DISCUSSION PAPER SERIES

No. 5754

ESTIMATION OF CLASS-SIZE EFFECTS, USING 'MAIMONIDES' RULE': THE CASE OF FRENCH JUNIOR HIGH SCHOOLS

Robert J. Gary-Bobo and
Mohamed-Badrane Mahjoub

LABOUR ECONOMICS and PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP5754.asp

ISSN 0265-8003

ESTIMATION OF CLASS-SIZE EFFECTS, USING 'MAIMONIDES' RULE': THE CASE OF FRENCH JUNIOR HIGH SCHOOLS

Robert J. Gary-Bobo, University Paris 1 and CEPR
Mohamed-Badrane Mahjoub, Université de Paris I

Discussion Paper No. 5754
July 2006

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5754

July 2006

ABSTRACT

Estimation of Class-Size Effects, Using 'Maimonides' Rule': The Case of French Junior High Schools*

Using a rich sample of students from French junior high schools with a panel structure, we obtain small but significant and negative effects of class size on probabilities of educational success, in grades 6 and 7. An 8 to 10 student reduction of class size puts the child of a non-educated mother on an equal footing with the child of a college-educated mother. These effects vanish in grades 8 and 9. We use Angrist and Lavy's (1999) theoretical class size (i.e., "Maimonides' rule") as an instrument for observed class size. This is possible, due to availability of total high school and total grade enrollment in each year, in our exceptional data set. We control for father occupation, mother education and other variables. Using a Probit framework to model transitions from one grade to another (and thus grade repetitions), we simultaneously estimate the student's probabilities of success over 4 years in junior high school. This is done while allowing a general covariance structure of the error terms that affect latent student-performance variables and class-size auxiliary equations.

JEL Classification: C33, C35 and I20
Keywords: class size, econometrics, education, instrumental variables and junior high school

Robert J. Gary-Bobo
Université Paris 1 Panthéon-Sorbonne
CES
106 Boulevard de l'Hôpital
75647 Paris Cedex 13
FRANCE
Tel: (33 1) 4407 8257
Fax: (33 1) 4407 8247
Email: garybobo@univ-paris1.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=125284

Mohamed Badrane Mahjoub
Université Paris 1 Panthéon-Sorbonne
CES
106 Boulevard de l'Hôpital
75647 Paris Cedex 13
FRANCE
Tel: (33 1) 44 07 82 52
Fax: (33 1) 44 07 82 47
Email: gbadrane@univ-paris1.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=164289

* We thank Thomas Brodaty, Denis Fougère, Francis Kramarz, Ana Prieto, Jean-Marc Robin, Fabienne Rosenwald and Daniel Vitry, for their help and helpful conversations. We thank the French Ministry of Education's *Direction de l'Evaluation et de la Prospective*, for having made the data used in the present research available to us.

Submitted 12 June 2006

1. Introduction

It seems that the debate about school resources, school quality and class-size effects[1] has not lost of its intensity. In the recent years, the class-size debate has been fueled by the availability of new data, in the form of controlled field experiments, such as the Student-Teacher Achievement Ratio (STAR) study, conducted in Tennessee[2], and the rise of Instrumental Variables estimation techniques, based on richer data sets[3]. The debate is complex, but, if we bear the risk of some over-simplification, seems to oppose two broad categories of researchers. On one side are those who think that the available evidence is partial, limited to some experiments, at best weak, and would in any case not warrant a general, uniform policy of further class-size reductions, if the magnitude of its social costs is taken into consideration. These authors insist on teachers' incentives and accountability. On this line of argument, see the important surveys of Hanushek (2002), (2003). On the other side, we find the scholars who think that there is indeed enough evidence on the fact that reduced class-size improves educational achievement and life chances, at least for disadvantaged children, and that the social benefits of class-size reduction programs should be more broadly assessed, in particular by paying attention to additional crime-reduction effects. On this side, see the vigorous synthesis of Krueger[4] (2002), (2003). A more "agnostic" point of view is expressed in the work of Carneiro and Heckman (2003); these authors emphasize the heterogeneity of individuals, the fact that the ability and educational achievement of children are the results of long-term factors, among which family background, during the early stages of life, plays a major role. Carneiro and Heckman underline the potential efficiency gains of properly targeting public educational policies. Among the "agnostics", see also Vignoles *et al.* (2000).

The econometric identification of class-size effects is difficult, because class-size is a highly endogenous variable. Naive, ordinary least squares estimations of the impact of class-size on test scores, or on other measures of educational achievement, usually yield positive coefficients: it seems that increasing class size helps students. This causal interpretation is of course not justified. Many student characteristics, including various aspects of student "ability", are imperfectly observed; but the teachers and headmasters are better informed than the econometrician. Based on student characteristics that the econometricians do not usually observe, headmasters seem to allocate weaker students to smaller classes, thus generating a positive correlation of class-size and student performance. This phenomenon is clearly mitigated by adding controls in the regression of test scores (or achievement) on class size. But the likely effect of better controls is to drive the coefficient of class size towards zero (or to increase the standard deviation of the estimate, so that it is no longer significantly different from zero). Instrumental variable estimation is then a way of dealing with endogeneity, and of uncovering the true value — or at least, the true sign — of the class-size coefficients. The problem is to find a statistical instrument, a source of

[1] which goes back at least to the Coleman report in the U.S, see Coleman et al.(1966).

[2] See e.g. Krueger (1999)

[3] Interesting work on education in developing countries has also brought about a wealth of new results; see e.g., Duflo (2001), Banerjee et al. (2005).

[4] See also Card and Krueger (1996).

variation of class size that is not correlated with the student-achievement equation's error term. This requires both an "exogenous" source of variation of class size, and appropriate controls.

Hoxby (2000) uses local population variation to identify class-size effects, and finds that even modest effects can be ruled out. She uses two different instrumentation strategies. In the first strategy, long enrollment series are regressed on a fourth-degree polynomial function of time. This regression's residuals isolate a "pure" random component in population variation, that is then used as an instrument for class size. Hoxby's second identification strategy — in the same spirit as Angrist and Lavy's (1999) method—, is to use the discontinuous changes in class size that are triggered by population variation, when local (i.e., district) class opening thresholds are reached.

Angrist and Lavy (1999) find negative and significant class-size effects (i.e., increasing class size reduces students' test scores). Their striking results are obtained with Israeli data, and with the help of a "theoretical class-size" variable, that, for short, we shall call Angrist-Lavy's instrument in the following. This instrument is computed as total enrollment in a given grade and school, divided by the theoretical number of classes in this given grade (and school). The theoretical number of classes is itself that which results from the application of a given threshold for opening new classes when enrollment grows. Formally, the theoretical number of classes is $\kappa = int((N-1)/\tau) + 1$, where N is total grade-enrollment, τ is the class-opening threshold, and $int(x)$ is the integer part of x. This is Maimonides' Talmudic rule, which commands that a new class be opened if there are more than 40 students in a class. In Hoxby's (2000) work, the threshold is around 25 students per class. In the present article, which studies the case of French junior high schools, a value of 30 students per class seems appropriate.

In Angrist and Lavy's (1999) study, the crucial assumption is that unobservable factors, correlated with total enrollment, can be controlled for in the test score regression, if only by using total school enrollment itself as a regressor, so that this latter regression's error term can reasonably be assumed without correlation with the instrument. These problems are likely to be important if school-choice phenomena cause total enrollment in a given school to reflect unobservable school characteristics such as "peer groups" or good teachers. Total enrollment can also signal that the school is located in a big city, in which parental human capital and incomes are higher than in rural areas — but some of these effects can in principle be controlled for. Even if one cannot be sure that the desired exogeneity property is always fully satisfied, it is reasonable to believe that it can be approximately true, for some data sets at least. The results obtained by IV estimation procedures also bear the risk of reflecting the effect of class size reductions on certain subgroups of the population under study, instead of the average effect, but in any case, they will provide more interesting insights than naive OLS estimates.

There are of course other approaches to IV estimation of class-size effects. Among these, see *e.g.*, Case and Deaton (1999), Boozer and Rouse (2001), Dearden *et al.* (2000), Dobbelsteen *et al.* (2002), Dustmann *et al.* (2003), Wössmann and West (2006).

To the best of our knowledge, Maimonide's rule has been used to construct an instrument for class-size by a limited number of authors only: Bonesrøning (2003) uses the method to study Norwegian data and finds significant effects in lower secondary schools;

Wössmann (2005) proposes international comparisons within Europe, using the TIMMS database and finds zero or negligible effects. The application of Angrist-Lavy's method to official data from the French Ministry of Education yields interesting results, with significant and negative effects of increased class-sizes. In France, this has been done by Piketty (2004), Piketty and Valdenaire (2006), Bressoux *et al.* (2005), and us, to the best of our knowledge. Piketty (2004) has studied the impact of class-size on test scores in French primary schools (grade 3), using Angrist-Lavy's instrument in a standard linear regression setting, and a rich data set. Piketty and Valdenaire (2006) apply the same methods to test scores recorded at the end of junior high school, using the 1995 Panel of the Ministry of Education. Bressoux *et al.* (2005) study the same data as Piketty (2004), but exploit a different quasi-experimental situation based on the existence of trained and untrained "novice" teachers; they however also employ Angrist and Lavy's method as a point of comparison. We use the Ministry of Education junior high-school Panel started in 1989, and study grade promotions instead of test scores.

Piketty's (2004) striking results show that decreasing class size by ten students in grade 3 would yield a 7 point increase in test scores, when Mathematics tests, with grades ranging from 0 to 100, and disadvantaged students are considered. These, and other similar results lead him to conclude that a reduction of class size in elementary school, and in disadvantaged areas, would substantially reduce the test-score gap with the average student. Bressoux *et al.* (2005) find negative class-size effects on mathematics test scores: a ten student reduction in class-size would increase the average test score by 4.4 points in maths. With an average score of 100, this figure amounts to 30% of the standard deviation of test scores, which is equal to 15. Using junior high-school test scores, at the end of grade 9, Piketty and Valdenaire (2006) find that a 10 student reduction of class size in grade 9 would improve the normalized grade point average by 2.16, which represents 20% of the standard deviation of these GPAs.

In the present paper, our sample contains more than 12,000 observations of individuals enrolled in French public junior high schools, scattered on the whole French territory. The panel, started in 1989, has been made available to us by the Ministry of Education. An important advantage of our data, when compared with those of Angrist and Lavy (1999), is that it is individual micro-data, instead of class-level data (i.e., class averages); another advantage of our data is the higher number of available control variables. We do not use test scores or examination results, but qualitative tracking, grade-promotion (or grade-repetition) decisions instead. These decisions are made by teachers' committees (i.e., the *conseils de classe*), at the end of each school year. To be more precise, we use class size, and a whole list of control variables, to explain the probabilities of being promoted to the next grade, of being held back (and repeat a grade), or of leaving general education for a vocational program (i.e., tracking, or "steering"). The best students are promoted, the weakly performing students repeat the grade, and those with results sufficiently below the average are "steered" towards vocational programs, and leave general education. We use an Ordered Probit structure to estimate these probabilities as functions of observed class size, individual and high school characteristics. Angrist-Lavy's theoretical class size is used as an instrument for observed class size, and we control for total school enrollment. Rigorous estimation of such a model requires the use of maximum likelihood techniques.

More precisely, we jointly estimate the probabilities of promotion, retention, and tracking, in grades 6 to 9 (i.e. 4 transitions), and 4 auxiliary (or first-stage) class-size equations, for a cohort of students who were enrolled in grade 6 in September 1989. This is done under the assumption that student-performance error terms, which depend on school-year, are *conditionally independent*, and more precisely, that they are independent, conditional on class-size. This does not mean that grade transitions are statistically independent of course, but allows us to write the probability of a given student grade-transition record as a product of probabilities — i.e., as a Markov chain. The 8-dimensional (unconditional) covariance structure of error terms is otherwise generic, and we provide estimates of the full covariance matrix, with, we think, interesting interpretations. Estimating the grade transitions separately would in fact impose more constraints than our joint estimation procedure under conditional independence, because we do not assume that class-size error terms are serially independent, and, of course, we do not assume that class-size error terms are independent of student-performance error-terms.

We find that class-size coefficients are significant and negative — that class-size reductions lead to higher student promotion rates — in grade 6 and grade 7 only. Our (conservative) summary of this finding is that a 8 to 10 student reduction of class size puts the child of an uneducated mother on an equal footing with the child of a college-educated mother, in these beginning of high-school grades. This means that class-size effects are moderate, given that the average class size is around 25 in the sample — but are not negligible. In contrast, class-size effects are not significant in grades 8 and 9, they seem to fade away in the higher grades. A possible explanation is that the intensity of class-size effects decreases when grade increases: the size of classes is less important for more mature, or more advanced students. We also find an interesting pattern of cross-correlation between student-performance and class-size in different years, as a by-product of our maximum likelihood estimation. These correlations reflect the impact of unobserved variables; they are always positive when significant (and almost always significant), which seems to confirm the intuition that unobserved factors do increase student performance and class size at the same time. In addition, there is a strong correlation of class-size residuals: the unobserved factors which determine a student's class-size are positively correlated through time, reflecting the presence of permanent individual effects. A possible interpretation is that a student which is categorized as "weak" by his teachers, conditional on observed family background factors, but for reasons that we do not observe, tends to be assigned to smaller classes during his (her) entire career in junior high school.

In the following, Section 2 is devoted to a description of the data, section 3 presents the model and estimation method, and Section 4 presents the estimation results.

2. The Data

Estimations have been realized with the help of a matched data set, merging two sources: a file called *Panel 89*, which is a large-scale survey conducted among high schools by the French Ministry of Education —and more precisely, the *Direction de l'Evaluation et de la Prospective* (DEP) within the education department—, and another administrative source from the same Ministry, the *établissements* files, providing yearly enrollment data and other information on the high schools themselves. Panel 89 is a sample of high school students, observed during several years. In the following, we use the observations during the first 5 years of high school, that include the French junior high-school years[5]. One fifth of all (junior) high schools are sampled; then, 1/30th of the junior high school students enrolled in grade 6 (i.e., *classe de sixième*) are sampled, and to be more precise, the students born on a particular day of each month are chosen. Each sampled student has an identification number. The headmasters in sampled high schools had to fill forms about each sampled student, providing a number of family background characteristics, and recording the grade or program attended by the student every year. In addition, a questionnaire has been sent to sampled students' parents. The family questionnaire has been used to fill missing observations in the headmasters' records. When the headmasters' records and the family's answers concerning family background information did not agree, the family answers have been used. For each student, we know a number of time-invariant characteristics: parental occupation and education, number of siblings, birth order, gender, and age at grade 6 entry. In addition, we observe answers to a number of questions asked to parents such as "is one of your children already enrolled in vocational education?", or questions about parents' preferences concerning the student's school-leaving age. For a sub-sample of the students only, we observe the results of tests in Math and French, taken at the end of primary school (thus before entering junior high school), with a grade ranging between 0 and 100. The tests are not compulsory, and therefore, students with a primary school test record are self-selected, with a potential bias of course. Finally — and crucially —, for each year of the observation period, we know the student's class-size, the grade followed (and thus observe grade repetitions), and if the student was "steered" towards vocational high schools or vocational programs at the end of the school year. In French high schools, "steering" is quite often an euphemistic expression for the headmaster's or principal's decision to expel the student from general education programs, essentially because the student's performances are not judged good enough — these decisions can as well be called "tracking". We now turn to the matched data set.

To construct the instruments for class-size, we have linked *Panel 89* with administrative files giving enrollment measures in each high school and each year (the so-called *établissements* files). From these files, we get total grade-enrollment for each grade and each year in each school, and total junior high-school enrollment in each high school. Enrollment observations are beginning-of-the-year (i.e., September) figures for each school year, starting in 1989. In addition to these, we observe some statistics concerning the teaching staff. We know the proportion of teachers above 50, the proportion of teachers below 30, the proportion of men, and the proportion of *professeurs agrégés*. There are several categories of secondary education teachers in France, with a widely recognized hierarchy. At the top of this hierarchy are the *professeurs agrégés*, the most prestigious high-school teachers, who successfully passed the *agrégation* contest, a competitive examination. These teachers have better pay, lighter teaching loads, and better careers than the others. The second, and most numerous category, is that of *professeurs certifiés*, considered less prestigious. The percentage of *agrégés* in a school might be a quality indicator. The other indicators are precious too, because of the organization of teachers' careers in

[5] The secondary *collège* years in France, correspond to junior high school years. They have nothing to do with the British or American higher-education college years, of course.

France. Pay is the same everywhere in the public sector, and essentially depends on seniority (within each category). But high schools are far from being equal. There are of course problem schools, in problem urban zones, etc. It follows that non-monetary factors are essential elements of reward. Young French high-school teachers thus form a giant waiting line, expecting to be appointed to "better" schools, in better places. It takes a lot of time to be appointed to a new school — this is the result of a central bureaucratic process, in which the teachers' unions play a major role. A better school is likely to be one with better working conditions, that is, essentially, better student "peer groups", less students from disadvantaged and immigrant families, etc. From the point of view of the average teacher, a better school is likely to be located in a rich city center[6], somewhere to the south of the Loire river (because of the sunnier weather) and probably in a small to mid-sized town (because of lower house prices). These well-known facts explain why the percentage of old teachers might convey information about teaching "quality", not only because older teachers have more experience, but because they are likely to cluster in — and thus signal — the "good schools".

Another important indication is the ZEP classification of schools. ZEP is the acronym of *Zone d'éducation prioritaire*, meaning a geographical area benefiting from some kind of redistribution, in the form of increased educational resources from the government, aimed at mitigating a number of handicaps[7].

Finally, each public-sector high school has an identifier (*code d'établissement*), that can be used to link the student observations in Panel 1989 to corresponding high school enrollment and teaching-staff environment statistics in the administrative files. The linked data will be used to perform IV estimations of class-size effects. We link individual student observations with corresponding public sector high-school enrollment data only (because private sector enrollment figures are not available), which is a reason why we lose a number of observations from the initial panel. There are also some observations lost because of missing data. To sum up, we obtained a final linked sample with 12,854 observations.

3. The Model. Estimation Method

Our first and most important assumption is to model the probabilities of promotion, grade retention and "steering" (or tracking) at the end of each school-year as a discrete Markov process, conditional on class-size history and student characteristics. All students start in grade $g = 6$ in year $t = 1$ (i.e., 1989). We observe student records over 5 years, if they are not "steered" towards vocational schools or programs. We thus observe at most 4 transitions, and any "steering" decision of the teachers leads to vocational high school, which is modeled as an absorbing state. Student i's record is by definition an array $s_i = (s_{it})_{t \geq 2} = (s_{i2}, s_{i3}, s_{i4}, s_{i5})$. Since all observed students start in grade 6, we set $s_{i1} = 6$ for all i. Possible values of the state s_{it} are in the set $G = \{6, 7, 8, 9, 10, v\}$,

[6] In France, city centers are likely to be inhabited by the richest (as suburbs in the US), and suburban schools are more likely to be "problem schools", because the working class (and immigrants) are more and more "relegated" to suburban towns.

[7] On the ZEP classification of schools in France, see Benabou, Kramarz and Prost (2004).

where v stands for vocational school. Promotion to grade 10 in general-education high schools being the last event observed for an individual who was promoted each year in the observation period, we conventionally treat $g = 10$ as an absorbing state too. Transition probabilities (from grade or state g to grade or state h) are denoted p_{gh}. For a given individual i, that is, conditional on the individual's environment and date t, we denote

$$p_{gh} = P(s_{i,t+1} \mid s_{it}), \tag{1}$$

if $s_{it} = g$ and $s_{i,t+1} = h$, where the probabilities P are specified as functions of i's observable characteristics.

The matrix of transition probabilities, denoted $\mathbf{P}$, has a special structure, because many transitions are not allowed. Formally,

$$\mathbf{P} = \begin{pmatrix} p_{66} & p_{67} & 0 & 0 & 0 & p_{6v} \\ 0 & p_{77} & p_{78} & 0 & 0 & p_{7v} \\ 0 & 0 & p_{88} & p_{89} & 0 & p_{8v} \\ 0 & 0 & 0 & p_{99} & p_{9,10} & p_{9v} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

To understand the meaning of this matrix, note that a student in grade 8 can be either promoted (with probability p_{89}), or held back and repeat grade 8 (with probability p_{88}), or "steered" (with probability p_{8v}). The initial distribution has all the students in grade 6.

We assume that state transitions are conditionally independent, that is, independent conditional on individual characteristics and class-size history. The conditional probability of observing record s_i, can thus be written as follows,

$$\Pr(s_i) = \prod_{t=2}^{t=5} P(s_{it} \mid s_{i,t-1}), \tag{3}$$

where $s_{i1} = 6$ for all i.

Now, our next important assumption is that P functions are Ordered Probit probabilities. To fully specify these functions, define the latent variable y_{it}, which represents student i's "performance" in year t, as follows,

$$y_{it} = X_i\beta_t + \sum_{g=6}^{g=9} \mathbf{1}_{itg}\alpha_{tg}n_{it} + \nu_{it}, \tag{4}$$

where X_i is vector of observed characteristics, β_t a vector of parameters, $\mathbf{1}_{itg}$ is an indicator, the value of which is 1 if i is in grade g in year t and 0 otherwise, n_{it} is the size of i's class in year t, and ν_{it} is a random, normally distributed error term. So, we assume that the β parameters can vary with t, (but not with g here), and that the crucial class-size parameters α_{tg} can vary with grade g and year t. Again, the matrix $\mathbf{A} = (\alpha_{tg})$ has a special structure, because in a given year, students cannot be observed in every grade. Formally,

$$\mathbf{A} = \begin{pmatrix} \alpha_{16} & 0 & 0 & 0 \\ \alpha_{26} & \alpha_{27} & 0 & 0 \\ 0 & \alpha_{37} & \alpha_{38} & 0 \\ 0 & \alpha_{47} & \alpha_{48} & \alpha_{49} \end{pmatrix}. \tag{5}$$

This formulation is equivalent to saying that $\mathbf{1}_{i1g} = 1$ if and only if $g = 6$, that $\mathbf{1}_{i2g} = 1$ if and only if $g = 6$ or $g = 7$, and so on.

We now specify the Ordered Probit structure. The conditional probability of being promoted is defined as follows,

$$P(s_{i,t+1} = g + 1 \mid s_{it} = g) = \Pr(y_{it} \geq \delta_t); \tag{6}$$

the conditional probability of grade retention is defined as,

$$P(s_{i,t+1} = g \mid s_{it} = g) = \Pr(\delta_t > y_{it} \geq \gamma_t). \tag{7}$$

and the conditional probability of "steering" is defined as,

$$P(s_{i,t+1} = v \mid s_{it}) = \Pr(y_{it} < \gamma_t); \tag{8}$$

where γ_t and δ_t are ordered thresholds or "cuts", with $\gamma_t < \delta_t$. Note that the cuts are allowed to vary with the year t (but not with the grade g here). The vector of error terms $\nu_i = (\nu_{i1}, \nu_{i2}, \nu_{i3}, \nu_{i4})$ is multivariate normal with mean 0 and (unconditional) covariance matrix $\Omega_{\nu\nu}$.

Until now, probabilities have been specified as conditional on individual i's class-size history, i.e., the vector $(n_{it})_{t=1,\ldots,4}$. We specify "first-stage" class-size regressions as follows,

$$n_{it} = a_t n^*_{it} + Z_i b_t + \epsilon_{it}, \tag{9}$$

where n^*_{it} is the theoretical class-size of individual i during year t, i.e. Angrist-Lavy's instrument, Z_i is a vector of other exogenous controls, ϵ_{it} is a normally distributed error term, and (a_t, b_t) are parameters to be estimated. The parameters a and b are constrained not to depend on grade g for simplicity, but a variant in which a at least varies with grade g could be estimated. Equation (9) forms a 4 dimensional system of related regressions, and the random term vector $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}, \epsilon_{i4})$ has a generic multivariate normal distribution with mean 0 and (unconditional) covariance matrix $\Omega_{\epsilon\epsilon}$.

Angrist-Lavy's instrument, or theoretical class-size n^*_{it} is the class size that would be experienced by student i in year t and grade $g(i,t) \in G$, on average, if the headmaster's rule was to open a new class each time total enrollment in grade $g(i,t)$ is strictly greater than $30q$, where q is an integer, and to minimize the class-size differences. We take $n = 30$ as being the contemporary French norm. Let N_{it} be the beginning-of-the-year total enrollment, in year t, and in student i's grade $g = g(i,t)$. Given this definition, the theoretical number of classes in grade $g(i,t)$, denoted κ^*_{it}, is by definition,

$$\kappa^*_{it} = int\left[\frac{N_{it} - 1}{30}\right] + 1, \tag{10}$$

where $int[x]$ is the largest integer q such that $q \leq x$. The theoretical number of students per class is simply

$$n^*_{it} = \frac{N_{it}}{\kappa^*_{it}}. \tag{11}$$

Thus, if there are 31 students in a given grade and in a given high school, there are, in principle, 2 classes with an average number of students equal to 15.5, if there are 61 students, there are 3 classes with 20.33 students, and so on; n^* is a discontinuous function of total grade enrollment N; it is an increasing function of N between its points of discontinuity. The philosophy behind the use of n^* as an instrument for class-size is developed in Angrist and Lavy (1999)– see also the reservations in Hoxby (2000). The main argument in favor of this procedure is that total enrollment fluctuations are essentially driven by demographic shocks that cause discontinuous changes in class-size, when combined with the class opening rules. Angrist-Lavy's instrument has better chances of being independent of the performance-equation error-term, if good controls are introduced in the performance equation. It is then essential that total school-enrollment be used as a control, because a number of unobserved factors can be correlated with school-size, itself correlated with total grade-enrollment N, and affect student performance.

To complete the specification of our model, we need to define the covariance structure. Given that our data have a panel structure, it is of course important to allow for correlation of class-size perturbations ϵ_{it} across time, and for correlation with the non-observable performance terms ν_{it}. Let $\nu_i = (\nu_{i1}, \nu_{i2}, \nu_{i3}, \nu_{i4})$, and let Ω be the 8-dimensional covariance matrix of (ν, ϵ). We partition the matrix as follows,

$$\Omega = \begin{pmatrix} \Omega_{\nu\nu} & \Omega_{\nu\epsilon} \\ \Omega_{\epsilon\nu} & \Omega_{\epsilon\epsilon} \end{pmatrix}, \tag{12}$$

where $\Omega_{\epsilon\nu}$ is the transposition of $\Omega_{\nu\epsilon}$. The conditional covariance matrix $\Omega_{\nu|\epsilon} = Cov(\nu \mid \epsilon)$ is given by the following formula, using a well-known result on Gaussian vectors,

$$\Omega_{\nu|\epsilon} = \Omega_{\nu\nu} - \Omega_{\nu\epsilon}\Omega^{-1}_{\epsilon\epsilon}\Omega'_{\nu\epsilon}. \tag{13}$$

Our conditional independence assumption translates into the property that the conditional covariance matrix $\Omega_{\nu|\epsilon}$ is diagonal. For identification purposes, given that we model 4 transitions by means of a Probit structure, we add 4 additional constraints on the 4 variances of the ν_{it} error terms. An easily tractable and natural choice is to set the conditional variances $Var(\nu_{it} \mid \epsilon) = 1$, which is equivalent to saying that $Cov(\nu \mid \epsilon)$ is the identity matrix, or

$$\Omega_{\nu\nu} - \Omega_{\nu\epsilon}\Omega^{-1}_{\epsilon\epsilon}\Omega'_{\nu\epsilon} = I. \tag{14}$$

These normalization constraints could have been imposed on the diagonal terms of the unconditional covariance matrix $\Omega_{\nu\nu}$ instead. The covariance matrix $\Omega_{\nu|\epsilon}$ would then have been a general diagonal matrix instead of an identity matrix. Given that the y_{it} are latent variables, the variance of ν terms is purely conventional, and this alternative normalization procedure would not change anything essential. With the current normalization, the diagonal terms of $\Omega_{\nu\nu}$ will end up greater than 1, while the alternative rule would have produced diagonal terms of $\Omega_{\nu|\epsilon}$ smaller than one.

We estimate this model by maximum likelihood, under the normality assumption for all perturbations. The model directly identifies all the parameters, a, b, α, β, γ, δ and the covariance parameters $\Omega_{\epsilon\epsilon}$, $\Omega_{\nu\epsilon}$. The matrix $\Omega_{\nu\nu}$ is then computed, using (14) above.

See the appendix, for a derivation of the likelihood function.

4. Results. Discussion

We start with a preliminary analysis of the data. Our first mode of analysis is "graphical": Figures 1-4 plot the probability of being promoted to the next grade and class-size, against the average total grade enrollment in deciles of the total grade-enrollment distribution. Figures 1 and 2, corresponding to grades 6 and 7, clearly show some negative correlation of class-size and grade-promotion probability. The negative correlation is much less visible on Figures 3 and 4, that correspond to grades 8 and 9.

Figures 5-8 plot the theoretical class size $n^* = N[int((N-1)/30)+1]^{-1}$ and the observed class-size n as a function of grade enrollment N, in grades 6, 7, 8, and 9, respectively, but with some preliminary averaging. To be more precise, we first group all student-year observations (i,t) for which total grade enrollment N_{it} is the same. These observations also have the same theoretical class-size values n^*_{it}, but different actual class-size values n_{it}. We take the average of the actual class-size values in each group. This yields the empirical curves on Fig. 1a, 2a, 3a, 4a. The figures show that Angrist and Lavy's instrument works reasonably well, and particularly when total grade enrollment is below 120. It seems that some headmasters do not really apply the 30 students-per-class rule. If we drop all observations for which $\mid n_{it} - n^*_{it} \mid > 6$, we loose 15% of all observations, and the averaging procedure yields Figures 1b, 2b, 3b and 4b, in which the goodness of fit is remarkable.

Table 1 displays elementary descriptive statistics for a number of variables used in the regressions. Table 2a describes the numbers of promoted, held back, and steered students in each year (where year 1 is school-year 1989-90, year 2 is school-year 1990-91, etc...). Recall that everybody is in grade $g = 6$ in year $t = 1$. Table 2b shows the distribution of individual school records $s_i = (s_{i2}, ..., s_{i5})$. We see that 52% of the students only have the normal record $(7, 8, 9, 10)$. We see that 11 students only (i.e., less than 0.1%) have repeated grade 7 twice! Once "steered" towards vocational programs, there is no further information on a student's record.

Table 3 gives the result of the "first-stage" regression of actual class size n on Angrist-Lavy's instrument n^* and, either total grade-enrollment N, or total school-enrollment. The first sub-column of each column in Table 3 gives the estimated coefficient, and the second sub-column gives the corresponding t statistic. The three variables work very well in pairs, for each grade, as shown by columns (1) and (2), which propose two variants of the regression. The education of the mother is a significant control for class-size. The less educated the mother, as compared to the reference (3-years-of-college and more), the smaller the size of the class attended by the child. It is clear that children from more educated backgrounds attend larger classes. In rural areas presumably, the children of farmers are in significantly smaller classes. It might seem strange, but the data show that, in grades 6 and 7, younger teachers teach significantly larger classes. The *agrégés* teach significantly larger classes in grades 8 and 9.

CAPBEP indicates that parents think that "a vocational degree is the most useful degree on the job market"; ASFE18 indicates that the parent's preferred school-leaving age is 18; FSENPRO indicates that the student has a sibling enrolled in vocational education. These three variables capture a number of unobserved characteristics of family background, and of the student's environment; they are significant in the first-stage regression. If we remove these variables from the regression, the results remain essentially the same.

As a preliminary test of the impact of class size on the qualitative dependent variable indicating grade promotion, we have estimated a linear probability model, i.e., a simple ordinary least squares regression of the indicator of grade promotion (vs repetition and steering taken together). Tables 4a to 4d present the results for grades 6 to 9 respectively. Each Table presents three regressions. The first is the naive OLS regression of the promotion dummy on class size, total school-enrollment, and controls; the other two are 2SLS estimates in which class-size is first regressed on Angrist-Lavy's instrument, total school-enrollment and other controls (IV1), or Angrist-Lavy's instrument, total grade-enrollment and other controls (IV2). The first sub-column of each column in Table 4 gives the estimated coefficient, and the second sub-column gives the corresponding t statistic. Tables 4a and 4b show that IV estimates of the class-size coefficient are negative and significant — they correspond to grades 6 and 7 respectively. The same coefficients are significant but positive in Tables 4c and 4d, which correspond to grades 8 and 9. The coefficients are small and weakly significant in the grade 9 regression of Table 4d. Total school-enrollment is everywhere added as a control, but its coefficient is uniformly small and non-significant. These results are an anticipation of the complete model's maximum likelihood estimations.

Table 5a-5b present the maximum likelihood estimates for the achievement and class-size equations, respectively. The top of Table 5a gives the α_{tg} coefficients with their corresponding t-statistics. We find stronger effects than with the linear probability model, with coefficients -0.122 for grade 6 in year 1 ($t = 2.43$), and -0.090 for grade 7 in year 2 ($t = 2.13$). Other α coefficients, corresponding to grade 8 and 9, are non-significant. The lower lines of Table 5a list the β_t coefficients. The class-size coefficients must be compared with some of these β coefficients. For instance, the highly educated mother and the executive or professional father being the reference, the coefficients of the least-educated mother dummy is -0.712 in grade 6, that is, roughly 7 times the class-size coefficient. The coefficient of the blue-collar father dummy is -0.44 in grade 6, meaning that a 4-student reduction of class-size would compensate for the detrimental effect of a blue-collar father. These ratios vary between grades 6 and 7. In grade 7, the class-size reduction needed to make up for the negative effect of a non-educated mother is 8, and it is 6 to make up for the negative effect of a blue-collar father. We conclude that class-size effects are moderate, yet non-negligible. These effects seem to vanish in grades higher than grade 7. It might be that we lack a crucial control variable and (or) an instrument to identify class-size effects in higher grades, presumably because they become weaker as students become more mature. An explanation is the specific character of grade 8 (*i.e., classe de quatrième*) in French high schools. according to some headmasters, there are only a few repetitions of grade 8, the important exam is at the end of grade 9. Teachers and headmasters (that we know) also seem to believe that class-size matters less in grades higher than grade 8, because an essential problem is to impose discipline in the class (and smaller classes are of course an advantage in this respect), but the discipline problem would tend to become less acute in higher grades. A disappointing result is that *observed* teacher characteristics in the school do not seem to explain performance well. The percentage of (supposedly) "high quality" teachers (*i.e., the agrégés*) is correlated with larger classes in grades 8 and 9. But their effects on performance are not clear. One difficulty is that these percentages can only indicate general characteristics of the high school, but do not tell us if the student was in the class of an older teacher, or of an *agrégé* in year t. The effect of observed teacher characteristics is difficult to capture (see Rivkin *et al.* (2005)).

Table 5b displays the maximum likelihood estimates of the auxiliary (i.e., first-stage) class-size regressions. These results closely parallel the corresponding OLS regressions, with differences due to the non-trivial covariance structure of error terms Ω.

Tables 5c to 5e give the estimated entries of the covariance matrix Ω. Table 5c lists the parameters of sub-matrix $\Omega_{\epsilon\epsilon}$, the class-size error-terms covariance matrix. These error terms exhibit a high degree of positive correlation. A child attending a large class in year $t = 1$ is more likely to attend a large class in years $t \geq 2$, which confirms our intuitions, that persistent unobservable factors are class-size determinants. The correlation can be explained by the presence of substantial student and school-related factors. If a student stayed in the same school with, for some unobservable reason, class sizes above the average in year t, it is likely that these class-sizes will remain high in year $t + 1$. At the same time, positive correlation can be generated by individual student effects: a student identified as weak by the teachers (for reasons that we do not observe) will tend to remain in a smaller classes during his (her) entire career in junior high school. Both effects are combined, and the variance of ϵ terms is large. Note that the correlation induced by $cov(\epsilon_t, \epsilon_u)$ decreases with the distance $\mid t - u \mid$.

Table 5d shows the cross-correlations between the ν and the ϵ, i.e., the off-diagonal blocks $\Omega_{\epsilon\nu}$. These cross-correlations are essentially all positive — almost all of them are significant. So, ν and ϵ are not independent. It is easy to see that $corr(\nu_u, \epsilon_t)$ tends to decrease with u when $t \leq u$ and t is fixed. There is some "memory" of past class sizes, which is positively correlated with later student performance.

Finally, Table 5e gives the complete estimated Ω matrix (under conditional independence constraints (14)). It is easy to see that the estimated unconditional $\Omega_{\nu\nu}$ sub-matrix has many small off-diagonal terms; the correlation terms $corr(\nu_t, \nu_u)$, $t \neq u$ are all below 5%. This is driven by the fact that the variances of ϵ terms (i.e., the diagonal terms of $\Omega_{\epsilon\epsilon}$) are large. The computed unconditional variances of ν_1 and ν_2 are greater than 1, as expected, given the adopted normalization of conditional variances. Finally, we have checked that the covariance matrix is positive definite. This had not been imposed during the estimation algorithm, but the 8 eigenvalues of Ω are all positive, ranging from .97 to 14.

5. Conclusion

We found small but significant and non-negligible class-size effects on the probabilities of grade promotion, grade repetition, and tracking, in the first two grades of French junior high schools (grades 6 and 7). These results are obtained by means of an IV estimation strategy in which class-size discontinuities, due to the combined effects of total grade enrollment fluctuations and class-opening rules (i.e., "Maimonide's rule") plays an essential role. Our method has been to estimate the probabilities of 4 transitions (from grade 6 to grade 10) and 4 class-size equations jointly by maximum likelihood, under the assumption that grade transitions are conditionally independent, viz., independent, conditional on the entire class-size history. This eight dimensional system generates an 8-dimensional covariance structure, and class-size coefficients. This method imposes much weaker constraints than an ordinary year-by-year or grade-by-grade estimation of the same model. The instrument for class-size, based on Maimonide's rule performs well in auxiliary class-size regressions, and has an important effect on class-size coefficients in student-performance equations. Our method provides estimates of an 8-dimensional covariance matrix of errors that reflects the presence of positive correlations between class-sizes in different years, as well as positive correlations between class-size error-terms and performance equation error-terms: in spite of family-background variables and other controls, some student characteristics are observed by the teachers, but not by the econometrician, and explain that conditionally weaker students are assigned to smaller classes. A substantial amount of unobserved remedial education by means of smaller classes is taking place, and ordinary estimates of class-size coefficients are strongly biased upwards. Class-size coefficients are in fact negative and significant, but in grades 6 and 7 only. In these grades, an 8 student reduction would compensate the handicaps generated by a non-educated mother, as compared to a college-educated mother. Class-size effects vanish in higher grades.

6. Appendix. Details of the Estimation Method

Let $k(i)$ be the number of years during which an individual i is observed; then, $k \in \{1, 2, 3, 4\}$. Formally, $s_{it} = v$ if and only if $k(i) = t-1$, for $t \geq 2$. Define $y_i = (y_{i1}, ..., y_{ik(i)})$ as a column vector and rewrite (4), the latent performance equations, in matrix notation, as follows,

$$y_i = A_i(X, \alpha, \beta) + \nu_i, \tag{A1}$$

where, to keep notations simple, we view ν_i as being the appropriate error term vector, i.e., $dim(\nu_i) = k(i)$, and A_i is the regression function specified by (4). Define also $n_i = (n_{i1}, ..., n_{ik(i)})$, and likewise, rewrite (9), the class-size equations, as

$$n_i = B_i(Z, a, b) + \epsilon_i, \tag{A2}$$

where, again, ϵ_i is the appropriate vector, with $dim(\epsilon_i) = k(i)$, and B_i is the regression function specified by (9).

To be fully rigorous, define Ω_k as the $2k$-dimensional covariance matrix of the Gaussian vector (ν_i, ϵ_i), when $k = k(i)$. We use the same partition as above for Ω, that is, define

$$\Omega_k = \begin{pmatrix} \Omega_{k\nu\nu} & \Omega_{k\nu\epsilon} \\ \Omega_{k\epsilon\nu} & \Omega_{k\epsilon\epsilon} \end{pmatrix}, \tag{A3}$$

where each element of the partitioned matrix is itself a k-dimensional matrix. We view ν_i and ϵ_i as the appropriate projections of the Gaussian vector (ν, ϵ), the covariance matrix of which is Ω, and for each k, Ω_k is obtained from Ω by deleting the appropriate columns and lines in each sub-matrix. Denote then f_k the Gaussian density of the k-dimensional vector n_i when $k(i) = k$. Formally, the density of n_i can be written as follows,

$$f_{k(i)}(n_i) = (2\pi)^{-\frac{k(i)}{2}} (\det \Omega_{k\epsilon\epsilon})^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}(n_i - B_i(.))'\Omega^{-1}_{k(i)\epsilon\epsilon}(n_i - B_i(.))\Big\} \qquad (A4)$$

Under normality, the error terms ν_i can be decomposed as follows,

$$\nu_i = \Omega_{k\nu\epsilon}\Omega^{-1}_{k\epsilon\epsilon}\epsilon_i + \xi_i, \qquad (A5a)$$

where $k = k(i)$, and

$$E(\nu_i \mid \epsilon_i) = \Omega_{k\nu\epsilon}\Omega^{-1}_{k\epsilon\epsilon}\epsilon_i, \qquad (A5b)$$

and the vector ξ_i is normal, with a zero mean, and independent from ϵ_i. We can then write,

$$\nu_i = \Omega_{\nu\epsilon}\Omega^{-1}_{\epsilon\epsilon}(n_i - B_i(Z, a, b)) + \xi_i, \qquad (A6a)$$

and for each ν_{it},

$$\nu_{it} = (\sigma_{\nu_t\epsilon_1}, ..., \sigma_{\nu_t\epsilon_{k(i)}})\Omega^{-1}_{\epsilon\epsilon}(n_i - B_i(Z, a, b)) + \xi_{it}, \qquad (A6b)$$

where $\sigma_{\nu_t\epsilon_u} = cov(\nu_{it}, \epsilon_{iu})$. Let $\sigma_{\nu_t\epsilon} = (\sigma_{\nu_t\epsilon_1}, ..., \sigma_{\nu_t\epsilon_{k(i)}})$

We now rewrite the transition probabilities P. From (6) above, using the fact that $var(\xi_{it} \mid \epsilon) = 1$, we derive,

$$\begin{aligned} &P(s_{i,t+1} = g + 1 \mid s_{it} = g) \\ &\quad = \Pr(y_{it} \geq \delta_t) \\ &\quad = \Pr[\nu_{it} \geq \delta_t - A_i(X, \alpha, \beta)] \\ &\quad = \Pr[\xi_{it} \geq \delta_t - A_i(X, \alpha, \beta) - \sigma_{\nu_t\epsilon}\Omega^{-1}_{\epsilon\epsilon}(n_i - B_i(Z, a, b))] \\ &\quad = 1 - \Phi[\delta_t - A_i(X, \alpha, \beta) - \sigma_{\nu_t\epsilon}\Omega^{-1}_{\epsilon\epsilon}(n_i - B_i(Z, a, b))]], \qquad (A7) \end{aligned}$$

where Φ is the cumulative distribution function of the standard $\mathcal{N}(0, 1)$ distribution. Using (7), the same reasoning yields,

$$\begin{aligned} &P(s_{i,t+1} = g \mid s_{it} = g) = \Pr(\delta_t > y_{it} \geq \gamma_t) \\ &\quad = \Phi[\delta_t - A_i(.) - \sigma_{\nu_t\epsilon}\Omega^{-1}_{\epsilon\epsilon}(n_i - B_i(.))] - \Phi[\gamma_t - A_i(.) - \sigma_{\nu_t\epsilon}\Omega^{-1}_{\epsilon\epsilon}(n_i - B_i(.))], \qquad (A8) \end{aligned}$$

and,

$$P(s_{i,t+1} = v \mid s_{it}) = \Pr(y_{it} < \gamma_t) = \Phi[\gamma_t - A_i(.) - \sigma_{\nu_t\epsilon}\Omega^{-1}_{\epsilon\epsilon}(n_i - B_i(.))]. \qquad (A9)$$

With the use of these expressions, we can compute individual i's contribution to likelihood L_i as follows,

$$L_i = \Pr(s_i \mid n_i)f_{k(i)}(n_i) = f_{k(i)}(n_i) \prod_{t=2}^{k(i)} P(s_{i,t} \mid s_{i,t-1}). \qquad (A10)$$

Finally, the likelihood is simply $L = \prod_i L_i$.

7. References

Angrist, J. D. and V. Lavy (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*,114, 533-575.

Banerjee, A., Cole, S., Duflo, E. and L. Linden (2005), "Remedying Education: Evidence from Two Randomized Experiments in India", manuscript, MIT, Cambridge, Massachusetts.

Benabou, R., Kramarz, F., and C. Prost (2004), "Zones d'éducation prioritaires: quels moyens pour quels résultats?", *Economie et Statistique*, 380, 3-30.

Bonesrøning, H. (2003), "Class size Effects on Student Achievement in Norway: Patterns and Explanations", *Southern Economic Journal*, 69, 952-965.

Boozer, M. and C. Rouse (2001), "Intraschool Variation in Class Size: Patterns and Implications", *Journal of Urban Economics*, 50, 163-189.

Bressoux, P., Kramarz, F. and C. Prost (2005),"Teachers' Training, Class Size, and Students' Outcomes: Evidence from Third Grade Classes in France", manuscript, CREST, INSEE, France.

Card, D. and A. Krueger (1996), "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina", *Journal of Economic Perspectives*, 10, 31-50.

Carneiro, P. and, J.J. Heckman (2003), "Human Capital Policy", in Heckman and Krueger, *Inequality in America; What Role for Human Capital Policies?*, MIT Press, Cambridge, Massachusetts.

Case, A. and A. Deaton (1999), "School Inputs and Educational Outcomes in South Africa",*Quarterly Journal of Economics*, 114, 1047-1084.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., and R. L. York. (1966), *Equality of Educational Opportunity*, US Government Printing Office, Washington D.C.

Dearden, L., Ferri, J. and C. Meghir (2002), "The Effect of School Quality on Educational Attainment and Wages", *Review of Economics and Statistics*, 84, 1-20.

Duflo, E. (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", *American Economic Review*, 91, 795-813.

Dobbelsteen, S., Levin, J., and H. Oosterbeek, (2002),“The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class-Size Effect From the Effect of Changes in Class Composition”, *Oxford Bulletin of Economics and Statistics*, 64, 17-38.

Dustmann, C., Rajah, N. and A. van Soest (2003), ”Class Size, Education and Wages”, *Economic Journal*, 113, F99-F120.

Finn, J. D. and C. M. Achilles (1990), ”Answers and Questions about Class Size: A Statewide Experiment”, *American Educational Research Journal*, 27, 557-577.

Hanushek, E. A. (2002), ”Evidence, Politics, and the Class-Size Debate”, chapter 2 in, Mishel and Rothstein, eds. *The Class Size Debate*, Economic Policy Institute, Washington, D.C.

Hanushek, E. A. (2003), ”The Failure of Input-Based Schooling Policies”, *Economic Journal*, 113, F64-F98.

Hoxby, C. M. (2000), ”The Effects of Class Size on Student Achievement: New Evidence from Population Variation”, *Quarterly Journal of Economics*, 115, 1239-1285.

Krueger, A. B. (1999), ”Experimental Estimates of Education Production Functions”, *Quarterly Journal of Economics*, 114, 497-532.

Krueger, A. B. (2002), ”Understanding the Magnitude and Effect of Class Size on Student Achievement”, chapter 1 in, Mishel and Rothstein, eds. *The Class Size Debate*, Economic Policy Institute, Washington, D.C.

Krueger, A. B. (2003), ”Inequality, Too Much of a Good Thing”, in Heckman and Krueger, *Inequality in America; What Role for Human Capital Policies?*, MIT Press, Cambridge, Massachusetts.

Piketty, T. (2004), ”L’impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises: une estimation à partir du panel primaire 1997”, manuscript; presentation in English: ”Should We Reduce Class Size or School Segregation? Theory and Evidence from France”, *http://www.jourdan.ens.fr/piketty*, PSE Jourdan, ENS, Paris, France.

Piketty, T. and M. Valdenaire (2006), “L’impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français : Estimations à partir du panel primaire 1997 et du panel secondaire 1995”, Ministère de l’Education Nationale, Les Dossiers n173.

Rivkin, S.G., Hanushek, E. A., and J.F. Kain (2005), “Teachers, Schools, and Academic Achievement”, *Econometrica*, 73, 417-458.

Vignoles, A., Levacic, R., Walker, J., Machin, S. and D. Reynolds, (2000), "The Relationship Between Resource Allocation and Pupil Attainment: A Review", *Center for Economics of Education*, London School of Economics, London, UK.

Wössmann, L. (2005), "Educational Production in Europe", *Economic Policy*, 20, issue 43, 445-504.

Wössmann, L., and M. West (2006),"Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS", *European Economic Review*, 50, 695-736.

TABLE 1 : GENERAL DESCRIPTIVE STATISTICS

SOCIOECONOMIC VARIABLES	Mean
Male	0.486
Mother's education	
Elementary school	0.179
Some Junior high school education	0.119
End of JHS certificate	0.092
2 or 3 years in a vocational high school	0.126
Some general HS education or vocational degree	0.110
End of HS certificate or 1year of college	0.122
Two years of college	0.066
Three years of college or more	0.074
Missing	0.112
Father's occupation	
Farmer	0.029
Craftsmen, Shopkeepers, owners-managers	0.083
Executives, Doctors, Lawyers, Engineers	0.144
Middle managers, technicians	0.205
White collars	0.102
Blue collars	0.368
Inactive	0.014
Missing	0.055
Age at grade 6 entry	
Ten	0.040
Eleven	0.736
Twelve	0.184
Thirteen	0.041
Birth order	
First	0.479
Second	0.348
Third and more	0.173
Number of siblings	
One	10.060
Two	41.900
Three	29.120
four	9.710
five or more	9.210
The child has some siblings in vocational school	0.106
Parents think that a vocational degree is the most useful degree in the job market	0.107
Parents' preferred school leaving age : 18	0.158
Repeated a grade in elementary school	0.215
Children in a ZEP School	0.140
First year	0.140
Second year	0.133
Third year	0.131
Fourth year	

INSTITUTIONAL VARIABLES	Mean	Std. D
Mean class size		
Grade6	25.30	3.29
Grade 6, repeaters	25.13	2.72
Grade 7	25.54	2.72
Grade 7, repeaters	25.10	2.98
Grade 8	25.51	3.19
Grade 8, repeaters	24.80	3.49
Grade 9	25.54	3.23
Total grade enrollment		
Grade6	156.50	56.41
Grade 6, repeaters	156.91	57.58
Grade 7	157.12	56.81
Grade 7, repeaters	159.25	57.12
Grade 8	127.22	47.31
Grade 8, repeaters	128.08	47.14
Grade 9	127.13	48.53
% teachers younger than 30		
Grade6	0.112	0.085
Grade 6, repeaters	0.108	0.086
Grade 7	0.114	0.087
Grade 7, repeaters	0.115	0.089
Grade 8	0.118	0.088
Grade 8, repeaters	0.142	0.135
Grade 9	0.145	0.136
% teachers older than 50		
Grade6	0.167	0.102
Grade 6, repeaters	0.181	0.106
Grade 7	0.184	0.106
Grade 7, repeaters	0.195	0.109
Grade 8	0.199	0.109
Grade 8, repeaters	0.218	0.129
Grade 9	0.224	0.132
% teachers *agrégés*		
Grade6	0.028	0.033
Grade 6, repeaters	0.027	0.031
Grade 7	0.030	0.033
Grade 7, repeaters	0.032	0.032
Grade 8	0.033	0.035
Grade 8, repeaters	0.035	0.034
Grade 9	0.037	0.037
% teachers men		
Grade6	0.393	0.097
Grade 6, repeaters	0.395	0.098
Grade 7	0.390	0.098
Grade 7, repeaters	0.386	0.096
Grade 8	0.386	0.096
Grade 8, repeaters	0.381	0.094
Grade 9	0.382	0.093

TABLE 2a : SCHOOL RECORDS

	YEAR 1	YEAR 2		YEAR 3		YEAR 4		
Enrolled in grade 6 in September 89	**12 854**							
promoted to grade 7	11 936							
repeated grade 6	918							
Enrolled in general education in September t :		**12 854**		**12 150**			**11 486**	
		Grade 6	**Grade 7**	**Grade 7**	**Grade 8**	**Grade 7**	**Grade 8**	**Grade 9**
Result in June t+1 :		**918**	**11 936**	**2 115**	**10 037**	**86**	**2 110**	**9 290**
promoted to next grade		918	10 037	1 450	9 290	46	1 936	6 776
repeated the grade		0	1 195	86	660	0	62	1 091
"Steered" towards vocational programs		0	704	577	87	40	112	1 423

TABLE 2b : SCHOOL RECORDS

YEAR 1	YEAR 2	YEAR 3	YEAR 4	YEAR 5	Nb observations
6	7	8	9	10	6 776
6	7	8	9	9	1 091
6	7	8	9	v	1 423
6	7	8	8	9	638
6	7	8	8	v	19
6	7	8	v	_	87
6	7	7	8	9	833
6	7	7	8	8	27
6	7	7	8	v	48
6	7	7	7	8	11
6	7	7	v	_	278
6	7	v	_	_	704
6	6	7	7	8	35
6	6	7	7	v	40
6	6	7	8	9	462
6	6	7	8	8	36
6	6	7	8	v	45
6	6	7	v	_	301
					12 854

TABLE 3: FIRST-STAGE REGRESSION

	GRADE 6				GRADE 7			
	(1)		(2)		(1)		(2)	
Angrist-Lavy's instrument	0.2032	12.42	0.1906	11.46	0.2401	17.37	0.2397	17.19
Total school enrollment	0.0019	11.19	_	_	0.0016	11.04	_	_
Total grade enrollment	_	_	0.0072	11.93	_	_	0.0054	10.73
ZEP School	-0.4758	-5.68	-0.5168	-6.17	-0.4661	-6.59	-0.4913	-6.93
Teachers								
% younger than 30	1.1731	3.04	1.1730	3.04	1.3631	4.27	1.3685	4.28
% older than 50	0.4100	1.26	0.5334	1.65	1.3277	5.05	1.3829	5.26
% *agrégés*	1.5666	1.75	1.8840	2.12	-0.4216	-0.57	-0.2915	-0.39
% men	-0.8308	-2.73	-0.8132	-2.67	0.2527	1.00	0.2544	1.01
Mother's education								
Ref : 3 years of college or more								
Elementary school	-0.4481	-3.16	-0.4572	-3.23	-0.3454	-2.94	-0.3529	-3.00
Some junior high school education	-0.4217	-2.87	-0.4388	-2.99	-0.4436	-3.63	-0.4510	-3.69
End of junior high school certificate	-0.3834	-2.55	-0.3955	-2.63	-0.2610	-2.10	-0.2673	-2.15
2 or 3 years in a vocational high school	-0.3673	-2.59	-0.3774	-2.66	-0.4228	-3.59	-0.4271	-3.63
Some general HS educ. or vocational degree	-0.3052	-2.15	-0.3186	-2.25	-0.4191	-3.59	-0.4276	-3.66
End of HS certificate or 1 year of college	-0.2858	-2.08	-0.2952	-2.15	-0.2750	-2.44	-0.2807	-2.49
Two years of college	-0.1329	-0.87	-0.1396	-0.91	-0.0306	-0.24	-0.0345	-0.28
Missing	-0.5383	-3.58	-0.5531	-3.68	-0.5092	-4.06	-0.5153	-4.11
Father's occupation								
Ref : Executives, Doctors, Lawyers, Engineers								
Farmers	-0.4486	-2.36	-0.4581	-2.41	-0.5378	-3.36	-0.5632	-3.52
Craftsmen, Shopkeepers, Owners-managers	-0.3009	-2.34	-0.3137	-2.44	-0.3728	-3.46	-0.3853	-3.58
Middle managers, technicians,	-0.1564	-1.52	-0.1665	-1.62	-0.2626	-3.09	-0.2718	-3.20
White collars	-0.1642	-1.32	-0.1773	-1.43	-0.4235	-4.09	-0.4322	-4.17
Blue collars	-0.1969	-1.87	-0.2078	-1.98	-0.4394	-5.03	-0.4482	-5.13
Inactive	-0.5113	-1.97	-0.5151	-1.98	-0.7849	-3.52	-0.7918	-3.55
Missing	-0.2610	-1.75	-0.2717	-1.82	-0.4923	-3.90	-0.5046	-4.00
Quarter of birth								
First	0.1646	2.00	0.1569	1.91	-0.0152	-0.22	-0.0209	-0.30
Second	0.0774	0.99	0.0736	0.94	0.0277	0.42	0.0260	0.39
Third	0.0422	0.52	0.0420	0.52	-0.0867	-1.25	-0.0878	-1.27
Birth order								
Second	0.0931	1.43	0.0974	1.50	0.0001	0.00	0.0024	0.04
Third or more	0.0118	0.12	0.0188	0.19	0.0091	0.11	0.0176	0.21
Number of siblings	-0.0345	-1.08	-0.0363	-1.14	-0.0807	-2.97	-0.0840	-3.09
CAP_BEP	-0.3323	-3.37	-0.3383	-3.43	-0.2803	-3.22	-0.2800	-3.22
FS_enPRO	-0.1547	-1.55	-0.1599	-1.61	-0.2196	-2.57	-0.2201	-2.57
ASFE_18	-0.1801	-2.11	-0.1757	-2.06	-0.1718	-2.31	-0.1715	-2.30
Constant	19.4974	42.59	19.7635	42.81	18.7198	48.85	18.7777	48.70
R^2	0.0644		0.0656		0.0991		0.0986	
Root MSE	3.1851		3.1830		2.5894		2.5901	
Number of observations	12854		12854		11936		11936	

	GRADE 8				GRADE 9			
	(1)		(2)		(1)		(2)	
Angrist-Lavy's instrument	0.2622	17.53	0.2490	16.54	0.3262	21.57	0.3168	20.78
Total school enrollment	0.0021	12.10	_	_	0.0020	11.38	_	_
Total grade enrollment	_	_	0.0102	13.44	_	_	0.0096	12.20
ZEP School	-0.3527	-3.84	-0.3196	-3.49	-0.4547	-4.82	-0.3945	-4.18
Teachers								
% younger than 30	-0.2694	-0.67	-0.3898	-0.97	-0.0538	-0.21	-0.0714	-0.27
% older than 50	-0.3843	-1.19	-0.5576	-1.72	0.2772	1.01	0.1030	0.38
% *agrégés*	4.2759	4.83	4.1832	4.74	3.8928	4.47	3.2906	3.76
% men	-0.6353	-1.94	-0.6524	-2.00	-1.4006	-4.04	-1.3118	-3.79
Mother's education								
Ref : 3 years of college or more								
Elementary school	-0.6547	-4.58	-0.6312	-4.42	-0.5807	-4.00	-0.5515	-3.80
Some junior high school education	-0.6212	-4.15	-0.5951	-3.98	-0.5380	-3.52	-0.5026	-3.29
End of junior high school certificate	-0.2696	-1.80	-0.2433	-1.63	-0.2526	-1.66	-0.2203	-1.44
2 or 3 years in a vocational high school	-0.2390	-1.69	-0.2120	-1.50	-0.3945	-2.74	-0.3582	-2.49
Some general HS educ. or vocational degree	-0.2595	-1.88	-0.2372	-1.72	-0.3359	-2.39	-0.3085	-2.20
End of HS certificate or 1 year of college	-0.2210	-1.66	-0.1995	-1.50	-0.1789	-1.33	-0.1566	-1.17
Two years of college	-0.0183	-0.12	0.0108	0.07	-0.0442	-0.30	-0.0191	-0.13
Missing	-0.6164	-4.00	-0.5919	-3.84	-0.6820	-4.32	-0.6575	-4.17
Father's occupation								
Ref : Executives, Doctors, Lawyers, Engineers								
Farmers	-0.6369	-3.29	-0.6447	-3.34	-0.7969	-4.06	-0.7902	-4.03
Craftsmen, Shopkeepers, Owners-managers	-0.2751	-2.11	-0.2628	-2.02	-0.2280	-1.71	-0.2063	-1.55
Middle managers, technicians,	-0.0970	-0.96	-0.0977	-0.97	-0.1917	-1.88	-0.1738	-1.70
White collars	-0.3858	-3.06	-0.3764	-2.99	-0.4646	-3.59	-0.4381	-3.39
Blue collars	-0.4040	-3.81	-0.4003	-3.78	-0.5077	-4.70	-0.4868	-4.51
Inactive	-0.7881	-2.77	-0.8017	-2.83	-0.3148	-1.06	-0.2938	-0.99
Missing	-0.2398	-1.48	-0.2532	-1.56	-0.1746	-1.04	-0.1753	-1.04
Quarter of birth								
First	0.2481	2.85	0.2473	2.84	-0.0590	-0.66	-0.0484	-0.54
Second	0.1595	1.91	0.1527	1.83	0.0643	0.75	0.0656	0.77
Third	0.0093	0.11	0.0070	0.08	-0.0823	-0.92	-0.0770	-0.86
Birth order								
Second	-0.0480	-0.70	-0.0531	-0.78	-0.0692	-0.99	-0.0777	-1.11
Third or more	0.1386	1.31	0.1380	1.31	0.0094	0.09	0.0069	0.06
Number of siblings	-0.0325	-0.93	-0.0306	-0.88	-0.0423	-1.17	-0.0409	-1.14
CAP_BEP	-0.6206	-4.92	-0.6230	-4.95	-0.3885	-2.92	-0.3964	-2.98
FS_enPRO	-0.4275	-3.72	-0.4250	-3.70	-0.3425	-2.86	-0.3424	-2.86
ASFE_18	-0.3530	-3.33	-0.3548	-3.35	-0.4360	-3.91	-0.4334	-3.89
Constant	18.3036	43.02	18.5998	43.49	16.9936	39.51	17.2120	39.81
R^2	0.1314		0.1343		0.1698		0.1715	
Root MSE	2.9792		2.9742		2.9457		2.9427	
Number of observations	10037		10037		9305		9305	

TABLE 4a : LINEAR PROBABILITY MODEL - GRADE 6

	GRADE 6					
	OLS		IV 1		IV 2	
Class size	0.0003	0.50	-0.0136	-2.10	-0.0209	-3.33
Total school enrollment . 10^{-3}	-0.0054	-0.47	0.0357	1.59	0.0573	2.60
Male	-0.0183	-4.10	-0.0197	-4.30	-0.0204	-4.38
Mother's education						
Ref : 3 years of college or more						
Elementary school	-0.0428	-3.82	-0.0491	-4.18	-0.0524	-4.39
Some junior high school education	-0.0627	-5.40	-0.0689	-5.68	-0.0721	-5.84
End of junior high school certificate	-0.0323	-2.72	-0.0381	-3.08	-0.0411	-3.27
2 or 3 years in a vocational high school	-0.0515	-4.59	-0.0568	-4.87	-0.0596	-5.02
Some general HS educ. or vocational degree	-0.0089	-0.79	-0.0136	-1.17	-0.0161	-1.36
End of HS certificate or 1 year of college	-0.0100	-0.92	-0.0143	-1.28	-0.0166	-1.45
Two years of college	-0.0038	-0.31	-0.0059	-0.48	-0.0069	-0.55
Missing	-0.0500	-4.20	-0.0577	-4.58	-0.0617	-4.82
Father's occupation						
Ref : Executives, Doctors, Lawyers, Engineers						
Farmers	-0.0302	-2.01	-0.0375	-2.40	-0.0414	-2.60
Craftsmen, Shopkeepers, Owners-managers	-0.0326	-3.20	-0.0369	-3.50	-0.0392	-3.65
Middle managers, technicians,	-0.0173	-2.13	-0.0192	-2.31	-0.0202	-2.38
White collars	-0.0281	-2.86	-0.0304	-3.03	-0.0316	-3.09
Blue collars	-0.0309	-3.71	-0.0336	-3.93	-0.0351	-4.03
Inactive	-0.0347	-1.69	-0.0416	-1.97	-0.0451	-2.10
Missing	-0.0452	-3.83	-0.0488	-4.03	-0.0506	-4.10
Quarter of birth						
First	0.0201	3.09	0.0223	3.34	0.0235	3.44
Second	0.0189	3.04	0.0200	3.17	0.0206	3.20
Third	0.0064	0.99	0.0068	1.04	0.0071	1.06
Birth order						
Second	-0.0005	-0.10	0.0007	0.14	0.0014	0.25
Third or more	-0.0224	-2.91	-0.0222	-2.84	-0.0221	-2.77
Number of siblings	0.0013	0.53	0.0009	0.34	0.0006	0.24
CAP_BEP	-0.0666	-8.53	-0.0711	-8.68	-0.0735	-8.82
FS_enPRO	-0.0169	-2.15	-0.0191	-2.37	-0.0203	-2.47
ASFE_18	-0.0613	-9.09	-0.0640	-9.19	-0.0654	-9.22
Teachers						
% younger than 30	0.1433	4.69	0.1606	5.02	0.1697	5.21
% older than 50	0.0078	0.30	0.0131	0.50	0.0159	0.59
% *agrégés*	0.0092	0.13	0.0292	0.40	0.0396	0.54
% men	0.0323	1.34	0.0181	0.71	0.0106	0.41
ZEP School	0.0054	0.81	-0.0006	-0.08	-0.0038	-0.51
Constant	0.9651	39.16	1.3065	8.18	1.4855	9.59
R^2	0.04590		0.01590		0.01580	
Root MSE	0.25187		0.25579		0.26089	
Number of observations	12854		12854		12854	

TABLE 4b : LINEAR PROBABILITY MODEL - GRADE 7

	GRADE 7					
	OLS		IV 1		IV 2	
Class size	0.0074	6.32	-0.0150	-1.98	-0.0154	-2.03
Total school enrollment . 10^{-3}	-0.0469	-2.88	0.0165	0.61	0.0176	0.65
Male	-0.0846	-13.69	-0.0860	-13.67	-0.0861	-13.67
Mother's education						
Ref : 3 years of college or more						
Elementary school	-0.1102	-7.22	-0.1175	-7.49	-0.1176	-7.49
Some junior high school education	-0.0986	-6.20	-0.1084	-6.58	-0.1086	-6.59
End of junior high school certificate	-0.0512	-3.17	-0.0569	-3.45	-0.0570	-3.45
2 or 3 years in a vocational high school	-0.0597	-3.90	-0.0694	-4.37	-0.0695	-4.38
Some general HS educ. or vocational degree	-0.0172	-1.14	-0.0266	-1.70	-0.0268	-1.71
End of HS certificate or 1 year of college	-0.0070	-0.48	-0.0131	-0.87	-0.0132	-0.88
Two years of college	-0.0130	-0.80	-0.0133	-0.80	-0.0133	-0.80
Missing	-0.1154	-7.09	-0.1270	-7.48	-0.1272	-7.48
Father's occupation						
Ref : Executives, Doctors, Lawyers, Engineers						
Farmers	0.0233	1.12	0.0093	0.43	0.0091	0.42
Craftsmen, Shopkeepers, Owners-managers	-0.0292	-2.09	-0.0374	-2.59	-0.0376	-2.59
Middle managers, technicians,	-0.0162	-1.47	-0.0220	-1.93	-0.0221	-1.94
White collars	-0.0345	-2.56	-0.0444	-3.16	-0.0446	-3.17
Blue collars	-0.0782	-6.88	-0.0881	-7.34	-0.0883	-7.35
Inactive	-0.0011	-0.04	-0.0196	-0.65	-0.0199	-0.66
Missing	-0.1283	-7.82	-0.1394	-8.17	-0.1396	-8.18
Quarter of birth						
First	0.0207	2.30	0.0203	2.22	0.0203	2.22
Second	0.0110	1.27	0.0120	1.37	0.0120	1.37
Third	0.0088	0.98	0.0069	0.76	0.0069	0.76
Birth order						
Second	0.0143	2.01	0.0139	1.93	0.0139	1.93
Third or more	0.0201	1.87	0.0199	1.83	0.0199	1.83
Number of siblings	-0.0185	-5.23	-0.0200	-5.52	-0.0200	-5.53
CAP_BEP	-0.1486	-13.13	-0.1553	-13.27	-0.1555	-13.27
FS_enPRO	-0.0644	-5.79	-0.0694	-6.08	-0.0695	-6.09
ASFE_18	-0.1875	-19.36	-0.1916	-19.30	-0.1917	-19.30
Teachers						
% younger than 30	0.0988	2.38	0.1261	2.92	0.1266	2.93
% older than 50	-0.0328	-0.96	-0.0061	-0.17	-0.0056	-0.16
% *agrégés*	-0.2916	-3.01	-0.2964	-3.02	-0.2965	-3.02
% men	0.0779	2.37	0.0810	2.43	0.0811	2.43
ZEP School	-0.0052	-0.57	-0.0151	-1.53	-0.0153	-1.54
Constant	0.8722	23.06	1.4241	7.57	1.4337	7.62
R^2	0.1567		0.1308		0.1299	
Root MSE	0.3364		0.3415		0.3417	
Number of observations	11936		11936		11936	

TABLE 4c : LINEAR PROBABILITY MODEL - GRADE 8

	GRADE 8					
	OLS		IV 1		IV 2	
Class size	0.0028	3.21	0.0104	2.08	0.0104	2.19
Total school enrollment . 10^{-3}	-0.0083	-0.61	-0.0360	-1.60	-0.0360	-1.66
Male	-0.0312	-5.96	-0.0301	-5.69	-0.0301	-5.69
Mother's education						
Ref : 3 years of college or more						
Elementary school	-0.0092	-0.74	-0.0044	-0.34	-0.0044	-0.34
Some junior high school education	-0.0461	-3.53	-0.0414	-3.08	-0.0414	-3.09
End of junior high school certificate	-0.0316	-2.42	-0.0295	-2.23	-0.0295	-2.24
2 or 3 years in a vocational high school	-0.0449	-3.64	-0.0432	-3.48	-0.0432	-3.48
Some general HS educ. or vocational degree	-0.0289	-2.4	-0.0272	-2.24	-0.0272	-2.24
End of HS certificate or 1 year of college	-0.0062	-0.53	-0.0045	-0.38	-0.0045	-0.38
Two years of college	0.0045	0.35	0.0048	0.37	0.0048	0.37
Missing	-0.0457	-3.4	-0.0410	-2.96	-0.0410	-2.97
Father's occupation						
Ref : Executives, Doctors, Lawyers, Engineers						
Farmers	-0.0069	-0.41	-0.0010	-0.06	-0.0010	-0.06
Craftsmen, Shopkeepers, Owners-managers	-0.0214	-1.88	-0.0189	-1.63	-0.0189	-1.64
Middle managers, technicians,	-0.0146	-1.66	-0.0137	-1.55	-0.0137	-1.55
White collars	-0.0361	-3.28	-0.0331	-2.96	-0.0331	-2.96
Blue collars	-0.0353	-3.81	-0.0321	-3.37	-0.0321	-3.38
Inactive	-0.0529	-2.13	-0.0469	-1.86	-0.0469	-1.87
Missing	-0.0475	-3.35	-0.0455	-3.18	-0.0455	-3.18
Quarter of birth						
First	0.0076	0.99	0.0058	0.75	0.0058	0.75
Second	0.0078	1.06	0.0066	0.90	0.0066	0.90
Third	0.0080	1.05	0.0081	1.07	0.0081	1.07
Birth order						
Second	0.0104	1.74	0.0108	1.81	0.0108	1.81
Third or more	0.0235	2.56	0.0228	2.46	0.0228	2.46
Number of siblings	-0.0006	-0.2	-0.0004	-0.13	-0.0004	-0.13
CAP_BEP	-0.0326	-2.96	-0.0279	-2.43	-0.0279	-2.44
FS_enPRO	-0.0273	-2.72	-0.0240	-2.33	-0.0240	-2.34
ASFE_18	-0.0402	-4.34	-0.0372	-3.92	-0.0372	-3.93
Teachers						
% younger than 30	0.0591	1.69	0.0631	1.79	0.0631	1.79
% men	0.0241	0.85	0.0310	1.07	0.0310	1.07
% *agrégés*	-0.1393	-1.8	-0.1756	-2.17	-0.1756	-2.18
% older than 50	0.0267	0.95	0.0308	1.08	0.0308	1.08
ZEP School	0.0149	1.86	0.0179	2.16	0.0179	2.17
Constant	0.8983	30.38	0.7112	5.71	0.7113	6.02
R^2	0.0236		0.0159		0.0159	
Root MSE	0.2598		0.2608		0.2608	
Number of observations	10037		10037		10037	

TABLE 4d : LINEAR PROBABILITY MODEL - GRADE 9

	GRADE 9					
	OLS		IV 1		IV 2	
Class size	0.0058	4.9	0.0062	1.15	0.0074	1.41
Total school enrollment . 10^{-3}	-0.0433	-2.34	-0.0451	-1.60	-0.0500	-1.80
Male	-0.0363	-5.05	-0.0362	-5.00	-0.0360	-4.97
Mother's education						
Ref : 3 years of college or more						
Elementary school	-0.0776	-4.59	-0.0773	-4.47	-0.0765	-4.43
Some junior high school education	-0.0862	-4.84	-0.0859	-4.76	-0.0852	-4.72
End of junior high school certificate	-0.0700	-3.94	-0.0699	-3.92	-0.0695	-3.90
2 or 3 years in a vocational high school	-0.0880	-5.25	-0.0878	-5.19	-0.0873	-5.16
Some general HS educ. or vocational degree	-0.0713	-4.36	-0.0712	-4.32	-0.0707	-4.29
End of HS certificate or 1 year of college	-0.0309	-1.98	-0.0308	-1.96	-0.0304	-1.94
Two years of college	-0.0192	-1.11	-0.0192	-1.11	-0.0191	-1.10
Missing	-0.0955	-5.19	-0.0951	-5.05	-0.0942	-5.01
Father's occupation						
Ref : Executives, Doctors, Lawyers, Engineers						
Farmers	-0.0375	-1.64	-0.0371	-1.58	-0.0359	-1.53
Craftsmen, Shopkeepers, Owners-managers	-0.0815	-5.25	-0.0813	-5.22	-0.0810	-5.19
Middle managers, technicians,	-0.0268	-2.25	-0.0267	-2.24	-0.0264	-2.22
White collars	-0.0505	-3.35	-0.0503	-3.30	-0.0498	-3.26
Blue collars	-0.0541	-4.29	-0.0539	-4.17	-0.0532	-4.13
Inactive	-0.0238	-0.69	-0.0237	-0.68	-0.0233	-0.67
Missing	-0.0792	-4.05	-0.0791	-4.04	-0.0790	-4.03
Quarter of birth						
First	0.0235	2.25	0.0235	2.25	0.0236	2.26
Second	0.0253	2.53	0.0252	2.52	0.0252	2.52
Third	0.0011	0.1	0.0011	0.11	0.0012	0.12
Birth order						
Second	-0.0141	-1.73	-0.0141	-1.73	-0.0140	-1.72
Third or more	-0.0017	-0.13	-0.0017	-0.13	-0.0017	-0.13
Number of siblings	0.0074	1.75	0.0074	1.75	0.0074	1.76
CAP_BEP	-0.0313	-2.02	-0.0312	-1.99	-0.0307	-1.96
FS_enPRO	-0.0165	-1.19	-0.0164	-1.17	-0.0160	-1.14
ASFE_18	-0.0552	-4.24	-0.0550	-4.16	-0.0544	-4.12
Teachers						
% younger than 30	-0.0168	-0.55	-0.0167	-0.55	-0.0166	-0.55
% older than 50	0.1111	2.75	0.1118	2.71	0.1139	2.76
% *agrégés*	-0.2528	-2.49	-0.2548	-2.45	-0.2603	-2.50
% men	-0.0752	-2.36	-0.0753	-2.36	-0.0756	-2.37
ZEP School	0.0077	0.7	0.0080	0.70	0.0085	0.76
Constant	0.8213	20.12	0.8105	5.95	0.7802	5.84
R^2	0.0317		0.0317		0.0315	
Root MSE	0.3428		0.3428		0.3428	
Number of observations	9290		9290		9290	

Table 5a : NONLINEAR MODEL

	YEAR 1		YEAR 2		YEAR 3		YEAR 4	
<u>**PERFORMANCE EQUATION**</u>								
Grade 6 class-size	**-0.1221**	**2.43**	0.2491	0.05				
Grade 7 class-size			**-0.0909**	**2.13**	-0.0034	0.08	-0.0286	1.08
Grade 8 class-size					0.0323	0.80	0.0310	1.19
Grade 9 class-size							0.0141	0.54
ZEP school	0.0141	0.24	-0.0663	1.47	0.0910	1.79	0.0421	0.91
Total School Enrollment $.10^{-3}$	0.2985	1.32	0.1143	0.81	-0.1161	0.71	-0.1090	0.85
% Teachers older than 50	-0.4212	2.31	-0.2488	1.69	-0.0877	0.59	-0.3328	2.87
% Teachers agrégés	0.3636	0.61	-1.4476	3.08	-0.6018	1.20	-0.8846	2.01
Male	-0.1413	4.00	-0.4204	14.15	-0.2960	9.21	-0.1630	5.43
Mother's education								
Ref : 3 years of college or more								
Elementary school	-0.7121	5.02	-0.7304	7.35	-0.2045	1.93	-0.4709	5.43
Some junior high school education	-0.8099	5.68	-0.7188	7.01	-0.3754	3.56	-0.5031	5.67
End of junior high school certificate	-0.6186	4.26	-0.4815	4.65	-0.2190	2.07	-0.3980	4.46
2 or 3 years in vocational high school	-0.7499	5.34	-0.5404	5.33	-0.3194	3.18	-0.4844	5.68
Some general HS educ. or vocational degree	-0.3628	2.50	-0.3242	3.12	-0.2391	2.35	-0.3795	4.46
End of HS certificate, 1 year of college	-0.3652	2.55	-0.1879	1.82	-0.0539	0.53	-0.1987	2.35
Two years of college	-0.1930	1.20	-0.1491	1.30	0.0071	0.06	-0.1585	1.69
Missing	-0.7589	5.19	-0.7883	7.59	-0.3402	3.17	-0.5164	5.67
Father's occupation : Ref : Executives. Doctors. Lawyers. Engineers								
Farmers	-0.4863	3.49	-0.1387	1.15	-0.1306	1.08	-0.3064	3.01
Craftsmen. Shopkeepers. Owners-managers	-0.4940	4.76	-0.3561	4.32	-0.2534	3.02	-0.3815	5.46
Middle managers. technicians.	-0.3197	3.48	-0.2360	3.32	-0.1861	2.63	-0.1629	2.76
White collars	-0.4334	4.36	-0.3868	4.85	-0.2652	3.30	-0.2318	3.36
Blue collars	-0.4466	4.90	-0.5723	8.13	-0.3221	4.44	-0.2747	4.49
Inactive	-0.5302	3.25	-0.3386	2.44	-0.2793	1.95	-0.3095	2.31
Missing	-0.5513	5.04	-0.7641	8.83	-0.4495	5.03	-0.3689	4.53
Number of siblings	-0.0198	1.13	-0.0795	5.48	-0.0086	0.55	0.0145	0.95
CAP_BEP	-0.3266	6.67	-0.4685	10.96	-0.2659	5.52	-0.2446	4.69
FS_ENPRO	-0.1221	2.35	-0.2314	5.34	-0.1974	4.05	-0.0943	1.88
ASFE_18	-0.3264	7.42	-0.6126	16.48	-0.2966	6.97	-0.2582	5.74
Constant	5.7210	3.45						
Cut 1			-5.6393	5.29	-1.9816	2.00	-2.3861	3.76
Cut 2			-4.9495	4.65	-1.4710	1.49	-1.5954	2.52

Table 5b : NONLINEAR MODEL ctd.

	YEAR 1		YEAR 2		YEAR 3		YEAR 4	
<u>CLASS-SIZE EQUATION</u>								
Instrument	0.1635	10.45	0.2036	16.20	0.2179	16.83	0.2860	21.64
Total School Enrollment .10-3	2.1940	13.31	1.7494	13.06	2.1878	14.23	2.2369	13.95
ZEP school	-0.4325	5.22	-0.4311	6.39	-0.4476	5.53	-0.4436	5.24
% Teachers older than 50	-0.0236	0.14	0.5946	2.77	-0.1629	0.67	0.4052	1.89
% Teachers agrégés	1.2906	1.48	-0.6030	0.87	2.8238	3.64	3.5822	4.62
Mother's education								
Ref : 3 years of college or more								
Elementary school	-0.5515	3.96	-0.4600	4.05	-0.7939	6.02	-0.8156	6.00
Some junior high school education	-0.5264	3.62	-0.5792	4.89	-0.7063	5.13	-0.8011	5.63
End of junior high school certificate	-0.4577	3.06	-0.3364	2.76	-0.4108	2.92	-0.5285	3.66
2 or 3 years in vocational high school	-0.4485	3.18	-0.4760	4.14	-0.3720	2.81	-0.6205	4.56
Some general HS educ. or vocational degree	-0.3545	2.50	-0.4768	4.13	-0.3086	2.33	-0.4398	3.24
End of HS certificate or 1 year of college	-0.3249	2.37	-0.2993	2.68	-0.2732	2.13	-0.2590	1.98
Two years of college	-0.1507	0.99	-0.0612	0.49	-0.0409	0.29	-0.0421	0.29
Missing	-0.6169	4.13	-0.6098	5.01	-0.7025	4.94	-0.8727	5.95
Father's occupation : Ref : Executives. Doctors. Lawyers. Engineers								
Farmers	-0.5629	2.97	-0.5376	3.49	-0.7903	4.40	-0.9069	4.89
Craftsmen. Shopkeepers. Owners-managers	-0.3561	2.77	-0.3799	3.63	-0.4539	3.73	-0.4192	3.34
Middle managers. technicians.	-0.1712	1.66	-0.2461	2.94	-0.1735	1.79	-0.2661	2.69
White collars	-0.2183	1.76	-0.4572	4.53	-0.4930	4.20	-0.4852	4.02
Blue collars	-0.2733	2.62	-0.4516	5.32	-0.5461	5.54	-0.6457	6.36
Inactive	-0.5862	2.26	-0.7631	3.61	-0.8484	3.40	-0.5787	2.21
Missing	-0.3168	2.13	-0.5396	4.45	-0.4927	3.43	-0.4869	3.23
Number of siblings	-0.0437	1.57	-0.0793	3.51	-0.0643	2.39	-0.0811	2.86
Constant	20.4399	49.88	19.9641	60.44	19.3806	57.10	17.4878	50.50

TABLE 5c : $\Omega_{\varepsilon\varepsilon}$

	coef	std-error	t-statistic	Correlation
var(eps1)	10.165	0.127	80.12	1.000
cov(eps1 , eps2)	2.285	0.076	30.32	0.276
cov(eps1 , eps3)	1.310	0.089	14.22	0.138
cov(eps1 , eps4)	1.073	0.092	11.37	0.111
var(eps2)	6.746	0.086	78.31	1.000
cov(eps2 , eps3)	1.586	0.075	20.34	0.205
cov(eps2 , eps4)	1.222	0.076	15.32	0.158
var(eps3)	8.873	0.115	76.51	1.000
cov(eps3 , eps4)	3.059	0.089	34.12	0.339
var(eps4)	9.152	0.120	75.57	1.000

TABLE 5d : $\Omega_{\varepsilon\upsilon}$

	coef	std-error	t-statistic	Correlation
cov(nu1 , eps1)	1.305	0.690	1.89	0.377
cov(nu1 , eps2)	0.377	0.164	2.29	0.112
cov(nu1 , eps3)	0.052	0.108	0.48	0.016
cov(nu1 , eps4)	0.353	0.095	3.73	0.102
cov(nu2 , eps1)	0.431	0.112	3.83	0.153
cov(nu2 , eps2)	0.863	0.290	2.97	0.315
cov(nu2 , eps3)	0.276	0.084	3.30	0.106
cov(nu2 , eps4)	0.272	0.074	3.65	0.104
cov(nu3 , eps1)	0.091	0.075	1.21	0.028
cov(nu3 , eps2)	0.166	0.077	2.15	0.053
cov(nu3 , eps3)	-0.061	0.362	0.17	-0.020
cov(nu3 , eps4)	-0.100	0.136	0.73	-0.033
cov(nu4 , eps1)	0.110	0.059	1.85	0.033
cov(nu4 , eps2)	0.187	0.052	3.59	0.059
cov(nu4 , eps3)	0.246	0.094	2.61	0.081
cov(nu4 , eps4)	0.133	0.245	0.54	0.044

TABLE 5e : Ω

	eps1	eps2	eps3	eps4	nu1	nu2	nu3	nu4
eps1	10.16	2.285	1.310	1.073	_	_	_	_
eps2	_	6.746	1.586	1.222	_	_	_	_
eps3	_	_	8.875	3.059	_	_	_	_
eps4	_	_	_	9.149	_	_	_	_
nu1	1.305	0.377	0.052	0.353	1.178	0.067	0.012	0.014
nu2	0.431	0.863	0.276	0.272	_	1.114	0.020	0.026
nu3	0.091	0.166	-0.061	-0.100	_	_	1.007	0.002
nu4	0.110	0.187	0.246	0.133	_	_	_	1.011

FIGURE 1

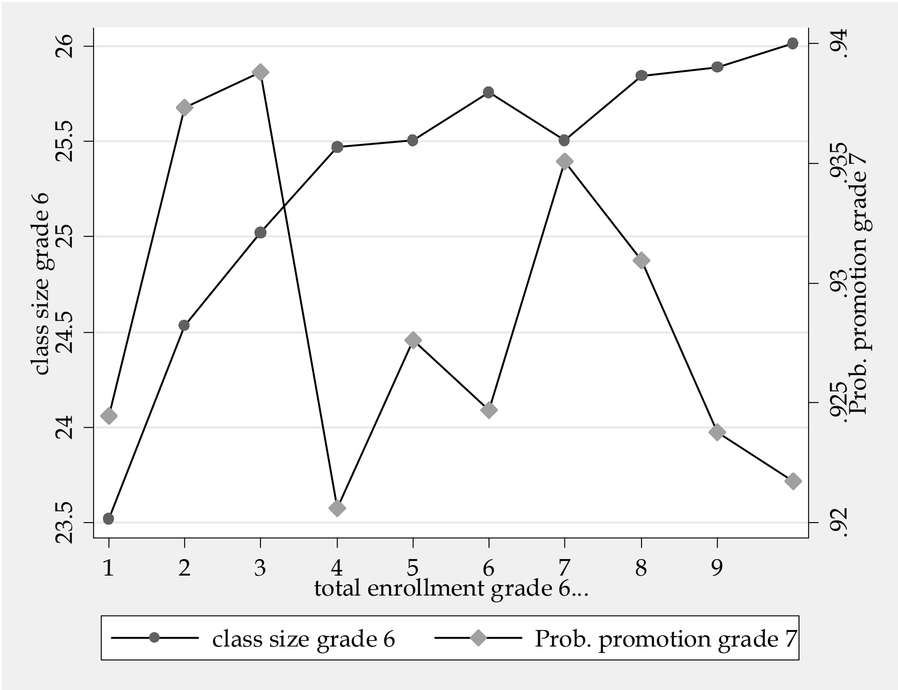


FIGURE 2

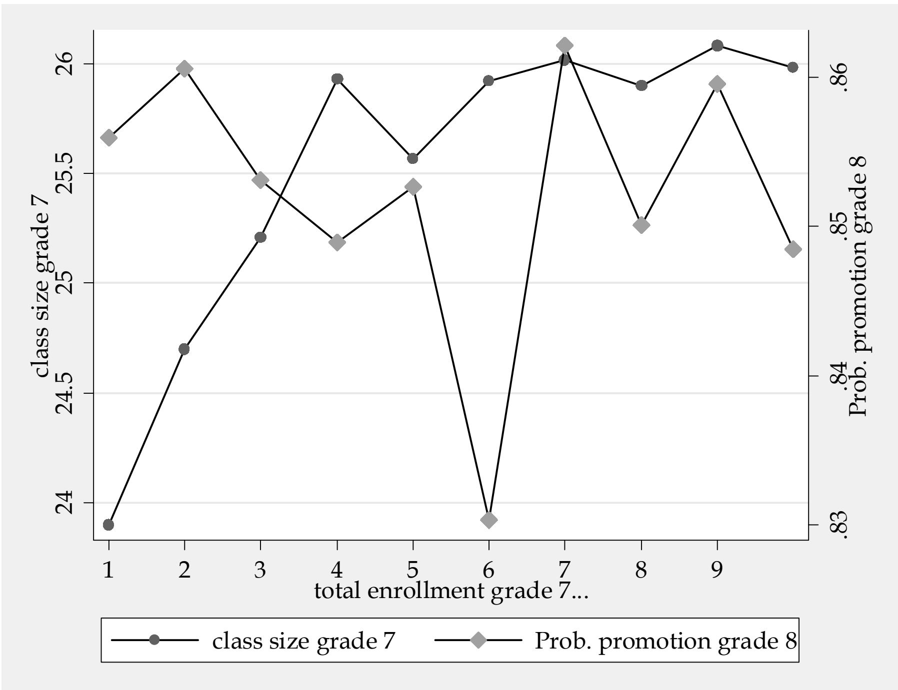


FIGURE 3

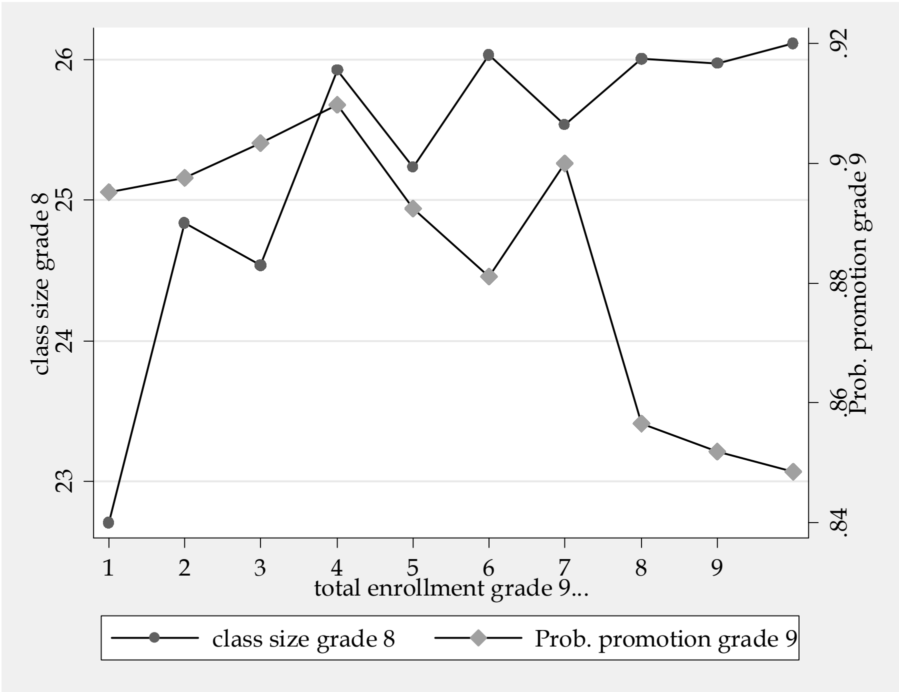


FIGURE 4

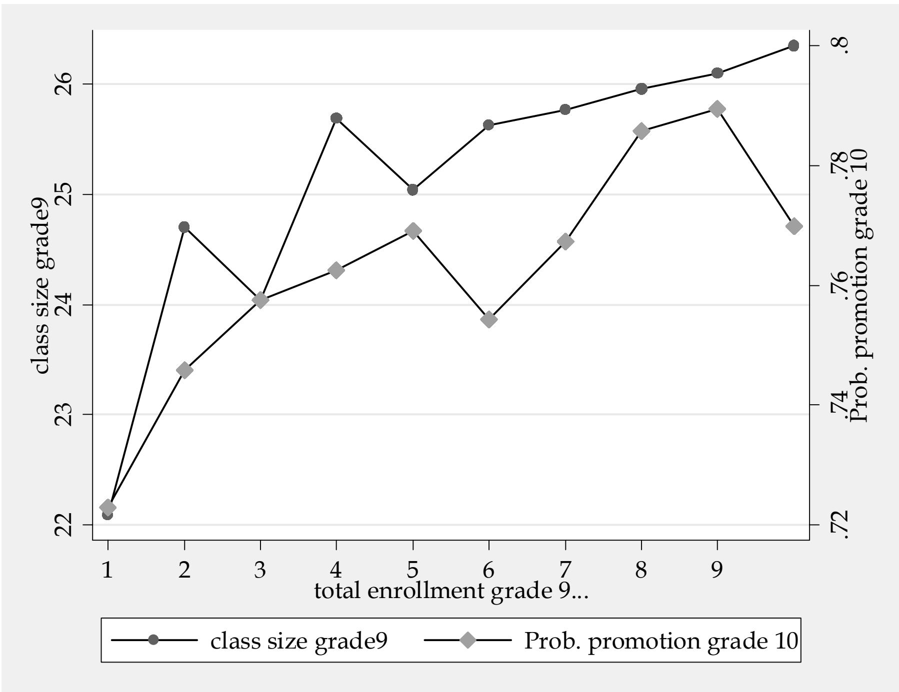
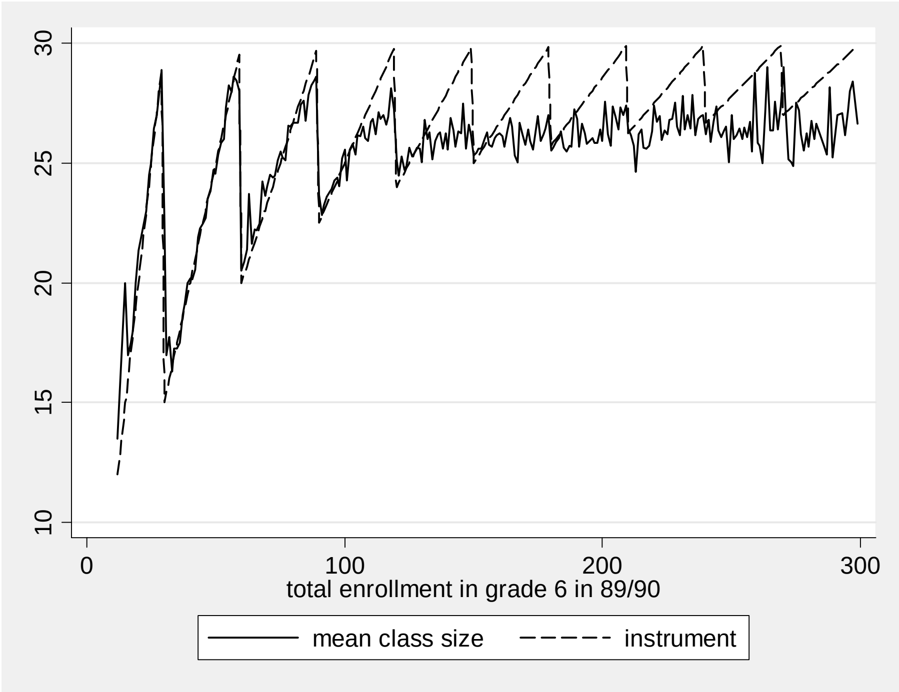


FIGURE 5 :

a. With all observations



b. With screened sample

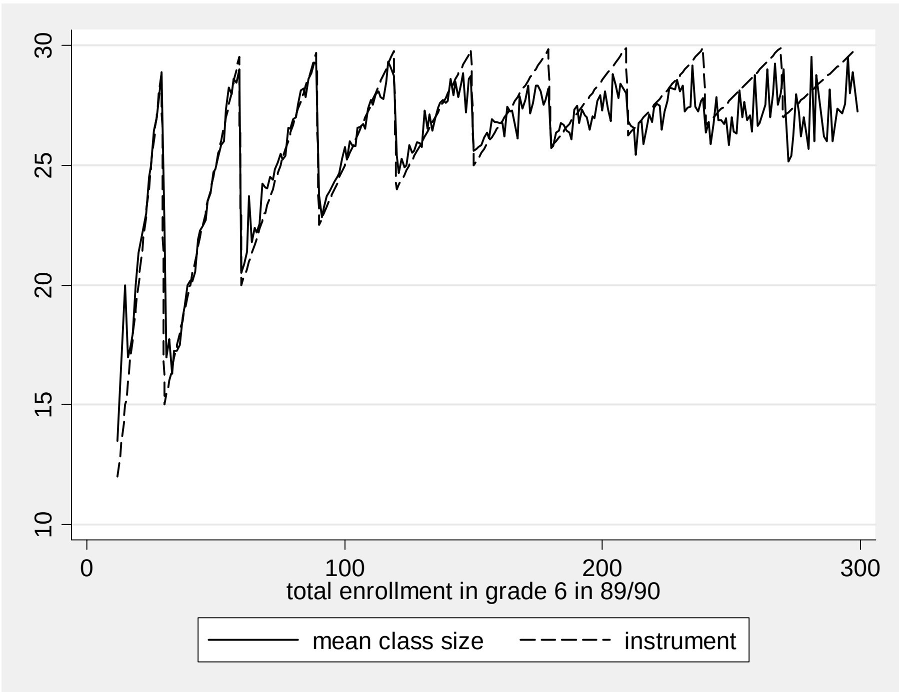
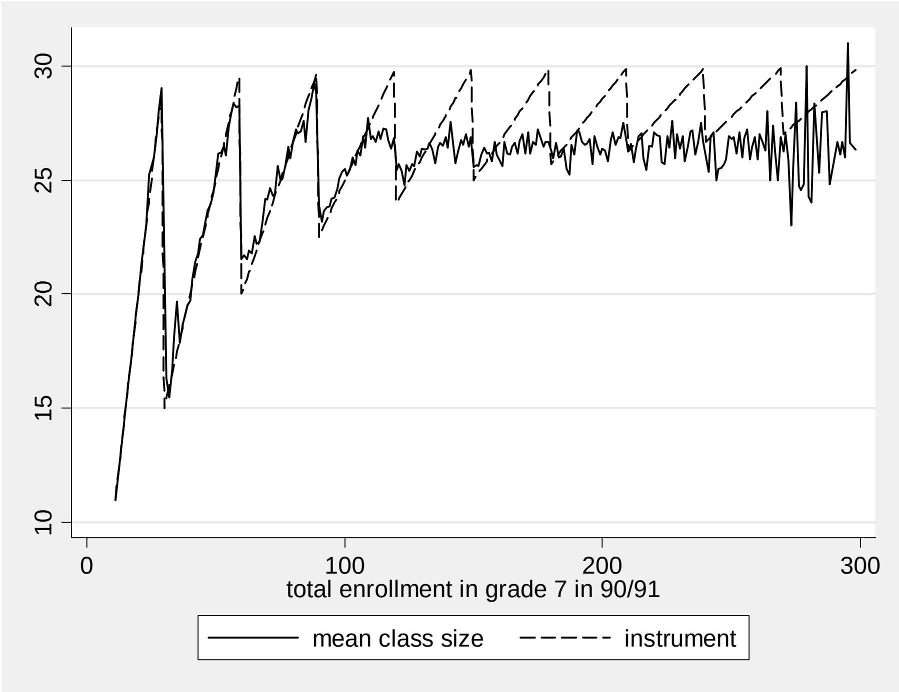


FIGURE 6 :

a. With all observations



b. With screened sample

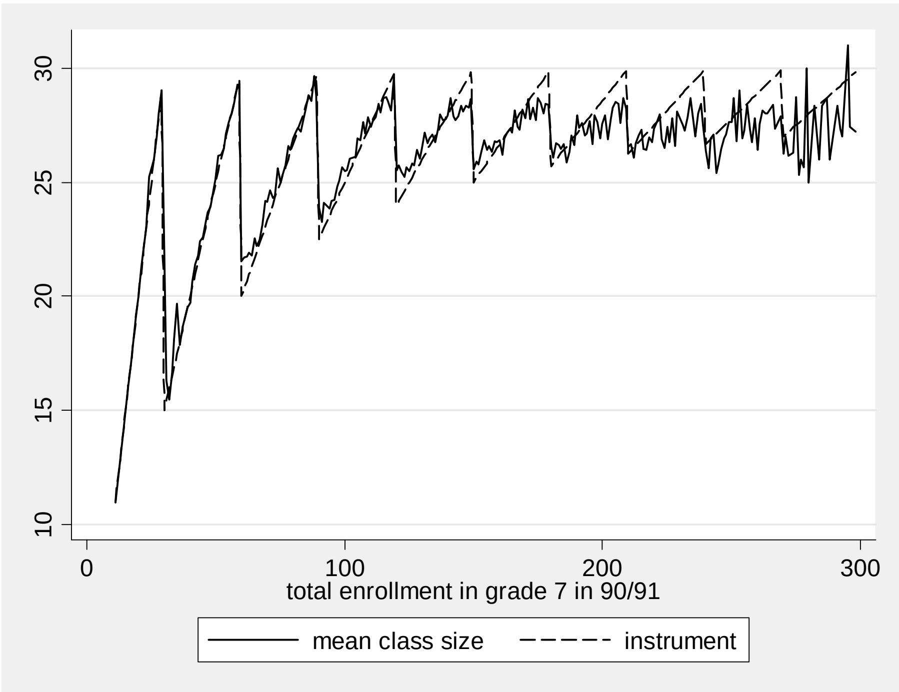
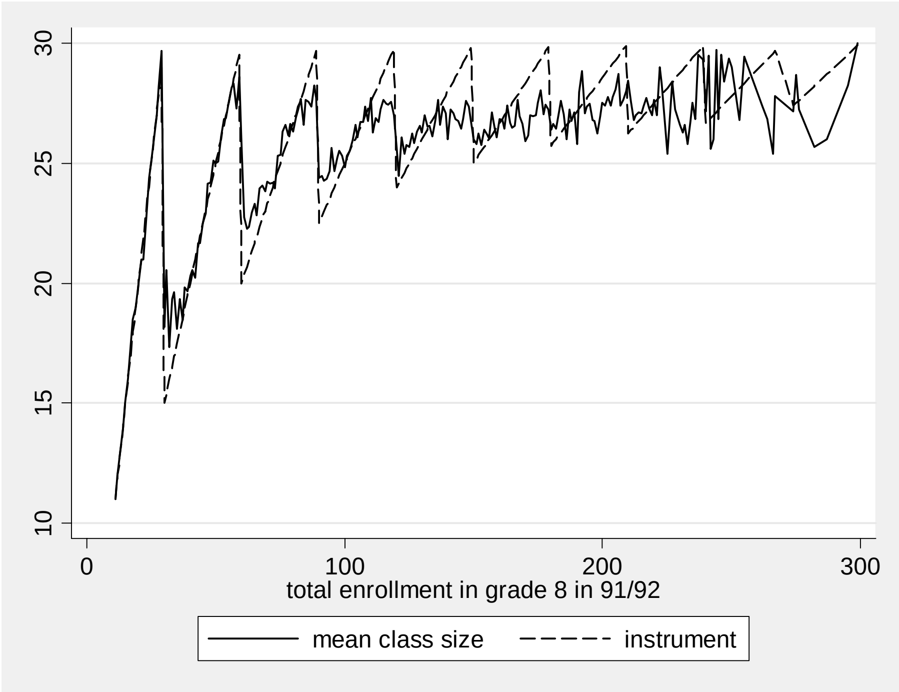


FIGURE 7 :

a. With all observations



b. With screened sample

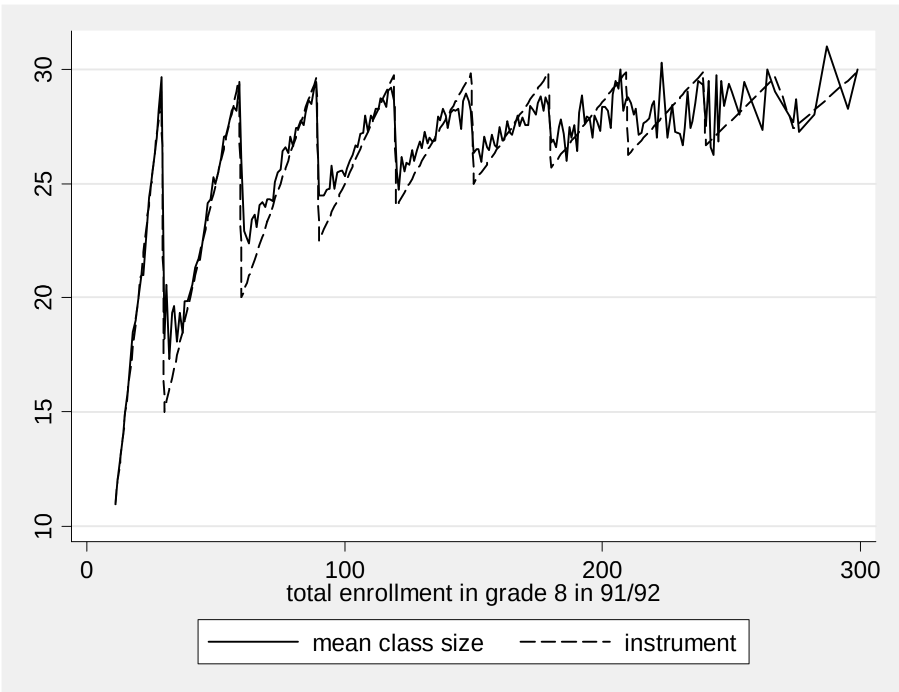


FIGURE 8 :

a. With all observations

b. With screened sample