we find the results to be a strong indication that the effects of changes in attendance boundaries are higher for those likely affected by the change.

Figure 8 demonstrates why we find little effect in the simple regression of test scores on mechanical SES. The effect is heterogeneous in SES, low SES children benefit while high SES children do not. However, this may not be the whole story. While all low SES children may benefit from stronger peers, we have no way of affirming this as they are not all exposed to the “treatment”. In other words, the majority of the sample shows no effect of the changes induced by the boundary changes, either because they are insensitive to peer compositions or because they never experience a change

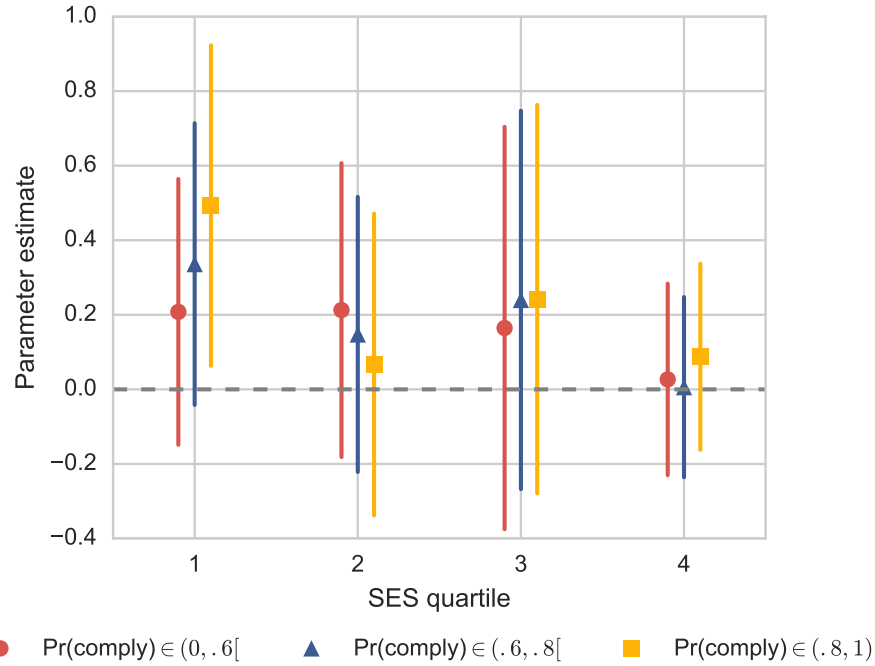


Figure 8: Heterogeneity in effects of MSS on test scores, interaction

The parameters are estimated in a model with three-way interactions between prediction group dummies, SES quartile dummies and MSS as a continuous measure. Two-way interactions between prediction ventiles and SES ventiles and interactions between a gender dummy are included as controls. The model is estimated with a fixed effect for all SES-quartile-SAB-year combinations. The dependent variable is test scores in the Danish language taken in second grade.

likely due to non-compliance.

7.2 Exposure

If the lack of effect for high SES-groups is due to non-compliance, we should see little effects of the mechanical changes on actual cohort-SES of the children. To investigate this, we perform the same estimation as in Figure 7 but exchange the dependent variable for the average cohort-SES in the school in which the child ends up enrolling. In this regression, we have not conditioned on the availability of test scores, and we, therefore, include all chil-

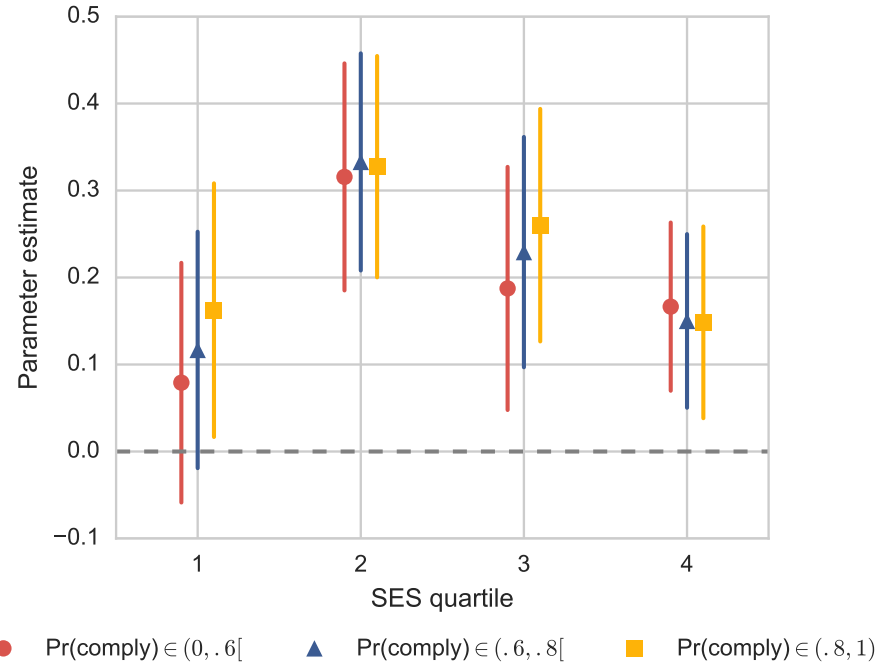


Figure 9: Effects of MSS on actual peer composition, split by expected compliance and household SES

The parameters are estimated in a model with three way interactions between prediction group dummies, SES quartile dummies and MSS as a continuous measure. Two-way interactions between prediction ventiles and SES ventiles and interactions between non-Western and gender dummies are included as controls. The dependent variable is the average SES in the cohort belonging to the school in which the child enrolls.

dren enrolled in primary school.

The results of this estimation are displayed in Figure 9 where the interactions between the mechanical SES, household SES-quartile and prediction quartile are plotted. All interactions but for the combination of low SES and low predicted compliance are positively significant, meaning that the first stage in an IV would have predictive power. In other words, the policy of boundary changes actually affect all subgroups to some degree.

There is heterogeneity across both predicted compliance and household SES. The general pattern along the SES dimension follows the shape of the out-of-sample prediction in Figure 4a.

The highest prediction group within the lowest SES-quartile has a higher estimate than the other groups within the same SES-quartile. This pattern provides supporting evidence that the heterogeneity of our estimated ITT-effects is actually due to exposure to the policy of boundary changes.

For the highest SES quartile, we see relatively smaller first stage coefficients, compared to the middle of the SES distribution. However, these are significant and positive, which imply that the zero effect of MSS cannot readily be explained by non-compliance. This implies that, even though some households may choose to opt out, the policy is actually effective in manipulating the actual peer composition of children, regardless of socioeconomic status. This mirrors our conclusions in Bjerre-Nielsen and Gandil (2018), that overall there is a degree of compliance when attendance boundaries change.

The large “first stages” in Figure 9 combined with the very small reduced-form estimates for the high-SES children in Figure 8 indicate that, even though high-SES groups are exposed to changes in the socioeconomic makeup of their peers, this does not meaningfully affect their performance on the Danish language tests.

The insignificant first stages for the lower prediction groups in the bottom of the SES distribution, however, point to the limits of SAB changes as a policy tool. We here identify a disadvantaged group, which is not affected by a policy, for which they may be the primary target. Though we cannot estimate it, we conjecture that these children may gain from being exposed to stronger peers, but this policy may not effectuate the intended exposure due to low compliance.

7.3 Missing outcomes

As mentioned in section 3, we only obtain test scores on the children enrolled in public schools. In other words, for all children enrolled in private schools we do not observe an outcome. If some children move to the private sector as a response to changes in the expected sociodemographic compo-

	(1)	(2)	(3)	(4)	(5)
MSS	0.142 (0.168)	0.226* (0.104)	0.213** (0.0711)	0.231** (0.0723)	0.230** (0.0728)
MSS x SES Q2	-0.214* (0.0942)	-0.132 (0.0813)	-0.111 [†] (0.0608)	-0.129* (0.0609)	-0.129* (0.0603)
MSS x SES Q3	-0.349*** (0.0956)	-0.130 (0.0789)	-0.138* (0.0587)	-0.183** (0.0612)	-0.172** (0.0573)
MSS x SES Q4	-0.330** (0.106)	-0.160* (0.0807)	-0.207*** (0.0576)	-0.208*** (0.0590)	-0.247*** (0.0573)
Outcome	TS missing	TS	TS	TS	TS
Imputation			Predicted TS	Pr(comply)=0	Pr(comply)=1
N	15648	11669	15648	15648	15648

Standard errors in parentheses

[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Effect of mechanical SES changes on Danish test scores, imputed outcomes

This table present robustness tests. The dependent variable in column one is a dummy for whether the test score is missing. Column 2 repeats the results of column 2 in Table 2. Column 3-5 present the results when using different imputation techniques. Column 3 use the random forest prediction model to impute unobserved outcomes. Column 4 and 5 also impute using the prediction model but synthetically setting the probability of complying by the school assignment to 0 and 1 respectively.

sition, then our results could be biased. In the present framework, we can illustrate this by regressing an indicator for missing outcome on the independent variables. The result can be read from column 1 in Table 3. We see that sensitivity is increasing in household-SES. For the lowest SES-quartile, we find an insignificant effect of 0.142. However, the sum of the base and interaction parameters become negative for the higher SES groups. In other words, the more positive a change in MSS the lower the probability of a missing test score. There is, therefore, indications of non-random attrition and we need to test whether our results are sensitive to the missing outcomes. We impute the missing test-scores for pupils in a number of ways.

First, we use a similar approach as when we predict baseline compliance. We use a Random Forrest regression on the training sample with the same background variables as for predicting compliance to predict test scores. However, as we never observe test scores for children enrolled in private

school we can only train the model on the children enrolled in public school. We include the predicted compliance rate in the model. Thus, the model may choose to let predicted test score be a function of the predicted compliance. We once again test our model against the test dataset and find that the model fits the mean well. We then use the model to predict the missing outcomes in our causal sample and include these in the regression of test scores on mechanical changes in SES.

Doing this, we assume that the children going to private school would do as well as the observably equivalent children in the public school system absent a change in characteristics. For the households who would always choose private school, imputation should matter little. For the “defiers”, which are moved to private school because of the change in MSS, we assume that the choice of private school is able to compensate fully for the effect of a change in SAB on the performance of the child.

The results of this exercise are seen in column 3 of Table 3, where we also repeat the baseline estimation in column 2 for comparison. We see almost no change in the results. To further test the sensitivity, we estimate the model two times, but now we artificially change the probability of compliance to zero and one respectively. We then use the predicted test scores with this artificial probability for imputation. The results barely budge. We also estimate a two-step Heckman correction model, where we include the baseline prediction in the first stage. Appendix Table A1 shows a wholly insignificant Mills ratio and estimates completely in line with the baseline results.

Using our three imputed test scores we perform our three-way interactions of the effects corresponding to Figure 7 in Appendix Figure B.2. We find that our conclusions are robust to the imputation. Overall, the robustness results provide no indication that our results should be biased due to non-random attrition. A reason could be, that we are fully able to control for the selection by including covariates. In other words, we do not find evidence that households sort into private school on unobservables.

8 Conclusion

Peer effects can potentially be important for understanding how inequality is transmitted from one generation to the next. In this paper, we provide evidence that children from disadvantaged backgrounds benefit from exposure to children from more advantaged backgrounds. Importantly, privileged children do not seem to be affected at all. This implies that associational redistribution may increase equality and economic efficiency at the same time. However, the analysis also elucidates the limitations of such policies. Compliance is crucial for the intended effect to materialize. In other words, the institutional context wherein authorities seek to manipulate student bodies matter for the expected effects.

As compliance is not random, it also underlines the difficulties going beyond the reduced form when estimating peer effects. A contribution of this paper is therefore also to show the limits to estimating causal peer effects in a context where compliance is not ensured.

References

- Angrist, J. D. 2014. The perils of peer effects. *Labour Economics*, 30, 98–108.
- Baum-Snow, N., Lutz, B. F. 2011. School desegregation, school choice, and changes in residential location patterns by race. *American Economic Review*, 101, 3019–46.
- Bjerre-Nielsen, A., Gandil, M. H. 2018. Defying attendance boundary policies and the limits to combating school segregation.
- Blume, L. E., Brock, W. A., Durlauf, S. N., Jayaraman, R. 2015. Linear social interactions models. *Journal of Political Economy*, 123, 444–496.
- Bramoullé, Y., Djebbari, H., Fortin, B. 2009. Identification of peer effects through social networks. *Journal of econometrics*, 150, 41–55.

- Breiman, L. 1984. Classification and regression trees. Wadsworth statistics/probability series, Wadsworth International Group.
- Breiman, L. 2001. Random forests. *Machine learning*, 45, 5–32.
- Carrell, S. E., Sacerdote, B. I., West, J. E. 2013. From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica*, 81, 855–882.
- Chetty, R., Hendren, N., Kline, P., Saez, E. 2014. Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129, 1553–1623.
- Coleman, J. S. et al. 1966. Equality of educational opportunity..
- Durlauf, S. N. 1996a. Associational redistribution: A defense. *Politics & Society*, 24, 391–410.
- Durlauf, S. N. 1996b. A theory of persistent income inequality. *Journal of Economic growth*, 1, 75–93.
- Epple, D., Romano, R. E. 1998. Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 33–62.
- Graham, B. S. 2018. Identifying and estimating neighborhood effects. *Journal of Economic Literature*, 56, 450–500.
- Gruber, J., Mullainathan, S. 2006. Do cigarette taxes make smokers happier? In *Happiness and Public Policy*, Springer, 109–146.
- Hanushek, E. A., Machin, S. J., Woessmann, L. 2016. *Handbook of the Economics of Education*. Elsevier.
- Hoxby, C. M., Weingarth, G. 2005. Taking race out of the equation: School reassignment and the structure of peer effects. Technical report, Working paper.

- Imberman, S. A., Kugler, A. D., Sacerdote, B. I. 2012. Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102, 2048–82.
- Manski, C. F. 1993. Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60, 531–542.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Sacerdote, B. 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, 3, Elsevier, 249–277.
- Wager, S., Athey, S. forthcoming. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*.

A Why not IV?

An implication of monotonicity is stochastic dominance of distributions of the endogenous treatment for different values of the instrument. In our context, this would imply that the distribution of values of the actual peer composition for a value of mechanical peer composition either dominates or are dominated by the distribution for another mechanical peer composition. We can check for this in our data. We group values of MSS (after demeaning out the fixed effects) into quartiles and for each quartile plot the empirical cumulative distribution functions (cdf) of cohort compositions. The result of this exercise is displayed in Figure A.1. It is immediately evident that the cdf of the two lowest quartiles cross the other cdfs, which mean that stochastic dominance is not satisfied. This implies that for sufficiently low values of the instrument some children will actually experience *higher* values of peer compositions. This could be the case if parents for a sufficiently low expected peer composition opt out of their designated school into another school with higher peer composition than baseline. We show in a Bjerre-Nielsen and Gandil (2018), that this is indeed the case. From this follows that some weights in the calculated average treatment effect will be negative. The usual result, that the IV measures an average of a heterogeneous effect on compliers, does therefore not hold in our context.

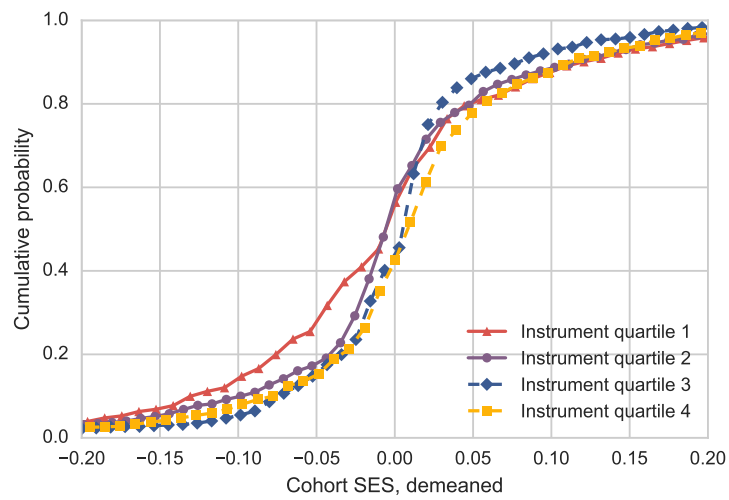


Figure A.1: CDF of treatment for quartiles of instrument variables

The figure presents the cumulative distribution functions for demeaned treatment for quartiles of demeaned mechanical SES. A necessary condition for monotonicity to be valid is stochastic dominance. This is clearly not fulfilled as the red density intersects the other densities.

B Additional tables and graphs

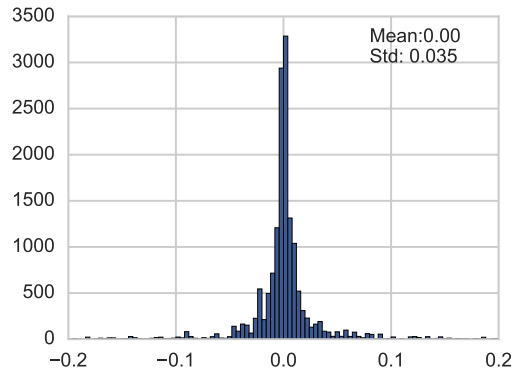


Figure B.1: Demeaned mechanical SES

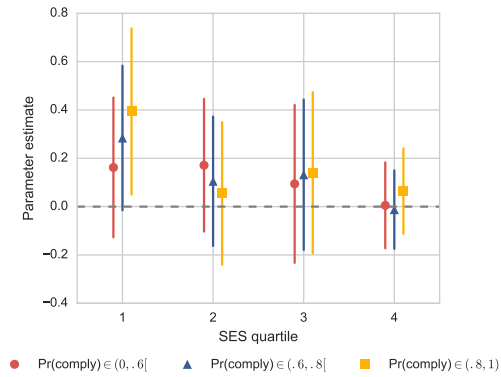
	Test score	Missing
MSS	0.229* (0.0900)	-0.477 (0.440)
MSS x SES Q2	-0.137 [†] (0.0794)	0.877* (0.391)
MSS x SES Q3	-0.137 (0.0841)	1.390*** (0.388)
MSS x SES Q4	-0.167* (0.0819)	0.893* (0.368)
Mills	-.013	
Mills SE	.07	
Obs.	15648	

Standard errors in parentheses

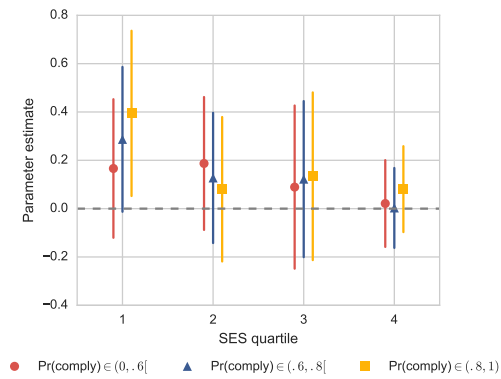
[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A1: Heckman 2-stage sample correction model

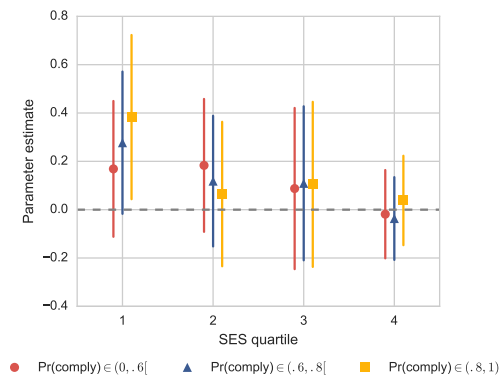
The model is estimated correspondingly to the models in Table 3. SAB-year fixed effects are included in the estimation.



(a) Imputed outcomes



(b) Imputed with $P(\text{comply})=0$



(c) Imputed with $P(\text{comply})=0$

Figure B.2: Robustness of estimates

The parameters are estimate in a model with three way interactions between prediction group dummies, SES quartile dummies and MSS as a continuous measure. Two-way interactions between prediction vintiles and SES vintiles and SES-SAB-year fixed effects are included as controls. The dependent variable is test scores in Danish language taken in second grade. The imputation method used can be read from the captions of the subfigures.