# The impact of teacher's wages on students' performance in the presence of heterogeneity and endogeneity. Evidence for Brazil

M. Sprietsma and F. Waltenberg

Discussion Paper 2005-8

Département des Sciences Économiques
de l'Université catholique de Louvain

UNIVERSITAS · CATHOLICA · LOVANIENSIS · SEDES · SAPIENTIAE · 1425

UCL

# The impact of teachers' wages on students' performance in the presence of heterogeneity and endogeneity. Evidence from Brazil.[ζ]

Maresa Sprietsma[φ] and Fábio D. Waltenberg[ϕ]

**Abstract**

In this paper we estimate the effect of teachers' wages on students' achievement in a developing country. We use test scores of pupils enrolled in the 8th grade of primary school, surveyed in 2001 in Brazil. We regress individual student test scores on gross monthly teacher wages allowing for nonlinearities. Given the strong heterogeneity of Brazilian pupils and teachers, we estimate quantile regressions (QR), which provide, instead of a constant mean coefficient, a detailed characterization of the effect of teachers' wages on conditional pupils' scores. For the same reason, we also run separate regressions for private and public schools. We then account for potential endogeneity of teachers' wages through the estimation of instrumental variables models (IV). Finally, we estimate two-stage least absolute deviation models (2SLAD), that allow us to cope simultaneously with the heterogeneity of the student-teacher relationship and with the endogeneity of teachers' wages. Our results show that wages of language teachers have a small, but positive and significant effect, on student test scores in private schools, controlling for endogeneity, but that they are insignificant, or even negative, in public schools. We also observe that teacher wages show a decreasing effect as we move along the conditional distribution of scores. The same effects are found for mathematics teachers, but the results are less robust and the coefficients are smaller.

**JEL Classification**: I2; J24, J31
**Theme:** Education and Training
**Keywords**: economics of education, human capital, resource allocation, education production functions, instrumental variables, two-stage least-squares, quantile regression, two-stage least absolute deviation.

[ζ] We would like to thank Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) for giving us permission to use the SAEB dataset.

[φ] Economics Department, Université Catholique de Louvain.
We are grateful for the financial support from the Belgian French Community's program 'Action de Recherches Concertée' ARC 03/08-302.

[ϕ] Economics Department, GIRSEF and Hoover Chair (associate member), Université Catholique de Louvain. I am grateful for the financial support from the Belgian French Community programs 'Action de Recherches Concertée' ARC 02/07-204 (main source of funding, GIRSEF).

## 1. Introduction: teachers wages as inputs in the education production function

Despite all recent technological advances, schooling remains essentially a labor-intensive sector. Indeed, wages still represent the most important share of educational budgets in many countries. For example, in 2001, wages of teachers and other schooling staff together represented 77% of total expenditures on primary and secondary education in Brazil (OECD, 2001). The efficacy of these expenditures in improving student achievement is therefore of high policy relevance.

In the economics of education literature, a teacher's wage is often taken as a proxy for his level of competence. It is usually assumed that better workers are better rewarded so that the wage is an indicator of their productivity. The most commonly used justification for this assumption is that more qualified and more capable individuals face a choice between working as a teacher or working in another sector (Southwick & Indermit, 1997; Dolton & Van der Klaauw, 1999; Angrist & Guryan, 2003). Too low relative wages in the educational sector would create an adverse selection problem, leading less-able graduates to choose to become teachers and lowering the overall level of teacher quality (Hoxby & Leigh, 2003). Alternatively, one could assume there exists some form of reward to merit in the teaching profession that would tend to increase the wages of better teachers. The latter assumption requires the existence of a competitive market for teachers. This is certainly not obvious in many settings, especially when we recall that a considerable fraction of many schooling systems is run by the state. In public schools, teachers' wages are typically determined by age, tenure, political indications and other factors not necessarily related to merit or productivity. Therefore, there is no financial incentive for the best teachers to stay in the educational sector since they may have better career prospects in other economic sectors (Lenkford & Wyckoff, 1997).

In Brazil, a private schooling system co-exists with a public one. In the former, there is a 'free market' for teachers: recruitment procedures and wage settings are decided on a decentralized basis, subject to some constraints imposed by unions and collective bargaining rules. Each school is considerably free to make decisions related to teacher hiring and performance rewarding. Contrary to what happens in the private system, in the public system there is no free market for teachers. Recruitment should in principle be done by means of public contests, but many teachers are indicated to their jobs by politicians or by other means. Wages are determined according to general guidelines stated by the federal authority (ministry of education), but mainly by state-level and/or municipal-level decision-makers. Public schools cannot decide autonomously to pay higher salaries in order to attract better teachers, nor to reward such teachers, who are supposedly the ones who could lead pupils to attain better performance. Thus we expect to find stronger correlations between teacher wages and pupils achievement in the private system than in the public one.

It is a much-debated topic in the empirical literature whether investing in resources such as teacher wages can improve student performance. A common finding is that resources do not have a significant effect on student test scores (Hanushek, 1986; 1997; 2002), but such a conclusion has frequently been questioned (Card & Krueger, 1992; Figlio, 1999). Particularly, the effect of relative wages is usually more significant than the absolute teacher wage (Southwick & Indermit, 1997) but both variables have a weak and not very robust effect on student scores (Dewey et al., 2000).

The purpose of our paper is to assess the effect of teachers' wages on student performance, exploiting the features of Brazil, a developing country in which the variation in teacher wages is particularly marked, as we will show later. The underlying reasoning is that it could be the case that in industrialized countries teachers' wages have reached a threshold above which there is not enough variation as to measure an effect of a change in wages on student performance. In a poorer country where the dispersion in wages is much higher, the level of teachers' wages might have a more significant effect, at least in the private schooling system (Case and Deaton, 1999).

We essentially test three hypotheses in this paper: (i) that teachers wages matter for students achievement, (ii) that the conditional correlation of teachers wages and scores is stronger in private than in public schools, given that the latter are imposed a greater number of constraints on their recruitment and payment policies, and (iii) that there are variations in the conditional correlation of teachers wages with students test scores, indicating that there is heterogeneity in the pupil-teacher relationship.

The paper is organized as follows. **Section 2** is devoted to the presentation of our data and a discussion of the variables we use. We then estimate an education production function, first by ordinary least-squares (OLS), which serves as our baseline reference (**section 3**). Given the high heterogeneity of Brazilian pupils and teachers, in **section 4** we turn to the estimation of quantile regressions (QR), which provide a detailed characterization of the effect of teachers wages (explanatory variable) along the distribution of pupils scores (dependent variable). This technique contrasts with OLS and IV estimations, which only provide mean conditional correlations of the explanatory and the dependent variables. Even including a series of controls, OLS coefficients risk being biased, since it is difficult to defend the assumption that students of different levels of performance are randomly assigned to teachers of different wages. In order to address this likely endogeneity issue, we use a set of variables, including teachers' experience and teachers' gender, as instruments for the teachers' wage (IV: **section 5**). Finally, we try to cope simultaneously with heterogeneity in the teacher-student relationship and endogeneity of teachers' wages through a combination of two-stage least squares and QR, in the so-called two-stage least absolute deviation estimation (2SLAD**: section 6**). **Section 7** contains a summary and the conclusions.

## 2. Data set and choice of variables

### *2.1. The SAEB database*

The data we use come from the 2001 Brazilian survey on pupils' achievement, the so-called SAEB, which stands for Basic Education Assessment System. SAEB is organized by INEP, a research institute subordinated to the Brazilian Ministry of Education[1]. SAEB consists of countrywide cognitive ability exams in language (Portuguese) and mathematics, coupled with a collection of data on relevant features of students, teachers, principals, and schools. It focuses on the evaluation of pupils at three key stages of their formal education: 4th year of primary school, 8th year of primary school, and 3rd year of secondary school. Each of these grades corresponds to the last year of a stage in the Brazilian schooling system. These are the end of first half of primary school (during which students have one teacher for all subjects), the end of second half of primary school (during which students have one teacher for each subject), and the end of secondary school (after which students can do college admission exams).

Schooling is mandatory in Brazil for children up to 14 years, regardless of the grade they are attending. The 8$^{th}$ grade sample constitutes the best available approximation to the end of compulsory schooling. Moreover, the 8$^{th}$ grade sample is also less exposed to dropping out than the 3$^{rd}$ grade of secondary school. Finally, the 8$^{th}$ grade datasets have fewer missing data in key questions (e.g. pupil’s age, mother’s education, number of books at home etc.) as compared to the 4$^{th}$ grade dataset. For these reasons, in this paper, we decided to focus exclusively on the 8$^{th}$ grade sample. It should be noted, however, that while the recommended age for pupils enrolled in the 8$^{th}$ grade is 14 years, the range of pupils’ ages in the sample is actually quite wide.

The SAEB database surveys random samples. For each grade and subject, the sample is representative of the students of the whole country, and of each of the 26 Brazilian states and the Federal District (in which is located country’s capital, Brasilia). In a first step, schools that took part in SAEB have been randomly chosen. In a second step, one class has then been randomly chosen inside each of these schools. All students of a given class have been submitted to the SAEB

[1] In Portuguese, SAEB stands for Sistema de Avaliação do Ensino Básico. INEP stands for Instituto Nacional de Estudos e Pesquisas Educacionais (In English: National Institute for Educational Studies and Research).

exam on only one of the two subjects, which means that a particular pupil could not do both language and mathematics tests. Pupils' test scores correspond to subject-specific pedagogical scales elaborated by INEP staff together with teachers, researchers and experts. Possible scores range from 0 to 500, and are supposed to evaluate skills and abilities of students. Scores are not comparable across subjects.

### 2.2. Choice of control variables

Tables 1 (language) and 2 (mathematics) contain descriptive statistics for all variables used in the estimations for all schools taken together and for each type of school: private, municipal public schools, and state-level public schools. We recall that they refer to the 2001 wave of SAEB, to both language and mathematics tests, and to the 8$^{th}$ grade.

Tables 1 and 2 show descriptive statistics of the dependent and the explanatory variables expressed both in their original units (*ntscore* and *ntwageprof*), that is before unit transformations undertaken for ease of interpretation (explained below). By looking at 'ntscore' and 'ntwageprof', we verify the high variability of both scores and teachers wages. In both subjects, while the average scores are not far from 250, the range of scores is quite wide (e.g. minimum score of 78, 21 and maximum score of 399,03 in mathematics). Teachers' wages range from 0,5 to 15 s.m., the monetary unit used here[2], and standard deviations are substantial[3]. Standard deviations of both dependent and explanatory variables are higher in private with respect to public schools.

The remaining rows show descriptive statistics for all the variables used in our estimations: the dependent variable (*score*), the explanatory variable (*wageprof*), and the control variables. Appendix 2 explains in detail how the explanatory variable has been constructed. The important point is that it has been created out of a transformation of the original information we had, such that the new variable has mean equal to 0 and standard deviation equal to 1. The dependent variable, i.e. the student's test score, was standardized in order to have a mean of 500, and a standard deviation of 100.

Following Carroll (1963) and subsequent literature[4], there are five factors that determine students learning rate: "(i) aptitude, (ii) ability to understand instructions, (iii) perseverance, (iv) opportunity, and (v) the quality of instruction". We use these categories as a reference for the choice of our

[2] The monetary unit used here is that one in which wages are expressed in Brazilian administration, namely 's.m.', which stands for 'salários mínimos', (literally, 'minimum wages'). One unit of s.m. was approximately equivalent to US$68,00 in October 2001, when SAEB exams took place. See Appendix 2 for further explanations.

[3] The range is the same for both subjects because this variable has been composed out of the same categorical question concerning their wages that has been posed to all teachers regardless of the discipline that they teach. See Appendix 2 for further explanations

[4] Creemer, 1994; Scheerens, 1997; Creemers *et al.*, 2000

control variables among all the information available on the pupil itself, on its family, its teacher and its school.

Firstly, to account for observable **individual characteristics** (potentially related to points i to iv), we include age, gender and race of the pupil. There is a broad range of ages inside the considered grade (pupils are aged between 8 and 14 in the 8th grade). We can therefore expect age to affect motivation, self-confidence and maturity of the pupils. We include dummies for self-reported pupil's race (black and mixed). Since, on average, mixed individuals are poorer than whites in Brazil, and blacks are the poorest race group, this dummy not only plays the role of a control for unobserved variables related to pupils race, but particularly, it is also a control for its socio-economic status.

Secondly, we control for pupils **family environment.** We use measures of mother education (misced), family wealth (as measured by the number of employees at home: nmaids), the number of books at home (as a proxy for the family interest in learning and as a home educational resource: nbbooks), and the type of family structure the pupil lives in (with both parents or not: nonnuclear). These variables might affect the motivation, level of effort and opportunities of the pupils as well as their ability to understand instructions.

Thirdly, to account specifically for **pupil effort** (point iii), we use information on the frequency with which pupils does their homework when asked to do it (hmwk)[5]. We also have information as to whether the pupil repeated grades (retention) and how often. We use this variable as an imperfect control for **past effort**, while we are aware that this complex variable also reflects innate talent and family background.

Fourthly, we include measures of the **quality of schooling** (point v) received by pupils. We have information on the size of classes, (the student/teacher ratio: stratio), the availability of a library in the school (library) and the number of computers available for pupil use (ncomp)[6]. Moreover, we include the gross monthly wage of the principal as a control for his overall level of competence (wageprinc).

[5] One could fear that the quantity of homework given by the teacher is either endogenous (unobserved characteristics influence both pupils homework records and their scores, simultaneously), or even as an outcome variable (well-paid teachers possibly have an impact, not only the score of pupils, but also on their homework activities). It should be noted that the variable measures the frequency with which pupils do their homework, and not simply the frequency with they are given homework. That is, the value this variable assumes depends primarily on a choice made by pupils, so classifying this variable as an outcome variable is not so straightforward. Simultaneous determination might be of concern, but we decided to keep this variable in the model for two reasons: (i) because it is just a control and not our main focus, (ii) it is the only variable indicating effort that is available in the SAEB dataset.

[6] We don't have information on the number of students in each school, so we were not able to generate the ideal variable that would be the number of computers per head in each school.

Finally, we add controls for regional specificities of the environment. We use indicator variables for each Brazilian state (uf11 to uf53), aimed at capturing all state-specific heterogeneity. We include municipal information on tax revenues per head (taxpercap) to take into account **variations in economic resources** that could influence both student performance and teachers wages, independently of their actual productivity level. For instance, wealthier municipalities could give access to public libraries, health services or other aid to families and on average parents' earnings would tend to be higher than in poorer municipalities.

As for **teachers characteristics** we assume that, conditional on having or not a university degree (univprof), all otherwise relevant characteristics related to teachers quality are synthesized by their wages. Other observed teachers variables are used as predetermined instruments in IV and 2SLAD estimations: teachers' gender (gendprof), teachers' experience in teaching the tested discipline (expprof), monthly hours of work (hoursprof) and a dummy indicating whether the teacher has another job besides teaching (otherjob) .

## 3. Do teachers' wages matter? The baseline model (OLS)

### *3.1. Model and explanatory variables*

Our basic model estimates an education production function for scores that 8th grade pupils obtained in SAEB tests, in both subjects: language and mathematics. Test scores are a function of teachers' gross monthly wages, controlling for the factors described in the previous section. Our OLS (benchmark) model is as follows:

$$\mathbf{Score_i = \alpha + \beta W_i + \gamma X_i + \delta X^2_i + \varepsilon_i} \qquad \textbf{(Equation 1)}$$

Where: Score is the performance of pupil i in SAEB test, W stands for each teacher's wage, X is a vector of control variables and ε is the error term.

Note that by including in the equation a vector ($X^2$) containing the squares of the continuously valued right-hand side variables, we avoid imposing the restrictive assumption of linearity in the relationship between right-hand side variables and the outcome variable (score), a procedure inspired by Figlio (1999). In subsection 3.1 we explain why we use the specification stated in equation 1 and not a more general one, which would include a squared teachers' wages term ($W^2$).

Test score observations have been standardized such that the mean for this variable is 500, with a standard deviation of 100 (variable name: score). And teachers wages observations have been set to have a mean equal to 0 and a standard deviation equal to 1 (variable name: wageprof). Thanks to these transformations, we can interpret the estimated effects of teacher wages on scores in an intuitive way. For example, a marginal effect of 10 would mean that a change of 1 unit of teachers wages (that is, 1 standard deviation of teachers wages) corresponds with a change of 10 points of pupils score (that is, 10% of standard deviation of pupils scores).

In addition to the estimations for the overall sample, we also estimate our models for private, municipal, and state schools separately. There are several reasons for doing so. First of all, it appears that there are important **differences between public and private schools**. In the descriptive statistics by type of school (tables 1 and 2) we observe that average scores of private schools students are between 15 to 20 percent higher than those of public schools. We also notice that, on average, private schools teachers earn 30% more than their public counterparts. When we turn to control variables, the differences are also striking. For example, pupils in private schools live much less often in non-nuclear families (27 compared to 40 per cent in public schools), have highly educated mothers (on average, mothers of students in private schools have been to university, mothers of students in public schools have not), live in wealthier families with more books and do more homework. Private school infrastructure and wages are also much better than those of the public sector. Private schools directors earn, on average, 30% more than their public schools counterparts and private schools have, on average, 23 computers for pupil use compared to less than 5 for public schools.

Of course, these differences in average values of dependent, explanatory and control variables do not make the case for partitioning the sample – indeed, these differences are controlled for in our estimations –, but they constitute evidences that private and public schools function in **very contrasted environments**. We believe that, beyond these observable variables, there might also exist unobservable heterogeneity of pupil's characteristics from one type of school to another. This leads us to think that in Brazil, private and public schools function in completely different environments, and are allocated completely different mixes of inputs (particularly, of inputs such as pupils and teachers characteristics).

Another reason why we think it is interesting to look at private and public schools separately relates to their respective **funding and managing characteristics**. Private schools are neither financed nor managed by the public authority, which significantly modifies their decision-making environment and the nature of their budget constraints. As mentioned previously, private schools function in competition with other schools in the educational market, whereas public schools are managed by the state and respond to bureaucratic rules.

As a consequence, we believe that the effect on student scores of an important input like teachers' wages should not be estimated (only) by taking the whole sample and assuming a constant coefficient for all types of schools. It seems worthwhile to gain some insights about this kind of heterogeneity, of both regressors and scores, by separately estimating private and public schools coefficients.

Although they are not as strong as those between private and public schools, there are also some relevant **differences among public schools**. Municipal and state schools differ mainly in their degree of decision-making autonomy (decentralization). Whereas public schools are funded and managed mainly at the local level (there are 5562 municipalities in Brazil), state schools are funded and managed mainly at a higher level (there are 27 states in the country). Consequently, to be consistent we decided to partition the public schools sample as well.

Therefore we estimate our models for all schools taken together, but also separately for the three types of schools – private, municipal and state schools. In order to keep results comparable when we use partitioned samples, we standardize the dependent variable (test score: 'score') and the explanatory variable (teacher's wage: wageprof) by subset, such that means and standard deviations are set, respectively, to 500 and 100, and to 0 and 1 for each type of school.

### ***3.2. OLS results***

F tests of structural change[7] inform us that we can reasonably reject the null hypothesis that the coefficient $\delta$, associated with the vector of squared control variables, is equal to zero. This conclusion holds for both subjects when all schools are taken together. However, we cannot reject the null hypotheses that the coefficients associated with the square of teacher's wage ($W^2$) are equal to zero. These tests suggest that the specification we use here (linear in W and nonlinear in X) is more adequate than the linear functional form usually employed in education production functions. These results are in line with those obtained by Figlio (1999). In the remaining of the paper, we will make use of this specification, stated in equation 1 above.[8]

Tables 3 and 4 show, for language and mathematics respectively, the results of OLS estimations of the impact of teachers' wages on scores, undertaken according to equation 1, namely, conditional on a series of controls. For each of the two subjects – language and mathematics – we present results for all schools taken together, and also for each type of school.

In this baseline (OLS) model, teacher wages have a *small but positive and significant (at 1% level) effect* on student test scores, when all schools are taken together. In the language sample, the coefficient is of 3,30, which means that a change of 1 unit in teachers wages (that is, 1 standard

[7] See table 5.

[8] When each type of school is treated separately, the most flexible model, namely the one which is nonlinear in both W and X, in some cases is preferred to the one we chose (linear in W, nonlinear in X), especially in municipal schools. However, we decided to stick to the same specification across models in order to make our comparisons clearer.

deviation of teachers wages) corresponds to a change of 3,30 points of pupils score (that is, 3,30% of standard deviation of pupils scores). A similar result is obtained in mathematics, but the coefficient is smaller (2,38). The coefficients associated with the private school dummy[9] (*private*) are positive and very high in both subjects.

Tables 3 and 4 also contain the results for the partitioned samples. The picture we obtain is quite different depending on the type of school. For both subjects, teachers' wages coefficients in private schools are considerably higher than those of the whole sample (6,99 in language and 4,02 in mathematics) and both are significant. For *municipal schools*, the more decentralized type of public school, teacher wage coefficients are not statistically different from zero in both subjects. The same is true for *state schools* in language, while in mathematics the coefficient is positive (3,14), statistically significant, and higher than the one associated with the all the schools taken together. So the effect of teacher wages on student test scores for the whole sample seems lower than that for private schools due to both types of public schools coefficients in the language exam, and only because of municipal schools in the mathematics case.

Our control variables yield the expected signs, in line with standard results found in the literature.

Boys perform better than girls do in mathematics and the reverse is true in language. Age and grade retention both affect test scores negatively. Self-reported black and mixed pupils perform worse on average, as well as individuals coming from non-nuclear families, and those that have fewer books at home. Having a teacher holding a university degree (only for language scores) or the availability of a school library (only for public schools) affect test-scores positively as well as having access to computers at school (in private schools only). Note that most public schools have very few computers and less often a library than private schools. This could explain the differences in significance of the measured effects between school types.

Our results suggest that teachers' wages do matter for students' achievement in Brazil, although the estimated coefficients are not extremely high. Moreover, the hypothesis that, in private schools, the conditional effects of teachers wages on scores are likely to be stronger than in public schools, given that the latter are imposed a greater number of constraints on their recruitment and payment policies, is largely supported by our OLS results.

[9] Private takes value one if the school is private, 0 otherwise

## 4. To whom do teachers' wages matter? The rationale for using quantile regressions (QR)

Education production function studies typically report average effects on student achievement (typically, the outcome variable) of school resources, family resources, or other relevant inputs. Widely used methods such as least squares and instrumental variables have the drawback of not allowing one to assess the effect of resources on achievement at different points of the conditional distribution of educational achievement (scores), since they only estimate a constant coefficient, the conditional mean.

However, understanding the *effect of teachers' wages along the distribution of scores* may be relevant for various reasons. It is particularly important when there are good reasons to believe that 'weak' and 'strong' students function in extremely different contexts. This is the case in the Brazilian schooling system. Indeed Brazilian students and teachers are heterogeneous in various respects. Firstly, this country presents one of the most unequal income distributions in the world, which is attested, for example, by a very high Gini coefficient throughout the 1980s and the 1990s – 0,59 – against an average of 0,50 for other Latin American countries (Barros et al., 2000). Secondly, inequality of student achievement is particularly strong in Brazil, in comparison with developed countries and even some developing countries. According to usual measures of inequality such as the ratio between the first and the ninth deciles of scores or educational Gini, Brazil ranks last (i.e. most unequal country) in an international student assessment exam, the so-called 'PISA 2000', recently organized by OECD. To sum up, the heterogeneity of pupils in Brazil is so strong that we suspect that coefficients calculated as averages, such as those provided by OLS and IV estimations, potentially hide insightful information. More importantly, they are likely to mislead policy-makers by giving an incorrect diagnosis of the relative impact of some inputs, and especially of teachers' wages, in the production of education. Indeed, we have no reason to impose, a priori, that the coefficient associated with the explanatory variable be constant along the distribution of scores.

One way of dealing with potential heterogeneity of the relationship between Brazilian pupils and teachers is to estimate coefficients of samples partitioned by type of school, as we do throughout this paper. However, such a procedure has a drawback: it forces us to drop, in each estimation, a considerable amount of variation in both scores and teachers' wages as we only consider the observations related to roughly one third of the overall sample. Another strategy for dealing with heterogeneity consists of using the quantile regression technique. Eide and Showalter (1998) advocated the use of quantile regression by saying that they "not only addressed the question 'does money matter?", but also "for whom does money matter?". Accordingly, in this paper, while least squares and instrumental variables estimations allow us to answer whether teachers' wages matter in Brazil, quantile regressions allow us to determine to whom they really matter.

Technically, quantile regressions consist of a generalization of the conditional *median* estimation (or least absolute deviation), which is in fact an old statistical technique. It was left aside for many years or only presented as a curiosity in statistical textbooks because of computational difficulties among other reasons (Koenker, 2000). It was then "rediscovered", developed and introduced in the economic literature by Koenker & Bassett (1978). It has been generalized as a method of estimation of conditional quantile functions for any quantile θ of the dependent variable. When estimating quantiles, absolute deviations are given positive and negative weights, in such a way that a fraction θ of the observations will lie below the fitted line while a fraction (1 – θ) will lie above it. As θ goes from 0 to 1, the entire distribution of test scores, conditional on teachers wages (W) and covariates (X and $X^2$), is described.

The $\theta^{th}$ quantile coefficients are obtained as a solution to the expression below:

$$\underset{\beta \in R^k}{Min} \left\{ \sum_{i:Score \geq x_i \beta} \theta \left| Score_i - x_i \beta_\theta \right| + \sum_{i:Score < x_i \beta} (1-\theta) \left| Score_i - x_i \beta_\theta \right| \right\} \qquad \text{(Equation 2)}$$

We thus estimate equation 2 for each quantile we are interested in so as to obtain a set of coefficients for each quantile: $\beta_\theta$.

### *4.1. Related literature and the procedure we adopted*

Some recent papers have been published which use quantile regression to assess the effect of resources on student achievement. Eide and Showalter (1998) estimate the effect of different types of school resources (pupil-teacher ratio, school year length, qualification of teachers, peer effects and per pupil expenditures) on the conditional distribution of performance (test score gains) both by ordinary least squares and by quantile regressions. Their results show that most of the coefficients are not statistically significant *on average* (i.e. by least squares estimation). However, some of them turn out to be statistically significant for some specific quantiles when the quantile regression method is used (e.g. school year length becomes significant for upper tail of distribution).

Later on, other papers have used the quantile regression technique in economics of education. Levin (2001) studies mainly the effect of class size, but also of peer effects, on achievement of Dutch pupils. His results show a strong downward trend in the effect of having more pupils of the same IQ in one's class on achievement as one moves up the achievement distribution. That is, low performing pupils benefit more from ability grouping than average or high performing ones.

Rangvid (2003) estimates peer effects along the conditional distribution of scores for Danish pupils, finding that peer effects are stronger for weak students, and that they decrease over the distribution of scores. In a setting very close to ours, Billger (2002) first uses ordinary least squares and quantile regressions separately, and then combines instrumental variables with quantile regressions in two stages, in order to estimate the effect of teacher pay on student performance in private schools in the US. In the latter formulation, she finds that higher maximum salaries have no significant impact on the measure of student performance she uses.

To our knowledge, our paper is the first to estimate education production functions through the quantile regression technique using Brazilian data. We estimate equation 2 for 5 quantiles of the test score distribution. We estimate the quartiles, including the median (0.25, 0.5, and 0.75) in order to obtain a clear picture of the changing pattern of the effect of wage variables on the distribution of scores. It is also useful to compare the median and the mean coefficients[10]. Additionally, we estimate two extreme quantiles (0.05 and 0.95) to give us an idea of the effect of our explanatory variables on scores for very weak and very strong students.

### *4.2. QR results*

Tables 6 (language) and 7 (mathematics) show QR estimations results for 5 quantiles, including a series of control variables. For each subject, we present results for all schools taken together and for each type of school, and we reproduce OLS coefficients for comparison. Recall that the dependent variable is the students' test score.

When we look at *all schools together*, we have quite different results in language (table 6) and mathematics (table 7). In language, the effect of teachers wages decreases (though not monotonically) with the performance of students. The coefficient drops from a positive and significant level (5,14) for the first estimated quantile ($\theta$=0,05) to a low coefficient (0,74) which is not significantly different from zero for the last estimated quantile ($\theta$=0,95). According to these results, teachers' wages are conditionally correlated to a higher extent with the scores of low-performing students than with those of high-performing students. In mathematics, such a clear decreasing pattern is not observed. Coefficients are small, positive, and significant (at 10% level) for all estimated quantiles, but they go up and down, without substantial shifts, even between the extreme quantiles (2,21 for $\theta$=0,05 against 2,97 for $\theta$=0,95).

[10] The median (QR with $\theta$ =0,5) and the mean (OLS) coefficients may differ for two reasons: (i) the conditional distribution of scores may be skewed: either there are relatively too many observations for which teachers wages are low (implying a low median coefficient), or there are relatively too few observations for which teachers wages are low (implying a high median coefficient), (ii) if there is a considerable number of outliers.

There is a gap between the median coefficient (2,95) and the mean coefficient (2,38) in both disciplines, suggesting that the OLS coefficient is unsuitable to express the conditional correlation of teachers' wages with the scores of weak students. Indeed, as the conditional distribution of scores is skewed (to the left in the case of mathematics, to the right for language), quantile regression should provide more precision for the estimation of the effects at different points of the distribution.

When we turn our attention to particular types of school, the most interesting result is found for *private schools*. In the language test, all estimated coefficients are positive. They are also statistically significant, except for the last one ($\theta$=0,95). The overall decreasing pattern verified for all schools taken together is reproduced here, but the intensity of the estimated effect is stronger, especially for $\theta$=0,05, where the coefficient is 14,39. Recall that this means that a marginal change of 1 standard deviation of teachers' wages is related to a change of 14,39 points (14,39% of standard deviation of scores). Even for the third quartile ($\theta$=0,75), the coefficient is much higher than the all-schools-taken-together analogous result (5,49 versus 2,95). We notice a considerable difference between the median (5,35) and the mean coefficient (6,99). In the mathematics test, none of the extreme quantiles coefficients are statistically different from zero, while the three coefficients corresponding to the quartiles are positive, statistically significant, and presenting a slight overall decreasing trend.

For *public schools*, only 4 out of 20 coefficients are significant at the 10% level, only 2 at the 1% level, and the results are much less revealing. None of the *municipal schools* coefficients is different from zero and no clear increasing or decreasing pattern is observed. In *state schools*, in both subjects, the coefficients for the high-performing students ($\theta$=0,95) are statistically significant, but while it is positive in mathematics (5,83), it is negative in language (-5,96). A very slight decreasing pattern can be found in language, but in mathematics no clear-cut picture is obtained.

Detention of a university degree by the teacher improves pupil performance in language, especially that of the better students. This result seems to indicate that a higher level of skills makes the students more able to reap the benefits of having a university trained teacher. These results do not hold for math test scores, which do not seem affected by the teacher's university education.

All other unobserved aspects of teacher quality, captured by the wages, mainly benefit the low-performing students pointing to some form of merit pay or efficient selection of teacher candidates especially in the private sector. We conclude that, either with the whole sample or with the private schools sample, the hypothesis that there are variations in the conditional correlation of teachers wages with students test scores (the heterogeneity hypothesis) is supported by the coefficients of language estimations. In mathematics, this is true as well but to a smaller extent. So, the achievement impact of an additional unit of teacher wage is not the same for all types of student.

## 5. Correcting for endogeneity: instrumental variables (IV)

### *5.1. Strategy used to take endogeneity of teacher wages into account*

We can not ignore the possibility that good students are assigned essentially better-paid teachers. Individuals whose unobservable characteristics make them high-performing students may well be those who are taught by well-paid teachers. For instance, well-paid teachers might live and work in richer neighborhoods where students would on average perform relatively well in SAEB exams, regardless of the quality of the teaching they are given. This could happen, for example, due to the provision of family support working as a substitute for school inputs. If this is the case, even after controlling for the available observable variables, the coefficients we estimate would not capture the true effect of teachers wages (and the teachers quality they are assumed to represent) on scores. They would measure the combined effect of teacher wages with these unobserved factors, biasing OLS coefficients, possibly upwards.

It may also be the case, instead, that rich parents whose children are not brilliant at school make an effort to offer their children the best available schooling. In order to do so, they could enroll their children in expensive private schools so that well-paid teachers working there could help these not-so-brilliant-but-rich pupils in the endeavor of acquiring a better level of academic skills. Since they are not very talented, however, these students might obtain test scores which are far from outstanding, even though their teachers earn a relatively high wage (and have a relatively high level of skills). In this case, teachers' wages coefficients would be excessively low because of a selection problem, regardless of the quality of teaching that is provided.

There are certainly other possible sources of selection bias in the relationship between teachers' wages and students' scores. Whatever the reason, one may suspect that *teachers' wages are not randomly assigned to different types of students*, even conditioning on all available covariates.

Trying to account for potential endogeneity of teachers' wages, we have estimated an instrumental variable model. In the first stage, we estimate teachers' wages as a function of the same set of variables used in OLS and QR estimations, but also with a set of instruments that are excluded from the main, second-stage, equation. The first step takes the following form:

$$\mathbf{W_i = \eta + \theta X_i + \kappa X^2_i + \lambda I + \mu_i} \qquad \textbf{(Equation 3)}$$

Where W stands for the teacher's wage, X is a vector of control variables, I represents the set of instrumental variables that are excluded in the second step (assumed to be orthogonal to scores) and

$\mu$ is the error term.

In the main equation, we exclude the instruments we assume to be predetermined with respect to students' score, especially the gender of the teacher (gendprof) and his or her years of experience as a teacher (expprof). Firstly, we believe that whether a teacher is a man or a woman is not a factor that is likely to influence students' score directly. On the other hand, there is typically a gender gap in wages, even when experience, age and other factors are controlled for, so that, on average, gender is correlated with wages. Secondly, education production functions frequently provide evidences that a teacher's experience, conditional on covariates, has no systematic significant impact on scores[11]. But experience can be expected to be strongly correlated with wages, especially in the public sector. So, in principle, these two variables seem to fill the required conditions for them to be valid instruments: they are correlated with the variable we suspect to be endogenous and they are correlated with the dependent variable of the main equation only through the channel of the endogenous variable. The set of instruments we use also include the monthly hours of work of a teacher (hoursprof), whether he or she has another job (otherjob), and the squares of the number of years of experience (experience2) and of the number of hours taught per month (hoursprof2). We use the same specification for the two subjects, all of these six instruments being used as predetermined variables.

The second step is exactly the same as the OLS specification, except for the fact that we replace the potentially endogenous variable (wageprof) by an instrumented variable (wagehat).

$$\mathbf{Score_i = \alpha + \beta WHat_i + \delta X_i + \zeta X^2_i + \varepsilon_i} \qquad \textbf{(Equation 4)}$$

Where Score is the performance of pupil i in SAEB test, WHat stands for the predicted value of teacher wage (based on the first stage), X is a vector of control variables and $\varepsilon$ is the error term.

### *5.2. IV results*

Results from the second stage of 2SLS estimations for the complete sample are found in tables 8 and 9, and they can be compared with OLS results (tables 3 and 4). As a first check of the validity of our instruments, we look at the results of first stage estimations. In the first stage, most coefficients of predetermined instruments are statistically significant. More interesting, though, is to look at partial R-squared, as well as the F tests of the predetermined instruments of the first stage regressions. The R-squared is reasonably high for all subjects and types of school (ranging from

[11] Vignoles et al.(2000)

0,13 to 0,18) and all F tests allow us to reject the hypotheses that the coefficients of excluded instruments are not different from zero. As a second stage check, we have computed Sargan's statistic for a test of over-identification. The test yields good results (p-values ranging from 0,28 to 0,40) in language, for all schools and for both private and municipal schools, but not for state schools. In mathematics, p-values are low for overall and municipal schools, and very low for state and private schools, casting doubts on the validity of the set of instruments. For that reason, we concentrate our analysis on the results of the language exam. We have tested some subsets of instruments for each sample, some of which pass the instrument validity checks and yield qualitatively comparable results. However we decided to report here only the results using the complete set for the sake of comparison.

In the language exam, when all schools are taken together, second stage estimation yields a positive and significant coefficient of teacher wages on test scores, as in OLS. In fact, *the coefficient is slightly higher than its OLS counterpart* (4,52 versus 3,30). The same pattern is verified for private schools, with a coefficient that is higher than the OLS one (8,30 versus 6,99), and also statistically significant. The state schools' (more decentralized public schools) and the municipal school's teacher wage coefficients are both insignificant. Detention of a university degree by the teacher has a significant positive effect on language test scores of a magnitude similar to the OLS specification. In mathematics, none of the coefficients is significant. The main qualitative conclusions we had drawn in the OLS section are thus preserved when we instrument the teacher's wage.

## 6. Quantile regressions combined with two-stage-least squares: two-stage absolute deviation (2SLAD)

In subsection 4, we have extended the estimation of conditional mean (OLS coefficients) to the estimation of coefficients that vary along the distribution of conditional scores, through the use of quantile regressions. It is now natural to proceed in an analogous way, extending the IV model of section 5 in a similar manner. The idea underlying this combination of two techniques (quantile regression and two-stage least squares) is to cope simultaneously with both problems that are likely to bias our OLS coefficients, namely the heterogeneity of students and possible endogeneity of teachers' wages. So our aim now is to determine to whom teachers' wages matter in Brazil (QR), but using as explanatory variable a "corrected" regressor (2SLS/IV approach).

This combination of quantile regression with two-stage least squares is called two-stage least absolute deviation (2SLAD). Levin (2001) and Billger (2002) have recently applied it in education production functions, for example. The consistency and asymptotic normality of the 2SLAD estimator has been shown by Amemiya (1982), Powell (1983) and Chen (1996).

The procedure consists of regressing the endogenous variables (teachers wages) on all instruments (including the predetermined ones) in the first stage through ordinary least squares, and then use the fitted values from the first stage as a regressor in the second-stage quantile regression estimation.

So our next step is to estimate the effects of our instrument (wagehat) on student achievement for different points in the test score distribution using 5 quantiles of the test score distribution (0.05, 0.25, 0.5, 0.75 and 0.95) in the second step of our model.

### *6.1. 2SLAD results (combining QR with 2SLS)*

Tables 10 (language) and 11 (mathematics) show 2SLAD estimations results for 5 quantiles, including a series of control variables. For each subject, we present results for all schools taken together and for each type of school and we reproduce 2SLS coefficients for comparison.

The results for mathematics are not quite reliable for the reasons outlined in section 5.2. Moreover, few coefficients can be considered statistically different from zero at usual levels of confidence.

As for the language exam, the results we obtain are similar to those obtained using plain quantile regression: the effect of teachers wages decreases (though not monotonically) with the performance of students, both when all schools are taken together, and when only private schools are picked. This strengthens our former conclusion that teacher quality excluding university education mostly benefits lower performing pupils. Detention of a university degree by the teacher again only affects student test scores in language of the best pupils in private schools. The picture is not very clear for public schools, since there are no significant coefficients.

## 7. Summary and conclusions

Using a more flexible functional form with respect to the ones that are commonly used in the economics of education literature, we investigated whether there is a correlation between teachers' wages and students' achievement in a developing country, Brazil.

We essentially tested three hypotheses in this paper:

(i) that teachers wages, an important and expensive educational input, matter for students achievement,

(ii) that the conditional correlation of teachers wages and scores is stronger in private than in public schools, given that the latter are imposed a greater number of constraints on their recruitment and payment policies, and

(iii) that there are variations in the conditional correlation of teachers wages with students test scores, suggesting the existence of heterogeneity in the pupil-teacher relationship in Brazil.

All our results using the language sample support all three hypotheses:

(i) on average, teachers wages do have a significant effect on students test scores, although the coefficients are quite small,

(ii) private schools (in which, typically rich families enroll their children), coefficients for teachers wages are positive, significant, and higher than public schools and all-schools coefficients,

(iii) the achievement impact of an additional unit of teacher wage is not the same for all types of student: the positive effect of teacher quality (as measured by teacher wage) is most important for low-performing pupils. These results are preserved when we correct for potential endogeneity of teacher wages using instrumental variables and two-stage least absolute deviation, reinforcing the credibility of the results. Nonetheless, the coefficients of IV and 2SLAD generally turn out to be less precise than their OLS and QR counterparts.

In mathematics, our OLS results suggest that teachers' wages do matter for student performance, and they are more pronounced in private schools. QR results also provide evidences of heterogeneity, since estimated coefficients are different across quantiles (coefficients for extreme quantiles are not significant while coefficients for intermediary quantiles are). However, we can not identify a decreasing pattern as clear as in language. OLS and QR results are not repeated in the 2SLS and 2SLAD estimations.

To the contrary, detention of a university degree by the teacher mainly improves the performance of the best pupils. This result seems to indicate that only the best pupils are able to reap the benefits of having a university trained teacher. All other unobserved aspects of teacher quality, captured by the wages, mainly benefit the low-performing students pointing to some form of merit pay or efficient selection of teacher candidates especially in the private sector.

It is not clear from our evidences which particular characteristic of private schools is responsible for the link found between teacher wage and pupil achievement in Brazil. The relation could point to a better functioning of the market in the private school sector, but also to the use of merit pay to a larger extent as compared to the public sector. The latter explanation has been investigated in the literature on merit pay in the US and it appears that failure to use merit pay in the public sector is mainly due to specific circumstances such as the opposition of teacher unions (Ballou, 2001).

A puzzling issue is why the results we obtain for teachers wages become so different from one subject to the other when we instrument teachers wages. These differences may be due to differences in the nature of pupil-teacher relationship or due to different labor market characteristics, such as the gender composition of teachers' population. In fact, an instrument such as teachers' gender was likely to be endogenous in mathematics, while being exogenous in language. The gender of the mathematics teacher seems to have a direct impact on scores (not through the channel of his or her wages). Understanding why this is so, as well as investigating whether this result is also true for other Brazilian data sets or, more generally, for other countries, are topics that require further research.

Finally, it should be mentioned that we do not have the ambition to claim that our results provide a strong causal relationship between teachers' wages and students' scores. But we do claim that our results contain some insightful descriptive static (cross-section) evidence of the relationship between these two variables, since we used a great number of controls across the estimations, and different econometric techniques. In any case, a possible path for further research is to test our three hypotheses again by using different strategies of identification, one of which could consist of exploiting pseudo-panel features of the SAEB datasets, such as Menezes-Filho & Pazzello (2004).

## References


Amemiya T.(1982),'Two stage least absolute deviations estimators', *Econometrica*, Vol.50, pp.689-711.

Angrist J. and Guryan J. (2003),"Does teacher testing raise teacher quality? Evidence from state certification requirements", *NBER Working Paper 9545*.

Ballou D.(2001), 'Pay for performance in public and private schools', *Economics of Education Review* 20, pp.51-61.

Barros, R. P., R. Henriques & R. Mendonça,(2000), "A estabilidade inaceitável : desigualdade e pobreza no Brasil.", Chapter 2 in: Henriques R. (org.) *Desigualdade e pobreza no Brasil*. Instituto de Pesquisa Econômica Aplicada, Rio de Janeiro, Brazil,

Billger S.(2002), "Heterogeneity and Endogeneity in the Teacher Pay and Performance Relationship", *unpublished* , Union College.

Card, D. & A. B. Krueger (1992), "Does school quality matter? Return to education and the characteristics of public schools in the United States", *Journal of Political Economy*, vol. 100 (1), pp. 1-40.

Caroll J.B. (1963), "A model of school learning", *Teachers College Record*, vol. 64, pp.723-733

Case, A., & Deaton, A. (1998). "School Quality and Educational Outcomes in South Africa", *Woodrow Wilson School - Development Studies - Discussion Paper 184*, Princeton.

Chen L-A, and Portnoy S.(1996)," Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equations models.", *Communications in Statistics-theory and Methods* 25, pp.1005-1032.

Creemer, Bert P.M., Scheerens, Jaap, and David Reynolds (2000), Theory development in school effectiveness research, in: Charles Teddlie and David Reynolds (Eds) *The International Handbook of School Effectiveness Research (*Fulmer, 411 p).

Dewey J., T.A. Husted & L.W. Kenny (2000), 'The ineffectiveness of school inputs: a product of misspecification?', *Economics of Education Review*, vol.19(1), pp 27-45

Dolton P. & W. Van der Klaauw (1999), "The turnover of teachers: a competing risks explanation", *The Review of Economics and Statistics*", vol. 81(3), pp.543-552

Eide E. & M. Showalter (1998), "The effect of school quality on student performance: A quantile regression approach", *Economics of Labour* vol. 58, pp.345-350

Figlio, D. N. (1999), 'Functional form and the estimated effects of school resources' *Economics of education review*, vol. 18, pp. 241-252.

Hanushek, Eric E. (1986), 'The economics of schooling: production and efficiency in public schools', *Journal of Economic Literature*, Vol.24, pp. 1141-1177.

______________ (1997), 'Assessing the effects of school resources on student performance: an update', *Educational Evaluation and Policy Analysis*, vol. 19 (2), pp. 141-164.

______________ (2002), 'The failure of input-based schooling policies', *National Bureau of Economic Research Working Paper 9040*.

Hanushek, E., J.F. Kain & S. G. Rivkin (1999), "Do Higher Salaries Buy Better Teachers?", *NBER Working Paper 7082*.

Hayashi, Fumio (2000), *Econometrics*, Princeton University Press: Princeton, USA.

Hoxby C. & A. Leigh (2004), 'Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States', *American Economic Review*, Vol. 94(2).

INEP(2002), "SAEB 2001: Relatório Nacional", Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), Ministério da Educação (MEC), Brazil.

_____(2004), SAEB English webpage - http://www.inep.gov.br/basica/saeb/ingles.htm,

Koenker R.(2000), "Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics", *Journal of Econometrics*, vol.95, pp 347-374

Koenker R. & Bassett G., Jr. (1978), "Regression quantiles", *Econometrica*, vol. 46(1), pp. 33-50.

Lankford H. & J. Wyckoff (1997), "The changing structure of teacher compensation, 1970-1994", *Economics of Education Review*, Vol.16 (4), pp.371-384.

Levin, J. (2001), “For whom the reductions count: A quantile regression analysis of class size aand peer effects on scholastic achievement”, *Empirical Economics*, Vol. 26 (1): 221-246.

Menezes-Filho N.A. & E. Pazzello (2004), "Does money in schools matter? Evaluating the effects of a funding reform on wages and test scores in Brazil", *mimeo*, Universidade de São Paulo, Brazil,

MEC (1999) [Brazilian Ministry of Education] “Balanço do primeiro ano do Fundef”. Report (in Portuguese) available on-line: <http://www.mec.gov.br/sef/fundef/pdf/Aval1998.pdf>.

Murnane R.J. & Olsen R.J.(1990), "The effect of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina", *The Journal of Human Resources*, vol. 25(1), pp. 106-124.

OECD (2001): *Education at a glance, OECD indicators 2003*, OECD. Available on line at: http://www.oecd.org .

Powell J. 1983),'The asymptotic normality of two-stage least absolute deviations estimators', *Econometrica*, Vol.51, pp.1569-1576.

Rangvid, B. S. (2003). “Educational peer effects: quantile regression evidence from Denmark with PISA 2000 data”, unpublished, The Aarhus Schools of Business, Copenhagen, Denmark.

Scheerens, Jaap (1997), Conceptual models and theory-embedded principles on effective schools, *School effectiveness and school improvement*, 8(3), pp. 269-310.

Southwick L. & S. Indermit (1997), "Unified salary schedule and student SAT scores: Adverse effects of adverse selection in the market for secondary teachers.", *Economics of Education Review*, vol. 16(2).

Staiger D. and Stock J.H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica*, Vol.65(3), pp.557-586.

Stinebrickner T.(2001), "A Dynamic Model of Teacher Labor Supply", *Journal of Labor Economics*, vol. 19(1).

Vignoles Anna, Levacic Rosalind, Walker James, Machin Stephen and Reynolds David (2000), "The Relationship between Resource Allocation and Pupil Attainment : A Review", DfEE Research Report. 228.

Wooldrige, J. M. (2002), *Econometric analysis of cross section and panel data*. MIT Press: Cambridge, Massachussetts.

## Appendix 1: Descriptive statistics and estimation results

**Table 1: Descriptive statistics, Language, 8th grade, 2001.**

| | | All schools | | Private Schools | | Municipal schools | | State Schools | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std, Dev, | Mean | Std, Dev, | Mean | Std, Dev, | Mean | Std, Dev, |
| Dependent variables | Ntscore | 244,84 | 51,39 | 276,12 | 46,37 | 227,00 | 45,78 | 227,91 | 45,24 |
| | Score | 500 | 100 | | | | | | |
| | Ntwageprof | 5,04 | 3,16 | 5,99 | 3,66 | 4,74 | 3,05 | 4,32 | 2,33 |
| Teacher Variables | Genderprof | 0,17 | 0,38 | | | | | | |
| | Hoursprof | 116,00 | 46,13 | 118,00 | 47,27 | 120,40 | 46,21 | 110,27 | 44,24 |
| | Expprof | 3,36 | 1,18 | | | | | | |
| | Otherwork | 0,16 | 0,37 | | | | | | |
| | Univprof | 0,90 | 0,30 | 0,95 | 0,21 | 0,89 | 0,31 | 0,86 | 0,35 |
| | Alterwage | 8,17 | 3,31 | | | | | | |
| Pupil characteristics | Gender | 0,46 | 0,50 | | | | | | |
| | Age | 15,04 | 1,66 | | | | | | |
| | Black | 0,07 | 0,26 | 0,03 | 0,18 | 0,10 | 0,30 | 0,08 | 0,28 |
| | Mixed | 0,38 | 0,48 | | | | | | |
| | Retention | 0,57 | 0,85 | 0,24 | 0,58 | 0,77 | 0,93 | 0,75 | 0,92 |
| | Hmwk | 1,93 | 0,98 | | | | | | |
| Family variables | Misced | 3,25 | 1,22 | 4,17 | 0,92 | 2,64 | 1,03 | 2,81 | 1,07 |
| | Nonnuclear | 0,36 | 0,48 | 0,27 | 0,44 | 0,41 | 0,49 | 0,41 | 0,49 |
| | N_maids | 0,59 | 1,00 | | | | | | |
| | Nbooks | 1,49 | 0,68 | | | | | | |
| School variables | Wageprinc | 0,00 | 1,00 | | | | | | |
| | Library | 0,81 | 0,39 | 0,91 | 0,29 | 0,73 | 0,44 | 0,79 | 0,41 |
| | Air | 0,83 | 0,38 | | | | | | |
| | Light | 0,92 | 0,27 | | | | | | |
| | Ncomp | 10,94 | 17,26 | 22,75 | 22,19 | 4,57 | 7,57 | 3,27 | 6,30 |
| | Stratio | 35,73 | 9,90 | | | | | | |
| | Private | 0,36 | 0,48 | | | | | | |
| Local tax variable | Taxpercap | 101,15 | 99,87 | 113,12 | 102,65 | 101,91 | 100,99 | 88,40 | 94,37 |
| Number of observations | | 50492 | | 18015 | | 14776 | | 17701 | |

Short description of variables: ntscore non transformed pupil test scores, ntwageprof: non transformed teacher wages, hoursprof: number of hours working as a teacher per month, expprof: number of years of experience, otherwork: dummy value 1 if teacher has another job on the side, univprof:dummy indicating whether teacher went to university, alterwage: potential alternative wage given gender, education and state, Age:pupil's age in years, black and mixed: self declared ethnic origin, retention: number of times pupil repeated a class, Hmwk: frequency with which the pupil does his homework, misced: level of education attained by the mother, nonnuclear: dummy of value 1 if family if parents are separated, n_maids: number of maids, Nbooks: number of books at home, wageprinc: wage of school director, library: dummy indicating presence of a library in school, air/light: dummies indicating whether the classroom is light and airy enough, ncomp:number of computers available for pupil use, stratio: student/teacher ratio, private : dummy indicating private management of schools, taxpercap: tex per capita perceived in the municipality.

**Table 2: Descriptive statistics, Mathematics, 8th grade, 2001.**

| | | All schools | | State schools | | Municipal Schools | | Private Schools | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std, Dev | Mean | Std, Dev | Mean | Std, Dev | Mean | Std, Dev |
| Dependent variables | Ntscore | 253,86 | 53,76 | 232,34 | 41,82 | 232,80 | 42,97 | 292,24 | 50,74 |
| | Score | 500 | 100 | | | | | | |
| | Ntwageprof | 5,27 | 3,44 | 4,37 | 2,53 | 4,73 | 3,29 | 6,54 | 3,89 |
| Teacher variables | Genderprof | 0,57 | 0,50 | | | | | | |
| | Hoursprof | 122,76 | 46,11 | 115,84 | 43,78 | 124,08 | 45,80 | 128,20 | 47,63 |
| | Expprof | 3,28 | 1,21 | | | | | | |
| | Otherwork | 0,24 | 0,43 | | | | | | |
| | Univprof | 0,87 | 0,34 | | | | | | |
| | Alterwage | 8,17 | 3,31 | | | | | | |
| Pupil characteristics | Gender | 0,48 | 0,50 | | | | | | |
| | Idade | 15,03 | 1,66 | | | | | | |
| | Black | 0,07 | 0,26 | 0,09 | 0,29 | 0,11 | 0,31 | 0,04 | 0,18 |
| | Mixed | 0,38 | 0,48 | | | | | | |
| | Retention | 0,58 | 0,87 | 0,77 | 0,94 | 0,77 | 0,94 | 0,24 | 0,58 |
| | Hmwk1 | 1,88 | 0,99 | | | | | | |
| Family variables | Misced | 3,26 | 1,22 | 2,82 | 1,07 | 2,64 | 1,02 | 4,18 | 0,92 |
| | Nonnuclear | 0,36 | 0,48 | 0,40 | 0,49 | 0,40 | 0,49 | 0,27 | 0,44 |
| | N_maids | 0,59 | 1,00 | | | | | | |
| | Nbooks | 1,49 | 0,68 | | | | | | |
| School Variables | Wageprinc | 0,00 | 1,00 | | | | | | |
| | Library | 0,82 | 0,39 | 0,79 | 0,41 | 0,73 | 0,44 | 0,91 | 0,29 |
| | Air | 0,83 | 0,38 | | | | | | |
| | Light | 0,92 | 0,27 | | | | | | |
| | Ncomp | 10,95 | 17,25 | 3,29 | 6,33 | 4,58 | 7,58 | 22,74 | 22,15 |
| | Stratio | 35,74 | 9,89 | 37,43 | 8,97 | 35,25 | 9,34 | 34,48 | 10,91 |
| | Private | 0,36 | 0,48 | | | | | | |
| Local tax variable | Taxpercap | 101,27 | 100,00 | 88,51 | 94,59 | 101,98 | 101,071 | 113,28 | 102,71 |
| Number of observations | | 50300 | | 17630 | | 14709 | | 17961 | |

**Table 3: OLS results, Language, 8th grade, by type of school.**

**Dependent variable: Student test scores**

| | All Schools | | Private Schools | | Municipal Schools | | State Schools | |
|---|---|---|---|---|---|---|---|---|
| | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, |
| Wageprof | 3,30*** | 0,58 | 6,99*** | 1,12 | 0,53 | 1,19 | 0,79 | 1,12 |
| Univprof | 7,93*** | 1,76 | 13,20** | 4,24 | 16,18*** | 3,68 | 2,11 | 2,98 |
| Gender | -15,24*** | 0,95 | -20,89*** | 1,73 | -12,55*** | 1,94 | -16,39*** | 1,86 |
| Idade | 15,27** | 5,51 | 70,44*** | 13,84 | 7,02 | 10,75 | 7,54 | 10,52 |
| Idade2 | -0,72*** | 0,17 | -2,60*** | 0,46 | -0,53 | 0,33 | -0,51 | 0,33 |
| Black | -15,55*** | 1,93 | -21,56*** | 4,91 | -18,53*** | 3,40 | -11,89** | 3,49 |
| Mixed | -4,13*** | 1,03 | -5,74** | 1,99 | -5,63** | 2,05 | -1,38 | 1,93 |
| Retention | -31,24*** | 1,76 | -50,98*** | 4,34 | -22,71*** | 3,24 | -32,18*** | 3,13 |
| Retention2 | 8,53*** | 0,67 | 11,36*** | 1,94 | 5,70*** | 1,19 | 9,61*** | 1,16 |
| Hmwk1 | 23,43*** | 1,90 | 21,46*** | 3,81 | 25,38*** | 3,73 | 27,75*** | 3,55 |
| Hmwk12 | -4,68*** | 0,53 | -4,24*** | 1,03 | -4,93*** | 1,05 | -5,55*** | 1,00 |
| Misced | -1,85 | 2,25 | 5,57 | 6,03 | -0,29 | 4,63 | -0,27 | 4,37 |
| Misced2 | 1,32*** | 0,35 | 0,37 | 0,82 | 0,92 | 0,79 | 1,29* | 0,72 |
| Nonnuclear | -13,84*** | 1,00 | -18,17*** | 1,97 | -11,23*** | 1,96 | -15,17*** | 1,86 |
| N_maids | -1,87 | 2,05 | 4,59* | 2,78 | -19,45** | 6,14 | -20,11*** | 5,74 |
| N_maids2 | 1,75* | 0,77 | 0,08 | 1,02 | 7,51** | 2,37 | 7,73** | 2,26 |
| Nbooks | 24,10*** | 4,40 | 21,46** | 6,88 | 29,94* | 11,89 | 46,20*** | 10,26 |
| Nbooks2 | -3,86** | 1,16 | -2,86* | 1,73 | -5,67* | 3,36 | -10,14*** | 2,86 |
| Wageprinc | 4,13*** | 0,70 | 5,58*** | 1,23 | 6,21*** | 1,76 | 4,98** | 1,80 |
| Wageprinc2 | -0,49 | 0,56 | 0,15 | 0,97 | -6,37*** | 1,25 | 2,55* | 1,54 |
| Library | 3,45** | 1,28 | 2,43 | 3,05 | 5,36* | 2,39 | 6,86** | 2,40 |
| Private | 37,88*** | 1,53 | | | | | | |
| Air | -1,85 | 1,45 | 1,09 | 3,76 | -3,82 | 2,76 | -0,99 | 2,56 |
| Light | 3,55* | 1,97 | 3,30 | 6,16 | 5,34 | 3,92 | 2,72 | 3,12 |
| Ncomp | 0,49*** | 0,07 | 0,34** | 0,12 | 0,55* | 0,32 | 0,73* | 0,38 |
| Ncomp2 | 0,00*** | 0,00 | 0,00* | 0,00 | 0,01 | 0,01 | -0,03 | 0,02 |
| Stratio | -0,13 | 0,23 | -0,16 | 0,36 | -0,05 | 0,52 | 0,35 | 0,56 |
| Stratio2 | 0,00 | 0,00 | 0,00 | 0,01 | 0,00 | 0,01 | -0,01 | 0,01 |
| Taxpercap | 0,15*** | 0,02 | 0,18*** | 0,04 | 0,02 | 0,04 | 0,20*** | 0,03 |
| Taxpercap2 | 0,00*** | 0,00 | 0,00*** | 0,00 | 0,00 | 0,00 | 0,00** | 0,00 |
| Constant | 371,01*** | 44,15 | -47,20 | 106,34 | 442,68*** | 87,05 | 452,15*** | 84,86 |
| State dummies? | Yes | | Yes | | Yes | | Yes | |
| R squared | 0,34 | | 0,22 | | 0,20 | | 0,17 | |
| Number of obs | 29015 | | 10400 | | 8500 | | 10115 | |

## Table 4: OLS results, Mathematics, 8th grade, by type of school.

**Dependent variable: Student test scores**

| | All schools | | Private Schools | | Municipal Schools | | State Schools | |
|---|---|---|---|---|---|---|---|---|
| Score | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. |
| Wageprof | 2,38*** | 0,59 | 4,02*** | 1,11 | -0,73 | 1,27 | 3,14** | 1,18 |
| Univprof | 1,62 | 1,51 | 8,52* | 3,53 | -0,10 | 3,19 | -0,12 | 2,84 |
| Gender | 26,49*** | 0,90 | 24,85*** | 1,70 | 36,87*** | 1,95 | 36,54*** | 1,86 |
| Idade | -10,36* | 5,28 | 31,31* | 13,59 | -13,60 | 11,02 | -21,26* | 10,58 |
| Idade2 | 0,07 | 0,17 | -1,41** | 0,45 | 0,07 | 0,34 | 0,33 | 0,33 |
| Black | -21,04*** | 1,80 | -29,30*** | 4,79 | -23,43*** | 3,37 | -22,73*** | 3,41 |
| Mixed | -7,81*** | 0,98 | -8,94*** | 1,93 | -7,11** | 2,08 | -8,97*** | 1,96 |
| Retention | -29,99*** | 1,68 | -54,00*** | 4,26 | -29,12*** | 3,26 | -26,53*** | 3,16 |
| Retention2 | 8,28*** | 0,64 | 12,68*** | 1,89 | 8,66*** | 1,21 | 7,17*** | 1,17 |
| Hmwk1 | 17,70*** | 1,79 | 14,01*** | 3,60 | 21,89*** | 3,77 | 18,71*** | 3,57 |
| Hmwk12 | -3,24*** | 0,50 | -1,84* | 0,98 | -3,91*** | 1,06 | -3,73*** | 1,01 |
| Misced | -8,27*** | 2,16 | 3,48 | 6,02 | -8,63* | 4,69 | 3,61 | 4,48 |
| Misced2 | 2,45*** | 0,34 | 1,02 | 0,81 | 2,52** | 0,80 | 0,56 | 0,74 |
| Nonnuclear | -15,95*** | 0,95 | -22,73*** | 1,93 | -15,42*** | 1,99 | -15,50*** | 1,88 |
| N_maids | 8,52*** | 1,96 | 13,05*** | 2,74 | -1,64 | 6,18 | -7,85 | 5,73 |
| N_maids2 | -1,72* | 0,74 | -3,23** | 1,01 | 1,97 | 2,41 | 4,30* | 2,24 |
| Nbooks | 30,18*** | 4,20 | 32,78*** | 6,73 | 43,32*** | 12,17 | 51,56*** | 10,37 |
| Nbooks2 | -5,44*** | 1,11 | -5,81** | 1,70 | -9,08** | 3,46 | -11,32*** | 2,89 |
| Wageprinc | 4,62*** | 0,68 | 7,82*** | 1,21 | 4,10* | 1,78 | 3,98* | 1,87 |
| Wageprinc2 | 0,50 | 0,54 | 0,22 | 0,94 | -4,54*** | 1,26 | 3,46* | 1,60 |
| Library | 2,23* | 1,23 | 2,23 | 3,08 | 7,95** | 2,42 | 4,27* | 2,44 |
| Private | 45,48*** | 1,45 | | | | | | |
| Air | -2,84* | 1,40 | -6,89* | 3,75 | 0,42 | 2,90 | -0,84 | 2,56 |
| Light | 5,35** | 1,93 | 14,68* | 6,16 | 6,23 | 4,17 | 4,21 | 3,18 |
| Ncomp | 0,75*** | 0,07 | 0,79*** | 0,11 | 0,25 | 0,32 | -0,05 | 0,39 |
| Ncomp2 | 0,00*** | 0,00 | -0,01*** | 0,00 | 0,02 | 0,01 | 0,03 | 0,02 |
| Stratio | -0,27 | 0,21 | -0,67* | 0,32 | 0,28 | 0,54 | 0,24 | 0,56 |
| Stratio2 | 0,00 | 0,00 | 0,01 | 0,00 | -0,01 | 0,01 | -0,01 | 0,01 |
| Taxpercap | 0,09*** | 0,02 | 0,15*** | 0,04 | 0,05 | 0,04 | 0,12*** | 0,03 |
| Taxpercap2 | 0,00*** | 0,00 | 0,00** | 0,00 | 0,00* | 0,00 | 0,00 | 0,00 |
| Constant | 554,6 | | 229,75 | | 595,74 | | 653,26 | |
| State dummies? | Yes | | Yes | | Yes | | Yes | |
| R squared | 0,44 | | 0,29 | | 0,24 | | 0,21 | |
| Number of Obs | 28484 | | 10279 | | 8500 | | 9705 | |

## Table 5: Chow test of structural change

."Sometimes called Chow test for structural change" (cf. Hayashi, 2000: 79).

**(Ho: coefficients of additional variables on each unrestricted model = 0)**

| | Language | | Mathematics | |
|---|---|---|---|---|
| | Test 1 | Test 2 | Test 1 | Test 2 |
| unrestricted | score = f(W, X, X²) | score = f(W, W², X, X²) | score = f(W, X, X²) | score = f(W, W², X, X²) |
| restricted | score = f(W, X) | score = f(W, X, X²) | score = f(W, X) | score = f(W, X, X²) |
| F ratio | 32,31 | 0,02 | 35,72 | 1,17 |
| p-value | 0 | 0,88 | 0 | 0,28 |
| Decision | Ho rejected | Ho not rejected | Ho rejected | Ho not rejected |
| Conclusion | Unrestricted is better | Restricted is better | Unrestricted is better | Restricted is better |
| Chosen specification: | Score = f(W, X²) | | Score = f(W, X²) | |

W = teachers' wages
X = all other variables
W² = the square of teachers' wages
X² = the square of all other variables

## Table 6: Quantile regression and OLS, Language, 8th grade.

| | OLS | | Q.05 | | Q.25 | | Q.50 | | Q.75 | | Q.95 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **All schools** | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, |
| Wageprof | 3,30*** | 0,58 | 5,14*** | 1,39 | 3,52*** | 0,79 | 2,84*** | 0,67 | 2,95*** | 0,77 | 0,74 | 1,04 |
| Univprof | 7,93*** | 1,76 | 7,32* | 4,07 | 8,43*** | 2,34 | 6,96** | 2,01 | 7,09** | 2,33 | 11,55*** | 3,22 |
| **Private Schools** | | | | | | | | | | | | |
| Zwprofpriv | 6,99*** | 1,12 | 14,39*** | 3,00 | 6,44*** | 1,47 | 5,35*** | 1,38 | 5,49*** | 1,41 | 2,85 | 1,94 |
| Univprof | 13,20** | 4,24 | 17,60 | 10,91 | 12,58* | 5,48 | 7,17 | 5,22 | 7,31 | 5,39 | 16,61* | 7,89 |
| **Municipal Schools** | | | | | | | | | | | | |
| Zwprofmun | 0,53 | 1,19 | -1,66 | 2,72 | -0,05 | 1,76 | -0,05 | 1,50 | 2,16 | 1,37 | -0,71 | 2,17 |
| Univprof | 16,18*** | 3,68 | 16,41* | 8,28 | 17,47** | 5,44 | 16,31*** | 4,63 | 10,60* | 4,19 | 9,45 | 7,10 |
| **State Schools** | | | | | | | | | | | | |
| Zwprofest | 0,79 | 1,12 | 3,15 | 2,33 | 3,10* | 1,41 | 2,16 | 1,55 | -1,87 | 1,56 | -5,96** | 2,26 |
| Univprof | 2,11 | 2,98 | -2,83 | 6,47 | 3,47 | 3,75 | 1,78 | 4,13 | 7,79* | 4,19 | 5,79 | 6,15 |

## Table 7: Quantile regression and OLS, Mathematics, 8th grade.

| | OLS | | Q.05 | | Q.25 | | Q.50 | | Q.75 | | Q.95 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **All schools** | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, |
| Wageprof | 2,38*** | 0,59 | 2,21* | 1,24 | 2,16** | 0,77 | 2,95*** | 0,71 | 2,23** | 0,71 | 2,97* | 1,22 |
| Univprof | 1,62 | 1,51 | -1,91 | 2,86 | 2,68 | 1,92 | 2,05 | 1,83 | 1,5 | 1,87 | 1,34 | 3,3 |
| **Private Schools** | | | | | | | | | | | | |
| Zwprofpriv | 4,02*** | 1,11 | 4,57 | 2,95 | 5** | 1,67 | 4,78*** | 1,31 | 3,69** | 1,34 | 0,65 | 1,99 |
| Univprof | 8,52* | 3,53 | 1,81 | 8,51 | 12,09* | 5,1 | 7,38* | 4,16 | 9,61* | 4,35 | 16,46** | 6,27 |
| **Municipal Schools** | | | | | | | | | | | | |
| Zwprofmun | -0,73 | 1,27 | 0,54 | 2,39 | 1,28 | 1,76 | -1,12 | 1,54 | -2,19 | 2,05 | -1,23 | 2,61 |
| Univprof | -0,10 | 3,19 | 3,29 | 5,4 | 1,32 | 4,27 | -3,09 | 3,87 | -3,52 | 5,35 | 7,28 | 7,18 |
| **State Schools** | | | | | | | | | | | | |
| Zwprofest | 3,14** | 1,18 | 0,05 | 2,02 | 1,69 | 1,69 | 4,01** | 1,23 | 1,69 | 1,58 | 5,83* | 3,2 |
| Univprof | -0,12 | 2,84 | -4,1 | 4,57 | -0,92 | 4 | -0,42 | 2,97 | 2,97 | 3,86 | 2,57 | 7,99 |

**Table 8: 2SLS results, Language, 8th grade.**

| | All Schools | | Private Schools | | Municipal Schools | | State Schools | |
|---|---|---|---|---|---|---|---|---|
| **First stage (teachers wages)** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** |
| Gendprof | 0,18*** | 0,01 | 0,19*** | 0,02 | 0,15*** | 0,02 | 0,08*** | 0,02 |
| Hoursprof | 0,007*** | 0,0005 | 0,01*** | 0,00 | 0,0005 | 0,001 | -0,004*** | 0,00 |
| Hoursprof2 | -10e-5*** | 0,00 | 3x10e-5*** | 0,00 | 0,0005* | 0,00 | 4x10e-5*** | 0,00 |
| Expprof | 0,15*** | 0,02 | 0,02 | 0,03 | 0,42*** | 0,04 | -0,02 | 0,04 |
| Expprof2 | 0,005 | 0,003 | 0,02** | 0,005 | -0,03*** | 0,01 | 0,04*** | 0,005 |
| Otherwork | -0,09*** | 0,01 | -0,14*** | 0,02 | -0,08*** | 0,02 | -0,09*** | 0,02 |
| Adjusted R-squared | 0,4742 | | 0,5256 | | 0,4721 | | 0,4779 | |
| Partial R-squared | 0,1513 | | 0,1606 | | 0,1455 | | 0,1821 | |
| F, p-value | 848,07 | 0,000 | 324,66 | 0,000 | 236,710 | 0,000 | 367,77 | 0,000 |
| | | | | | | | | |
| **Main equation (test scores)** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** |
| Wageprof | 4,20** | 1,52 | 8,30** | 2,81 | 3,44 | 3,14 | -0,61 | 2,64 |
| Univprof | 7,63*** | 1,90 | 11,8** | 4,37 | 15,83*** | 3,95 | 2,49 | 3,44 |
| R-squared | 0,34 | | 0,22 | | 0,20 | | 0,17 | |
| Overid test: Sargan statistic and p-value | 5,121 | 0,40 | 6,245 | 0,28 | 5,941 | 0,31 | 21,673 | 0,0006 |
| Number of observations | 28605 | | 10238 | | 8397 | | 9970 | |

**Table 9: 2SLS results, Mathematics, 8th grade.**

| | All Schools | | Private Schools | | Municipal Schools | | State Schools | |
|---|---|---|---|---|---|---|---|---|
| **First stage (teachers wages)** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** |
| Genderprof | 0,11*** | 0,009 | 0,17*** | 0,02 | 0,12*** | 0,02 | 0,035* | 0,015 |
| Hoursprof | 0,004*** | 0,0005 | 0,004*** | 0,00 | 0,005*** | 0,0009 | 0,003** | 0,0009 |
| Hoursprof2 | 3x10e-6* | 0,00 | 2x10e-6 | 0,00 | 5x10e-6 | 4x10e-6 | 8x10e-6* | 4x10e-6 |
| Expprof | 0,035* | 0,02 | 0,07* | 0,035 | -0,08* | 0,04 | -0,016 | 0,034 |
| Expprof2 | 0,02*** | 0,003 | 0,006 | 0,005 | 0,04*** | 0,005 | 0,03*** | 0,005 |
| Otherwork | -0,12*** | 0,01 | -0,05** | 0,02 | -0,15*** | 0,02 | -0,2*** | 0,02 |
| Adjusted R-squared | 0,5180 | | 0,5013 | | 0,5657 | | 0,49 | |
| Partial R-squared | 0,1527 | | 0,1246 | | 0,1751 | | 0,1844 | |
| F, p-value | 837,23 | 0,000 | 238,08 | 0,000 | 293,34 | 0,000 | 355,66 | 0,000 |
| | | | | | | | | |
| **Main equation (test scores)** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** | **Coef,** | **Std, Err,** |
| Wageprof | 0,30 | 1,51 | -2,64 | 3,17 | 0,44 | 3,13 | 2,14 | 2,75 |
| Univprof | 2,42 | 1,63 | 10,74** | 3,66 | 0,86 | 3,42 | -1,15 | 3,16 |
| R-squared | 0,44 | | 0,29 | | 0,24 | | 0,21 | |
| Overid test: Sargan statistic and p-value | 9,623 | 0,09 | 11,495 | 0,04 | 8,985 | 0,11 | 18,393 | 0,002 |
| Number of observations | 27942 | | 10095 | | 8350 | | 9497 | |

**Table 10: Quantile regression and 2SLS (2SLAD), Language, 8th grade.**

| | 2SLS | | Q.05 | | Q.25 | | Q.50 | | Q.75 | | Q.95 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **All schools** | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, |
| Wageprof | 4,20** | 1,52 | 10,89** | 3,63 | 5,08* | 2,22 | 3,67* | 1,87 | 2,77 | 1,89 | -2,1 | 2,78 |
| Univprof | 7,63*** | 1,90 | 6,53 | 4,47 | 7,48** | 2,77 | 7,1** | 2,33 | 8** | 2,35 | 13,8*** | 3,48 |
| **Private Schools** | | | | | | | | | | | | |
| Zwprofpriv | 8,30** | 2,81 | 17,89* | 8 | 10,92** | 3,91 | 5,08 | 3,86 | 4,50 | 3,54 | -4,16 | 4,85 |
| Univprof | 11,8** | 4,37 | 18,37 | 12,65 | 10,11* | 6,2 | 4,33 | 5,98 | 5,55 | 5,45 | 19,55* | 7,61 |
| **Municipal Schools** | | | | | | | | | | | | |
| Zwprofmun | 3,44 | 3,14 | 10,25 | 7,12 | 4,22 | 4,36 | 0,50 | 3,87 | 4,64 | 4,06 | 0,18 | 6,29 |
| Univprof | 15,83*** | 3,95 | 8,18 | 9,04 | 15,98** | 5,54 | 15,22** | 4,83 | 12,53* | 5 | 9,19 | 7,77 |
| **State Schools** | | | | | | | | | | | | |
| Zwprofest | -0,61 | 2,64 | -5,48 | 5,38 | 1,45 | 3,7 | 3,49 | 3,76 | -1,59 | 3,59 | -3,82 | 5,82 |
| Univprof | 2,49 | 3,44 | -0,76 | 7,19 | 4,38 | 4,74 | 0,06 | 4,89 | 7,24 | 4,65 | 3,63 | 7,46 |

**Table 11: Quantile regression and 2SLS (2SLAD), Mathematics, 8th grade.**

| | 2SLS | | Q.05 | | Q.25 | | Q.50 | | Q.75 | | Q.95 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **All schools** | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, | Coef, | Std, Err, |
| Wageprof | 0,30 | 1,51 | -4,48 | 2,85 | -1 | 2,08 | 3,76* | 1,93 | 1,98 | 1,85 | -2,79 | 3,26 |
| Univprof | 2,42 | 1,63 | 0,59 | 2,98 | 3,73* | 2,21 | 1,88 | 2,08 | 2,03 | 2 | 3,23 | 3,6 |
| **Private Schools** | | | | | | | | | | | | |
| Zwprofpriv | -2,64 | 3,17 | -9,81 | 8,61 | -5,58 | 4,75 | -0,42 | 3,9 | 2,1 | 4,03 | -5,25 | 5,59 |
| Univprof | 10,74** | 3,66 | 13,44 | 9,29 | 14,57** | 5,51 | 9,57* | 4,50 | 10,32* | 4,72 | 17,6** | 6,6 |
| **Municipal Schools** | | | | | | | | | | | | |
| Zwprofmun | 0,44 | 3,13 | -0,4 | 5,69 | 0,70 | 4,33 | 4,70 | 4,16 | 2,68 | 4,76 | -5,19 | 7,39 |
| Univprof | 0,86 | 3,42 | 7,75 | 5,85 | 3,09 | 4,77 | -4,11 | 4,68 | -6,79 | 5,645 | 7,36 | 8,65 |
| **State Schools** | | | | | | | | | | | | |
| Zwprofest | 2,14 | 2,75 | -8,28* | 4,53 | 0,63 | 3,86 | 7,13* | 3,43 | 2,37 | 3,39 | -6,26 | 6,75 |
| Univprof | -1,15 | 3,16 | -0,33 | 5,05 | -2,17 | 4,36 | -3,67 | 3,93 | 1,06 | 3,91 | 3,04 | 7,86 |

## Appendix 2: Computation of wages variables

*1.From 'reais' to an index*: Categories of gross monthly wages of principals and teachers, which were originally expressed in Brazilian currency (1 real, 2 reais...), have been transformed by us into categories of an index used by Brazilian administration, the so-called 'salário mínimo' (sm used as shorthand; in English, it means ‘minimum wage’). We used the ratio 1sm = 180,00 reais, which corresponded in October 2001, when data was collected, to about 68 US dollars.

*2. Attributing values to the categories*: We have then attributed a value to each category of wage. For example, category 1 was 'up to R180,00 reais'. In step 1, this has been converted into 'up to 1sm'. In step 2, the value 0,5 has been attributed to every observation in this category, such that the wage is now 0,5 sm. The variables created have been called ‘ntwageprof’ (teachers’ wages, present in tables 1 and 2) and ‘ntwageprinc’ (principals’ wages).

*3. Standardization of gross monthly wages*: Finally, in order to make the interpretation of the coefficients straightforward, we standardized the variables created in step 2 above, with a mean of 0 and a standard deviation of 1, and called them wageprof (teachers’ wage) and wageprinc (principals’ wage). These variables are used in our estimations.

Département des Sciences Économiques
de l'Université catholique de Louvain
Institut de Recherches Économiques et Sociales

Place Montesquieu, 3
1348 Louvain-la-Neuve, Belgique

ISSN 1379-244X D/2005/3082/008