JMS Journées de méthodologie statistique de l'Insee 2018

Efficiency or equality: why should we choose? The differentiated effect of high schools on student achievement

Milena Suarez(*), Pauline Givord(*)

(*) INSEE-CREST

milena.suarez-castillo@insee.fr - pauline.givord@gmail.com

Key words Value added; quantile regression; Student Growth Percentiles

Abstract

This article proposes a new value-added statistical model, which encompasses the ability of schools to mitigate or intensify the expected differences in grades between students of similar background. While usual value-added indicators measure inter school performance, they do not account for differences in intra school achievements. From parents' and students' points of view, the expected gain when attending a school may matter as much as the distribution of gains. Some schools may improve mostly the achievements of already high-performing students, while others help the students the most likely to drop out. We thus propose to estimate value-added indicators at different levels of the distribution of final achievements. We apply this method to exhaustive data of the 2015 French high-school diploma. We find evidence that almost one-sixth of the French high schools behave either in an equal or on the contrary unequal way, meaning that they significantly reduce, or on the contrary increase, the dispersion in final grades which were expected given the school population. We observe that the highest performing high schools on average are over-represented among the egalitarian high schools. Beside performance, equality is found in small high-schools that welcome priorly good students.

Introduction

Assessing the performance of schools has become standard in most developed countries. A first objective is to provide tools for public action, for example by evaluating the effectiveness of resource allocation between schools. These evaluations may also aim at providing useful information to families when they can choose the school of their children. Several performance metrics have been proposed in the literature. Most of them suffer an important limitation: they focus on the *average* school impact which may mask significant disparities in achievement outcomes among students. In this paper, we propose going beyond the classic *average* school value-added measures by modelling whether a school contributes to decreasing or on the contrary increasing the inequalities in academic outcomes. The very same average performance may be the reflection of either a combination of significant academic progress of only a few pupils (possibly at the cost of a smaller investment in other pupils at the same school), or a modest but homogeneous gain for all. We aim at measuring how, depending on the school, students can make equal progress or on the contrary face unequal achievements because of either school behaviour (e.g. priority given to some) or school environment (e.g. peers effects). By doing so, we study how school inputs are related to achievements in a broader sense than is usually done in the literature.

Evaluating school performance raises a number of issues concerning the definition, the estimation but also the interpretation of school performance indicators. First, to measure school efficiency one must dissociate the impact of school on its student achievements from the many factors influencing academic achievement (initial academic level, family background, etc.). Students are not randomly assigned to schools, and differences in school raw performances also reflect this unequal assignment of pre-existing potentials.

Several types of indicators have been proposed to isolate school effects from selection ones. The most classic indicators are based on the so-called value-added models (Koedel et al. [2015], Raudenbush and Willms [1995]). In practice, they come from models where the schooling performance is regressed on a school effects while controlling for individual observable characteristics of students (such as familial background or past scores). Measuring the academic pay-off of these characteristics over the entire population indicates whether the average performance in this school exceeds or conversely reduces what is expected, given the school enrollment. However, such indicators do not inform on the potential heterogeneity in schooling progress among the students within this school. Conversely, the so-called Student Growth Percentile (SGP) indicators (Guarino et al. [2015], Tzavidis and Brown [2010], Walsh and Isenberg [2013]) leverage heterogeneity in student academic progress. These indicators rely on quantile regressions of the test-scores obtained at a given year on past academic performance, and aim at measuring how each student's test-score growth ranks among academically similar students. SGP estimates allow one to look into the great heterogeneity in individual performances within the same school. Highlighting this within school heterogeneity is essential in order to get an accurate picture of school performance. However, as pointed out by Guarino et al. [2014] and Walsh and Isenberg [2013], controlling only for students' initial score may produce inequitable estimates of school efficiency. Students are not randomly distributed throughout schools and the differences in enrollment may also explain part of the observed final performances. For instance, if social background is correlated with schooling performance growth, schools that benefit from a privileged enrollment may display higher than average student score growth without any specific teaching investment. As a consequence, these indicators may induce schools to sort students according to their social or academic backgrounds. One should thus control for variables correlated for both the academic performance and student sorting among schools (Rothstein [2009]).

In this article, we propose a model that is inspired from both value-added and SGP models. This model limits the problem of selection into schools based on school performance highlighted by Guarino et al. [2014] in the SGP context. We estimate school-specific effects adjusting for the composition of its enrollement using quantile regressions. These estimates are derived at several quantiles of the distribution of final schooling achievements (second decile, median and eighth decile). We thus obtain indicators of school value-added that go beyond the measure of average performance. These estimates allow us to discriminate among similar value-added schools those with an homogeneous impact on pupils from those where the main achievements is limited to some students. We can construct a typology of schools, depending on whether they reduce or increase the dispersion of performance (compared to those predicted by the model). The more or less egalitarian behaviours of schools may also be confronted with their median effectiveness and other schools' characteristics.

We use an exhaustive database on the French non vocational high schools (*lycées*, corresponding to secondary education for children between the ages of 15 and 18) for the year 2015. Empirical estimates are based on the results at a national, standardized evaluation corresponding to the final exam of secondary school (the French *baccalauréat*, cf. Duclos and Murat [2014]).

These data provide detailed grades but also information on the individual characteristics of the pupils, such as familial background and gender. First and foremost, we use a detailed measure of the schooling level of pupils at the entry of high school, as provided by the final exam (*brevet des collèges*) that is taken by French pupils at the end of middle school (year 10/ninth grade).

By comparing school effects at the top and the bottom of final grade distribution, we discriminate between "egalitarian" high schools in the sense that they tend to reduce performance gaps between students compared to what would be expected given their enrollment, and unequal high schools, which on the contrary tend to increase achievement differences. We observe that these two categories account for almost a third of high schools, in an almost equal proportion. The other high schools have a "homogeneous" effect on all their students, meaning that they have an impact, positive or negative, that is not statistically different at the top of the score distribution than at its bottom. When accounting for multiple hypothesis testing, one-sixth of schools is found to have an heterogeneous impact, either equal or unequal. As a robustness check, we verify that this typology is robust to student mobilities, which may partly reflect strategic high-school behaviour (as they may exclude the less promising students).

This classification (defining a "egalitarian" behaviour) complements the information provided by the "classic" value-added indicator - which is proxied by a school-specific effect measured at the score distribution median. We observe an over-representation of positive value-added high schools (the fixed effect measured at the median level is significantly positive) among the egalitarian high schools, and an under-representation of these high schools among the unequal ones. In order to characterize the high schools with egalitarian behaviour, we also provide a descriptive analysis. Specifically, we rely on random forest models that allow us to classify highshools in one or another types according to observable characteristics (headcount, teachers seniority, etc.).

The following section presents an overview of the institutional context in France and the data. The second section discusses the econometric model, detailing in particular the issues related to fixed-effect quantile estimation and the interpretation that can be made from these estimates. The third part details the results obtained and their robustness to selection of students during the high-school years. A forth part details how school features are related to inequalities' estimates.

1 Institutional context and data

We focus on the French *lycées* (senior high schools) which constitute the second part of the French secondary education. These high schools enroll students from age 15 to 18 and follow the *collège* (junior high school) that enroll pupils from age 11 to 15. For the five first years of secondary education (junior high school and first year of senior high school) pupils follow the same curriculum. In what follows, we call high school the French "lycées". For the two final years, they are assigned to one track (see Figure 1): general, technological and vocational tracks. The assignment is made in theory according to student preferences, but in practice it depends mostly on their schooling achievement, the general track being considered as the more selective and the vocational track the least. Within each track, students must choose between different streams, called *séries*. For instance, in the general track, pupils may choose between three main streams: humanities, economics and social sciences and finally sciences. The technological track includes five streams: sciences and technologies in health and social field, hotel and restaurants management, industrial science and technologies and sustainable development, laboratory science, finally management technologies. The vocational track includes more than eighty streams and is proposed only in dedicated schools. On the other hand, many high schools propose both a general and technological curriculum. The curricula of all the streams and tracks are defined by the French Ministry of Education.

Notably and importantly for the measurement of the *lycée* performance, both junior and senior high schools end with a national mandatory examination, respectively the *diplôme national*

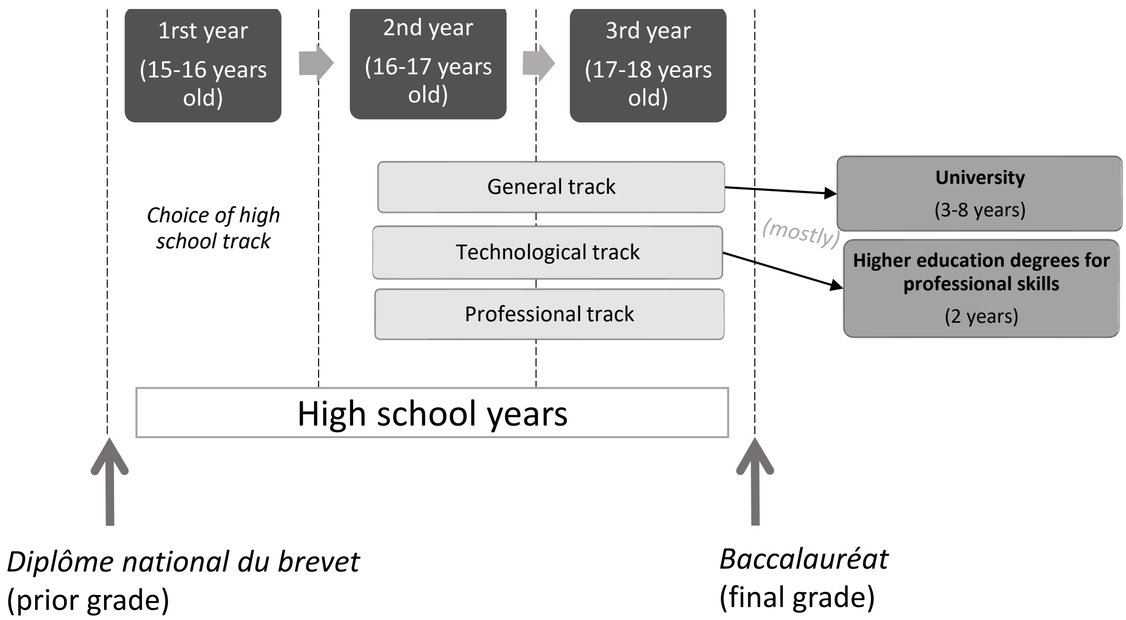


Figure 1: French secondary education and timing of national examinations

du brevet and the *baccalauréat*. Concerning the former, the final score is partly based on three written parts, respectively in Mathematics, French and finally "History, Geography and Civic Education", with a national examination test.[1] The grades obtained at this written examination provide a standardized proxy of pupil level before entering high school. At the end of high school, students sit for the *baccalauréat*, a national examination. The type of *baccalauréat* depends on the schooling track: it distinguishes general, technological and vocational ones. It relies on standardized tests that occur at the end of the final grade. In theory, the *baccalauréat* diploma qualifies for university entrance (with some restrictions depending on the type of *baccalauréat*). It is, inter alia, a standardized measure of pupils' achievement at the end of high school. The French *baccalauréat* is a national institution whose results are each year highly publicized. Newspapers usually provide a ranking of high schools depending on the proportion of their students who pass the exam.

Every pupil is assured of accessing a public high school according to a residence-based rule. They can however formulate a choice that is satisfied if there are still places available at the requested high school. If the requested derogations exceed the capacity of the school, a number of criteria are considered (for instance disability, means-based scholarships, etc.). The ranking of those criteria varies with the school district (*"académie"*, meaning 31 large administrative regions). Besides the public school sector, families can opt-out for a private school as these schools are not subject to restrictive zoning. These schools may also charge fees to families, but they are usually rather low as a large part of the private school costs are subsidized by the state (including the wages of teachers) or local authorities.[2] In our sample (restricted to general and technological tracks), 21% of pupils are enrolled in the private sector. In theory, state-run schools cannot operate selection amongst students, while private schools are free to do so. However, in both types of high schools, dynamic sorting into track may occur depending on the schooling achievements observed during the first year of high school. Certain high schools with high academic standards may direct low achieving students into streams outside the school's

[1] The final score also encompasses grades obtained at school, but these measures are not standardized - some schools may have different academic standards resulting in more or less generous grading policies.

[2] On the basis that these schools follow the state requirements in terms of curriculum, teaching hours, recruitments, etc.

curricula. In case of conflict with the parents on the stream favored by the school, the headmaster has the final say. We discuss this point later.

The analysis is based on several sources: the exhaustive database of the results at the *baccalauréat* national exam, matched with the FAERE database (*Fichier anonymisé d'élèves pour la recherche et les études*, i.e. anonymous file of students for research and studies), data obtained from the statistical entity of the ministry of Education (*DEPP: Direction de l'évalution, de la prospective et de la performance*).

The FAERE database contains detailed information on students. First of all, the scores at the test passed just before the entrance in senior high school, the DNB (*diplôme national du brevet*). The DNB grade used is the raw score, i.e. before any grading adjustment: it represents a homogeneous initial level for all students. The dataset also includes student characteristics such as gender, socioprofesional characteristics of parents and a dummy of repeating a class before entering high school. The social position index is that of Rocher [2016], who derive a continuous index from the discrete occupation and social category data of the parents (*PCS: profession et catégorie sociale*). Survey data on social, economic and cultural characteristics of parents have been analysed through a multiple correspondence analysis and the first axis appears a good summary of the degree of diploma, the amount of revenue of the parents as well as cultural capital (number of books at home for instance), underlying that the unidimensional approximation is valid. Averaging the projection on the first axis by social category of both parents (for example mother in the teacher category, father in the qualified worker category), one obtains the social position index associated with this combination of parent social backgrounds. This social position index has been made available by the *DEPP*.

The school of each student in the year of the *baccalauréat* is recorded in both bases, while the FAERE files allow us to recover the past schools history of students. Finally, the APAE[3] database contains information on the high school: sector, staff, teaching options, etc.

Table 1: Characteristics of students applying for the *baccalauréat* 2015

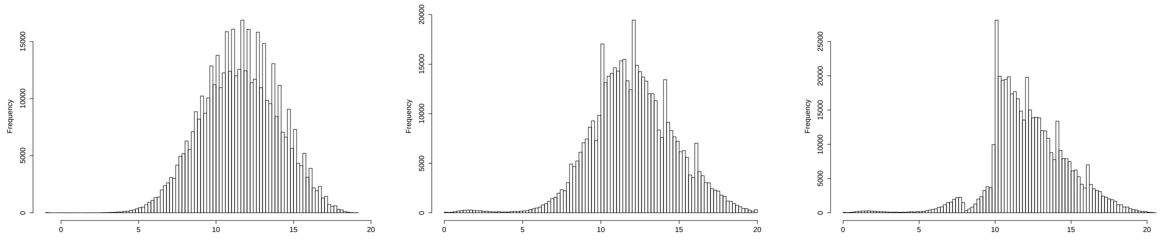
Track	General		Technological	
	Mean	sd	Mean	sd
DNB grade (past score, over 20)	12.3	2.3	9.7	2.0
Social position index	122.4	35.0	104.6	33.6
Repeaters	5.7%		18.3%	
Girls	56.5%		50.0%	

Source: FAERE and APEA database, Author calculations.
DNB corresponds to the pre high-school examination.

At the end of the first session of the *baccalauréat* exam, the candidate gets a grade, that results from the first set of tests which are graded by teachers. In order to guarantee uniform marks, teachers who mark the answers attend a standardisation jury. These jury may increase the score obtained at the first session in order to allow a candidate to obtain 10 out of 20, the pass mark, or 12, 14 or 16, the thresholds for particular distinctions. When a student gets an average mark below 10 out of 20 but above 8 out of 20, he or she may resit the exam in a second session. The final mark (after the second session) obtained at the *baccalauréat* shows a very strong threshold effect at 10, due to a large extent to the students who barely pass. We choose to work on the note at the first session, i.e. the average of the scores on the tests with their

[3] *Aide au Pilotage et à l'Auto-évaluation des Établissements*, the school entry on this database is made available to the head of school as administrative information and performance indicators.

respective weightings, which represents a relatively homogeneous test for the whole sample. The raw grades, before standardisation, are not available in the data. The grades are centered and standardized in what follows.



(a) DNB grades (raw prior score) (b) Baccalauréat grades (first session) (c) Baccalauréat grades (after second session, i.e. resit)

Figure 2: Empirical grades distributions (first session and final grades after resit)

4

We choose to exclude the vocational track, as it offers very heterogeneous and numerous streams. This makes the comparison between achievements difficult (within each stream, the number of students may be rather small). We analyze separately the general and technological *baccalauréats*, as they are different exams whose enrollment differs. Each *série* has a different weighting for subjects and students choose two years before taking the *baccalauréat* which exam path they will prepare. Thus, empirically, at given initial level and equal characteristics, a student taking the scientific *série*(S) will on average have a lower score than a student of taking the humanities *série* (L). This observation is partly due to the fact that students are graded and therefore compared between students within the same path. Students taking the scientific *série* have more often than those taking the humanities one characteristics related with academic success. In practice, when the grading is partly standardized, scientific students compete with students who are better at school and their results are on average worse compared to humanities students with equal characteristics. To account for the peculiarities in the grading in each *série* (stream) of the *baccalauréat* exam, a fixed "*série*" effect is introduced in the quantile regressions.

2 Estimation of high school value-added at different level of the distribution

2.1 Related literature

The identification of school (or teacher) efficiency is difficult because of selection effects. In case of selective enrollment in schools depending on previous schooling performances, the actual school action can be confounded with the enrollment of high-performing students. Several types of models have been proposed to deal with this selection issue.

Value-added models are the most classic way of dealing with these selection effects with observational data. In their simplest form, these models rely on a linear specification of final grade of students as a function of their observable characteristics plus a school-specific effect, aimed at characterizing the specific action of the school. This class of models is usually used for modeling the impact on the average test-score level, but may be extended to analyze the impact on the entire distribution of test score (see for instance Page et al. [2016]). In practice, these school-specific effects are modelized either by fixed or random effects. Random effects estimators are more efficient (especially when few observations are available). They may be biased and inconsistent when pupil characteristics are correlated with school-specific effects (which is expected to be the case). On the other hand, fixed-effects estimators are consistent even when selection on observables occurs. They may be biased however when only a few observations are available by cluster. This may be an issue for instance if the data are provided by survey with only a few observations by school in the sample, but it is less likely with exhaustive data.

The measures based on Student Growth Percentiles (SGP) are another popular way of assessing the school or teacher efficiency. SGP relies on repeated test scores, and compares one student's performance increase with those of other students having similar prior test scores. In practice, the calculation of the efficiency relies on several steps. First, quantile regression of the current student level on previous schooling level are performed for numerous quantile. Quantile regressions rely on local linear approximation of conditional quantiles. Formally:

$$q_\tau(Y_i|X_i) = m(\tau) + X_i\beta(\tau) \tag{1}$$

where Y_i stands for the current test score of student i and X_i individual characteristics - here the student past test scores. Such regressions can be run for several level of quantiles τ in order to approximate the test score distribution conditional on previous achievement. One may thus locate the final score of a student in the corresponding distribution conditional on his or her prior test score. In practice, the SGP is the student's grade percentile in the distribution of grades conditional on his or her past scores. In other words, it is a rank among peers comparable in terms of prior grades. Formally, the rank of student i is computed from

$$\widehat{SGP_i} = \max\{\tau : \widehat{q_\tau}(Y|X_i) \leq Y_i\}$$

School (or teacher) performance is usually defined by aggregating the SGPs of the students attending the school. The efficiency of the school j is then calculated as the average of its students' SGP, for example $\hat{\tau}_j =< \hat{S}GP_i >_{i\in j}$. The principle of the SGP can be simply summarized as "How well did a student compared to those who had the same grade than him or her in the state?" If he did better than say 80% of the others, the school efficiency indicator is increased by $\frac{0.80}{n}$ where n is the number of students in the school j. As far as we know, this methodology has not been used so far to measure the school efficiency beyond the average effect.

The SGP models provide an intuitive way of measuring school efficiency. However, controlling for previous test score only may provide biased estimates. First, as emphasized by Walsh and Isenberg [2013] and Guarino et al. [2014], some other student characteristics (social background for instance) may be correlated with both the past and current test scores. As students are not randomly assigned to school, one may wrongly attribute to the school action the simple consequence of a favorable enrollment. This is all the more serious if these student characteristics are also correlated with the expected high-school efficiency. For instance, if students from a privileged background have easier access to the "best" high schools or assign to the "best" teachers. Using simulations, Guarino et al. [2014] observe that when this selective match process operates, SGP is unable to take it into account. Specifically, they compare the ability of both SGP model and value-added model (with fixed-effects) to correctly assess the efficiency of "teachers", depending on the way students have been allocated in classes. When students are assigned to "teachers" based on their prior grade (dynamic grouping), SGP models perform in a similar way than a value-added model that control for past score. However, in case of selective match (teachers allocated to particular classes depending on their effectiveness), the SGP models underperform fixed effects models.

2.2 Econometric estimation of high school value-added models per quantile

Our objective is to propose indicators that offer a more complete picture of school action. Specifically, we want to model simultaneously whether a high school mitigate the differences in school achievements across students (equality) and if it increases achievements (efficiency). To that end, we rely on quantile regressions that allow us to model the value added of high school on different points of the distribution of test scores. These models rely on local linearization of the different quantiles of the conditional distribution. We can thus model both the median (which is close to the classic analysis of the average effect of the high school), but also a measure of the inequalities within school (defined by the difference in extreme quintiles). This modelisation is closed to both standard value-added models and SGP, but is expected to overcome the limitations of these models.

In practice, we model the distribution of test scores - and not only the mean - by quantile regressions. These regressions control for past scores - but also for other observable student characteristics. The specific effect of a school j is measured by fixed effects, which vary from one quantile to another. Formally, our model is the following:

$$q_\tau(Y_{ij}|X_{ij}, 1\{i \in j\}) = \underbrace{m(\tau) + X_{ij}\beta(\tau)}_{\text{Expectation}} + \underbrace{\alpha_j^*(\tau)}_{\text{School effect}} \tag{2}$$

where Y_{ij} corresponds to the grades obtained at the final test *baccalauréat*, X_{ij} are the observable covariates (initial test score and its square, social position index, gender and a repeating dummy). $\alpha_j^*(\tau)$ is the fixed effect specific to the high school j at the quantile $\tau \in [0, 1]$ of the distribution of score conditional on observables. No assumption is made about the distribution of the school specific effects $\alpha_j^*(\tau)$ among the population of schools (nor about their potential correlation with the characteristics of the enrollment of these schools). For each quantile, specific school effects $\alpha_j^*(\tau)$ are normalized across schools, with $\alpha_j^*(\tau) = \alpha_j(\tau) - \frac{1}{|J|}\sum_l \alpha_l(\tau)$ and $m(\tau) = \frac{1}{|J|}\sum_l \alpha_l(\tau) + \mu(\tau)$[5]. Our specification is thus very closed to the one proposed by Page et al. [2016]. However, while they assume a Gaussian distribution for the specific high-school effects, we do not have a priori on this distribution. Specifically, we do not assume that these effects are not correlated with other observable covariates used in the model. This is important in the French context, as one may assume that endogenous selection may occur in high-school enrollment (especially in private high schools that are free to choose their students unlike the state-run high schools).

Koenker [2005] has shown that estimates can be obtained as the solution of convex optimization program.[6] Estimates are run for three quantiles: the lowest quintile ($\tau = 0.2$), the median ($\tau = 0.5$) and the highest quintile ($\tau = 0.8$). In practice, we thus estimate each school specific effects, thus three coefficients for each J high schools. This specification can be related to the recent contributions to the econometric literature on quantile regressions on panel data (see for instance Koenker [2004], Canay [2011], Kato et al. [2012]). We have indeed students observations clustered by high school. While in panel data fixed effects are not interesting per se, specific high-school effects are the main parameter of interest here. Usually, we observe a higher number of observations per cluster (high school) than in standard panel data and thus do not face "small T" issues (see also Ponomareva [2011]). In some rare cases, some high schools enroll few students. As the estimates of value added may be not accurate, we restrict the estimation sample to high schools with more than 65 students in general tracks or 25 students in the technological tracks. These thresholds correspond to a trade-off between estimating specific high-school effects, that requires enough observations per high school and keeping the sample representative. Concerning

[5]Where $q_\tau(Y_{ij}|X_{ij}, 1\{i \in j\}) = \mu(\tau) + X_{ij}\beta(\tau) + \alpha_j(\tau)$, $\mu(\tau)$ is the intercept when the constraint for identification is $\alpha_1(\tau) = 0$

[6]Formally,

$$(\{\hat{\alpha}_j(\tau)\}_{1 \leq j \leq J}, \hat{\beta}(\tau)) = \text{argmin} \sum_{i,j} \rho_\tau(Y_{ij} - \mu(\tau) - \alpha_j(\tau) - X_{ij}\beta(\tau)) \tag{3}$$

where $\rho_\tau(u) = |u| (\tau 1\{u \geq 0\} + (1 - \tau)1\{u < 0\})$ et $\alpha_1(\tau) = 0$. where J is the number of high schools.

the latter, we keep 95% of students are kept in both curriculum, corresponding to respectively around 318,000 students enrolled in the general track in 1,759 high schools, and 123,000 students enrolled in the technological track in 1,549 high schools. This corresponds to exclude 15% of high schools from the initial sample (see Table 10 in the Appendix).[7]

We also estimate a more precise model that isolates not only the contribution of high schools on the achievement of their students as a whole, but also whether this contribution differs depending on the type of students (high ability or low ability for instance). Namely, we add high-school specific effects not only in the intercept, but also terms interacted with student characteristics. This lets us identify whether already high achieving or privileged students benefits more in terms of efficiency and equality from a particular school. In practice, we define different types of students depending on their prior schooling level at the beginning of high school and on their social background (measured by the social position index). These types are defined by the quartile of the distribution of past score (respectively the distribution of the social position index). Formally, we estimate:

$$q_\tau(Y_{ij}|X_{ij}, 1\{i \in j\}) = m(\tau) + X_{ij}\beta(\tau) + \gamma_q(\tau) + \underbrace{\alpha^*_{jq}(\tau)}_{\text{quartile-specific school effect}} \tag{4}$$

as before, X_{ij} coresponds to student observable characteristics. $q \in [1, 4]$ stands for the position of the student i in the distribution of prior test score and $\gamma_q(\tau)$ is a dummy for belonging to the type q ($q = 1$ when the student belongs to the lowest quartile of test score). $\alpha^*_{jq}(\tau)$ is a school and student-quartile specific effect. This specification is more demanding, as we estimate high-school specific effects by group of students within school. In order to have "enough" observations for the estimation of empirical distribution within clusters, for these specifications we keep only high schools which enroll at least 10 students in each quartile of past achievements in the general tracks (techological tracks in high school usually enroll less students and this condition is very stringent). In these specifications, we keep around 290,000 students enrolled in general track in around 1,500 high schools (see Table 10 in the appendix).

In practical terms, the numerical estimation of a large number of coefficients can be computationally complex. We use the adaptation of the Frish-Newton algorithm for sparse matrices proposed by Koenker and Ng [2005]. Estimates of the precision are made by bootstrap. The bootstrap sample is stratified per high school, with within high school sampling with replacement, the sampled unit being the student, i.e. the test score and his individual characteristics. Significance criteria are based on the $B = 500$ bootstrap b draws.[8]

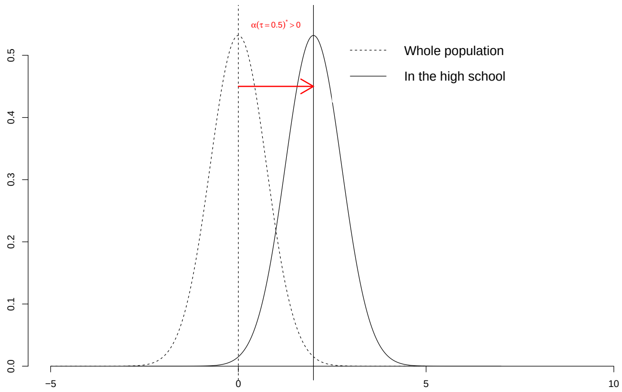
2.3 Efficient and unequal high schools: definitions

As in School Growth Percentile models, the quantile regression without school specific fixed effects measures the expected distribution of test scores conditional of observable characteristics. The school effect quantifies whether the performance reached by a given proportion of students outperform or underperform what is expected given the observable characteristics of students enrolled in this school. For example, a positive value of $\alpha^*(\tau = 0.2)$ means that the grade exceeded by 80% of students of this school is higher than what was expected given its enrollment.

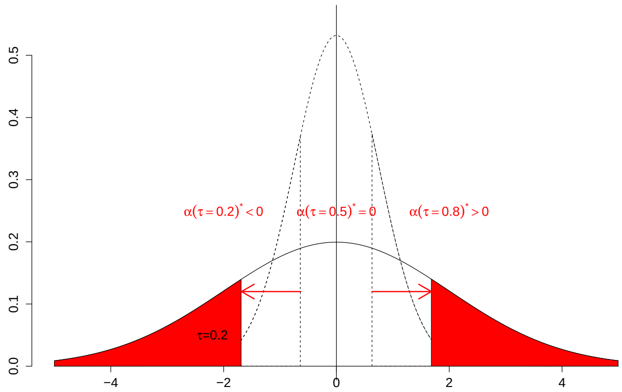
It is useful for the sake of interpretation to consider the function $\tau \to \alpha^*_j(\tau)$. It describes how the school-specific effect (for one high school j) evolves along the distribution. This function is constant when the high school operates a simple translation of the outcome conditional distribution (as illustrated in the panel (a) of the Figure 3). On the contrary, the panel (b) of the Figure

[7] We check that these sample restrictions do not alter our main results.

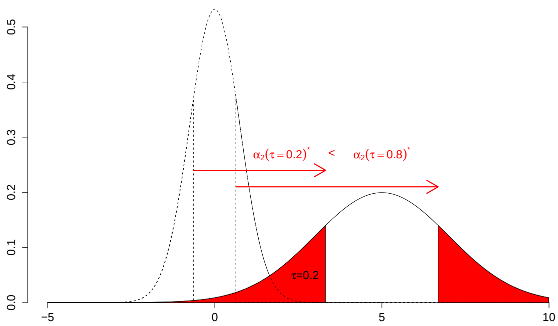
[8] The 95% confidence intervals computed with $[q_{0.025}(\{\alpha^{*,(b)}_\tau\}_{b\in B}), q_{0.975}(\{\alpha^{*,(b)}_\tau\}_{b\in B})]$. It does not exist closed-form estimates of the precision of the estimators for quantile regressions expect beyond strong assumptions on the underlying generating model. To our knowledge there are no results for inference on fixed effects by quantile.



(a) Homogeneous and positive effect : $\tau \to \alpha^*(\tau) = \alpha_0 > 0$



(b) Heterogeneous effect: $\tau \to \alpha^*(\tau)$ is increasing



(c) Heterogeneous and positive effect : $\tau \to \alpha^*(\tau)$ is positive increasing

Figure 3: Interpreting $\alpha^*(\tau)$ coefficients: type of school effects on the expected grade distribution

3 illustrate the case when the results obtained by the high school students are more scattered than expected considering their observable characteristics and this function is increasing. The opposite situation (less scattered grades results in a function decreasing in τ) may be also be observed. As illustrated in the panel (c), a high school may intensify inequalities in outcomes *and* increase the observed results at every quantile of the final test score distribution. Conversely, a high school may have a negative impact on the score (compared to what is expected) *and* reduce homogeneity. Both dimensions, that we respectively relate to efficiency and equality, are analyzed here.

In practice, the "average" efficiency of a high school j is measured by the specific fixed effect at the median $\hat{\alpha}_j^*(0.5)$. Besides, we approximate the monotonicity of the function $\alpha_j^*(\tau)$ for each high school j by the difference between the school-specific effects estimated respectively for the highest and lowest quintiles $\hat{\alpha}_j^*(0.8) - \hat{\alpha}_j^*(0.2)$. In the following, high schools for which this function is strictly monotonic are refered as "heterogeneous". These high schools tend to either widen or reduce inequalities in outcomes from what would be expected. In the former case, we refer to the school as "unequal" (since inequalities in outcomes are higher at the exit of the high school than expected), and in the latter case such a school as "egalitarian".

If the empirical probability of order conservation $\mathbb{P}_b(\text{order}(\alpha_{0.2}^{*,(b)}, \alpha_{0.8}^{*,(b)}) = \text{order}(\hat{\alpha}_{0.2}^*, \hat{\alpha}_{0.8}^*))$ is higher than 90%, the high school is assigned to the category of the corresponding heterogeneous effect, "unequal" or "egalitarian", otherwise, the effect is assumed homogeneous. As we perform thousands of test, we should account for multiple hypothesis testing (MHT). Under the full null hypothesis (no heterogenous effect in our sample), 10% of false positives are expected. The first type of corrections that have been proposed in the literature concentrates on the type one error (FWER, False Discovery Rate). It implies to compare the p-values to a threshold corrected by a factor inversely proportional to the number of tests N_H, hence $\frac{0.1}{N_H}$ for a 10% significance. However, these tests are very conservative as soon as the number of tests is high (Carvajal-Rodríguez et al. [2009]).[9] Several alternatives have been proposed in order to have tests with enough statistical power (for a discussion, see de Uña-Alvarez [2011] and Castro-Conde and de Uña-Álvarez [2015]). We follow Carvajal-Rodríguez et al. [2009] and de Uña-Alvarez [2012] and used their so called Sequential-Goodness-of-Fit (hereafter *SGoF*) method which has proven to provide enough power in cases where, as here, thousands of tests are performed.

3 Results: distributional effects

3.1 Estimation of quantile high-school effects and high-school classification

In our main specification, we estimate three quantile regressions with high-school fixed effects (first quintile, median and last quintile). The analysis is conducted separately for the general and technological tracks. For each regression, we derive a distribution of school-specific effects. The estimation requires an identification constraint and for interpretation reasons we choose to normalize all distributions (see Figure 4). We observe a high level of dispersion in the school-specific estimates. This dispersion is higher in the technological track than in the general one, and in both tracks it is higher for the lower quintile than the higher.

For each high school, we compare the estimates obtained at the lowest and highest quintiles in order to evaluate whether it has a "heterogeneous" effect, meaning whether it tends to modify the distribution in student outcomes compared to model prediction based on student observable characteristics. As explained above, a high school j is defined as unequal (respectively egalitarian) verifies $\alpha^j_{0.20} < \alpha^j_{0.80}$ (resp. $\alpha^j_{0.20} > \alpha^j_{0.80}$). After correction for multiple hypothesis testing, we find heterogenous effects in 16.7 % (13.6%) of high schools in the general (technological) track, with unequal and egalitarian schools equally represented (see Table 11).[10]

Regarding the estimates corresponding to other covariates, they are in line with previous results in related literature (see Table 3). The three analyzed quantiles of the final grades mainly depend on the student past scores, and the quadratic term is also significant and positive. The conditional distribution of final grades of students who had repeated a year during their past schooling is also strongly below that of non repeater students, the gap widening at the bottom of the distribution. Girls usually have better and less dispersed final results than boys, especially in the technology track. Higher social position index is correlated with better final grade. For the sake of comparison, estimates without high-school fixed effects are also shown. One may notice that most of the estimates vary between the two specifications (even if it does not correspond to a formal statistical test), especially the social position index and the prior score. This is consistent with the fact that the most favored students may be selected in the best schools. As suggested

[9] In practice it requires a very large number of bootstrap samples to so that $\frac{1}{B}$ is at least smaller than $\frac{0.1}{N_H}$. Otherwise, we will accept that order is conserved only if it is conserved on *all* the bootstrap draws.

[10] When we do not correct for multiple testing, these proportions are 29.0% and 28.8% in respectively general and technology tracks.

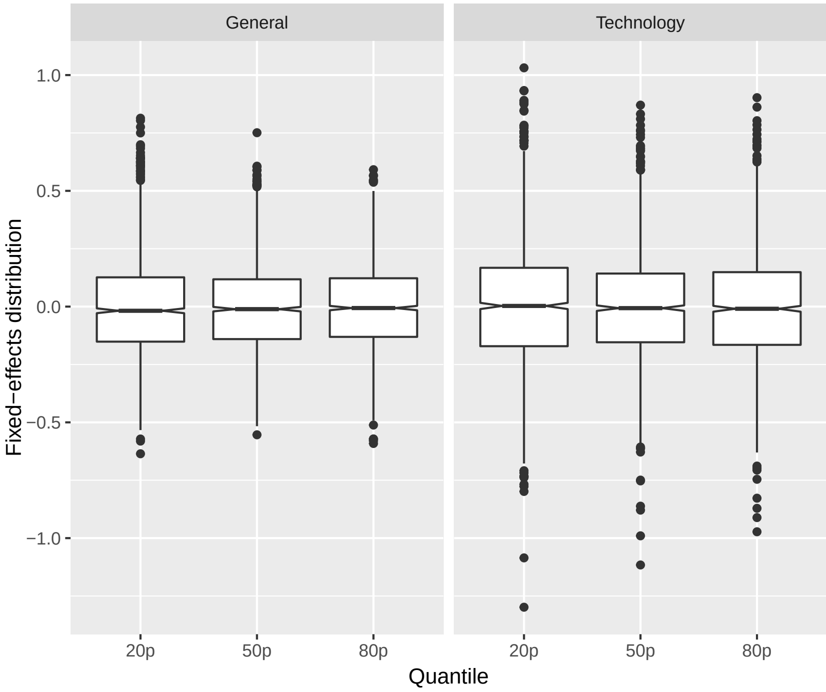


Figure 4: Characteristics of the distributions of the high-school specific effects

Table 2: Proportion of egalitarian and unequal high schools

	General	Technology
Unequal	13.6	12.5
Egalitarian	15.2	14.2

Restriction to high school with headcount $\geq$ 65 (resp. 25) in the general (technology) track. A high school j is defined as egalitarian (resp. unequal) when $\alpha^j_{0.20} > \alpha^j_{0.80}$ (resp. $\alpha^j_{0.20} < \alpha^j_{0.80}$)

by Guarino et al. [2014], neglecting these correlations would result in biased estimates.

Table 3: Quantile regression estimates - main covariates

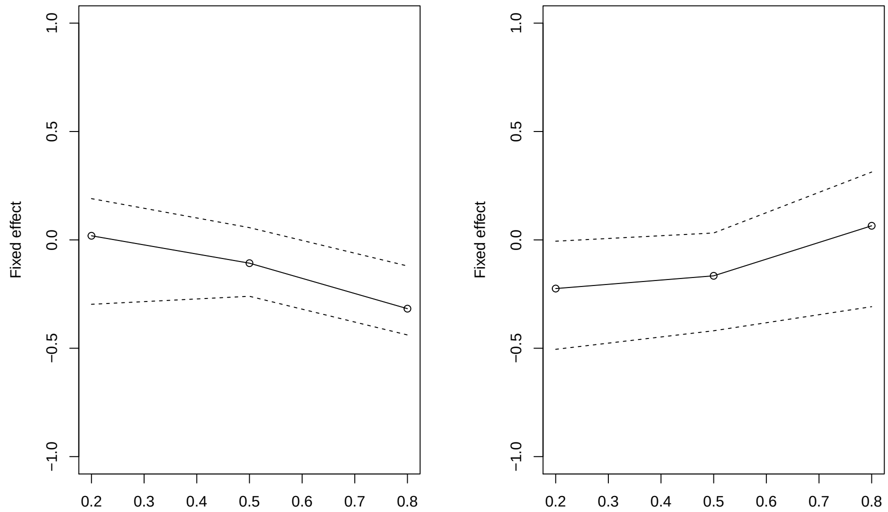
	With fixed effects						Without fixed effects					
Quantile	20%		50%		80%		20%		50%		80%	
General track												
Intercept	−0.705	(0.059)	−0.097	(0.082)	0.683	(0.143)	−0.586	(0.004)	−0.004	(0.004)	0.607	(0.004)
Prior score	0.593	(0.002)	0.632	(0.002)	0.646	(0.002)	0.622	(0.002)	0.655	(0.002)	0.661	(0.002)
Prior score squared	0.107	(0.001)	0.105	(0.001)	0.082	(0.001)	0.108	(0.001)	0.105	(0.001)	0.079	(0.001)
Social position	0.079	(0.002)	0.079	(0.002)	0.079	(0.002)	0.108	(0.002)	0.106	(0.002)	0.102	(0.002)
Repeaters	−0.271	(0.008)	−0.245	(0.007)	−0.193	(0.008)	−0.284	(0.008)	−0.253	(0.008)	−0.201	(0.008)
Girls	0.080	(0.004)	0.052	(0.003)	0.032	(0.004)	0.080	(0.004)	0.051	(0.003)	0.035	(0.004)
L (réf ES)	0.074	(0.005)	0.086	(0.005)	0.088	(0.006)	0.065	(0.005)	0.077	(0.005)	0.070	(0.005)
S	−0.194	(0.004)	−0.172	(0.004)	−0.147	(0.004)	−0.214	(0.004)	−0.184	(0.004)	−0.162	(0.004)
Technology track												
Intercept	−0.518	(0.171)	−0.139	(0.089)	0.329	(0.138)	−1.004	(0.022)	−0.423	(0.019)	0.176	(0.024)
Prior score	0.358	(0.004)	0.392	(0.003)	0.408	(0.004)	0.388	(0.003)	0.404	(0.003)	0.409	(0.003)
Prior score squared	0.018	(0.003)	0.025	(0.002)	0.034	(0.002)	0.009	(0.002)	0.021	(0.002)	0.029	(0.002)
Social position	0.034	(0.004)	0.027	(0.003)	0.027	(0.003)	0.060	(0.003)	0.049	(0.003)	0.043	(0.003)
Repeaters	−0.285	(0.009)	−0.258	(0.007)	−0.228	(0.009)	−0.299	(0.009)	−0.268	(0.007)	−0.236	(0.009)
Girls	0.242	(0.007)	0.211	(0.006)	0.189	(0.007)	0.235	(0.008)	0.206	(0.006)	0.189	(0.008)
ST2S (réf.HOT)	0.205	(0.057)	0.229	(0.039)	0.291	(0.056)	0.218	(0.023)	0.235	(0.020)	0.234	(0.025)
STD2A	0.362	(0.066)	0.385	(0.049)	0.511	(0.060)	0.339	(0.030)	0.368	(0.026)	0.451	(0.030)
STI2D	0.371	(0.059)	0.464	(0.039)	0.624	(0.054)	0.329	(0.023)	0.428	(0.020)	0.547	(0.025)
STL	0.500	(0.061)	0.604	(0.039)	0.717	(0.055)	0.450	(0.027)	0.579	(0.022)	0.637	(0.028)
STMG	0.360	(0.057)	0.397	(0.038)	0.456	(0.055)	0.354	(0.022)	0.407	(0.019)	0.430	(0.024)

Notes: Estimation of the $\hat{\beta}(\tau)$ in the equation 2. The variable of interest (final high-school grade), as well as continuous covariates are normalized and standardized. *Séries*: L: Humanities, ES: Economics and Social Sciences, S: Sciences. ST2S: Sciences and Technologies in Health and Social, HOT: Hotel and restaurants management, STD2A: Sciences and Technologies in Design and Applied Arts, STI2D:Industrial Science and Technologies and sustainable development, STL:Laboratory Science and Technologies, STMG: Management Sciences and Technologies. Restriction to high-school with headcount ≥ 65 (resp. 25) in the general (technology) track.

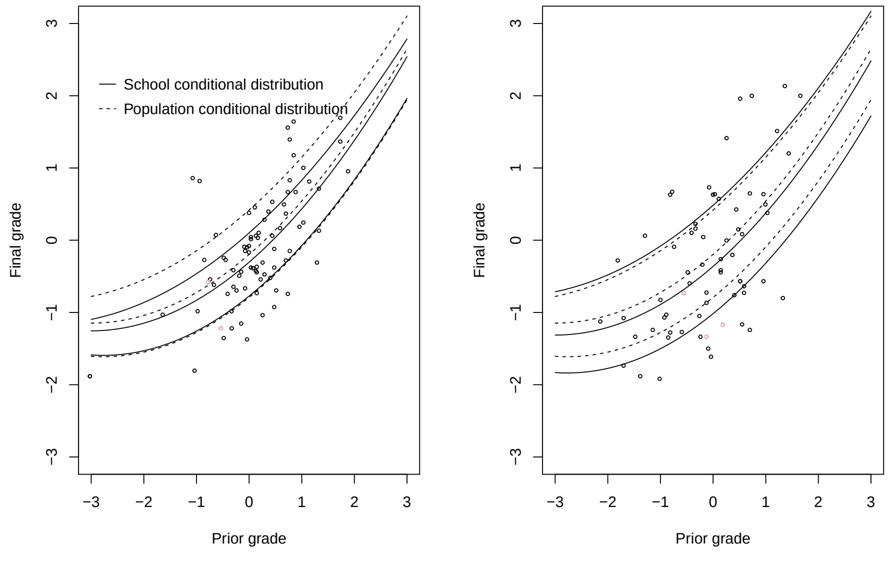
To illustrate how confining to the mean effect can be partially misleading, we present the cases of two high schools of opposite classification. Figure 5 represents the empirical distributions of the past and final grade Y_{ij} as a function of the prior grade x_{ij} as observed in two high schools j_1 and j_2 and compares it with the expected conditional quantiles in these high schools for $\tau \in \{0.2, 0.5, 0.8\}$ (from to the estimates of equation (2)). The dotted lines represent the prediction without the high-school effect while the solid lines take them into account. The shape of the conditional quantiles is driven by the quadratic dependance in prior grades. For the high school where the triplet $\alpha^*_{i,\tau}, \tau \in \{0.2, 0.5, 0.8\}$ is in descending order (left panel) - and thus classified as egalitarian, we expect a smaller dispersion of final grades than the one that would have been expected from the observable characteristics of these pupils. On the contrary, in the high school where the triplet is in ascending sequence (right panel in Figure 5) - and thus classifies as unequal- the empirical dispersion of the conditional distribution of final grade is higher than that would have been observed "on average".

These effects do not coincide with an average effect. The classification in the egalitarian group of the high school in the left panel of Figure 5 comes from the fact that it performs worse than average at the top of the final grade distribution.

The fact that a high school reduces the inequalities in academic level among its students may be an ambiguous indicator, though. For instance, in the previous example (Figure 5) for both the unequal and the egalitarian high schools, the median effect (which approximates the average value-added effect as usually measured) is not significantly different from zero. One may assume also that more egalitarian results are obtained by lowering the academic standards within the



(a) high-school specific effects



(b) Predicted distributions of final grades conditional on prior grade (quantile 20%,50%, 80%) before and after adding specific high-school effect

Figure 5: Empirical and predicted conditional distribution in two high-schools, and specific high-school effects

schools. It could be the case that the performances of the students within the school are equally worse than the average (race-to-the-bottom phenomena). We thus look more carefully to the correlation between equality, as defined by the comparison between the first and last quintiles and a measure of performance. For the latter, we use the estimate of the high-school specific effect at the median. A "performing" high school is one where the median estimate is significantly positive (and on the contrary a non-performing is a high school whose median value-added is negative). When correlating these two classifications, we observe a - not so expected - positive correlation between the classification to the performing class and to the egalitarian one. The egalitarian schools are less often under-performing and have more often than other categories positive value-added (see Table 4). If most positive value-added schools do not display heterogeneous effects, they are more often egalitarian than unequal (29 % vs 8% in general track) while the opposite holds for negative value added schools.

Table 4: Comparison of high schools in terms of dispersion in value-added vs median value-added (in %)

	Unequal	No het. effect	Equal	
General				
Negative median value added	15.6	76.1	8.3	100
	30.9	*29.2*	*15.6*	
Non significant median value-added	15.5	73.1	11.3	100
	55.1	*50.5*	*38.3*	
Positive median value added	8.4	62.6	29.1	100
	14	*20.3*	*46.1*	
	100	*100*	*100*	
Technological				
Negative median value added	16.4	72.5	11.1	100
	25.8	18.9	16.2	
Non significant median value added	11.8	75.8	12.4	100
	60.5	64.5	59.3	
Positive median value added	9.8	71.4	18.8	100
	13.7	16.6	24.5	
	100	*100*	*100*	

Source: FAERE and APAE database, author calculations.

We use a more complex specification that interacts the high-school effects with the position of the students either in the distribution of prior grades or social position index (as defined by the Equation 4). We restrict the sample to the general track in order to have enough observations in each high schools. The main parameters of the obtained empirical distribution of estimated fixed effects are of the same extent whatever the quartiles (see Figure 10 in the Appendix). We obtain rather similar proportion of unequal or egalitarian high schools when considering separately every types of students (see Table 5). We observe that around one-third of the high schools are classified as heterogeneous for at least one type of students. However, when a high school is classified as heterogeneous for one type of students, it is rarely heterogeneous on another. We also observe that the classification using all the sample or restricted to each quartile are loosely consistent, in the sense that a high school classified as unequal (respectively egalitarian) for one type of students is not classified as egalitarian (respectively unequal) considering the enrollment as a whole. In half of the cases, the global classification matches that found on a subpopulation, and in half of the cases a high school classified as heterogeneous for one quartile is classified as homogeneous when considering the whole sample (see Table 13 in the Appendix).

Table 5: Proportion of egalitarian and unequal high schools by quartile (general track)

	Overall	On at least		On quartile			
		a quartile	two quartile	Q1	Q2	Q3	Q4
Unequal	0.148	0.145	0.014	0.041	0.042	0.038	0.039
Egalitarian	0.141	0.191	0.023	0.055	0.052	0.052	0.058

Source: FAERE and APAE database, baccalaureat 2015, authors' calculation. Note: Restriction to high school with headcount ≥ 10 in each quartiles. A high school j is defined as egalitarian (resp. unequal) when $\alpha^j_{0.20} > \alpha^j_{0.80}$ (resp. $\alpha^j_{0.20} < \alpha^j_{0.80}$)

3.2 Selection of students and high-school classification

We observe that private high schools are more often classified as "egalitarian" than public (state-run) high schools. This may be due to a more individualized pedagogy that reduces inequalities. However, this may also due to selection effects - indeed, private high schools are completly free to select the students they enroll (while many restrictions apply for state-run high schools). While we control for the academic level of students just before the high schools, some characteristics of students may be observed during the recruitment process but not measured in the data (for instance the parent involment). Besides, students may be selected during the high school years. As the raw percentage of high school students that have passed the "baccalauréat" is highly publicized and scrutinized by parents, high schools may be tempted to get rid of previously promising students that have eventually obtained disappointing achievements on the first or second year of high school. One objective of high schools could be to enter only the best students at the final examination. If state-run high schools are in principle not allowed to choose their students as private high schools are, indirect selection may also occur. For instance, by directing students after the first year of high school to a stream that is not available in the school (as the headmaster of the school has the final decision regarding academic direction taken in secondary school).

This selection effect in French high schools is well known and for this reason the French ministry of education computed not only the average value-added on success rate at the final examination, but also on the "access rate" within high school. The latter measures the retention rate, compared to what would have been expected from the enrollment composition. The retention rate corresponds to the ratio of students that are still in the high schools from one year to another (they are either enrolled in the next level, or have repeated the same level *within* the same school). As retention rate may be also be due to volontary mobilities of students (because they are not satisfied with the school for instance, or simply because they have moved to a location too far from the previous school), the indicator compares the observed retention rate to the one that should have been expected when relying on the composition of the high-school enrollment. We observe that the distributions of this indicator vary depending on whether we consider private or state-run high schools (see Figure 6). In private high schools, the actual access rate is almost always smaller than the one that would have been expected. While in the public sector we observe a symetric distribution, centered around zero, it is much more dispersed and asymetric in the private sector.

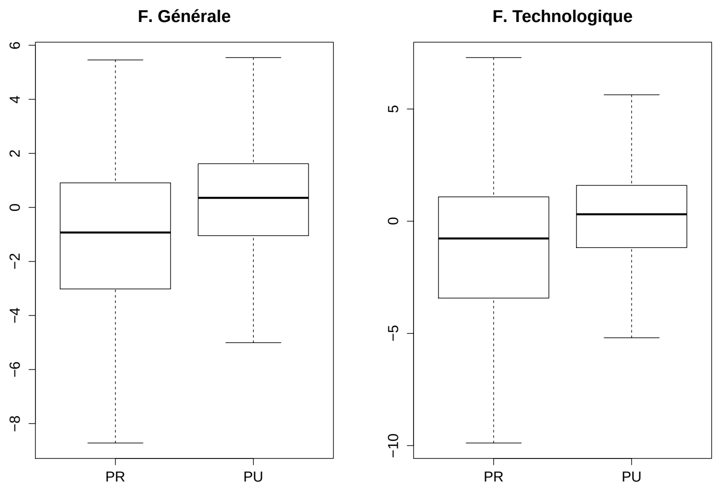


Figure 6: Access rate in public and private schools (Left panel: general track, right panel: technological track)

This selection process may affect the classification of high schools. For instance, if a high school chooses to get rid of poor performers (compared to what would have been expected from their previous results) in order to avoid being held accountable of their failure at the final exam, the estimated high-school specific effect at the lower quintile will be artificially high. In such a case of "cream skimming", the high school may be classified instead of unequal as homogeneous or even egalitarian, or instead of homogeneous as egalitarian.

As a first test of this mechanism, we reply the previous analysis but estimating the high-school fixed effect based on students enrolled in high schools the year before the final exam (even if they are not actually enrolled in the same high-shool at the date of the exam in 2015). The estimated high-school specific fixed-effect are thus partially biased (as they mixed the impact of this specific high school where all students were enrolled the year before, but also partially reflects the impact of the high schools where some students have been actually enrolled the year of the exam). However, they are expected to be less sensitive to the selection effects (even if a high school left behind the lowest achievers for the year of the exam they will be re-attributed to this high school for the estimation). We thus confront the classification obtained in this way (using the enrollment in 2014) with the one obtained using the enrollment in 2015. Both classifications appear highly correlated (see Table 6). Comfortingly, in no case we observe a high divergence in classification (from unequal to egalitarian), suggesting that the endogenous selection described below is not of sufficient magnitude to radically change the classification. We observe that a significant part of the high schools classified as unequal when using the enrollment the year before the exam in 2014 are classified as homogeneous using the 2015 enrollment, and some classified as homogeneous with 2014 enrollement are classified as egalitarian using the 2015 enrollment. We also observe similar changes in the other way (from homogeneous to unequal and from egalitarian to homogeneous) and it is thus difficult to analyze whether this only reflects the fact that the estimation of the specific high school effects is blurred when using the cohorts of the previous year.

As a second test, we use instead of the classification a more precise information provided by the difference between the estimated high-school specific effects at respectively the highest and lowest quintiles ($\hat{\alpha}_{0.8} - \hat{\alpha}_{0.2}$). We consider here the continuous variable, the classification relies only on its sign. The higher this variable, the more unequal the high school is. We test whether this indicator of dispersion within high school varies with the student mobilities (forced or chosen) or the sector (private or state-run). We indeed observe that in the general track, this dispersion measure is positively correlated with the access rate indicator (Table 6), meaning that high schools that select less students during high-school years are more likely to be classified as homogenenous or unequal than egalitarian. This effect is merely due to private high schools. While on average, the indicator of within-dispersion is not significantly different in both sectors, this positive correlation with the access rate indicator is significant in private sectors. In the technological track however, we observe the opposite pattern: the private sector has a strong negative impact on the indicator of dispersion, and this is amplified by a high access rate (meaning that it is not artifially created by sacking low performing students). All in all, this suggests that eventhough we find evidence of private-sector cream-skimming effects in the general track, the impact on the inequality measure is most likely weak.

Table 6: Transition between categories when taking into account the mobility of students

General track			
Students 2015/Students 2014	U	H	E
U	205	28	0
H	44	1161	52
E	0	40	220
technological track			
Students 2015/Students 2014	U	H	E
U	148	37	0
H	34	1067	43
E	0	51	171

Sources: FAERE and APAE database (baccalaureat 2014 and 2015), authors' calculation. Note: U stands for unequal, H for heterogeneous and E for egalitarian.

4 Characterization of egalitarian and unequal high schools

The previous analysis allows us to classify high schools depending on their educational outcomes. In order to better characterize "egalitarian" or "unequal" high schools, we then correlate these labels with observable characteristics. The underlying question is to know whether we could predict that a high school is more likely to be one or other type depending on its staff or enrollment characteristics. The objective here is not to identify causal relationships between these attributes and the high-school outcomes. It may represent the situation faced by parents who have to choose between different high schools and want to decide with more information than the "raw" average passing rate.

Specifically, we use a set of variables first related to the high-school level variables (for instance private or public sector or school district, etc.), second to its staff (proportion of respectively senior or junior teachers for example) and finally to its enrollment (median social position index, academic and social heterogeneities as measured by interquartile of these variables within school, etc.). Many of these variables are highly correlated, and we do not have a priori ideas of the way they main enter the model. We thus rely on random forests, which are by now a common tool for classification issues and do not rely on a predefined parametric specification.

Table 7: Inequality measures depending on access rate within the high school and sector.

	Dependent variable:			
	$\Delta\alpha$			
	General		Technological	
	(1)	(2)	(3)	(4)
$\Delta\alpha_{14}$	0.903***	0.905***	0.912***	0.904***
	(0.008)	(0.008)	(0.011)	(0.011)
Access rate to final grade	0.001***	0.0004	−0.001*	−0.0002
	(0.0003)	(0.0003)	(0.0005)	(0.001)
Private sector		0.002		−0.020***
		(0.002)		(0.006)
Access rate to final grade:Private sector		0.001*		−0.003***
		(0.001)		(0.001)
Constant	0.0005	0.0002	−0.001	0.002
	(0.001)	(0.001)	(0.002)	(0.002)
Observations	1,733	1,733	1,504	1,504
R^2	0.893	0.894	0.828	0.830

Source: FAERE and APAE database, authors' calculation. Note : $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Generally speaking, a decision tree associates rectangular regions of the explanatory variables space (here school features) to a classification (here "egalitarian", "unequal" or "homogeneous") by solving an iterative optimization problem. Namely, at each nodes it splits the corresponding sample on one single variable, the one that provides the best homogeneous subsets of population (depending on criteria corresponding to lowest entropy or Gini impurity). However, as classification trees are known for being prone to overfit (meaning they have low bias but very high variance), it is common to use random forests (Friedman et al. [2001]) that aggregate the outcomes of several classification trees grown from bootstrap samples of the original data in order to improve the performance of the classifier. In practice, each tree corresponds to a classification (a "classifier"), and for one observation it "votes" for one or the other class depending on the observations' features. The forest classifies this observation to the class that has obtained the higher number of "votes" from all the trees.

In random forests, randomness is introduced in two ways to obtain several decorrelated trees and to decrease the variance. First, by generating bootstrap samples (i.e. samples with replacement from the original data to get a sample of the same size as the original one). Second, at each node only a subset of the variables, selected at random from the entire set of predictors, is used to split the node. The overall performance of a random forest is obtained by aggregating "out-of-bag errors", measured for each bootstrap sample using the corresponding classifier (the tree grown on this training sample) on the subset of observations of the original sample that is not included in the training sample.

Main performance metrics are provided by the accuracy, which corresponds to the proportion of units correctly classified. However, this indicator may lead to spurious result as the data are imbalanced: the homogeneous class represents almost two-third of high schools. A classifier that classifies all high schools in this majority class would obtain rather high level of accuracy (but would not be very informative). A classic way to deal with this issue is to up-sample the minority classes in order to obtain a balanced panel. Besides, the Kappa statistic provides a performance indicator that is not sensitive to imbalancing. Specifically, it compares this observed accuracy with an expected accuracy (that would have been obtained by random allocation).

Table 8: Performance metrics of the random forests

Sample	Accuracy	Kappa	Class error (final model)		
	(OOB samples)		Homogeneous	Egalitarian	Unequal
General track	0.675	0.081	0.103	0.004	0.004
Technological track	0.727	0.058	0.078	0.000	0.003

Source: FAERE and APAE databases, authors' calculation.
Note: For both models, the final models comprise 500 trees and results from sampling 26 variables at each node of an individual tree (categorical variables count for the sum of their levels). In order to obtain balanced samples the minority classes (egalitarian and unequal classes) are over-represented using up-sampling. The accuracy statistic is computed by averaging the proportion of correctly classified observations obtained on the 10 out-of-the-bag samples. The Kappa statistics corresponds to the amount of correctly classified observations compared to what would have been expected by random classification. The class error is the rate of misclassification on the training sample.

One may also rank the importance of variables in the classifier provided by one random forest. This ranking depends on the loss of classification accuracy (as measured by the out-of-bag error) entailed by randomly permuting the value of one specific variable among the training sample. The higher this loss in accuracy, the higher the importance of the variable in the classification process. An illustration is provided in the Table 9, that provides the loss of accuracy for each variables used for the classification in one of the three types (unequal, homogeneous, egalitarian) for both tracks. Overall, these observable covariates appear to intervene more in the classifiers in the general track than in the technological one. In both cases however, the median prior grade in the high school and its headcount are among the variables that count the most in the classification.

However, such calculations provide limited information on the way variables impact the classification (for instance, whether a high value of one "important" variable results in a higher probability that the high school is classified as egalitarian rather than homogeneous). Partial dependence plots provide mode detailed insights in this respect. They can be interpreted as the marginal effect of each variable on the classification into a specific category (for instance, how the vote for the egalitarian category depends on the school size measured by headcount). For a given variable Z it defines the function $z \to f(z, x)$ that measures how the "votes" for a given class k depends on the value z of this variable (setting the vector of the values x of other covariates fixed). Formally it may written here:

$$f(z, x) = logp_k(z, x) - \frac{1}{3} \sum_{l \in \{u,e,h\}} logp_l(z, x)$$

where $p_k(z, x)$ corresponds to the fraction of trees that have voted for the class k for values (z, x), and e stands for "egalitarian", u for "unequal" and h for "homogeneous". In practice, for a given variable Z it sets a grid of value z, and take the mean of f over the observed values of other classification variables X in the sample (it may be viewed as the partial likelihood of the sample). As before, the variables are standardized and normalized,[11] so one may easily overlay the partial dependence plots for various classification variables. In practice, one should compute this value only on the empirical support of the analyzed variable - and we choose to exclude the extreme values of the support (typically either the 5% either the 1% of both side of the empirical support).

[11] Except when they are naturally in percents, for readibility.

Table 9: Importance of school levels variables for classification (Random Forest)

	Mean decrease accuracy	
	General	Technology
High-school characteristics		
Headcount	16.50	12.48
Class size	5.21	8.94
Hours per students	5.76	6.89
Median value-added	7.91	11.51
Sector (public or private)	9.51	8.63
Access rate from second to final year	5.38	7.92
Access rate from first to second year	6.36	5.92
High school paired with middle school	0.72	0.91
School districts (*Académies*)	.	.
Enrollment characteristics		
Proportion of girls	6.11	6.21
Proportion of repeaters	10.77	8.14
Median social position index	5.49	6.26
Social heterogeneity	6.69	5.68
Median prior grade	15.76	24.00
Academic heterogeneity	26.41	10.72
Prop. of students living in sensitive urban zone	5.10	4.88
Teacher characteristics		
Mean teacher age	4.32	4.26
Prop. part-time teachers	6.80	6.27
Prop. of junior teachers ($\leq$ 35)	4.49	7.16
Prop. of senior teachers ($\geq$ 50)	4.97	5.49
Prop. of female teachers	12.72	7.26

Note: Academic and social heterogeneity are the interquartile of respectively prior grades and social position index within the school. School districts are coded as 31 dummies variables and do not appear here: they all have a MDA smaller than 1 in both tracks, except one (Nantes, that has a MDA of about 3 in both tracks).

For both tracks, we distinguish in Figure 7 four variables among the ones that have been identified as the most important in the classification. In this case, we can distinguish whether a given variable impacts more the vote for the “egalitarian”, “unequal” or “homogeneous” class (we do not represent here the plot corresponding to the plot of the last category), and in what sense it does. In the general track, the high-school headcount appears to have a high impact on the votes for unequal categories. Potential effort granted to each student may decrease with size, as may unifying social interactions. This can be compared to a large literature on the link between school size and students’ achievements. In a recent literature survey on this issue, [Scheerens et al.] (see also [Leithwood and Jantzi, 2009]) conclude that if school size has probably no impact on cognitive outcomes, it may improve equity (as disadvantaged students are doing better in smaller schools).[12] This latter result may be consistent with the link observed here between the size of the high school and the egalitarian treatment of students. Indeed, when we use a more precise classification of high schools for the general track, based on whether they have an unequal, homogeneous or egalitarian behaviour by group of students (see Section 3), for students with low academic level at the beginning of the high school (first quartile at testscore of the pre-highschool exam DNB) we observe that the probability that a high school is in an egalitarian class varies dramatically with headcount (see Figure 9). The marginal effect of headcount on inequality appears less marked for the students at the top of the distribution of prior grades. In general tracks, the social backgrounds of students is also linked with the votes for unequal or egalitarian type of the high- school (pooling all students in one sample), the socially favored schools being predicted as more egalitarian and less unequal (Figure 7). When looking separately by types of student, this correlation appears due to the top quartile of prior achievements. It appears that in high schools characterized by a highly socially-advantaged environment, the grades of previously top-achieving students varies much less than expected (votes for equality increases). In high schools with high proportion of students from disadvantaged background, the grades of previously low-achieving students vary more than expected. This can be due to unobserved characteristics of these students (see Figure 9).

For both tracks, the composition of enrollment in terms of academic performance has a strong impact on the classification. The higher the previous academic achievement of pupils in the high school, the lower the relative probability of being classified as unequal, and the higher the relative probability of being classified as egalitarian. Previous analyses on French high schools already emphasize that the initial academic level of high-school enrollment is the main predictor of the expected average high-school outcomes (see for instance Duclos and Murat [2014]). The results here suggest that it impacts also the variance of these expected outcomes. A possible explanation could be that when a large share of students already performs well, it is easier to focus on the less performing ones. However, this pattern varies with the type of students. We observe a strong positive relationship between the enrollment academic level (measured by median prior level) and the probability that the high school is classified as egalitarian for the first two quartiles (that corresponds to the lowest performing students). For the first quartile however, this link is reversed as the very top of the support. Symetrically, we observe a U-shape correlation between the relative probability that the high school is classified as unequal and its academic median level. This correlation is consistent with different peer effects model (see e.g. [Sacerdote, 2011b])- for instance, the invidious comparison model assumes that students are harmed by the presence of better students in the same classroom or the boutique model suggests that a student performs better if it is surrounded by similar peers. [Hoxby and Weingarth, 2005] obtain for instance results in favor of the latter for the U.S. However, as far as we know little is known about the French situation.[13] These stylized facts do not imply causal links, and further analysis is required

[12] For France, Afsa [2014] (in French) notices that, once controlled for the composition of schools, smaller medium schools perform better, especially for students from disadvantaged backgrounds.

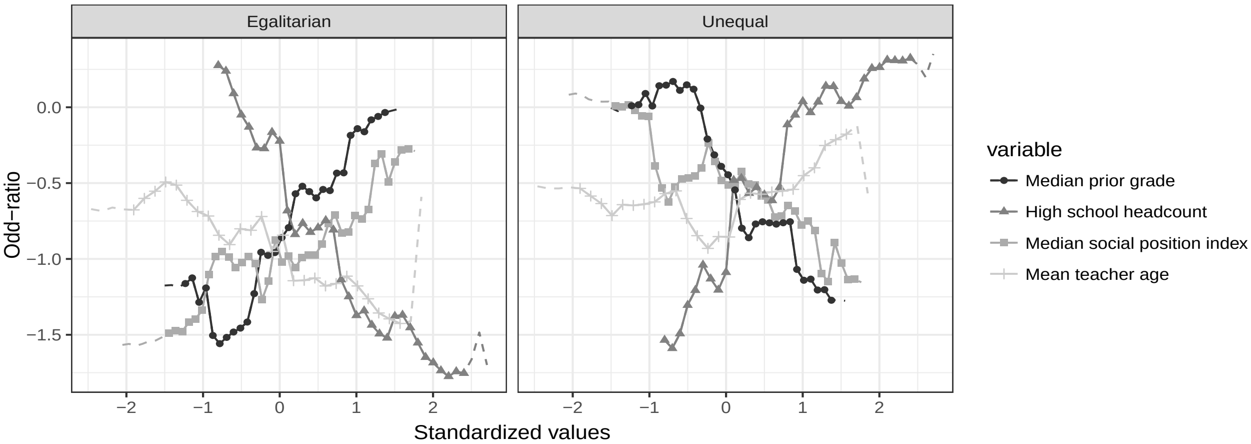
[13] Boutchenik and Maillard [2018] on similar data than those used here obtain preliminary evidence that suggests that the peer effects in French are highly heterogeneous depending on the previous level of

to validate this assumption.

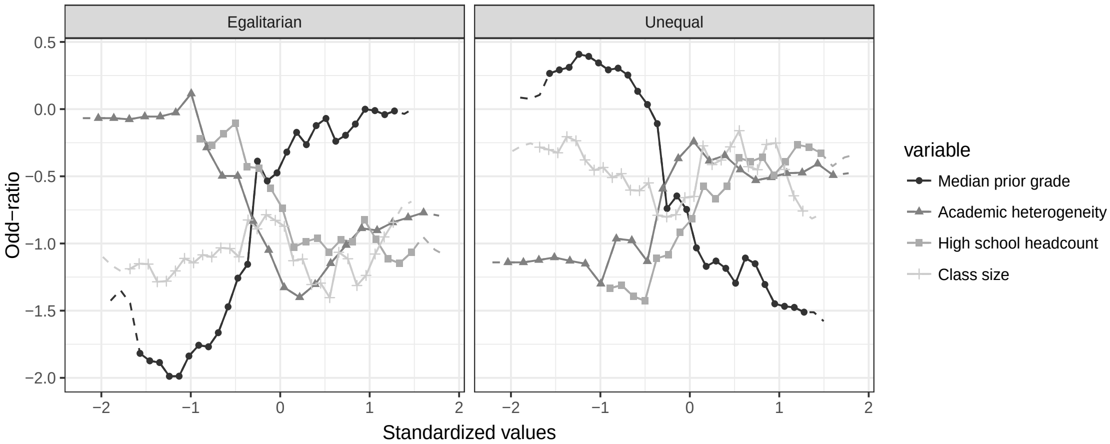
We can however relate the average effect on academic achievements and the inequalities outcome in the high school. To do so, the high school effect as measured at the median (meaning the $\hat{\alpha}_{0.5}$ as obtained in previous analysis) has been introduced as a classification variable. This indicator is closely related to classic measure of value-added (that corresponds to the high-school fixed effect measured at the mean). The dependancy of the classification in the "unequal" or "egalitarian" class appears non trivial, however (see Figure 8). In both tracks, we observe a U-shaped dependency for the detection of inequality. For most of the support of the high-school median value-added (a measure of the difference between the median achievements observed in this specific high school and what would have been expected considering its enrollment) we observe a negative link with the relative probability that this high school is classified as unequal. However, for the upper quarter of the support, we observe on the contrary a positive relationship. The higher the median value-added at the top of its support, the more likely this high school is classified as unequal rather than another category. Symetrically, for the top value-added schools egalitarian votes decrease while on the rest of the support a higher value-added would on the contrary increases the probability to be classified as egalitarian. Taken together, these links suggest that performance does not seem to be achieved at the expense of equity. However, we observe that the most performing high schools at the median are at the same time more likely to be either unequal (amplifying academic outcomes gap within school compared to what could have been expected considering the composition of enrollment) or egalitarian (reducing this gap)[14] in the general track. In the technological track, the relationship between egalitarian votes and median value-added is globally increasing, although the inversion rise toward unequal votes happen earlier in the support of value-added. This result may be compared to the vast literature on the impact of peer effects. For instance, in a comprenhensive review Sacerdote [2011a] emphasizes the considerable heterogeneity in peer effects and non-linear relationship between the academic level of a student and that of their peers. Again, this is not a causal analysis but only emphasizes that one may observe non trivial correlation between high-school average and distributional impacts. We cannot exclude for instance that the best schools attract a more homogeneous enrollment (in terms of unobservable variables that may also impact academic outcomes).

students.

[14]Or, equivalently, high performing schools in terms of median are less likely to be classified as homogeneous, see Figure 11 in the Appendix).



(a) General



(b) Technological

Figure 7: Partial dependence plots. Odd-ratios as function of standardized school-level variables, chosen as the most discriminating in terms of marginal effect on votes on their support. *For variable z, the panel "unequal" corresponds to the average over school characteristics x of* $logp_u(z,x) - \frac{1}{3}\sum_{l\in\{u,e,h\}} logp_l(z,x)$ *(e stand for egalitarian, h for homogeneous). The solid line is estimated on the grid of values of the empirical support of the variable, excluding 5% of schools on each side. The dotted lines is estimated on the support excluding 1% of schools on each side.*

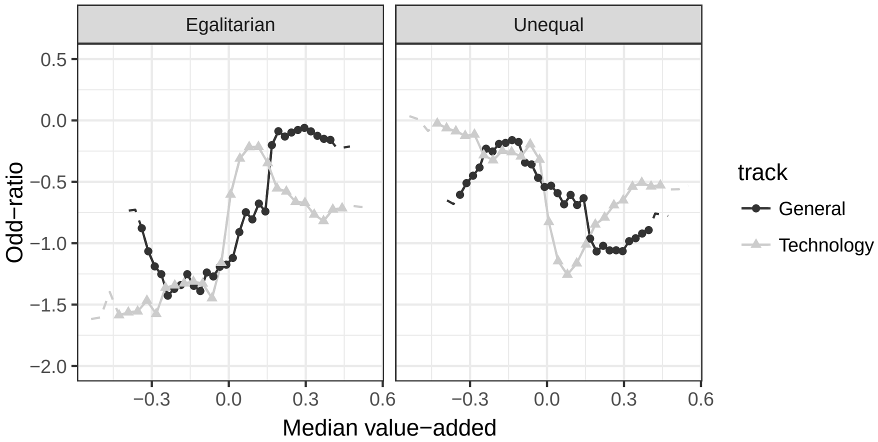
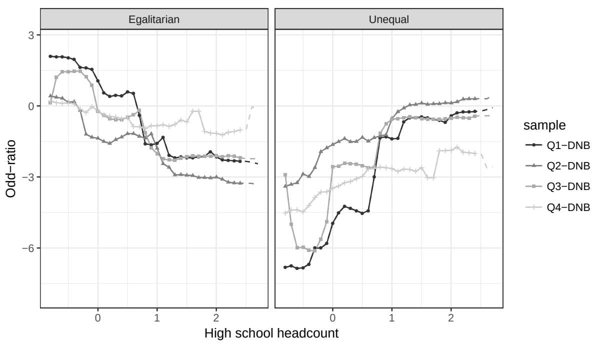
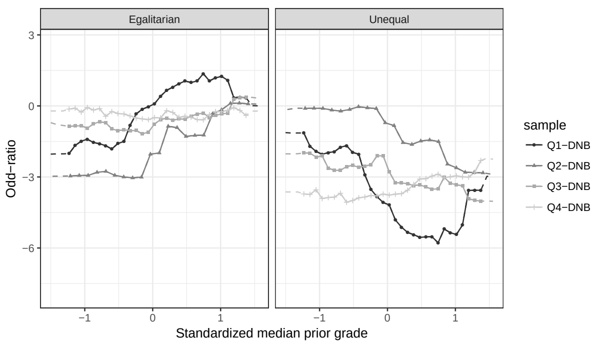


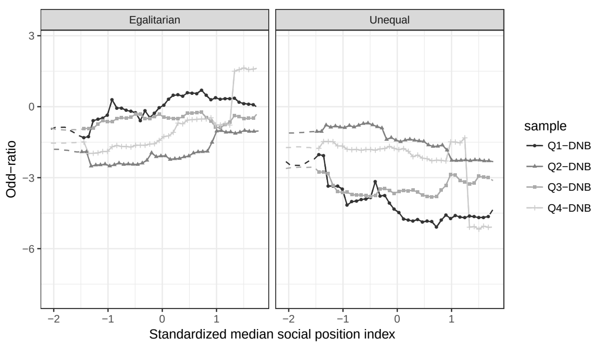
Figure 8: Partial dependence plot on performance (as measured by the median value-added). *Vote for u ="unequal" corresponds to the average over school characteristics x of* $log p_u(z, x) - \frac{1}{3}\sum_{l\in\{u,e,h\}} log p_l(z, x)$ *(e stand for egalitarian, h for homogeneous). The solid line is estimated on the grid of values of the empirical support of the variable, excluding 5% of schools on each side. The dotted lines is estimated on the support excluding 1% of schools on each side. See Figure 11 in the Appendix for the votes for the homogeneous category.*



(a) General track: Headcount



(b) General track: prior grade



(c) General track: social position

Figure 9: Partial dependence plots in the most distinctive variables by quartile of prior grades. *Vote for u ="unequal" is the average over school characteristics x of the function* $log p_u(z, x) - \frac{1}{3}\sum_{l\in\{u,e,h\}} log p_l(z, x)$ *(e stand for egalitarian, h for homogeneous). The solid line is on the 90% of median VA, excluding 5% of schools on each side. The dotted lines exclude only the extreme percentiles. See Figure 11 in the Appendix for the votes for the homogeneous category.*

5 Conclusion

All in all, we propose new indicators that complement the existing measures of high-school performance. We focus on the way high schools reduce or increase the inequalities in academic level of their students. According to our results, a non negligible proportion of French high schools have significant within-school distributional effects, either toward more homogeneity in final achievements (egalitarian schools) or toward spreading achievements of students (unequal schools). We do not find evidence that a more equal behaviour is obtained through a "race-to-the-bottom" phenomena whereby high-achieving students would be sacrificed, as the fact of being classified as egalitarian (respectively unequal) is also positively (respectively negatively) correlated with higher performance at the median. The methodology proposed here can be viewed as a synthesis between two popular models, SGP and mean value-added. It provides more detailed information on the value-added of the high schools, detailing the dispersion within school. We address a potential concern about selection effects that arises with SGP model by controlling for various observable covariates (such as academic level before enrollment but also socio-academic background). The estimation of the high school effects does not require independance of these specific effects and those covariates. We also compare the obtained classification with several observable features of high schools and observe several noticeable correlations. Specifically, these correlations suggest that egalitarian high schools are comparatively small. They also enroll students on average from more privileged background and who are on average more academically performing at the beginning of the high school. These correlations greatly vary with the type of students. In particular, they are the strongest for previously lowest performing students.

These indicators should be used with cautious, though. As pointed out by Raudenbush and Willms [1995], students are not randomly distributed among schools and one cannot thus isolate the impact attributable to pedagogical action provided by the school from those due to the social and schooling composition of its enrollment. Specifically, even when controlling for observable individual characteristics, the causal impact of the high school cannot be distinguished first from some unobservable characteristics of its enrollment (for instance parental involvement in the student scholarship) and second from the general composition of the enrollment that may impact student achievements through peer effects. The estimated high school specific effects measured here thus correspond to the type "B" school effect in the typology proposed by Raudenbush and Willms [1995]: it measures how a pupil actual performance differ from the one that would have been expected if he or she had attended an average school.[15] The type of indicators proposed here are thus still useful for families confronted to a choice between several high schools and in need of indicators more informative than the "raw" characteristics (as provided by the rate of success to the final exam for instance). In the same vein, the observed correlations correspond only to descriptive evidence and do not inform on potential causal relationship between the high-school specificities and its performance regarding inegalities in student performances. Providing empirical evidence of such correlations would require specific and detailed analysis that are much beyond the scope of the paper. The easily observable characteristics can however be used by families as indicators that one or another high school would tend to reduce inequalities among its student performances.

[15] The type "A" being the causal impact of the high school once controlling entirely from enrollment effect and thus relates to the required information for public authorities that would evaluate the high-school performance.

References

Cédric Afsa. Une question de taille. *Éducation & formations*, 85, 2014.

Beatric Boutchenik and Sophie Maillard. Peer effects with peer and student heterogeneity: An assessment for french baccalaureat. mimeo, Insee, 2018.

Ivan A Canay. A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386, 2011.

Antonio Carvajal-Rodríguez, Jacobo de Uña-Alvarez, and Emilio Rolán-Alvarez. A new multitest correction (sgof) that increases its statistical power when increasing the number of tests. *BMC bioinformatics*, 10(1):209, 2009.

Irene Castro-Conde and Jacobo de Uña-Álvarez. Power, fdr and conservativeness of bb-sgof method. *Computational Statistics*, 30(4):1143–1161, 2015.

Jacobo de Uña-Alvarez. On the statistical properties of sgof multitesting method. *Statistical Applications in Genetics and Molecular Biology*, 10(1), 2011.

Jacobo de Uña-Alvarez. The beta-binomial sgof method for multiple dependent tests. *Statistical Applications in Genetics and Molecular Biology*, 11(3):1–32, 2012.

Marc Duclos and Fabrice Murat. Comment évaluer la performance des lycées. *Éducation & formations*, (85):74, 2014.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

Cassandra Guarino, Mark Reckase, Brian Stacy, and Jeffrey Wooldridge. A comparison of student growth percentile and value-added models of teacher performance. *Statistics and Public Policy*, 2(1):1–11, 2015.

Cassandra M Guarino, Mark D Reckase, Brian W Stacy, and Jeffrey M Wooldridge. A comparison of growth percentile and value-added models of teacher performance. working paper# 39. *Education Policy Center, Michigan State University*, 2014.

Caroline M. Hoxby and Gretchen Weingarth. Taking race out of the equation: School reassignment and the structure of peer effects, 2005.

Kengo Kato, Antonio F Galvao, and Gabriel V Montes-Rojas. Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170(1):76–91, 2012.

Cory Koedel, Kata Mihaly, and Jonah E Rockoff. Value-added modeling: A review. *Economics of Education Review*, 47:180–195, 2015.

Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91 (1):74–89, 2004.

Roger Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.

Roger Koenker and Pin Ng. A frisch-newton algorithm for sparse quantile regression. *Acta Mathematicae Applicatae Sinica*, 21(2):225–236, 2005.

Kenneth Leithwood and Doris Jantzi. A review of empirical evidence about school size effects: A policy perspective. *Review of educational research*, 79(1):464–490, 2009.

Garritt L Page, Ernesto San Martín, Javiera Orellana, and Jorge González. Exploring complete school effectiveness via quantile value added. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2016.

Maria Ponomareva. Quantile regression for panel data models with fixed effects and small t: Identification and estimation. *University of Western Ontario*, 04 2011.

Stephen W Raudenbush and JDouglas Willms. The estimation of school effects. *Journal of educational and behavioral statistics*, 20(4):307–335, 1995.

Thierry Rocher. Construction d'un indice de position sociale des élèves. *Éducation & formations*, 90, 2016.

Jesse Rothstein. Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education*, 4(4):537–571, 2009.

Bruce Sacerdote. Chapter 4 - peer effects in education: How might they work, how big are they and how much do we know thus far? volume 3 of *Handbook of the Economics of Education*, pages 249 – 277. Elsevier, 2011a.

Bruce Sacerdote. Peer effects in education: How might they work, how big are they and how much do we know thus far? volume 3, chapter 04, pages 249–277. Elsevier, 1 edition, 2011b.

Jaap Scheerens, Maria Hendriks, and Hans Luyten. *School Size Effects Revisited*, chapter 2.

Nikos Tzavidis and James Brown. Using m-quantile models as an alternative to random effects to model the contextual value-added of schools in london. 2010.

Elias Walsh and Eric Isenberg. How does a value-added model compare to the colorado growth model? Technical report, Mathematica Policy Research, 2013.

Appendix

A Sample restriction

Table 10: Students, high schools and sample restriction

	High schools	Students	Students per school		
			Mean	Sd	Median
Candidates at the baccalauréat 2015					
General tracks	2248	340469	151.5	95.9	140
technological tracks	1895	130887	69.1	49.3	58
All	2521	471356	187.0	125.7	169
Sample (A) restricted to headcounts ≥ 65 (25) in general (technology)					
General tracks	1759	318222	180.9	82.3	166
technological tracks	1549	122286	78.9	46.3	67
All	2113	440508	208.5	118.2	195
Sample (B): (A) with more than 10 students within each quartile of prior grade					
General tracks	1534	290617	189.5	82.0	178
			Students per school per quartile		
			47.4	26.3	42
Sample (C): (A) + with more than 10 students within each quartile of social position index					
	1518	291400	226.3	86.4	218
			Students per school per quartile		
			48	29.1	42

B Multiple Hypothesis tests

Table 11: Presence of heterogeneous effects: tests results, accounting for multiple hypothesis tests (MHT)

	Proportion of high schools (%)		
	Unequal	Equal	Heterogenous
General track			
No correction	13.6	15.2	28.8
MHT correction: SGoF	8.4	8.9	17.3
MHT correction: BBSGoF	8.2	8.5	16.7
technological track			
No correction	12.5	14.5	27
MHT correction: SGoF	6.5	8.6	15.1
MHT correction: BBSGoF	6	7.6	13.6

C Per quantile high school fixed effects

Table 12: Quantile fixed effect per quartile of prior grade, in terms of standard devition of the baccalaureat grade

	Q1			Q2			Q3			Q4		
Quantile	20%	50%	80%	20%	50%	80%	20%	50%	80%	20%	50%	80%
Per quartile of prior grade												
Min.	−0.901	−0.745	−0.790	−0.827	−0.735	−0.786	−0.728	−0.787	−0.758	−1.310	−1.014	−0.768
1st Qu.	−0.177	−0.153	−0.154	−0.158	−0.154	−0.160	−0.172	−0.165	−0.173	−0.173	−0.147	−0.152
Median	−0.020	−0.017	−0.021	−0.010	−0.015	−0.016	−0.008	−0.013	−0.003	−0.006	0	0.002
3rd Qu.	0.154	0.130	0.135	0.141	0.138	0.150	0.165	0.166	0.168	0.175	0.161	0.163
Max.	0.994	0.912	0.946	0.966	0.962	1.001	0.797	0.781	0.931	1.088	0.844	0.840
Mean	0	0	0	0	0	0	0	0	0	0	0	0
Per quartile of social position index												
Min.	−1.045	−0.711	−0.769	−0.744	−0.663	−0.916	−1.219	−1.020	−0.918	−1.119	−0.902	−1.046
1st Qu.	−0.170	−0.151	−0.152	−0.159	−0.149	−0.148	−0.166	−0.150	−0.157	−0.153	−0.144	−0.138
Median	−0.018	−0.011	−0.016	−0.017	−0.015	−0.013	0.009	−0.002	−0.007	0.007	−0.002	0.009
3rd Qu.	0.148	0.136	0.152	0.139	0.138	0.138	0.170	0.154	0.162	0.160	0.149	0.151
Max.	1.150	0.907	0.901	0.849	0.925	0.886	0.871	0.658	0.736	0.965	0.770	0.826
Mean	0	0	0	0	0	0	0	0	0	0	0	0

Source: FAERE and APAE database, author calculations.

Table 13: Proportion of egalitarian and unequal high-schools by quartile (general track)

Heterogeneous on at least a quartile	0.326
Among which,	
no overall heterogeneous effect	0.470
consistent with overall effect	0.530
	1

Source: FAERE and APAE database, baccalaureat 2015, authors' calculation. Note: Restriction to high-school with headcount ≥ 10 in each quartiles. A high-school j is defined as egalitarian (resp. unequal) when $\alpha^j_{0.20} > \alpha^j_{0.80}$ (resp. $\alpha^j_{0.20} < \alpha^j_{0.80}$)

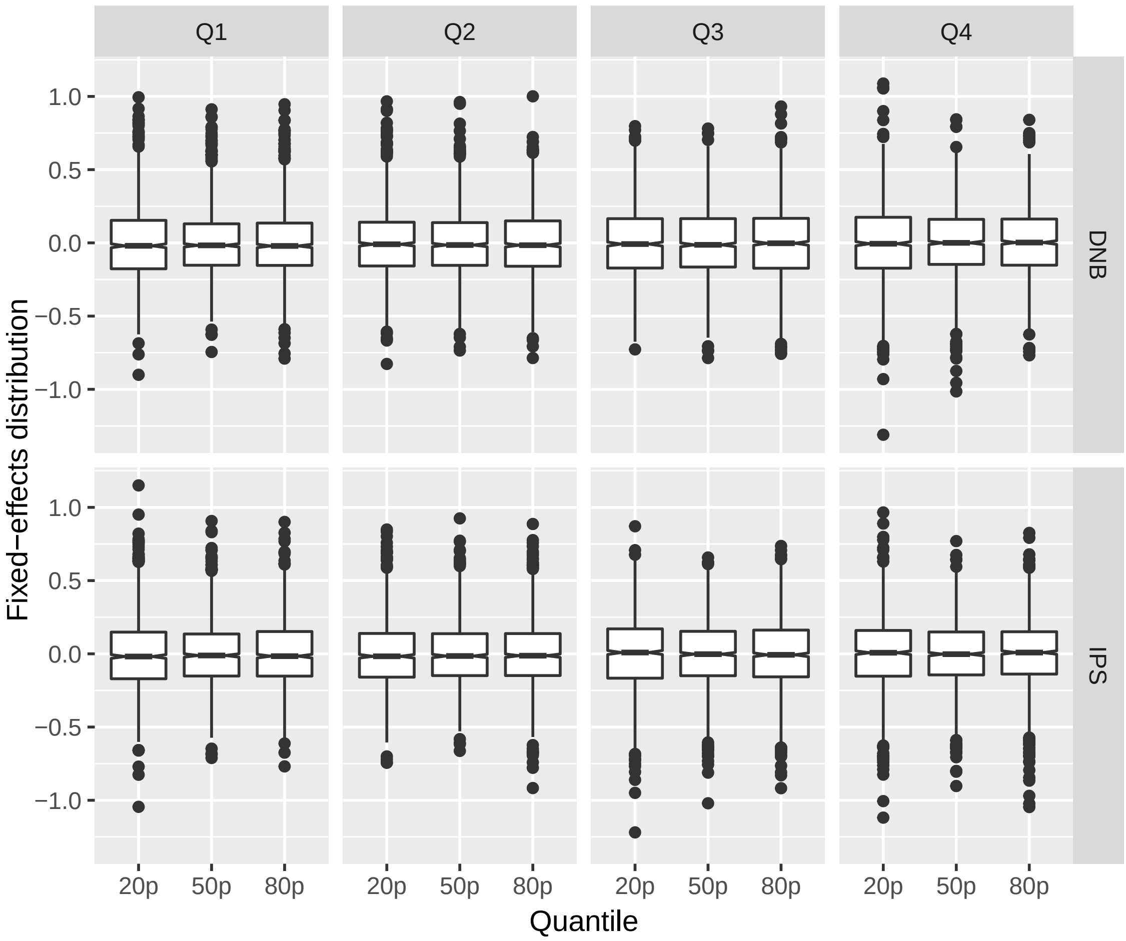


Figure 10: Distribution of quantile fixed effect per quartile of prior grade, scaled in terms of standard deviation of the baccalaureat grade

D Random forest: other dependence plots

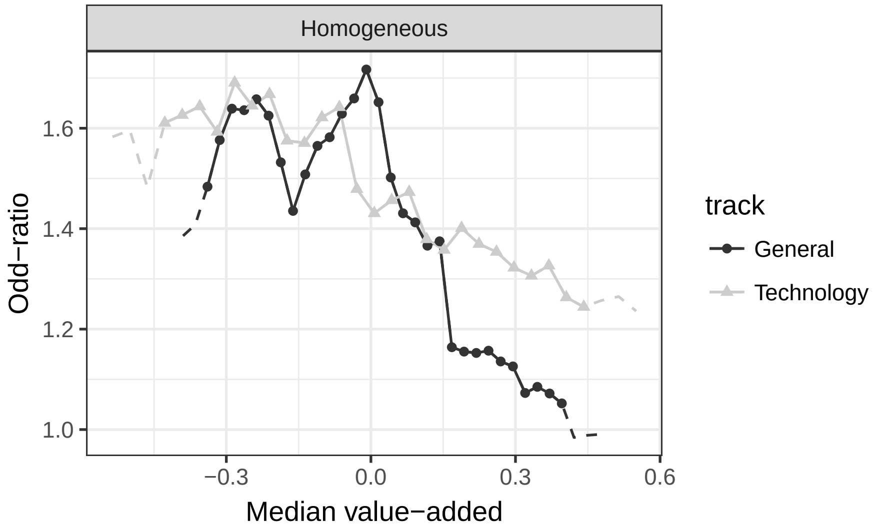
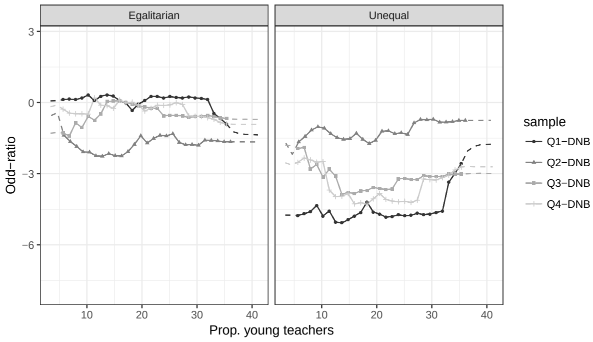
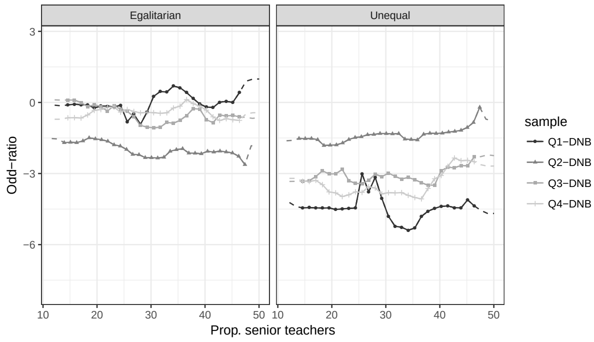


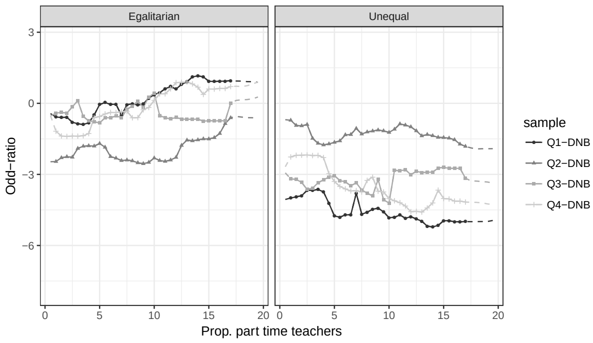
Figure 11: Partial prediction of inequality when performance is known, for the homogeneous category



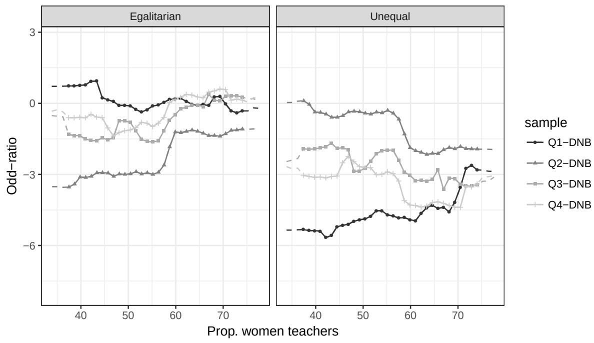
(a)



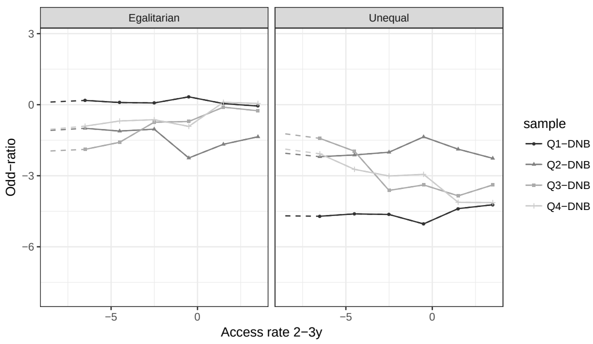
(b)



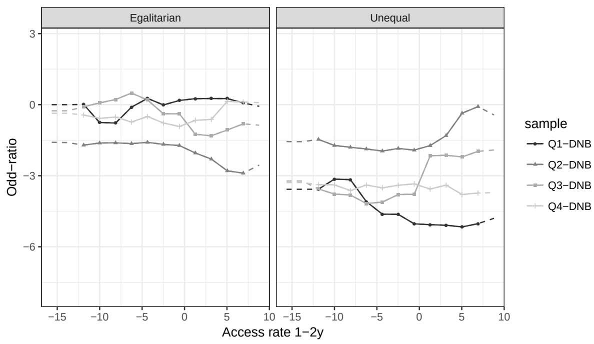
(c)



(d)



(e)



(f)

Figure 12: Prediction of inequalities as function of proportion of young (less than 35 years old), old teachers (over 50 years old), proportion of part-time teacher and proportion of women teachers. Bottom panel shows the same graphes for access rates. *Vote for $u =$"unequal" is the average over school characteristics x of the function $logp_u(z,x) - \frac{1}{3}\sum_{l\in\{u,e,h\}} logp_l(z,x)$, z is the variable under consideration). The solid line is on the 90% support of the variable, excluding 5% of schools on each side. The dotted lines excludes only the extreme percentiles of schools.*

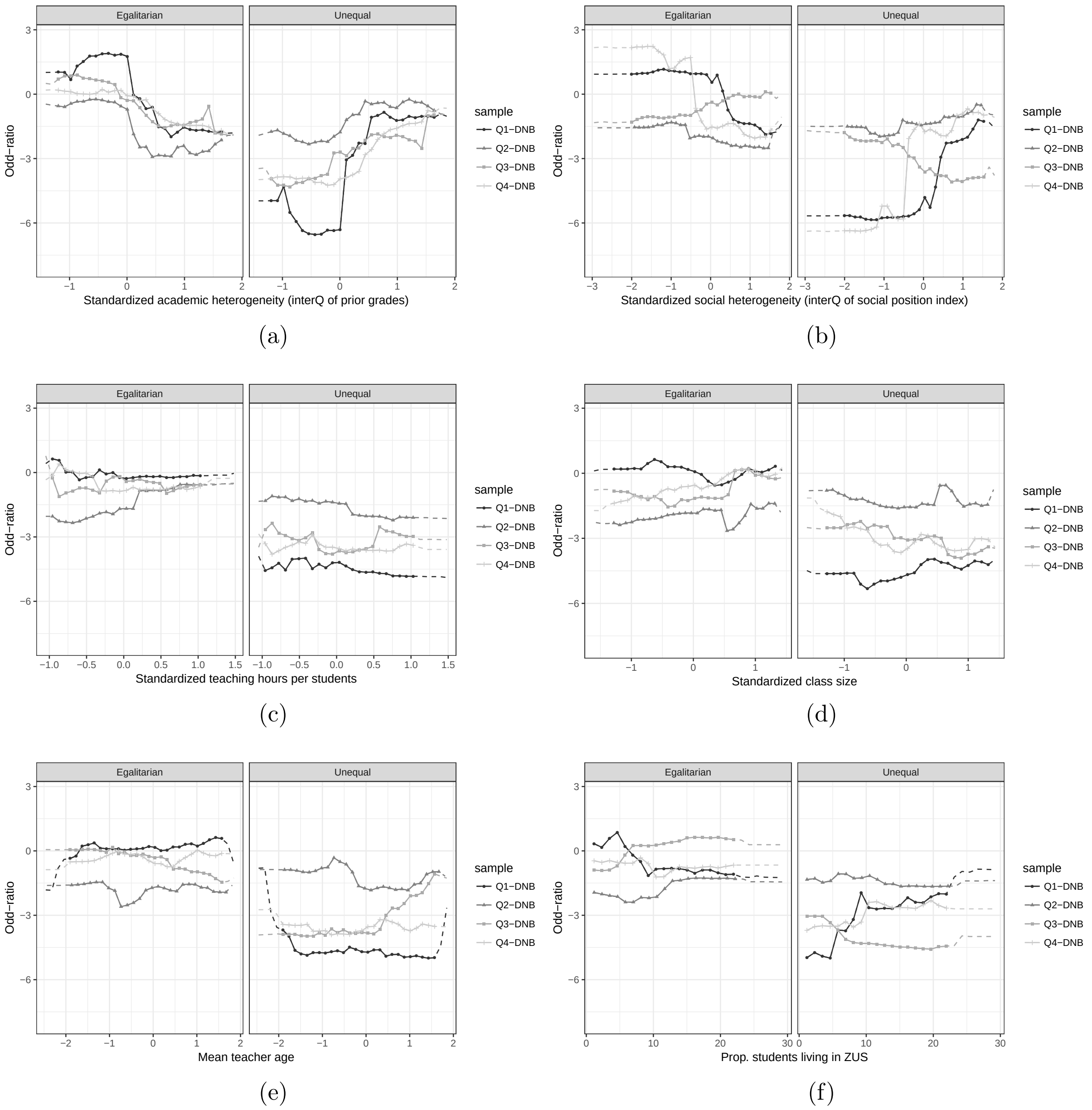


Figure 13: Prediction of inequalities as function of other variables. *Vote for* $u =$ *"unequal" is the average over school characteristics x of the function* $logp_u(z, x) - \frac{1}{3}\sum_{l\in\{u,e,h\}} logp_l(z, x)$*, z is the variable under consideration). The solid line is on the 90% support of the variable, excluding 5% of schools on each side. The dotted lines excludes only the extreme percentiles of schools.*