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

Robert J. Gary-Bobo**

Mohamed-Badrane Mahjoub***

5 May 2009

Abstract

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. A 10 student reduction of class-size would put the child of a blue collar on an equal footing with the child of an educated professional, in grade 6. 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 actual class size and total grade enrollment in our exceptional data set. We control for many family background variables and entry test grades. 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. The simultaneous equation model allows for estimation of a general covariance structure of the error terms affecting latent student-performance and class-size equations, which sheds light on the endogeneity of class-size.

* We thank Josh Angrist, Thomas Brodaty, Olivier Donni, Denis Fougère, Eric Hanushek, Thierry Magnac, Eric Maurin, Thomas Piketty, Jean-Marc Robin and Isabelle Robin for their help and comments. We also thank Francis Kramarz for having made some crucial data files available to us.

** CREST-INSEE, 15 boulevard Gabriel Peri, 92245 Malakoff, France, Paris School of Economics and CEPR. Email: garybobo@univ-paris1.fr

*** Paris School of Economics, Université Paris 1 Panthéon-Sorbonne, 106, boulevard de l'Hôpital, 75647 Paris cedex 13, France. Email: gbadrane@univ-paris1.fr

1. Introduction

It seems that the debate about school resources, school quality and class-size effects has not lost of its intensity. In the recent years, research on class-size effects 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 (see *e.g.*, Finn and Achilles (1990), Krueger (1999)), and the rise of Instrumental Variables estimation techniques, applied to richer data sets. A wealth of new results have been produced by randomized evaluations of education policy in developing countries; see *e.g.*, Banerjee *et al.* (2007). 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, if the magnitude of social costs is taken into consideration, would not warrant a general policy of further class-size reductions. On this line of argument, see the important surveys of Hanushek (2002), (2003). These authors emphasize teachers' incentives and accountability (see for instance Rivkin *et al.* (2005)). On the other side, we find scholars for whom the evidence that reduced class-size improves educational achievement and life chances is ample enough, , at least for disadvantaged children. They emphasize the social benefits of class-size reduction programs, in particular by paying attention to their additional side-effects, like reductions in crime. On this side, see the vigorous synthesis of Krueger (2002), (2003), (see also Card and Krueger (1996)). A more agnostic point of view is expressed in the work of Carneiro and Heckman (2003), Cunha and Heckman (2007); these authors focus on the heterogeneity of individuals, the fact that the ability and educational achievement of children result from 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 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. For some evidence of this phenomenon in the US, see Boozer and Rouse (2001). The estimation bias caused by unobservable sorting of weak students in smaller classes is mitigated by adding controls in the regression of test scores on class size. But the likely effect of better controls is to drive the coefficient of class size towards zero. Instrumental variable estimation is then a way of dealing with endogeneity and of uncovering the true value of the class-size coefficients. The problem is to find a valid statistical instrument, a source of variation of class size that is not correlated with unobserved factors of student performance.

Literature

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 triggered by population variation, when local (i.e., district) class-opening thresholds are reached.

Angrist and Lavy (1999) find a negative and significant effect of class-size on students' test scores. Their striking results are obtained with Israeli data, and with the help of a "theoretical class-size" variable, that 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 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. Theoretical class-size being a discontinuous function of grade enrollment, the discontinuous jumps of class size identify the class-size effect.

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 the regression's error term can reasonably be assumed without correlation with the instrument. Enrollment endogeneity problems are likely to be important if total enrollment in a given school reflects parental school-choice strategies based on unobserved school characteristics such as favorable "peer groups" or the presence of better 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 with sufficiently rich data sets, many effects of this kind can in principle be controlled for (see for instance Dearden *et al.* (2002), Dustmann *et al.* (2003)). The results obtained by IV estimation procedures also bear the risk of reflecting local treatment effects, *i.e.*, the effect of class-size reductions on certain sub-groups of the population under study, instead of the true average effect, but in any case, they will provide more interesting insights than naive OLS estimates.

There are other approaches to IV estimation of class-size effects, see *e.g.*, Akerhielm (1995), Case and Deaton (1999), Boozer and Rouse (2001), Dobbelsteen *et al.* (2002), 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. Leuven, Oosterbeek and Rønning (2008) apply the method to another data set from Norway and cannot reject that the class-size effect is equal to zero. Browning and Heinesen (2007) apply Angrist-Lavy's method to Danish data and explore the impact of class size on alternative outcomes: years of education and completion of upper secondary education. They find modest, marginally significant but negative effects of class-size. The study of Danish data is taken over by Bingley, Jensen and Walker (2005); they compare siblings who experienced different class sizes to difference out some unobserved family background effects; they find that class-size reductions in junior high school significantly increase the student's years of education, but the effects are too small to justify their public costs. Urquiola (2006) focuses on schools in rural Bolivia with only 1 or 2 classes and finds a negative effect of class size on test scores. Urquiola and Verhoogen (2009) raise some methodological difficulties associated with Angrist-Lavy's regression discontinuity approach. They develop a model of school choice in a competitive market with private, for-profit schools, which is applied to Chilean data. Higher-income households sort into higher quality schools and schools may increase tuition to avoid having to open an additional class. It follows that the assumptions of the regression discontinuity design may be violated if there is a substantial amount of student sorting on both sides of the discontinuities. In spite of the fact that these phenomena are unlikely to be important in the highly regulated public educational systems of European countries like France, we will check that our data is indeed exempt of such problems.

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.* (2009), 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. 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.* (2009) 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, but only as a point of comparison.

Piketty's (2004) results show that decreasing class size by 10 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 results lead him to conclude that a reduction of class size in the primary schools of disadvantaged areas would substantially reduce the test-score gap with the average student. Bressoux *et al.* (2009) also find negative class-size effects on mathematics test scores: a 10-student reduction in class-size would increase the average test score by 4.4 points. This figure amounts to 30% of the standard deviation of test scores. Using junior high-school GPAs 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 GPA by 2.16, which represents 20% of the standard deviation (using the 1995 panel).

Our contribution

We use another panel of the Ministry of Education, the junior high-school Panel started in 1989, and study grade promotions instead of test scores. These data have some remarkable features. In the present paper, our sample contains more than 16,000 observations of individuals enrolled in French public junior high schools, scattered on the whole French territory. An important advantage of this panel, when compared with Angrist and Lavy's (1999) data, is that it provides student-level instead of class-level observations (i.e., instead of class averages); another advantage is the higher number of available control variables; in addition, we observe *actual* class sizes, not simply the average. 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' staff meetings (i.e., the *conseils de classe*), at the end of each school year. In essence, these staff meetings base decisions on the student's grade-point average (hereafter GPA) at the end of each year, and decide wether to promote, to hold back, or to "steer" the student towards vocational education. Students with a GPA above a certain threshold are promoted; students with a low record are "steered"; students with a mediocre record repeat the grade, if the teachers' committee thinks that they can benefit from the repetition. It is therefore legitimate to use an Ordered

Probit model in which the student's GPA is the latent variable.

To be more precise, we use class size and a whole list of control variables to explain the probabilities of being promoted, of repeating a grade, or of leaving general education for a vocational program (*i.e.*, steering). We estimate these probabilities as functions of observed class size and control variables. Angrist-Lavy's theoretical class size is used as an instrument for observed class size. We control for family background, entry test scores in Math and French and total school enrollment.

The model is estimated by Maximum Likelihood. We jointly estimate the probabilities of promotion, retention, and tracking, in grades 6 to 9 (*i.e.*, 4 transitions), and 4 auxiliary class-size equations, for a cohort of 16,000 students who were enrolled in grade 6 in September 1989. Estimation is made under the assumption that student-performance error terms are conditionally independent, with conditioning on class-size. Thus, grade transitions are not assumed statistically independent, but the conditional independence restriction allows us to write the probability of a given student's grade-transition record as a product of probabilities. As a by-product, we obtain the 8-dimensional (unconditional) covariance matrix of error terms which sheds light on some dynamic aspects of the class-size endogeneity problem. Year-by-year estimation of grade transition models would not allow for a rigorous study of these correlations. Class-size error terms are not independent and are not independent of student-performance error-terms.

To convince ourselves (and the reader) that the results are robust, we have made a number of robustness checks, and present some estimates obtained by means of standard linear IV methods. These estimates fully confirm the results obtained with the nonlinear, simultaneous-equations model.

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. A possible (conservative) summary of this finding is that a 1 student reduction of class size increases the probability of promotion to the next grade by 0.3%. In other words, the probability of being promoted would be raised from $X\%$ to $X + 3\%$ for a 10 student reduction of class size. Another way of presenting the results would be to say that a 10 student reduction of class-size (thus a reduction of more than 30%) would put the child of a blue collar on an equal footing with the child of a highly educated professional, in grade 6. So, the effects of class size are modest, but not negligible. In contrast, class-size effects are not significant in grades 8 and 9. The intensity of these effects seems to fade away when grade increases: class size is less important for more advanced students.

We also find a pattern of cross-correlations between student-performance and class-size in different years. These correlations reflect the impact of unobserved variables; they are typically significant and positive. This confirms the intuition that unobserved factors increase student performance and class size at the same time. The strong correlation of class-size residuals indicates intertemporal persistence. It seems 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. To the best of our knowledge, this type of approach has not been used in the school-resource literature.

In the following, Section 2 is devoted to a description of the institutional context and of the data; Section 3 presents the model and estimation method; Section 4 presents results obtained by means of linear IV methods and some robustness checks. Finally, Section 5 presents the results obtained with the nonlinear, simultaneous-equations model.

2. Institutional Context and Data

In the late 1980s, the French secondary education system is highly centralized and highly regulated. To manage the educational system, the country has been divided into 26 regions, called *Académies*. Each of these regions is headed by a kind of governor general of educational affairs, who directly represents the central government, called the *recteur*. The rector himself is helped by lieutenants called inspectors ((*i.e., Inspecteurs d'académie*), each in charge of a "county" (*i.e.*, a French *département*), and high-school principals. In 1989, the French primary and secondary education is still characterized by a rigid district system. Each public high-school has a monopoly in its official catchment area. There are few exceptions to the rule of student allocation on the basis of residence, managed by the local inspectors. For the purpose of our research, it is interesting to note that, according to French law, there is a counterpart to rigid zoning: inspectors *must* welcome any resident of a catchment area in the corresponding public high-school. Inspectors also determine the maximum number of admissions in each high school placed under their supervision[1].

Private schools are the only really important alternative. French private high schools have a stable market share around 20% since more than 20 years. These schools are highly regulated and 90% are Catholic. They work under a contract with the State. The teacher's qualifications are the same as in the public sector, the curriculum is national and the teachers are paid by the government! The main difference between the two types of schools is that the proportion of students coming from relatively richer families is higher in the private sector, but not considerably higher. Given that fees are not high, the French private high schools are not particularly aristocratic. On the contrary, some public schools — the traditional urban *lycées* for instance — are more prestigious than the ordinary private junior high schools. The social stratification of schools induced by urban stratification is much more important than the selection due to the public-private division. For the present research, the proportion of students leaving the public sector to study in a private school (or coming to public schools from private primary schools) could be an important element. It happens that, in France, most of the switchover from public to private (or from private to public) institutions takes place at the beginning of junior high-school (*i.e.,* grade 6), and at the beginning of senior high school (*i.e.,* grade 10)[2]. A very small proportion of students, *i.e.* 3%, changes for a private school in our data.

The education system is also highly centralized because, if high-school buildings are now funded and maintained by local governments, the teachers and principals are civil servants paid by the central government. The career system is national and uniform, wages being mainly determined by seniority, 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 compensation. Young French high-school teachers form

[1] See for instance van Zanten and Obin (2008).

[2] See, for instance, DEP (2003).

a giant waiting line, expecting to be appointed to “better” schools, in better places. In the late 80s and early 90s, the matching of teachers to schools used to be the result of a central bureaucratic process, in which teachers’ unions played a major role. The matching process is nowadays somewhat decentralized within administrative regions. 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 also likely to be located in a rich city center[3], 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 older teachers tend to cluster in the “good schools”.

We should now make sure that the notion of class size has a meaning in French junior high schools. Class size is a well-determined notion because students of a given grade are typically divided into groups of 25-30 students who stay together during an entire school year. Different teachers, specialized in French, Math, English, History-and-Geography, Natural Sciences, and so on, lecture in front of the same group. These groups are sometimes divided for elective courses (for instance, for the teaching of a second foreign language), but classes are not fictive objects.

Now, when the count of students rises above a certain threshold, in a given school and in a given grade, the principal can, and typically will open a new class and ask his rector for the means of doing it. If the principal doesn’t do it, the school parents’ associations and the teachers’ unions will protest vehemently. Given that the rector is required by law to welcome any new resident of the catchment area, he or she will discuss with the principal the appointment of new teachers and release the necessary additional funds. The opening of a new class is a complex operation because a whole team of specialized teachers is required. The principal will pragmatically combine different tools to open her class. Teachers are typically in charge of several classes and spend a few hours per week with each class. So, the principal can require overtime hours from the personnel already

[3] In France, and in stark contrast with the US, 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.

allocated to her school — but there are strict limits. She can also share a new teacher with a neighboring school. In addition, each *académie* has a "reserve army force" of teachers[4], not tied to a particular school, that can be employed to fill the gaps during several months if needed. This means that some appropriate management tools permit a more or less systematic application of class-opening rules. The norm is currently 30 students per class in French junior high schools. It has been lower in the recent past and a 28 students-per-class threshold is suitable for our data from 1989-1994. Even if there are various exceptions and even if a principal can lobby her rector more or less efficiently, this rule is a prevailing norm that structures the allocation of resources in the entire school system. This is the main reason why Angrist-Lavy's approach works well with French data. However, it seems extremely unlikely that many parents can easily forecast class size in their children's school and outmaneuver the administration. As discussed in the introduction, the endogeneity of class size is mainly driven by the fact that weak students are allocated to smaller classes.

Our estimates are based on 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 (*Direction de l'Evaluation et de la Prospective*), and another administrative source from the same Ministry, the *établissements* files, providing yearly enrollment data and other information on high schools. *Panel 89* is a sample of high-school students observed during several consecutive years. In the following, we use observations made during the first 5 years of high school (including the French junior high-school years). One fifth of all junior high schools are sampled; then, the grade 6 students born on the 17th of each month are chosen[5]. The sampled high schools' principals have been asked to fill forms about each sampled student, providing a number of family-background characteristics and recording the grade or program attended by the student each year. For each student, we notably know: the gender, parental occupation and education, the number of siblings, birth order, the age at grade 6 entry, the month of birth and entry test scores. Grade 6 students take a test in French and Math at the beginning of the school year. The grades are

[4] The TZR *i.e., titulaires sur zone de remplacement.*

[5] Grade 6 is our translation of the French *classe de sixième*; grade 7 is the translation of *classe de cinquième*, and so on.

reported as A, B, C or D in both disciplines and will be used as controls in all equations.

Crucially, for each year of the observation period, we know the student's actual class-size, the grade followed, we know if the student repeated a grade or if he or she was "steered" towards vocational high schools at the end of the school year. In French high schools, "steering" is quite often an euphemistic expression for the principal's decision to expel the student from general education programs, essentially because the student's performances are not judged good enough by the teaching staff— these decisions could as well be called "tracking". As mentioned in the introduction, these decisions are mainly based on the student's GPA.

The administrative files called *établissements* files) give total grade enrollment for each grade and each year in each public high school. Enrollment observations are beginning-of-the-year (i.e., September) figures for each school year, starting in 1987. Another important indication is the so-called 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. ZEP classification aims at mitigating a number of social handicaps[6].

Each public-sector high school has an identifier, the (*code d'établissement*), that can be used to link student observations in *Panel 1989* to corresponding high-school enrollment figures in the administrative files. The panel contains observations of students registered in private sector schools, but we do not observe total grade enrollment in these schools. We therefore selected the sub-sample of students admitted in the public sector, in grade 6, in 1989 and who never left for the private sector during the 4 subsequent years. 19,100 out of 24,710 panel students who entered grade 6 in 1989 are registered in the public sector. The possible problems caused by this type of selection are limited because observed switchover from public to private (or private to public) sectors are mostly clustered at grade 6 entry or at the entry of senior high schools, as discussed above. Few students leave the public sector in grades 7, 8 and 9. Indeed, only 500 panel students have been lost because they opted out of the private sector during the observed junior high-school years, *i.e;*, between

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

1989 and 1994.

There are also some observations lost because of missing data. In particular, we lost 2,500 observations because of missing data in the linked high-school file. This is mainly due to the fact that we need grade enrollment to be observed during several consecutive years for each student. There is no reason why these missing administrative records should be correlated with the characteristics of students. In comparison, very few observations (*i.e.*, 36) had missing entry-test grades. To sum up, we finally obtained a sample with 16,055 observations.

3. The Model

We model the probabilities of promotion, grade retention and "steering" (or tracking) at the end of each school-year as a discrete 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 of states $G = \{6, 7, 8, 9, 10, v\}$, where v stands for vocational school. Promotion to grade 10 in general-education high schools is 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 g to grade 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.

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

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

where X_i is vector of observed characteristics, β_g 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 vary with g, (but not with t 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,

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

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.

The conditional probability of being promoted is defined as follows,

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

the conditional probability of grade retention is defined as,

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

and the conditional probability of steering is defined as,

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

where γ_g and δ_g are ordered thresholds or "cuts", with $\gamma_g < \delta_g$. 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,

$$n_{it} = \sum_{g=6}^{g=9} (a_g n^*_{it} + Z_i b_g) \mathbf{1}_{itg} + \epsilon_{it}, \tag{9}$$

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_g, b_g) are parameters to 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 multivariate normal distribution with mean 0 and (unconditional) covariance matrix $\Omega_{\epsilon\epsilon}$.

Angrist-Lavy's instrument 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 $28q$, where q is an integer, and to minimize class-size differences. We take $n = 28$ as the French norm in 1989. 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,

$$\kappa^*_{it} = int\left[\frac{N_{it}-1}{28}\right] + 1, \tag{10}$$

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 29 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 14.5, if there are 57 students, there are 3 classes with 19 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. Total enrollment fluctuations, driven by local demographic shocks, cause discontinuous changes in class-size when combined with the class opening rules. The discontinuous jumps of class-size due to applications of the administrative rule provide a quasi-experimental setting which identifies the impact of class-size. Angrist-Lavy's instrument and total grade enrollment are more likely to be exogenous if a sufficiently rich set of controls is introduced in the student-performance equation to capture characteristics of students and schools. It is in particular important to use total school size as a control, because a number of unobserved factors may be correlated with enrollment N and affect student performance.

To complete the specification of our model, we need to define the covariance structure of endogenous variables. Given that our data have a panel structure, it is of course important to allow for correlation of class-size error terms ϵ_{it} across time, and for correlation of these class-size errors with the non-observable performance error 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,

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

where $\Omega^T_{\nu\epsilon}$ is the transpose 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,

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

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. We set $Var(\nu_{it} \mid \epsilon) = 1$ as identification constraints, as usual in this type of latent-index specification. This is equivalent to saying that $Cov(\nu \mid \epsilon)$ is the identity matrix, or

$$\Omega_{\nu\nu} - \Omega_{\nu\epsilon}\Omega_{\epsilon\epsilon}^{-1}\Omega_{\nu\epsilon}^{T} = I. \tag{14}$$

We estimate this model by maximum likelihood, under the normality assumption for all perturbations, and under the conditional independence constraints imposed by (14) on Ω. Identification does not rely on functional form or normality in an essential way, even if normality is used to estimate the model. In fact, we have estimated all parameters, except the covariances, by means of standard linear IV methods, as a robustness check. The model naturally identifies all the parameters, a, b, α, β, γ, δ with the help of only one exclusion and the usual variance normalization. The matrix $\Omega_{\nu\nu}$ is then computed, using (14) above, with the help of covariance parameters $\Omega_{\epsilon\epsilon}$, $\Omega_{\nu\epsilon}$. In fact, Ω is expressed using its Cholesky decomposition, and we find a way of expressing the constraints (14) in terms of this decomposition. See the appendix for further technical details and a derivation of the likelihood function.

4. Standard IV Estimates. Robustness Checks

We start with a preliminary analysis of the data. Figures 1-4 plot the theoretical class size $n^* = N[int((N-1)/28)+1]^{-1}$ 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 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. 1-4. The figures show that Angrist and Lavy's instrument works well, and particularly when total grade enrollment is below 120.

Table 1 displays elementary descriptive statistics for a number of variables used in the regressions. The social background of students is captured by means of indicators of the father's occupation. The professionals' category includes the executives, doctors, lawyers, engineers and teachers. The rightmost column gives the means (*i.e.*, proportions) of the observed characteristics in the selected sub-sample. It is easy to compare them with the proportions in *Panel 89*, that also includes private-sector students. The family-background distributions are not very different. The grades are slightly better if we exclude the private sector. This might be due to the fact that some relatively weaker students have been registered in private high schools at the end of primary school. It is likely that many of the parents of relatively weaker students opt out of public schools at grade 6 entry to make sure that their children will benefit from better working conditions, and this would be reflected in the entry test grades. Table 2 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 3 shows the distribution of individual school records $s_i = (s_{i2}, ..., s_{i5})$. We see that 47.5% of the students only have the normal record $(7, 8, 9, 10)$. We see, for instance, that 27 students only repeated grade 8 and have finally been steered. Once steered towards vocational programs, there is no further information on a student's record.

Table 4 reports naive OLS regressions of the indicator of promotion to the next grade on class size and a list of controls. The impact of class size is either zero, or significant, but positive. The controls, and particularly the father's occupation and entry test grades do matter a lot. Note that each regression is based on a sub-sample of students enrolled in grade $6 + t - 1$ in year $t = 1, 2, 3, 4$. These sub-samples are made of selected students since by definition: they attended the same grade in the same year; they never repeated a grade; they survived the steering process. The same grade-based sub-samples are used below to compute IV estimates of the class-size coefficients in the same regressions.

Table 5 gives the result of the first-stage regression of actual class size n on Angrist-Lavy's instrument n^*, total grade-enrollment N, N^2, total school-enrollment and total school-enrollment squared. For a given student-year (i, t), total school-enrollment is the total sum of enrollment in grades 6, 7, 8 and 9, in i's junior high school, during year t. The first sub-column of each column in Table 5 gives the estimated coefficient, and the second sub-column gives the corresponding t statistic, in parentheses. There are separate first-stage regressions for each grade and we see at the top of the table that Angrist-Lavy's instrument is very significant. This is true and at the same time, total grade and total school enrollment are also significant. We will check below that Angrist-Lavy's variable and total grade-enrollment are jointly very strong instruments. In rural areas presumably, the children of farmers are in significantly smaller classes. But children from more educated backgrounds attend larger classes on average, since all the coefficients on father occupation dummies are negative, with a nice ranking of the estimated values, and professionals are the reference category.

Table 6 shows the second-stage 2SLS estimates of the coefficient on class size in the linear regression of the promotion decision. We now control for total school-enrollment and total school-enrollment squared. The top row of Table 6 gives the impact of class size on students who never repeated a grade before, in each grade. The coefficients on class size are clearly significant and negative in grade 6 and 7, but non-significant in grades 8 and 9. An additional student decreases the probability of promotion by less than 2%. We can check at this point that a 4 student reduction of class size would put the child of a blue collar on a equal footing with the child of a professional, in grade 7. So, the effects of class size are moderate, but non-negligible. These results are an anticipation of the complete model's maximum likelihood estimations.

Robustness checks

Table 7 explores how results vary when we vary the class-opening threshold τ, using the exact same specification. We report first-stage and 2SLS estimates for values of the threshold between 25 and 35, in grades 6 and 7. Extreme values 25 or 35 clearly do not work well. A threshold of 28 or 30 seems to fit the data well, but the best results are unambiguously obtained with $\tau = 28$, on all fronts. The F-test for the strength of instruments has very high values, greater than 120, far above the prevailing rule of thumb used to reject weak instruments.

Table 8 compares the results obtained by means of 2SLS with alternative specifications of the first-stage regression. Column A recalls the OLS results in Grade 6 and 7. Column E recalls the results of our benchmark model. Column B shows the results obtained when Angrist-Lavy's variable is the only instrument: we lose the significance of class size in grade 6 (but keep the negative sign). It follows that total grade enrollment plays a specific role (recall that total school enrollment is a control in all equations). Column C shows that the estimated impacts of class size are stronger if the only instruments are total grade-enrollment and total grade-enrollment squared. Given these results, fluctuations of the ratio of total grade to total school enrollment might be an important identifying source of variation of class size. Column D, in which the ratio of grade enrollment to school enrollment is the only instrument confirms this intuition. In column F, we add Angrist-Lavy's variable as a second instrument and obtain excellent results that are very close to the benchmark in column E.

We would like to understand better if total grade enrollment is endogenous. There is a risk that changes in this variable be correlated with unobserved student characteristics, instead of being driven by random demographic shocks. Unobserved characteristics have a good chance of being correlated with family background and entry test grades, so we regress total grade enrollment on controls. This is done in Table 9. Model 1 is a simple regression of total grade enrollment on controls. We see that being the son of a farmer significantly reduces class size (rural areas are correlated with smaller classes). But if we add total school-enrollment as a regressor, as in Model 2, the R^2 jumps upward to 90% and all controls lose their significance. We conclude that, in essence, total grade enrollment is equal to one fourth of total school enrollment plus some noise. In this context, total grade-enrollment is likely to be a valid instrument.

In Table 10, we explore the potential of an alternative instrument based on variations of total school enrollment only. The instrument is now the difference of total school-enrollment with the mean total school-enrollment over 6 years (1987-1992). Estimates are obtained with a slightly smaller sample because of a few missing enrollment observations in

1987 and 1988. Although the quality of the first stage presented in Table 10 is somewhat lower as that of the benchmark model, we still obtain negative and significant impacts of class size in grades 6 and 7, combined with no effects in grades 8 and 9. Class size coefficients are higher in absolute value, with the same order of magnitude as before. This set of results shows the robustness of our findings.

We finally examine the possibility of sorting on each side of the class size discontinuities. Table 11 compares some descriptive statistics in the sub-sample of observations for which total grade-enrollment is between $28k-4$ and $28k$ and the sub-sample for which total grade-enrollment is between $28k$ and $28k+4$, where k is an integer. The small differences found are not significant, so, we conclude that such self-selection problems are not present in our data.

5. Maximum Likelihood Estimates of the Simultaneous-Equations Model

Table 12 presents the maximum likelihood estimates for the complete simultaneous equations model described in Section 3 above. The top of Table 12 gives the α_{tg} coefficients with their corresponding t-statistics, *i.e.*, the estimate of matrix $\mathbf{A}$ defined by (5) above. The coefficients are significant in the grade 6 and grade 7 columns only. The last lines of Table 12's upper panel are the ordered probit cuts, γ_g and δ_g, defined by (6) and (7) above. These cuts are precisely estimated. It is remarkable to find that the δs and γs increase with grade g, we get $\delta_6 < \delta_7 < \delta_8 < \delta_9$ and $\gamma_7 < \gamma_8 < \gamma_9$. This reflects the increasing demands of the teaching staff. We computed marginal effects of an additional student on the probability of promotion to the next grade. The marginal impact is -0.004 in grade 6 and -0.003 in grade 7. A 10-student reduction in class size would yield an increase of 3 to 4% in the probability of promotion. The class-size coefficients must be compared with some of the β_g coefficients listed in Table 12. For instance, a 10-student reduction in class size would put the child of a blue collar on an equal footing with the child of a professional in grade 6, *ceteris paribus*. The class-size reductions needed to make up for the negative effects of a low entry-test grade are much higher than 10. 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, or simply that they become weaker as students become more mature[8].

Table 13 gives the estimated entries of the covariance matrix Ω and the associated t-statistics, estimated under constraint (14). These error terms exhibit a significant and substantial degree of positive correlation, as can be seen on Table 14. The terms corresponding to $\Omega_{\epsilon\epsilon}$, (the top 4 lines of Table 13 or 14) show that a student belonging to a large class in year $t = 1$ is more likely to belong to a large class in years $t \geq 2$. This confirms our intuition that some persistent unobservable individual characteristics are class-size determinants. 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 class 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 14 also 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 and almost always significant. So, ν and ϵ are not independent random vectors. It is easy to see that $corr(\nu_u, \epsilon_t)$ tends to decrease when the distance between t and u increases. There is some "memory" or persistence of the effects of past class sizes. More generally, the unobserved components of class-size are positively correlated with the unobserved factors of future (or past) student performance. Finally, the estimated unconditional $\Omega_{\nu\nu}$ sub-matrix has small off-diagonal terms; the correlation terms $corr(\nu_t, \nu_u)$, are all below 5%.

[8] An explanation is the specific character of grade 8 (*i.e., classe de quatrième*) in French high schools. According to some teachers, there are few repetitions of grade 8, the important exam being at the end of grade 9. Teachers 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, but this discipline problem becomes less acute in higher grades.

6. Conclusion

A substantial amount of unobserved remedial education by means of smaller classes is taking place and ordinary least squares estimates of class-size coefficients are biased upwards. We find 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 (*i.e.*, grades 6 and 7). The impact of class size on the probability of promotion to the next grade is negative in grades 6 and 7 but seems to vanish in higher grades. These results are obtained by means of an IV estimation strategy which exploits the changes in class-size induced by the combined effects of total grade-enrollment fluctuations and class-opening rules. We first obtain these results with the help of robust linear econometric methods. We then specify a simultaneous-equations model which takes care of class-size endogeneity problems and we jointly estimate the probabilities of 4 transitions (from grade 6 to grade 10) and 4 class-size equations by Maximum Likelihood. Estimates are obtained under normality and under the assumption that grade transitions are conditionally independent (viz., independent, conditional on the entire class-size history). The instrument for class size based on Maimonide's rule performs very well in "first-stage" class-size regressions. The estimated 8-dimensional covariance matrix of errors reflects the presence of positive correlations between class sizes in different years, as well as positive correlations between class-size error-terms and the error terms of student-performance equations. This confirms that, in spite of the presence of family-background variables and other controls, some student characteristics are observed by the teachers, but not by the econometrician and play a role in the matching of students to classes.

7. References

Akerhielm, Karen (1995), "Does Class Size Matter?", *Economics of Education Review*, 14, 229-241.

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

Banerjee, Ahbijit V., Cole, Shawn, Duflo, Esther and Leigh Linden (2007), "Remedying Education: Evidence from Two Randomized Experiments in India", *Quarterly Journal of Economics*,122, 1235-1264.

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

Bingley, Paul, Jensen, Vibeke M., and Ian Walker (2005),"The Effects of School class Size on Length of Post-Compulsory Education: Some cost-benefit Analysis", *IZA Discussion paper* No. 1605, Bonn, Germany.

Browning, Martin, and Eskil Heinesen (2007), "Class Size, Teacher Hours and Educational Attainment", *Scandinavian Journal of Economics*, 109, 415-438.

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

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

Bressoux, Pascal, Kramarz, Francis, and Corinne Prost (2009),"Teachers' Training, Class Size, and Students' Outcomes: Learning from Administrative Forecasting Mistakes", *Economic Journal*, 119, 540-561.

Card, David, and Alan B. 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, Pedro, and, James J. Heckman (2003), "Human Capital Policy", in Heckman and Krueger, *Inequality in America; What Role for Human Capital Policies?*, MIT Press, Cambridge, Massachusetts.

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

Cunha, Flavio and James Heckman (2007), "The Technology of Skill Formation", *American Economic Review, Papers and proceedings of the A.E.A.*, 97, 31-47.

Dearden, Lorraine M., Ferri, Javier, and Costas Meghir (2002), "The Effect of School Quality on Educational Attainment and Wages", *Review of Economics and Statistics*, 84, 1-20.

Direction de l'Evaluation et de la Prospective (2003), "Public-Privé: quelles différences?", *Education et Formations*, 66, 167-175.

Dobbelsteen, Simone, Levin, Jesse, and Hessel 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, Christian, Rajah, Najma, and Arthur H. van Soest (2003), "Class Size, Education and Wages", *Economic Journal*, 113, F99-F120.

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

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

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

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

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

Leuven Edwin, Oosterbeek, Hessel, and Marte Rønning (2008), "Quasi-Experimental Estimates of the Effect of Class Size on Achievement in Norway" *Scandinavian Journal of Economics*, 110, 663-693.

Piketty, Thomas (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, Thomas, and Mathieu 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, Steven G., Hanushek, Eric A., and John F. Kain (2005), "Teachers, Schools, and Academic Achievement", *Econometrica*, 73, 417-458.

Urquiola, Miguel (2006),"Identifying Class-Size effects in developing Countries: Evidence from Rural Bolivia", *Review of Economics and Statistics*,88,171-177.

Urquiola, Miguel, and Eric Verhoogen (2009), "Class-Size Caps, Sorting, and the Regression Discontinuity Design", *American Economic Review*, 99, 179-215.

van Zanten, Agnès, and Jean-Pierre Obin (2008), *La carte scolaire*, "Que sais-je?" no 3820, Presses Universitaires de France, Paris, France.

Vignoles, Anna, Levacic, Rosalind, Walker, James, Machin, Steve and David Reynolds, (2000), "The Relationship Between Resource Allocation and Pupil Attainment: A Review", *Center for Economics of Education*, London School of Economics, London, UK.

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

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

8. 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, for $t \geq 2$, we have $k(i) = t - 1$ if and only if $s_{it} = v$. Define $y_i = (y_{i1}, ..., y_{ik(i)})$ as a column vector and rewrite the latent performance equations (4) 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).

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

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

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,

$$f_{k(i)}(n_i) = (2\pi)^{-\frac{k(i)}{2}} (\det \Omega_{k(i)\epsilon\epsilon})^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(n_i - B_i(.))^T \Omega_{k(i)\epsilon\epsilon}^{-1} (n_i - B_i(.)) \right\} \tag{A4}$$

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

$$\nu_i = \Omega_{k\nu\epsilon} \Omega_{k\epsilon\epsilon}^{-1} \epsilon_i + \xi_i, \tag{A5a}$$

where $k = k(i)$, and

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

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

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

and for each ν_{it},

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

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,

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

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

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

and,

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

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

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

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

The covariance matrix Ω is estimated using its Cholesky decomposition, to ensure symmetry and positive definition. We have,

$$\begin{pmatrix} \Omega_{\epsilon\epsilon} & \Omega_{\nu\epsilon}^T \\ \Omega_{\nu\epsilon} & \Omega_{\nu\nu} \end{pmatrix} = \begin{pmatrix} U & 0 \\ V & W \end{pmatrix} \begin{pmatrix} U^T & V^T \\ 0 & W^T \end{pmatrix} = \begin{pmatrix} UU^T & UV^T \\ VU^T & VV^T + WW^T \end{pmatrix}. \qquad (A11)$$

Constraint (14) should be translated in terms of the $(4,4)$-blocks U, V and W. This yields,

$$VV^T + WW^T = I + VU^T(UU^T)^{-1}UV^T. \qquad (A12)$$

Using the property, $(UU^T)^{-1} = (U^T)^{-1}U^{-1}$, it is easy to check that $(A12)$ boils down to $I = WW^T$. It follows that the conditional independence constraint (14) can be expressed as follows:

$$\begin{pmatrix} \Omega_{\epsilon\epsilon} & \Omega_{\nu\epsilon}^T \\ \Omega_{\nu\epsilon} & \Omega_{\nu\nu} \end{pmatrix} = \begin{pmatrix} UU^T & UV^T \\ VU^T & VV^T + I \end{pmatrix}. \qquad (A13)$$

This shows that to estimate Ω by ML under the appropriate constraints, we in fact need to estimate only U, which is $(4,4)$ triangular, and the $(4,4)$ matrix V, the two blocks totalling $10 + 16 = 26$ parameters.

Figure 1

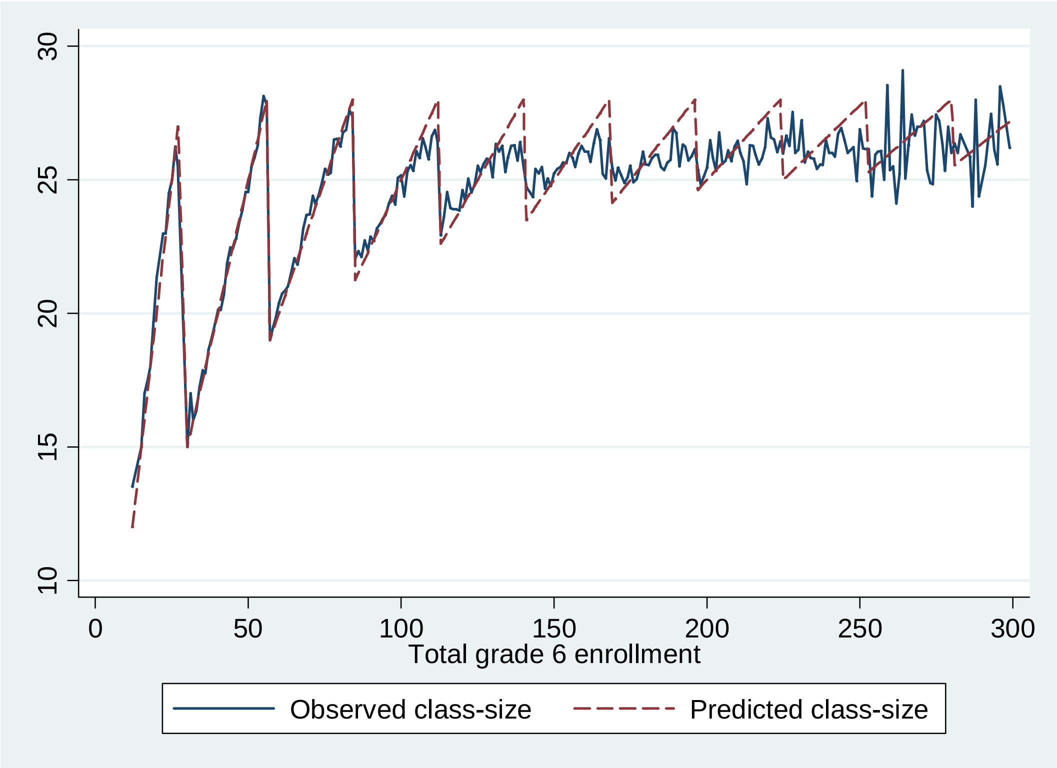

Figure 2

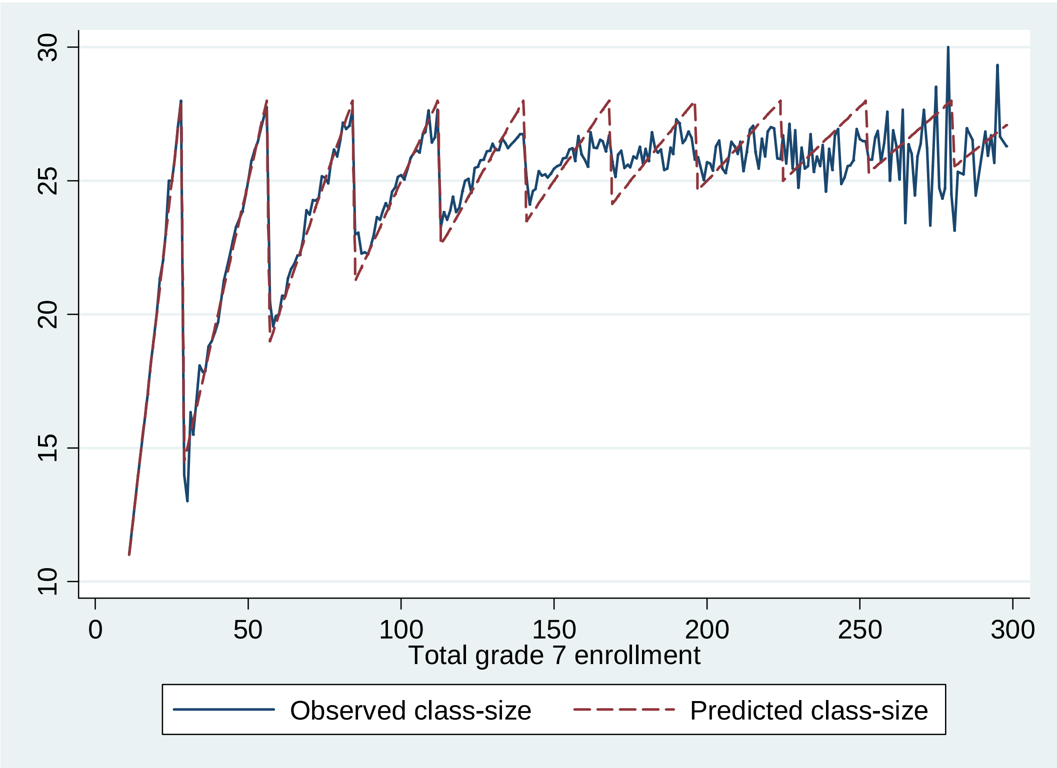

Figure 3

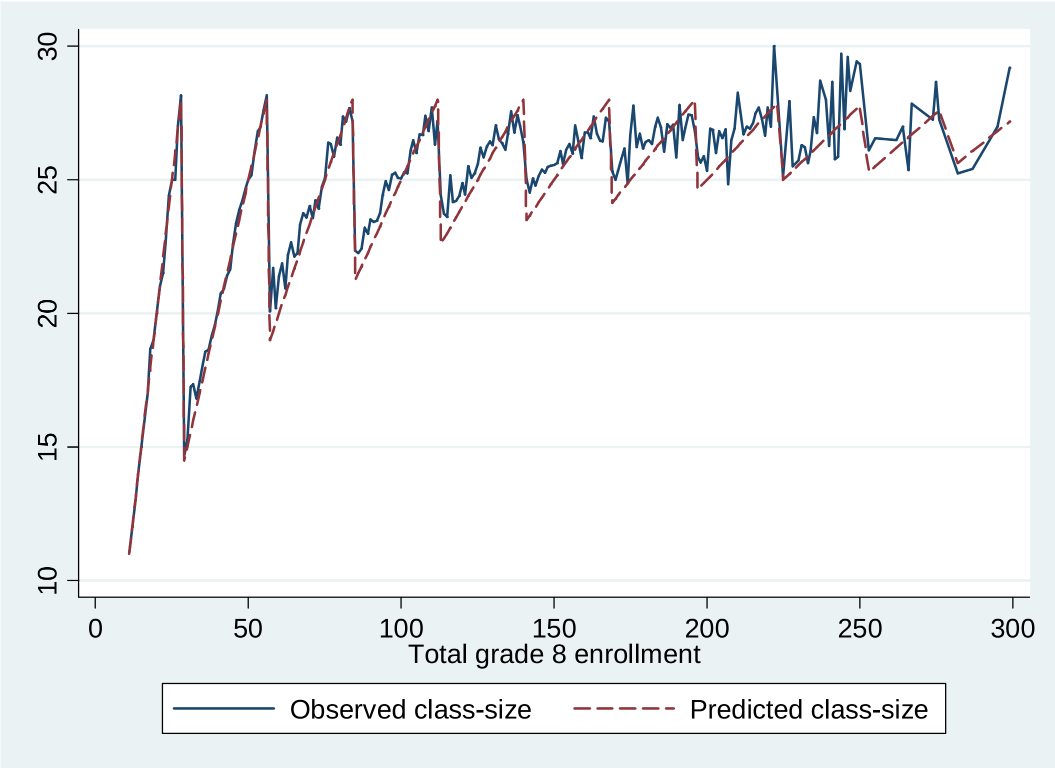

Figure 4

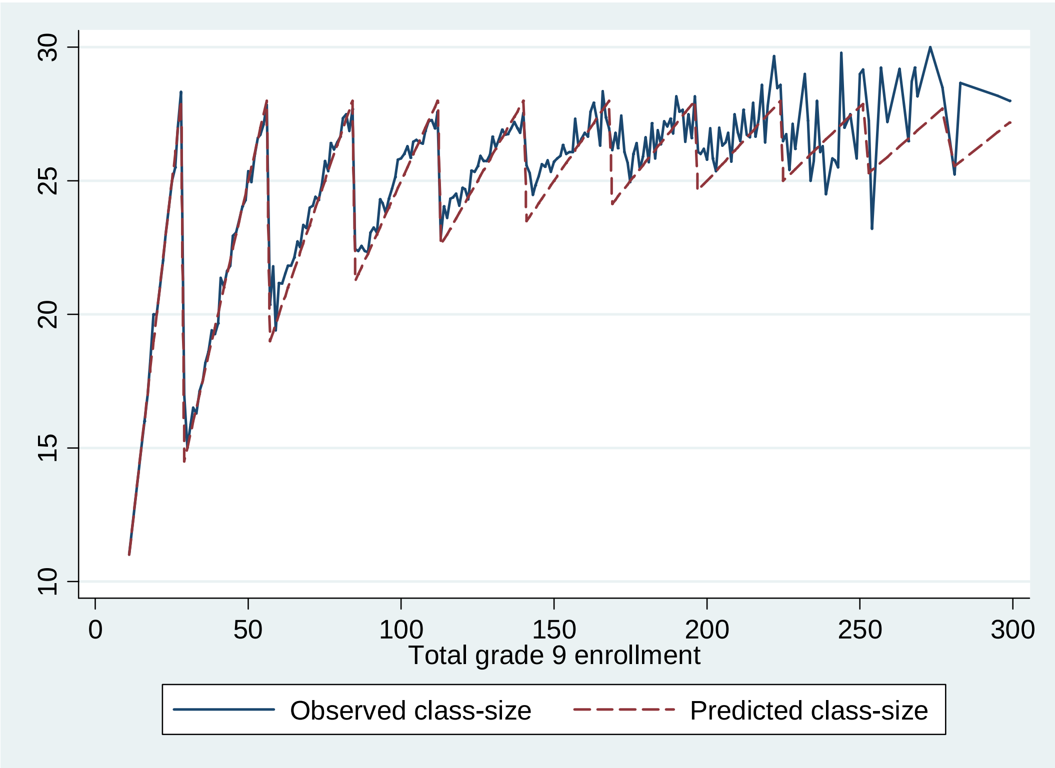

TABLE 1: DESCRIPTIVE STATISTICS

	Overall Mean	Our Sample Mean
Number of observations	24710	16055
12 years old or more at Junior High Scool Entry	0.3212	0.2790
Foreign Student	0.0894	0.0991
Father's Occcupation		
Farmer	0.0352	0.0301
Craftsmen, owners-managers	0.1002	0.0855
Professionals	0.1385	0.1327
Middle managers , technicians	0.1811	0.1973
White collars	0.1331	0.1306
Blue collars	0.3558	0.3812
Inactive or Missing	0.0562	0.0427
Student's grade at Junior High School entry:		
French grade A	0.3014	0.3380
French grade B	0.3555	0.3607
French grade C	0.2042	0.1889
French grade D	0.1373	0.1125
Math grade A	0.4093	0.4520
Math grade B	0.3259	0.3189
Math grade C	0.1468	0.1342
Math grade D	0.1162	0.0948

TABLE 2: SUMMARY OF PROMOTION, RETENTION AND STEERING DECISIONS

	89-90	90-91		91-92		92-93	
Enrolled in grade 6 in September 89	**16 055**						
promoted to grade 7	14 781						
repeated grade 6	1 274						
Enrolled in general education in September t :		**16 055**		**14 515**		**13 256**	
		Grade 6	**Grade 7**	**Grade 7**	**Grade 8**	**Grade 8**	**Grade 9**
Total in June t+1 :		**1 274**	**14 781**	**2 889**	**11 626**	**2 614**	**10 642**
promoted to next grade		1 227	11 626	1 804	10 642	2 462	7 613
repeated the grade		0	1 662	0	810	0	1 324
Steered towards vocational programs		47	1 493	1 085	174	152	1 705

TABLE 3: DISTRIBUTION OF RECORDS s

Grade 89-90	Grade 90-91	Grade 91-92	Grade 92-93	Grade 93-94	Number of observations	%
6	7	8	9	10	**7613**	**47.42%**
6	7	8	9	9	**1324**	**8.25%**
6	7	8	8	9	**783**	**4.88%**
6	7	7	8	9	**1077**	**6.71%**
6	6	7	8	9	**602**	**3.75%**
6	7	8	9	V	**1705**	**10.62%**
6	7	8	8	V	**27**	**0.17%**
6	7	7	8	V	**66**	**0.41%**
6	6	7	8	V	**59**	**0.37%**
6	7	8	V		**174**	**1.08%**
6	7	7	V		**519**	**3.23%**
6	6	7	V		**566**	**3.53%**
6	7	V			**1493**	**9.30%**
6	6	V			**47**	**0.29%**
					16055	**100%**

TABLE 4: OLS REGRESSION OF PROMOTION DECISION (LINEAR PROBABILITY MODEL)

	Grade 6		Grade 7		Grade 8		Grade 9	
Class-size	**-0.001**	**(1.69)**	**0.003**	**(3.60)**	**0.002**	**(2.28)**	**0.009**	**(7.06)**
Total school enrollment ($.10^{-2}$)	0.002	(0.44)	-0.007	(1.28)	0.001	(0.11)	-0.014	(1.80)
Total school enrollment squared ($.10^{-4}$)	-0.0002	(0.73)	0.0004	(0.91)	-0.0001	(0.21)	0.001	(1.51)
Male	-0.011	(2.70)	-0.068	(12.1)	-0.031	(6.11)	-0.050	(6.62)
Foreign student	0.047	(6.21)	0.070	(6.70)	0.057	(5.50)	0.075	(4.95)
Head of household's occupation :								
Reference: Professionals								
Farmer	-0.033	(2.52)	-0.044	(2.47)	-0.025	(1.54)	-0.088	(3.80)
Craftsmen, Shopkeepers, owners-managers	-0.028	(3.14)	-0.032	(2.64)	-0.023	(2.18)	-0.122	(7.96)
Middle managers , technicians	-0.020	(2.85)	-0.020	(2.05)	-0.019	(2.30)	-0.060	(5.12)
White collars	-0.034	(4.25)	-0.049	(4.40)	-0.036	(3.73)	-0.095	(6.73)
Blue collars	-0.029	(4.21)	-0.072	(7.50)	-0.043	(5.18)	-0.137	(11.4)
Inactive or Missing	-0.039	(3.29)	-0.087	(5.20)	-0.078	(4.70)	-0.169	(6.75)
Repeated a grade in elementary school	0.029	(2.80)	-0.058	(3.95)	0.020	(1.29)	-0.068	(2.99)
Two children in the family	0.012	(1.76)	0.003	(0.36)	-0.008	(0.97)	-0.002	(0.13)
Three children in the family	-0.003	(0.47)	0.004	(0.42)	-0.013	(1.40)	-0.018	(1.39)
Four children in the family	0.013	(1.40)	-0.014	(1.11)	-0.001	(0.06)	-0.006	(0.33)
Five children or more in the family	0.029	(2.96)	-0.017	(1.28)	-0.002	(0.18)	0.013	(0.64)
Single mother	0.002	(0.30)	-0.023	(2.48)	-0.024	(2.75)	-0.012	(0.96)
12 years old or more at grade 6	0.028	(2.79)	-0.128	(8.78)	-0.011	(0.73)	-0.236	(10.6)
Quarter of birth : First	0.020	(3.52)	-0.016	(1.97)	0.003	(0.34)	-0.023	(2.12)
Quarter of birth : Second	0.013	(2.33)	-0.012	(1.55)	0.002	(0.26)	-0.010	(0.94)
Quarter of birth : Fourth	0.010	(1.77)	-0.004	(0.54)	0.005	(0.68)	-0.013	(1.20)
Zep school	0.023	(2.14)	0.055	(3.61)	0.038	(2.71)	-0.008	(0.42)
Elective course : German	0.004	(0.72)	0.010	(1.29)	0.021	(3.08)	0.023	(2.42)
Scholarship	-0.017	(3.11)	-0.007	(0.98)	0.015	(2.07)	-0.049	(4.73)
Math grade B	-0.038	(7.35)	-0.097	(13.7)	-0.062	(10.0)	-0.143	(15.8)
Math grade C	-0.132	(18.6)	-0.239	(23.6)	-0.096	(9.28)	-0.245	(15.6)
Math grade D	-0.215	(25.6)	-0.365	(28.9)	-0.094	(5.95)	-0.277	(11.5)
French grade B	-0.019	(3.51)	-0.044	(6.15)	-0.039	(6.39)	-0.115	(12.9)
French grade C	-0.081	(11.9)	-0.169	(17.8)	-0.083	(9.31)	-0.208	(15.6)
French grade D	-0.161	(19.1)	-0.257	(20.8)	-0.104	(7.54)	-0.293	(13.9)
Constant	1.037	(45.3)	1.014	(33.2)	0.962	(36.0)	0.868	(22.2)
R-squared	0.14		0.34		0.06		0.30	
Number of observations	16 055		14 781		11 626		10 642	

Absolute value of t statistic in parentheses.

TABLE 5: FIRST-STAGE REGRESSIONS

Dependent variable: Class Size	Grade 6		Grade 7		Grade 8		Grade 9	
Angrist-Lavy's instrument	**0.225**	**(15.94)**	**0.294**	**(18.4)**	**0.290**	**(18.1)**	**0.338**	**(20.7)**
Total grade enrollment (.10^{-2})	**3.972**	**(10.7)**	**1.943**	**(4.02)**	**3.489**	**(6.20)**	**2.413**	**(4.74)**
Total grade enrollment squared (.10^{-4})	**-0.866**	**(8.80)**	**-0.435**	**(3.38)**	**-0.686**	**(3.82)**	**-0.293**	**(1.88)**
Total school enrollment (.10^{-2})	-0.290	(3.10)	0.156	(1.25)	0.099	(0.84)	0.476	(4.15)
Total school enrollment squared (.10^{-4})	0.015	(2.29)	-0.012	(1.35)	-0.016	(2.01)	-0.045	(5.78)
Male	-0.076	(1.82)	-0.027	(0.56)	-0.088	(1.56)	-0.089	(1.54)
Foreign student	-0.211	(2.72)	0.066	(0.74)	-0.081	(0.72)	0.050	(0.43)
Head of household's occupation :								
Reference: Professionals								
Farmers	-0.506	(3.76)	-0.468	(3.03)	-0.546	(3.12)	-0.701	(3.91)
Craftsmen, Shopkeepers, owners-managers	-0.381	(4.16)	-0.420	(4.04)	-0.269	(2.32)	-0.269	(2.27)
Middle managers , technicians	-0.239	(3.24)	-0.351	(4.26)	-0.163	(1.84)	-0.246	(2.73)
White collars	-0.413	(4.96)	-0.465	(4.92)	-0.305	(2.87)	-0.482	(4.42)
Blue collars	-0.454	(6.31)	-0.503	(6.17)	-0.381	(4.19)	-0.523	(5.62)
Inactive or Missing	-0.661	(5.38)	-0.542	(3.78)	-0.518	(2.83)	-0.435	(2.25)
Repeated a grade in elementary school	0.280	(2.65)	0.140	(1.12)	-0.432	(2.57)	-0.274	(1.55)
Two children in the family	0.023	(0.33)	0.007	(0.09)	-0.127	(1.37)	-0.193	(2.04)
Three children in the family	0.054	(0.74)	-0.047	(0.55)	0.032	(0.33)	-0.181	(1.78)
Four children in the family	0.126	(1.35)	-0.105	(0.97)	0.063	(0.48)	0.063	(0.47)
Five or more children in the family	-0.138	(1.37)	-0.200	(1.71)	0.008	(0.06)	-0.325	(2.14)
Single mother	0.045	(0.67)	-0.011	(0.14)	0.007	(0.08)	0.100	(1.01)
12 years old or more at grade 6	-0.596	(5.67)	-0.491	(3.93)	-0.369	(2.26)	-0.389	(2.26)
Quarter of birth : First	0.082	(1.36)	0.012	(0.18)	0.092	(1.13)	-0.066	(0.79)
Quarter of birth : Second	0.053	(0.92)	0.016	(0.25)	0.014	(0.18)	-0.034	(0.43)
Quarter of birth : Fourth	0.020	(0.34)	0.063	(0.93)	-0.085	(1.06)	-0.043	(0.51)
Zep school	0.171	(1.50)	-0.604	(5.60)	-0.336	(2.54)	-0.638	(4.62)
Elective course : German	0.224	(3.86)	0.232	(3.51)	0.246	(3.37)	0.229	(3.08)
Scholarship	-0.088	(1.59)	-0.156	(2.42)	-0.211	(2.70)	-0.200	(2.48)
Math grade B	0.081	(1.52)	-0.126	(2.09)	-0.247	(3.66)	-0.320	(4.57)
Math grade C	-0.010	(0.14)	-0.325	(3.74)	-0.595	(5.26)	-0.443	(3.66)
Math grade D	0.034	(0.39)	-0.390	(3.60)	-0.997	(5.76)	-0.486	(2.60)
French grade B	-0.125	(2.28)	-0.049	(0.80)	-0.297	(4.47)	-0.223	(3.28)
French grade C	-0.221	(3.12)	-0.106	(1.30)	-0.715	(7.32)	-0.664	(6.44)
French grade D	-0.311	(3.55)	-0.345	(3.27)	-0.570	(3.79)	-0.899	(5.53)
Constant	17.295	(51.5)	16.388	(42.8)	15.966	(41.7)	14.501	(37.5)
R-squared	0.11		0.10		0.17		0.20	
Number of observations	16 055		14 781		11 626		10 642	

Absolute value of t statistics in parentheses.

TABLE 6: INSTRUMENTAL VARIABLES ESTIMATES (2SLS) OF CLASS-SIZE EFFECTS

Dependent variable: Promotion decision	**Grade 6**		**Grade 7**		**Grade 8**		**Grade 9**	
Class-size	**-0.014**	**(3.13)**	**-0.020**	**(3.38)**	**0.006**	**(1.46)**	**0.004**	**(0.79)**
Total school enrollment (.10^{-2})	0.013	(2.30)	0.015	(1.94)	-0.004	(0.61)	-0.007	(0.69)
Total school enrollment squared (.10^{-4})	-0.001	(2.25)	-0.001	(1.74)	0.000	(0.39)	0.000	(0.70)
Male	-0.012	(2.93)	-0.068	(11.9)	-0.031	(6.03)	-0.050	(6.65)
Foreign student	0.044	(5.83)	0.073	(6.75)	0.057	(5.54)	0.075	(4.95)
Head of household's occupation :								
Reference: Professionals								
Farmer	-0.040	(2.98)	-0.055	(2.97)	-0.022	(1.34)	-0.093	(3.91)
Craftsmen, Shopkeepers, owners-managers	-0.033	(3.61)	-0.042	(3.31)	-0.022	(2.04)	-0.124	(8.02)
Middle managers , technicians	-0.023	(3.21)	-0.028	(2.79)	-0.018	(2.20)	-0.061	(5.21)
White collars	-0.039	(4.74)	-0.060	(5.12)	-0.035	(3.54)	-0.098	(6.78)
Blue collars	-0.035	(4.79)	-0.083	(8.19)	-0.041	(4.87)	-0.140	(11.2)
Inactive or Missing	-0.047	(3.82)	-0.099	(5.71)	-0.076	(4.52)	-0.172	(6.82)
Repeated a grade in elementary school	0.032	(3.07)	-0.056	(3.70)	0.021	(1.38)	-0.070	(3.03)
Two children in the family	0.012	(1.78)	0.004	(0.37)	-0.008	(0.89)	-0.002	(0.20)
Three children in the family	-0.003	(0.38)	0.003	(0.28)	-0.013	(1.40)	-0.019	(1.45)
Four children in the family	0.014	(1.53)	-0.016	(1.24)	-0.001	(0.05)	-0.006	(0.33)
Five or more children in the family	0.027	(2.73)	-0.022	(1.59)	-0.002	(0.17)	0.011	(0.56)
Single mother	0.003	(0.39)	-0.023	(2.45)	-0.024	(2.76)	-0.012	(0.91)
12 years old or more at grade 6	0.021	(1.99)	-0.139	(9.18)	-0.009	(0.60)	-0.238	(10.6)
Quarter of birth : First	0.022	(3.68)	-0.015	(1.87)	0.002	(0.28)	-0.023	(2.14)
Quarter of birth : Second	0.014	(2.42)	-0.012	(1.46)	0.002	(0.24)	-0.010	(0.96)
Quarter of birth : Fourth	0.010	(1.78)	-0.003	(0.36)	0.005	(0.72)	-0.013	(1.22)
Zep school	0.025	(2.23)	0.057	(3.72)	0.038	(2.71)	-0.008	(0.40)
Elective course : German	0.007	(1.22)	0.016	(1.98)	0.020	(2.89)	0.024	(2.45)
Scholarship	-0.018	(3.34)	-0.011	(1.47)	0.016	(2.19)	-0.050	(4.79)
Math grade B	-0.037	(7.06)	-0.100	(13.8)	-0.061	(9.71)	-0.145	(15.7)
Math grade C	-0.132	(18.5)	-0.247	(23.5)	-0.094	(8.85)	-0.247	(15.6)
Math grade D	-0.214	(25.2)	-0.374	(28.5)	-0.090	(5.55)	-0.279	(11.5)
French grade B	-0.020	(3.73)	-0.045	(6.13)	-0.038	(6.07)	-0.116	(12.9)
French grade C	-0.083	(12.0)	-0.171	(17.7)	-0.081	(8.60)	-0.211	(15.4)
French grade D	-0.165	(19.1)	-0.265	(20.8)	-0.101	(7.27)	-0.297	(13.8)
Constant	1.316	(13.3)	1.554	(11.2)	0.873	(9.48)	0.966	(8.49)
R-squared	0.13		0.31		0.06		0.30	
Number of observations	16 055		14 781		11 626		10 642	

Absolute value of t statistics in parentheses.

TABLE 7: VARYING THE CLASS-OPENING THRESHOLD

	τ=35		τ=32		τ=30		τ=28		τ=25	
	Grade 6	Grade 7	Grade 6	Grade 7	Grade 6	Grade 7	Grade 6	Grade 7	Grade 6	Grade 7
	First stage									
Angrist-Lavy's instrument	-0.009	-0.133	-0.027	-0.036	0.141	0.140	0.225	0.294	-0.050	-0.064
	(0.93)	(-1.24)	(2.57)	(2.91)	(11.3)	(9.79)	(15.9)	(18.4)	(2.96)	(3.32)
Total grade enrollment . 10^{-2}	5.110	3.167	5.170	3.273	4.364	2.324	3.972	1.943	5.178	3.300
	(13.7)	(6.49)	(13.9)	(6.70)	(11.7)	(4.75)	(10.7)	(4.02)	(13.9)	(6.75)
Total grade enrollment squared . 10^{-4}	-1.084	-0.668	-1.093	-0.685	-0.943	-0.506	-0.866	-0.435	-1.093	-0.694
	(10.9)	(5.14)	(11.1)	(5.28)	(9.55)	(3.90)	(8.80)	(3.38)	(11.1)	(5.34)
F-statistic for the instruments	79.86	18.62	81.80	20.94	122.88	50.14	165.38	131.32	82.52	21.78
R-squared	0.10	0.08	0.10	0.08	0.10	0.09	0.11	0.10	0.10	0.08
	Second stage									
Class-size	**-0.028**	**-0.034**	**-0.028**	**-0.048**	**-0.017**	**-0.026**	**-0.014**	**-0.020**	**-0.026**	**-0.041**
	(4.39)	(2.11)	(4.38)	(3.03)	(3.29)	(2.66)	(3.13)	(3.38)	(4.17)	(2.66)
R-squared	0.08	0.27	0.08	0.20	0.12	0.29	0.13	0.31	0.09	0.24

Absolute value of t statistics in parentheses.

The two stages of the 2SLS regression procedure include controls for gender, age, foreign student, head of household's occupation, number of siblings, single mother, quarter of birth, ZEP school, german as elective course, scholarship, total school enrollment, total school enrollment squared, and entry test grades.

TABLE 8: ALTERNATIVE SPECIFICATIONS OF FIRST STAGE REGRESSION

	A		B		C		D		E		F	
	Grade 6	Grade 7	Grade 6	Grade 7	Grade 6	Grade 7	Grade 6	Grade 7	Grade 6	Grade 7	Grade 6	Grade 7
	OLS		2SLS									
			First stage									
Angrist-Lavy's instrument	-	-	0.261	0.305	-	-			0.225	0.294	0.212	0.296
			(18.9)	(19.4)					(15.9)	(18.4)	(16.4)	(18.5)
Total grade enrollment $.10^{-2}$	-	-	-	-	5.053	3.096	-	-	3.972	1.943	-	-
					(13.7)	(6.39)			(10.7)	(4.02)		
Tot. grade enr. squared $. 10^{-4}$	-	-	-	-	-1.072	-0.654	-	-	-0.866	-0.435	-	-
					(10.9)	(5.05)			(8.80)	(3.38)		
Tot. grade enr. / tot. school enr.	-	-	-	-	-	-	7.988	5.193	-	-	5.820	2.757
							(13.4)	(6.53)			(9.64)	(3.46)
F-statistic for the instruments	-	-	-	-	119.35	27.16	-	-	165.38	131.32	226.75	193.52
R-squared	-	-	0.10	0.10	0.10	0.08	0.09	0.08	0.11	0.10	0.11	0.10
			Second stage									
Class-size coefficient	**-0.001**	**0.003**	**-0.005**	**-0.019**	**-0.028**	**-0.041**	**-0.028**	**-0.059**	**-0.014**	**-0.020**	**-0.012**	**-0.021**
	(1.69)	(3.60)	(1.05)	(3.07)	(4.37)	(2.44)	(3.84)	(2.94)	(3.13)	(3.38)	(2.66)	(3.50)
R-squared	0.14	0.34	0.14	0.31	0.08	0.24	0.07	0.14	0.13	0.31	0.13	0.31

Absolute value of t statistics in parentheses.

The two stages of the 2SLS estimation procedure include controls for gender, age, foreign students, head of household's occupation, number of siblings, single mothers, quarter of birth, ZEP schools, german as elective course, scholarships, total school enrollment, total school enrollment squared and entry test grades in Math and French.

TABLE 9: REGRESSION OF TOTAL GRADE ENROLLMENT

DEPENDENT VARIABLE: TOTAL GRADE ENROLLMENT	GRADE 6				GRADE 7			
	Model 1		Model 2		Model 1		Model 2	
Total School Enrollment	-	-	0.242	(359.56)	-	-	0.249	(390.91)
Head of household's occupation:								
Farmers	-0.366	(10.66)	-0.020	(1.77)	-0.359	(10.4)	-0.004	(0.34)
Craftsmen, Shopkeepers, owners-managers	-0.046	(1.66)	-0.003	(0.31)	-0.033	(1.20)	0.006	(0.68)
Professionals	0.066	(2.48)	-0.012	(1.37)	0.076	(2.84)	0.000	(0.06)
Middle managers , technicians	0.045	(1.78)	0.004	(0.51)	0.049	(1.93)	0.008	(1.09)
White collars	0.008	(0.31)	0.006	(0.72)	0.012	(0.47)	0.006	(0.78)
Blue collars	-0.050	(2.13)	0.000	(0.03)	-0.056	(2.37)	-0.002	(0.35)
Math grade A	-0.046	(2.45)	-0.003	(0.51)	-0.051	(2.69)	-0.020	(0.42)
Math grade B	-0.026	(1.45)	-0.002	(0.27)	-0.033	(1.85)	-0.017	(0.13)
Math grade C	0.000	(0.03)	-0.007	(1.09)	-0.008	(0.44)	-0.011	(1.94)
French grade A	0.044	(2.32)	-0.011	(1.75)	0.038	(2.00)	-0.004	(0.66)
French grade B	0.034	(2.03)	-0.006	(1.11)	0.023	(1.34)	-0.007	(1.44)
French grade C	0.048	(2.81)	-0.001	(0.11)	0.041	(2.37)	-0.003	(0.49)
R-squared	0.031		0.894		0.031		0.919	

Absolute value of t statistics are in parentheses.
Regressions include also controls used in the other tables.

TABLE 10: ALTERNATIVE INSTRUMENT BASED ON TOTAL SCHOOL ENROLLMENT

	First Stage			
	Grade 6	**Grade 7**	**Grade 8**	**Grade 9**
Instrument	0.0027 (6.83)	0.0023 (5.31)	0.0054 (10.6)	0.0046 (9.70)
R-squared	0.03	0.04	0.07	0.07
	Second Stage			
	Grade 6	**Grade 7**	**Grade 8**	**Grade 9**
Class-size coefficient	**-0.044** **(2.88)**	**-0.071** **(2.75)**	**0.000** **(0.02)**	**-0.017** **(1.40)**
R-squared	0.07	0.08	0.06	0.27
Nb of observations	15 565	14 309	11 344	10 370

Absolute value of t statistics are in parentheses.

The instrument is computed as a difference between contemporary total school-enrollment and the mean of total school-enrollment over 6 years (1987-1992)

TABLE 11: DISCONTINUITIES OF CLASS SIZE AND STUDENT SORTING

	Overall sample	Sample 1	Sample 2
	%	%	%
Father's Occcupation			
Farmer	3.01	2.57	2.65
Craftsmen, owners-managers	8.55	8.96	8.43
Professionals	13.27	13.43	13.91
Middle managers , technicians	19.73	19.93	19.96
White collars	13.06	13.46	12.94
Blue collars	38.12	37.78	37.48
Inactive or Missing	4.27	3.87	4.63
Student's entry-test grade:			
French grade A	45.20	45.55	45.02
French grade B	31.89	30.87	32.45
French grade C	13.42	14.12	12.68
French grade D	9.48	9.46	9.85
Math grade A	33.80	34.68	33.99
Math grade B	36.07	35.37	36.47
Math grade C	18.89	18.92	17.93
Math grade D	11.25	11.02	11.62
Sample size	16055	3774	4234

Sample 1 contains students enrolled in grades 6 and 7 with a total grade enrollment between (28k) and (28k+4), where k is an integer.
Sample 2 contains students enrolled in grades 6 and 7 with a total grade enrollment between (28k-4) and (28k), where k is an integer.

TABLE 12: ML ESTIMATES OF SIMULTANEOUS EQUATIONS MODEL

	Grade 6		Grade 7		Grade 8		Grade 9	
GRADE-PROMOTION EQUATIONS								
Class-size year 1	-0.052	**(2.23)**	-	-	-	-	-	-
Class-size year 2	-0.047	**(2.31)**	-0.033	**(2.16)**	-	-	-	-
Class-size year 3	-	-	-0.040	**(2.60)**	-0.021	**(1.47)**	-	-
Class-size year 4	-	-	-	-	-0.003	**(0.20)**	-0.005	**(0.31)**
Total school enrollment ($.10^{-2}$)	-0.002	**(0.25)**	0.002	**(0.35)**	0.008	**(0.88)**	0.014	**(1.61)**
Male	-0.113	**(3.31)**	-0.368	**(15.1)**	-0.229	**(6.97)**	-0.193	**(6.71)**
Foreign student	0.338	**(5.95)**	0.269	**(7.18)**	0.306	**(5.07)**	0.295	**(5.95)**
Head of household's occupation :								
Farmers	-0.597	**(4.59)**	-0.563	**(6.18)**	-0.272	**(2.47)**	-0.644	**(6.93)**
Craftsmen, owners-managers	-0.534	**(5.19)**	-0.492	**(7.08)**	-0.271	**(3.53)**	-0.676	**(10.5)**
Middle managers , technicians	-0.452	**(4.69)**	-0.380	**(5.99)**	-0.205	**(3.14)**	-0.416	**(7.56)**
White collars	-0.583	**(6.02)**	-0.560	**(8.73)**	-0.313	**(4.51)**	-0.590	**(9.76)**
Blue collars	-0.544	**(5.94)**	-0.636	**(10.6)**	-0.355	**(5.66)**	-0.758	**(14.0)**
Inactive or Missing	-0.585	**(5.23)**	-0.729	**(9.63)**	-0.532	**(5.61)**	-0.864	**(9.57)**
Repeated a grade in elementary school	0.167	**(2.39)**	-0.113	**(2.22)**	0.060	**(0.72)**	-0.182	**(2.39)**
12 years old or more at grade 6	0.061	**(0.86)**	-0.708	**(13.8)**	-0.280	**(3.46)**	-0.876	**(11.8)**
Zep school	0.060	**(1.01)**	0.011	**(0.27)**	0.033	**(0.56)**	-0.006	**(0.13)**
Elective course : German	0.057	**(1.03)**	0.140	**(3.39)**	0.135	**(2.68)**	0.091	**(2.25)**
Scholarship	-0.081	**(2.17)**	-0.070	**(2.60)**	0.045	**(1.16)**	-0.211	**(6.21)**
Math grade B	-0.591	**(10.1)**	-0.551	**(16.0)**	-0.402	**(9.95)**	-0.471	**(14.4)**
Math grade C	-1.075	**(17.2)**	-0.811	**(20.1)**	-0.529	**(9.60)**	-0.683	**(13.1)**
Math grade D	-1.291	**(19.6)**	-1.027	**(22.9)**	-0.597	**(8.24)**	-0.877	**(10.7)**
Language grade B	-0.610	**(7.81)**	-0.510	**(11.8)**	-0.363	**(8.04)**	-0.481	**(14.0)**
Language grade C	-0.993	**(12.3)**	-0.782	**(16.9)**	-0.494	**(8.98)**	-0.712	**(15.0)**
Language grade D	-1.260	**(14.9)**	-1.022	**(20.0)**	-0.581	**(8.32)**	-1.019	**(14.0)**
Promotion to grade (g+1) cut	-4.632	**(7.97)**	-4.163	**(10.8)**	-3.440	**(9.48)**	-2.665	**(7.24)**
Repetition of grade g cut			-3.650	**(9.48)**	-2.785	**(7.68)**	-2.060	**(5.60)**
CLASS-SIZE EQUATIONS								
Angrist-Lavy's instrument	0.243	**(19.8)**	0.308	**(25.1)**	0.278	**(21.2)**	0.346	**(22.7)**
Total grade enrollment ($.10^{-2}$)	1.008	**(9.64)**	0.504	**(4.46)**	1.406	**(10.1)**	1.371	**(9.54)**
Total school enrollment ($.10^{-2}$)	-0.075	**(2.82)**	0.002	**(0.07)**	-0.062	**(2.09)**	-0.066	**(2.13)**
Male	-0.051	**(1.27)**	-0.046	**(1.11)**	-0.088	**(1.73)**	-0.099	**(1.68)**
Foreign student	-0.234	**(3.28)**	-0.106	**(1.47)**	0.044	**(0.47)**	0.103	**(0.93)**
Head of household's occupation :								
Farmers	-0.630	**(4.86)**	-0.648	**(4.90)**	-0.738	**(4.61)**	-0.978	**(5.40)**
Craftsmen, owners-managers	-0.391	**(4.39)**	-0.549	**(6.09)**	-0.294	**(2.74)**	-0.379	**(3.13)**
Middle managers , technicians	-0.259	**(3.60)**	-0.363	**(5.00)**	-0.178	**(2.11)**	-0.261	**(2.84)**
White collars	-0.428	**(5.32)**	-0.541	**(6.65)**	-0.296	**(3.06)**	-0.481	**(4.36)**
Blue collars	-0.436	**(6.23)**	-0.544	**(7.67)**	-0.439	**(5.25)**	-0.599	**(6.34)**
Inactive or Missing	-0.636	**(5.44)**	-0.729	**(6.13)**	-0.431	**(2.74)**	-0.440	**(2.28)**
Repeated a grade in elementary school	0.244	**(2.40)**	0.104	**(1.00)**	-0.173	**(1.19)**	-0.196	**(1.10)**
12 years old or more at grade 6	-0.579	**(5.77)**	-0.472	**(4.59)**	-0.605	**(4.29)**	-0.504	**(2.90)**
Zep school	-0.346	**(4.83)**	-0.433	**(5.98)**	-0.209	**(2.27)**	-0.348	**(3.23)**
Elective course : German	0.261	**(4.66)**	0.307	**(5.35)**	0.277	**(4.06)**	0.282	**(3.73)**
Scholarship	-0.093	**(1.93)**	-0.087	**(1.79)**	-0.275	**(4.43)**	-0.222	**(2.99)**
Math grade B	0.069	**(1.32)**	-0.134	**(2.56)**	-0.311	**(4.98)**	-0.331	**(4.60)**
Math grade C	-0.033	**(0.46)**	-0.384	**(5.33)**	-0.564	**(6.04)**	-0.439	**(3.56)**
Math grade D	0.023	**(0.29)**	-0.332	**(3.87)**	-0.920	**(7.38)**	-0.482	**(2.55)**
Language grade B	-0.141	**(2.62)**	-0.074	**(1.37)**	-0.298	**(4.74)**	-0.252	**(3.62)**
Language grade C	-0.204	**(2.98)**	-0.128	**(1.84)**	-0.662	**(7.76)**	-0.724	**(6.88)**
Language grade D	-0.263	**(3.15)**	-0.278	**(3.22)**	-0.571	**(4.80)**	-0.999	**(6.07)**
Constant	18.509	**(62.6)**	17.555	**(59.2)**	17.936	**(57.6)**	16.367	**(45.7)**

TABLE 13: COVARIANCE MATRIX OF RESIDUALS

	epsilon1	epsilon2	epsilon3	epsilon4	nu1	nu2	nu3	nu4
epsilon1	6.4959 **(132.07)**							
epsilon2	2.2367 **(37.96)**	6.6542 **(128.14)**						
epsilon3	1.0957 **(17.81)**	1.4621 **(18.56)**	8.5256 **(115.85)**					
epsilon4	0.8885 **(11.12)**	1.0739 **(13.80)**	2.8453 **(31.11)**	9.0030 **(132.43)**				
nu1	0.3125 **(2.76)**	0.1117 **(1.15)**	-0.0570 **(2.05)**	0.1332 **(2.57)**	1.0185 **(2.03)**			
nu2	0.2108 **(3.44)**	0.3928 **(2.05)**	0.1288 **(1.09)**	0.1480 **(1.97)**	0.0109 **(2.94)**	1.0250 **(3.57)**		
nu3	0.1761 **(5.02)**	0.2549 **(5.70)**	0.3531 **(3.75)**	0.1270 **(1.05)**	0.0045 **(1.02)**	0.0174 **(1.99)**	1.0214 **(3.65)**	
nu4	0.1555 **(3.88)**	0.1826 **(3.39)**	0.3205 **(4.47)**	0.3586 **(0.87)**	0.0076 **(0.96)**	0.0152 **(1.67)**	0.0177 **(1.14)**	1.0225 **(4.21)**

TABLE 14: CORRELATION MATRIX OF RESIDUALS

	epsilon1	epsilon2	epsilon3	epsilon4	nu1	nu2	nu3	nu4
epsilon1	1.0000							
epsilon2	0.3402	1.0000						
epsilon3	0.1472	0.1941	1.0000					
epsilon4	0.1162	0.1387	0.3248	1.0000				
nu1	0.1215	0.0429	-0.0193	0.0440	1.0000			
nu2	0.0817	0.1504	0.0436	0.0487	0.0107	1.0000		
nu3	0.0684	0.0978	0.1197	0.0419	0.0044	0.0170	1.0000	
nu4	0.0603	0.0700	0.1086	0.1182	0.0074	0.0148	0.0173	1.0000