Teacher shortages, teacher contracts and their impact on education in Africa

Bourdon, J ; Frölich, M ; Michaelowa, Katharina

Abstract: Primary school enrolment rates are very low in francophone Africa. In order to enhance education supply, many countries have launched large teacher recruitment programmes in recent years, whereby teachers are no longer engaged on civil servant positions, but on the basis of (fixed-term) contracts typically implying considerably lower salaries and a sharply reduced duration of professional training. While this policy has led to a boost of primary enrolment, there is a concern about a loss in the quality of education. In this paper we analyse the impact on educational quality, by estimating nonparametrically the quantile treatment effects for Niger, Togo and Mali, based on very informative data, comparable across these countries. We find that contract teachers do relatively better for low ability children in low grades than for high ability children in higher grades. When positive treatment effects were found, they tended to be more positive at the low to medium quantiles; when negative effects were found they tended to be more pronounced at the high ability quantiles. Hence, overall it seems that contract teachers do a relatively better job for teaching students with learning difficulties than for teaching the 'more advanced' children. This implies that contract teachers tend to reduce inequalities in student outcomes. At the same time, we also observe clear differences between the countries. We find that, overall, effects are positive in Mali, somewhat mixed in Togo (with positive effects in 2nd and negative effects in 5th grade) and negative in Niger. This ordering is consistent with theoretical expectations derived from a closer examination of the different ways of implementation of the contract teacher programme in the three countries. In Mali and, to some extent, in Togo, the contract teacher system works more through the local communities. This may have led to closer monitoring and more effective hiring of contract teachers. In Niger, the system was changed in a centralized way with all contract teachers being public employees, so that there is no reason to expect much impact on local monitoring. In addition, the extremely fast hiring of huge numbers of contract teachers may also have contributed to relatively poor performance in Niger. These results are expected to be relevant for other sub-Saharan African countries, too, as well as for the design of new contract teacher programmes in the future.

DOI: https://doi.org/10.1111/j.1467-985X.2009.00601.x

Originally published at:
Bourdon, J; Frölich, M; Michaelowa, Katharina (2010). Teacher shortages, teacher contracts and their impact on education in Africa. Journal of the Royal Statistical Society: Series A, 173(1):93-116.
DOI: https://doi.org/10.1111/j.1467-985X.2009.00601.x

This is the authors' accepted manuscript including the supplementary appendix. The paper was published in the Journal of the Royal Statistical Society Series A, 2010, vol. 173, issue 1, 93-116, https://doi.org/10.1111/j.1467-985X.2009.00601.x

Teacher Shortages, Teacher Contracts, and their Impact on Education in Africa

Jean Bourdon

Institute for Research on the Sociology and Economics of Education (IREDU), jbourdon@u-bourgogne.fr

Markus Frölich

University of Mannheim, IZA, Bonn and IFAU, Uppsala, froelich@uni-mannheim.de

Katharina Michaelowa

University of Zurich, Center for Comparative and International Studies (CIS) and HWWI Hamburg, katja.michaelowa@pw.uzh.ch

Summary. In order to enhance primary enrolment rates, many African countries have launched large teacher recruitment programmes in recent years. Given tight budgetary constraints, teachers are no longer employed in civil service positions, but on the basis of (fixed-term) contracts typically implying considerably lower salaries and a sharply reduced amount of professional training.

In this paper we analyse the impact of this change on educational quality in Niger, Togo, and Mali, based on very informative data, comparable across these countries. We use a variety of estimation techniques, including a nonparametric estimation of quantile treatment effects. Our results demonstrate that contract teachers tend to reduce inequalities in student outcomes. Overall, effects are positive in Mali, somewhat mixed in Togo and negative in Niger. This ordering is consistent with theoretical expectations related to the manner in which contract teacher programmes were implemented differently in each of the three countries under study.

JEL classifications: O15, I 21, C14

Keywords: contract teachers, primary education, Africa, quantile treatment effects, matching estimators

1. Introduction

With a rapidly increasing youth population and swift increases in the proportion of children attending school, many developing countries are facing serious difficulties recruiting and financing qualified teachers. In response, many African countries have been experimenting with alternative teacher training and recruitment programmes, employing "*contract teachers"* instead of traditional civil servants. These new teachers are usually hired on fixed-term contracts with shorter training and lower remuneration. Similarly, contract teachers have also been employed in Latin America and South Asia (for an overview, see Duthilleul 2005). Despite the widespread introduction of contract teacher programmes in many countries, from the mid-1990s onwards, a detailed microeconometric evaluation concerning the effects of this policy change has not yet been conducted.

In this paper, we analyse the effect of the reform process on student achievement in three francophone sub-Saharan African countries - Mali, Niger, and Togo. To this end, we employ methods from the treatment evaluation literature. The three countries under examination show many social, economic, and cultural similarities and all introduced contract teacher programmes to cope with similar problems. However the way these programmes have been implemented, and also the characteristics of the actual contracts used, vary considerably.

Earlier studies of individual African countries have produced divergent results. This suggests that specific characteristics may determine the success or failure of these programmes. This calls for a direct country-comparison using comparable data and a uniform estimation approach. It is the objective of this paper to provide this comparison. In addition, this paper attempts to improve upon the estimation techniques used in earlier studies. In a first step, we estimate average treatment effects and compare the results of conventional regression analysis with the results of different instrumental variable approaches, and with the results using non-parametric estimation techniques. In a second step, we proceed by estimating quantile treatment effects. Nonparametric estimations avoid functional form assumptions. This feature is advantageous for current purposes, and produces detailed information on the effect of contract teachers on different segments of the student population.

This paper is strongly related to the literature on teacher incentives and working conditions. This literature covers performance pay, teacher monitoring, and monitoring by local communities. In addition, the impact of teacher quality and its relation to academic and professional qualifications has recently received increased attention. For a review of this literature with a focus on developing countries, see Glewwe and Kremer (2006) or Wößmann (2005).

This paper is also linked to an ongoing policy debate in many African countries. The introduction of contract teacher programmes often implies reduced pre-service training and lower salaries for new teachers. Stakeholders in the education system fear that these cuts will result in diminished education quality. This paper analyses to what extent these fears are justified, and provides evidence on which forms of contracts (or which types of programme implementation) exacerbate or mitigate potential problems.

Section 2 provides background information on the implementation of contract teacher programmes. Section 3 discusses the data set. Section 4 explains the estimation approach, and Section 5 presents the results. Section 6 provides conclusions and recommendations for education policy.

2. Teacher incentives, teacher contracts, and working conditions

This section provides background information on contract teacher programmes (for more details, see Bourdon, Frölich and Michaelowa 2007). In most countries, teachers are employed as civil servants. Exceptions include supply teachers, or teachers on probation. The broad-based introduction of *contract teacher* schemes in sub-Saharan Africa around the turn of the century breaks with this tradition. While contract teachers exist in other parts of the world, and were also introduced much earlier in many African countries, this occurred only on a limited scale and on the initiative of parents and local communities. In order to enhance the public supply of primary education while simultaneously coping with high population growth and tight budget constraints, contract teacher programmes were introduced rapidly in many countries. In the early years of the new century – the reference period for the empirical analysis that follows – contract teachers already constituted a sizeable share of the total teaching population (for current shares, see Table 1).

The contract teacher programme was introduced in Niger in 1998. From 1998 onwards, no new civil servants, but only contract teachers were employed in the teaching profession at the primary education level. Further, for many years prior to the programme's introduction, funding had been insufficient to train and recruit teachers according to earlier rules and regulations – including salaries amounting to almost 10 times GDP per capita (UNESCO-UIS 2006, p. 87).It was therefore necessary for newly engaged contract teachers to close the existing gaps (see the recent publication from the Programme on the Analysis of Education Systems, PASEC 2005).In the year 2000, contract teachers made up the majority of the primary teacher population. Indeed the long-term maintenance of teacher salaries far above market rates effectively paved the way for radical change in teacher employment policies at the moment the traditional system collapsed. Unsustainable salary levels triggered the contract teacher reform and created a situation in which contract teachers earn only one third of those in the civil service.

In Mali and Togo, changes were somewhat more moderate than in Niger and also began slightly earlier. In both countries, reform was initially driven by local communities who engaged (private) contract teachers in response to the failure of the state to provide the required staff. It should also be noted that, while the policy shift towards the employment of contract teachers was certainly less abrupt than in Niger, the salary differential was much greater. This is particularly the case in Mali, where community teacher salaries amount to only 15% of traditional teacher salaries.

In Togo, from the mid 1990s onwards, many community schools were formally recognised. In consequence, the government began financing community school teachers as contract teachers from the central budget and also recruiting public contract teachers for other schools (World Bank 2002).

Table 1 synthesises the aforementioned characteristics, and shows Mali, Niger, and Togo in their regional context. The table demonstrates that the engagement of contract teachers has indeed become a wide spread phenomenon throughout the region. Table 1 further illustrates that contract conditions in terms of remuneration, and also in terms of the authorities responsible for contract conditions and monitoring (public or private), vary considerably across countries. Privately employed contract teachers earn much lower salaries than traditional teachers in all countries, and often much less even when compared to new public contract teachers. In addition, privately engaged contract teachers do not only face the potential challenge of a fixed-term contract, which may or may not be renewed. These teachers must also expect much closer monitoring by parents in the local community who (directly or indirectly) finance their post.

Further cross-country differences in contract teacher programmes exist with respect to entry requirements in terms of educational attainment and professional training. Typically, professional training has been considerably reduced from several years in specialised teacher training institutes to a few months, or even weeks. Training is provided by diverse institutions or under the supervision of senior teachers who provide training ‘on-the-job.’ In Mali and Niger, pre-service training requirements were reduced to 3 months and 45 days respectively, while in Togo, contract teachers do not necessarily receive any professional training at all. Cross-country differences are less relevant with respect to the level of educational attainment required to enter the teaching profession. In Mali and, to some extent, in Togo, upper secondary education is required as a minimum, while lower secondary attainment is sufficient in Niger in addition to the successful completion of an entrance exam.

Table 1: Distribution and remuneration of primary teachers according to their employment status

	Breakdown by employment status (%)			**Wages relative to GDP per capita**				
Country	Civil servants	Contract teachers		Civil servants			Contract teachers	
		public[1]	private[2]	Total	Full	Assistants[3]	public[1]	private[2]
Benin (2005)	54,7	16,4	29,0	5,2	5,7	3,9	2,1	1,1
Burkina Faso (2002)	64,1	23,6	12,2	5,8	7,1	5,1	5,6	2,2
Cameroon (2002)	34,9	20,4	44,7	5,3	5,7	4,1	1,4	0,8
Chad (2003)	38,4	17,2	44,4	7,4	8,2	6,0	1,7	0,4
Congo, Rep. of (2005)[4]	55,0	14,0	31,0	2,8	2,9	2,62	1,3	na
Guinea (2003)	30,9	38,9	30,1	3,4	3,5	2,7	1,9	1,2
Ivory Coast (2001)	87,3	0,0	12,7	4,8	5,0	3,0	-	-
Madagascar (2003)	46,1	0,0	53,9	4,4	-	-	-	1,0
Mali (2004)	35,7	34,7	29,6	7,5	-	-	4,8	1,0
Niger (2003)	46,0	50,2	3,8	8,9	10,5	8,0	3,5	-
Senegal (2003)	43,6	41,5	15,0	5,7	6,2	4,9	2,6	na
Togo (2001)	35,0	30,5	34,6	6,4	7,8	5,4	3,3	1,3
Average (12 countries)	47,6	24,0	28,4	5,6	6,2	4,5	2,8	

[1] *Public*: under contract with public authorities
[2] *Private*: engaged by and/or under contract with parents or local communities. This does not always correspond to the local terminology. In Togo, for instance, community teachers tend to be classified officially as “public” while the term “private” refers exclusively to expensive and well equipped schools run by other external providers such as the church.
[3] *Assistants* are public employees engaged as a support of full teachers.
[4] In Congo, salaries are calculated relative to GDP in 2003 in order to avoid the artificial effect of the change in petroleum prices in 2005.
Source: World Bank (Africa Region) and Pôle de Dakar (2007, p. 66), slightly updated by the authors.

We now consider how the aforementioned features of contract teaching programmes may affect student learning. Theoretically, we must consider the following potential effects: (1) the effect of new educational and training requirements for entry into the teaching profession, (2) the incentive effect of the teaching contract, (3) a selection effect (changed demand for and supply of new teachers), and (4) a dynamic effect.

(1) The effect of new education and training requirements

Despite the rather ambiguous results of recent research (see e.g. Hanushek et al. 2005, and Michaelowa and Wechtler 2006), one would expect that general education as well as professional training prior to job entry has a positive impact on teacher performance. Togolese contract teachers, many of whom have not received any professional training at all, should be particularly disadvantaged.

(2) The incentive effect of the teaching contract

There are two possible directions of this effect. On the one hand, the unfavourable conditions of new teacher contracts could be regarded as unfair and demotivating, and short-term contracts could prevent personal investments in pedagogical training and school specific human capital. At the time of data collection, Mali exhibited the greatest salary differential. In this respect, the Malian programme may be the least conducive to teaching quality.

On the other hand, contract teacher programmes may also bring about incentives conducive to better teaching. For teachers on non-permanent contract positions, further employment prospects depend on performance. Among the three countries considered, non-permanent contracts exist for all new contract teachers in Niger, for community teachers in Mali, and for the majority of community teachers in Togo. Moreover, employment in the local community can be expected to induce an additional and particularly relevant incentive effect. Community teachers are usually selected and paid by parents so that the latter have a high incentive to monitor. In this respect, the teachers depend directly on parental satisfaction. This should ensure at least a minimum standard of performance, such as the regular appearance of teachers at their workplace.

If true, we should expect contract teacher programmes in Mali and Togo to be at an advantage over the programme in Niger, which relies on public contract teachers alone. In addition, comparing Mali and Togo, we might expect the Malian system to be at a relative advantage, both due to the higher share of community teachers among the contract teachers, and due to the greater autonomy of Malian community schools given the increasing involvement of the Togolese government.

(3) The selection effect

Changed employment conditions could lead to a different composition of candidates applying for teaching positions. Reduced entry requirements could reduce entry costs and increase the attractiveness of (temporary) teaching positions. However, inferior contract conditions may reduce the number of highly skilled candidates. In addition, higher demand for teachers would lead us to expect lower quality among marginal (newly employed) teachers. Given the strong acceleration of the recruitment process, the latter is likely to dominate.

This effect may be substantial because in all countries, newly engaged teachers represent a significant percentage of young adults with at least lower secondary education attainment. The annual increase in teacher recruitment currently represents around 10% of qualified graduates in Mali and Togo, and over 20% in Niger (World Bank 2002, p. 77, 2004, p. 97, and 2006, p. 141).

(4) The dynamic effect

Inferior contract conditions may reduce the length of time teaching staff are retained and act to increase turnover. This effect could lead to a different distribution of job experience before and after the reform, with a higher proportion of young and inexperienced teachers. This is likely to have an impact on teaching quality.

However, the relevance of this effect depends crucially on general labour market conditions. In all three countries considered here, alternative employment opportunities in the modern sector are extremely limited. Currently, only 25% of secondary graduates in Mali and Niger, and only 10% in Togo, find jobs in the formal, non-agricultural economy (World Bank 2002, p. 77, 2004, p. 97, and 2006, p. 141). The dynamic effect does not therefore seem to hold high relevance in any of the three countries at the present time.

A summary of the different effects expected in the countries under consideration is presented in Table 2.

Table 2: Expected effects of contract teacher programmes

	(1) Training effect	(2) Incentive effect	(3) Selection effect	(4) Dynamic effect
Mali	– reduction to 3-months pre-service	+++contract renewal may depend on performance; community monitoring	– limited teacher supply	Not relevant given current labour market conditions

		– – salary reductions leading to lowest salaries		
Niger	– reduction to 45-days pre-service	+ contract renewal may depend on performance – salary reduction	– – extremely limited teacher supply	Not relevant given current labour market conditions
Togo	– – no regular pre-service training	++ contract renewal may depend on performance; some community monitoring – salary reduction	– limited teacher supply	Not relevant given current labour market conditions

Note: The number of +/– signs reflects the expected strength of each positive (+) and negative (-) effect relative to the same effect in the other countries.

Taken together, the above arguments on education and training, incentives, and selection effects, lead us to expect that Niger will face the greatest difficulties with its contract teacher programme. This is in line with at least the latter two theoretical expectations discussed above. In Niger, positive incentives are limited as the programme is fully anchored in the public administration system. Moreover, the selection effect can be expected to be negative due to the limited supply of qualified candidates. In Mali, the most critical issue appears to be the incentive effect. If there does exist an important disincentive related to low salary levels, Malian contract teachers should perform very badly. However, if the positive incentive effect related to parental responsibility and community monitoring dominates, Mali should do rather well. The case of Togo can be expected to lie somewhere in between, with a marked disadvantage only regarding its failure to provide pre-service teacher training on a regular basis.

In the econometric part of this paper we examine whether these expectations are confirmed by the empirical evidence. One of the main concerns for the empirical analysis is that the schools and classes we observe as having a contract teacher may be different in relevant ways from those that we observe with regular teachers. These schools and classes may differ in observed and unobserved characteristics whose effects on education quality may confound those of teacher type. We follow a strategy of controlling for many of these characteristics by taking advantage of very rich PASEC data, discussed in the next section. This data provides detailed information on schools, teachers, and students including their test scores at the *beginning* and at the *end* of the school year.

3. Data and initial descriptive statistics

Empirical study of the effect of contract teacher programmes on education quality requires comprehensive information on teachers, schools, and students. The PASEC programme collects such data for the 2nd and 5th grade of primary schools in francophone sub-Saharan Africa. PASEC uses student, teacher, and director questionnaires that are uniform for a number of core questions. Therefore, results are comparable across countries. Education quality is measured in terms of student achievement in Mathematics and French, which is assessed using standardized tests for all three countries considered. The Mathematics test contains a wide variety of items ranging from numeracy, problem solving (application to situations of daily life), and simple geometry. The French test covers general understanding and orthography as well as grammar skills. Tests were administered in the classroom, item by item, following detailed instructions on the way to present each question and the time to be allocated to each response. Test results are coded in terms of the percentage of test items answered correctly in each of the two subjects, French and Maths. The tests and their results are not used for any official purpose, i.e. teacher assessments, and the final dataset preserves the anonymity of schools, teachers, and students. The tests comprise a majority of multiple choice items. The testing language is French.

Students are tested both at the beginning and at the end of the school year using a pre-test in autumn and a post-test in summer. This is particularly relevant for our study because the effect of a contract teacher who may have taught the students only in the year of assessment needs to be distinguished from the effect of various other teachers who taught the class before. The pre-test score can act as control for this. In francophone Africa, teachers tend to teach the same grade or Year and do not follow the cohort. In primary school, a single teacher usually teaches all relevant subjects in a given class. Consequently the subject matter that students have learned over the year can be attributed relatively well to this particular person.

PASEC surveys were carried out in Niger and Togo during the academic year 2000/2001, and in Mali in 2001/2002. In all cases, the sampling frame consisted of all primary school teachers included in the database available from the relevant national Ministry of Education. Table 3 shows the actual number of classes contained in the data set for each country, separately for 2nd and 5th grade. In addition to regular civil servant teachers and contract teachers, there are also some other types of teachers, which include teaching assistants and interns. These teachers are dropped in the analysis as we aim to compare the contract teachers with a well defined control group of regular civil servant teachers. In addition to this, we also drop teachers with more than 10 years of job experience for common support considerations. Common support refers to all values of the characteristics *X* that are observed among both the contract and the regular teachers. Nonparametric estimation of the treatment effect can only be done for those values of *X* that can be observed in both treatment states. If a particular value of *X* is observed only among the contract teachers, it would be impossible to infer the treatment effect since no comparable regular teacher exists.

Since the reforms were enacted only relatively recently there cannot be any new contract teachers with more than 10 years of experience. Hence, the support of the variable job experience differs between civil servant and regular teachers and we need to impose a common support restriction for the nonparametric estimation approach.

Table 3: Number of teachers in the datasets

	Niger		Togo		Mali	
	2nd grade	5th grade	2nd grade	5th grade	2nd grade	5th grade
No. of classes	125	140	116	119	139	140
Contract teachers	59	27	40	42	74	50
Civil servants	58	92	70	64	48	76
Other teachers	8	21	6	13	17	14
After deleting 'other teachers' and teachers with more than 10 years of job experience						
Contract teachers	59	27	38	42	73	50
Civil servants	33	45	23	28	10	12

More precisely, in Togo we observe contract teachers with up to eleven years of job experience. In Mali, we observe contract teachers with up to eight years of job experience. The situation is somewhat different in Niger, where the reform was enacted even more recently, in 1998. There, we observe contract teachers with up to only four years of job experience, except one with five years and three with eight years. As we attempt to adhere to the same sample and variable definitions across countries to keep results comparable, we use 10 years as the cut-off point in all countries for the main analyses. (We obtain very similar results when only teachers with at most 4 years of job experience are retained.) Due to this sample restriction only few regular teachers remain in Mali, which is likely to lead to rather imprecise estimates.

Table 4 provides an overview of selected characteristics regarding students, teachers, classrooms, and schools. This data was collected through the use of additional questionnaires in which students, teachers, and principals acted as respondents. Pupils were questioned in tandem with the pre-test at the beginning of the school year, whereas teachers and principals were interviewed at the end of the year. The questionnaires provide information on a wide range of the students' personal characteristics, including family background and prior educational history. The questionnaires also generate information on the personal characteristics and pedagogical methods of both teachers and principals, and on school and classroom equipment, inspections, and the interaction of school staff with the local community.

Table 4 shows averages for classes taught by contract teachers and civil servant teachers. In both 2nd and 5th grade, students taught by traditional teachers typically achieved a higher (or about equal) percentage of correct answers than students taught by contract teachers. Only in the case of Mali 2nd grade Mathematics, do we observe strongly higher scores with contract teachers. However, these differences may be related to factors other than the teachers' contract status. They may, for instance, be a consequence of the assignment of contract teachers to different learning environments, or to different characteristics of the teachers themselves. In particular, it might be that contract teachers work with children who show lower performance at the beginning of the school year. The data provide some evidence for this in Togo 5th grade and in Mali 2nd grade. Moreover, at least in Mali and Togo, our data show a more favourable context for traditional teachers in terms of the socio-economic background of their students and school equipment. In addition, we tend to find contract teachers more often in rural than in urban schools.

It should be noted here, that PASEC attempted to match schools with contract teachers by assigning comparable comparison schools and classes (see Bourdon, Frölich and Michaelowa 2007). This explains why we find only minor differences in terms of students' socio-economic background and educational resources. In addition, our modifications to the sample in order to ensure common support brought

Table 4: Selected characteristics of teachers, students and schools in our sample, by country, grade and teacher status

Variable (range)	Mali				Niger				Togo			
	Grade 2		Grade 5		Grade 2		Grade 5		Grade 2		Grade 5	
Type of teacher	Civil servants	Contract teachers	Civil servants	Contract teachers	Civil servants	Contract teachers	Civil servants	Contract teachers	Civil servants	Contract teachers	Civil servants	Contract teachers
Number of students	123	913	146	668	429	709	584	336	245	422	308	464
Number of classes	10	73	12	50	33	59	45	27	23	38	28	42
Test scores: post-test French	0.440	0.430	0.319	0.338	0.495	0.400	0.278	0.258	0.584	0.581	0.518	0.389
Test scores: post-test Maths	0.355	0.419	0.319	0.354	0.470	0.383	0.309	0.280	0.508	0.515	0.503	0.394
Test scores: pre-test French	0.292	0.192	0.33	0.308	0.182	0.157	0.241	0.233	0.387	0.385	0.551	0.413
Test scores: pre-test Maths	0.439	0.323	0.380	0.355	0.390	0.325	0.271	0.267	0.491	0.490	0.620	0.545
Socio-economic index based on family possessions (1-8)	0.90	0.62	0.96	0.60	2.50	2.37	2.34	2.87	2.65	1.95	3.71	1.96
Index of school equipment (0-16)	5.48	4.85	4.82	4.90	3.78	3.90	3.77	4.31	4.29	3.00	5.68	3.23
Class size	71.41	63.51	58.77	54.33	46.00	45.68	34.43	36.14	40.46	38.36	32.81	36.83
School located in rural area (0,1)	0.10	0.50	0.36	0.52	0.34	0.38	0.30	0.41	0.40	0.52	0.31	0.58
Teachers' age (in years)	35.01	29.83	36.53	31.93	30.08	27.79	30.07	28.74	34.62	32.52	36.01	34.09
Teachers' job experience (in years)	8.43	3.12	7.04	3.82	4.88	2.44	5.58	2.70	6.20	5.47	7.04	5.87
Teachers' educational attainment (0-6)	2.34	1.61	2.34	2.15	4.01	3.72	4.51	4.18	2.86	2.88	3.82	4.00
Teachers without pre-service training (0,1)	0.08	0.34	0.00	0.35	0.03	0.19	0.00	0.15	0.18	0.41	0.22	0.44

Note: The first column defines the variables and the values in brackets indicate the scale of the measurement. For example, Index of school equipment is measured on a scale from 0 to 16. Binary variables are indicated by (0,1). The subsequent columns provide the sample averages. For binary variables this represents the proportions with this attribute. In each school, one 2nd and one 5th grade were tested. The pre-test is conducted at the beginning of the school year (in autumn), the post-test is conducted at the end of the school year (in summer). Pre- and post-test are based on different items, so that the scores cannot be directly compared.

about a certain convergence of mean values for the age and job experience variables. Nevertheless, some differences may remain in our sample.

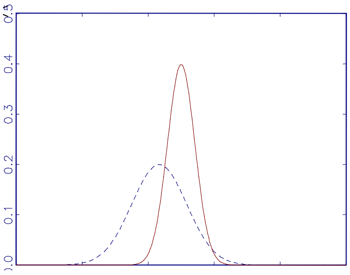
While providing insights on the weakly binding character of public rules and regulations in our three countries, the above discussion does not offer any clear explanations for the difference in student achievement reached by contract teachers and traditional teachers respectively. It appears worthwhile to compute the net effects by controlling for all confounding variables. In addition to the small set of variables presented here for illustrative purposes, a much larger number of variables can be used as a basis to select appropriate controls. Both the variable selection and the actual estimation procedure will be described in detail in the following section.

4. Econometric methodology

To evaluate the impact of the contract teacher status on education quality we adopt a nonparametric *treatment evaluation* framework, which allows us to get as close as possible to the concept of randomized trials when estimating our parameters. Let Y_i^0 denote the outcome, e.g. Mathematics or French proficiency at the end of the school year, of child *i* if being taught by a regular civil servant teacher. Y_i^1 represents the outcome of child *i* if being taught by a contract teacher instead, during the current school year. Thus, the treatment being evaluated is having a contract teacher during the current year as compared to having a civil servant teacher. Each child is being taught either by a contract teacher (D_i =1) or by a regular civil servant teacher (D_i =0), and we observe the outcome $Y_i = Y_i^1 D_i + Y_i^0 (1 - D_i)$.

Suppose for the moment that children were completely *randomly* (and with equal probability) allocated either to a contract teacher or to a regular teacher. Since allocation is at random, the two ensuing subpopulations are identical in (the distribution of) their observed and unobserved characteristics. The density function of the outcome variable in the former subpopulation thus represents the (unconditional) density function $f(Y^1)$, whereas the density function in the latter subpopulation represents $f(Y^0)$. The following figure shows a stylised example for $f(Y^0)$ (dashed line) and $f(Y^1)$ (solid line) in order to give a more intuitive understanding of the potential outcomes. Two effects are clearly discernible: the mean of Y^1 is larger than the mean of Y^0, and the variance of Y^1 is smaller than the variance of Y^0. In this example, the contract teacher programme has a positive effect on the average outcome and at the same time reduces educational inequality. To make this statement more precise, we introduce some additional notation.

Figure 1: Hypothetical example of $f(Y^0)$ (dashed line) and $f(Y^1)$



The average treatment effect (ATE) of being taught by a contract teacher for one year on a randomly drawn child is

$$E[Y^1 - Y^0].$$

While the average treatment effect provides an overall assessment of the average impact of the contract teacher programme, it is also of particular interest how the contract teachers change the distribution of the outcomes. We define the quantile treatment effects (QTE) as

$$Q^\tau(Y^1) - Q^\tau(Y^0),$$

where $Q^\tau(\cdot)$ refers to the τ-Quantile of the random variable in brackets. The quantile treatment effects show the impact of the contract teacher programme for children at different locations in the performance distribution. Contract teachers may perform poorly e.g. with weak or very strong students. In the example of Figure 1 we see that the treatment effect is positive and very large for low quantiles, still positive but smaller at the median, and is negative for very high quantiles.

The two distributions in Figure 1 reflect the unconditional distributions, i.e. in the entire population without conditioning on any observed characteristics. (These are the distributions we would expect if all teachers were either contract or regular teachers, everything else equal.) The lower quantiles represent the students performing poorly for a variety of reasons, e.g. living in rural areas, having less educated parents, having had less effective teachers in earlier grades, and/or lower innate ability.

Note that we do not need to assume rank invariance of individuals in making these comparisons of the two distributions, although it appears rather plausible in our context. Rank invariance would imply that an individual *i* who is at the quantile τ in the $f(Y^0)$

distribution would also be at quantile τ in the $f(Y^1)$ distribution. With rank invariance, low performing students would always be low performing students, no matter whether they were taught by a regular teacher or by a contract teacher. Our assumptions below do not allow us to test for this, and even with a perfect experiment, one could not test for rank invariance. Hence, we cannot rule out that the low performing students if taught by a regular teacher might actually be the high performers if taught by a contract teacher, and vice versa. This implies that their rank could change when moving from the dashed curve to the solid curve in Figure 1. In this case, the median student in the $f(Y^0)$ distribution might even lose when moving to the $f(Y^1)$ distribution. Nevertheless, rank invariance is not needed for any welfare comparisons of the contract teacher programme or to assess whether contract teachers compress or widen the outcome distribution.

Unfortunately the contract teacher programme was not implemented in the form of a randomized trial and the students being taught by a contract teacher differ from those being taught by a regular teacher in several characteristics, as witnessed in Section 3. We therefore have to control for differences in observed characteristics *X*. By the law of iterated expectations, we can always re-write the average treatment effect as

$$E[Y^1 - Y^0] = E[E[Y^1 \mid X]] - E[E[Y^0 \mid X]] .$$

To identify this expression, we need to express it in terms of observable variables. We assume in the following that

$$Y^1, Y^0 \amalg D \mid X , \qquad (1)$$

where the symbol $\amalg$ denotes statistical independence. In words, the marginal distributions of the potential outcomes are independent of *D*, given *X*. Consider a variable (e.g. rural versus urban) that determines treatment status *D* and also (at least one of) the potential outcomes. Such a variable would usually introduce a correlation between *D* and the potential outcomes, unless we control for it by including it in *X*.

The independence assumption (1) also implies mean independence and thus identifies the average treatment effect as

$$E[Y^1 - Y^0] = E[E[Y^1 \mid X]] - E[E[Y^0 \mid X]] = E[E[Y \mid X, D = 1]] - E[E[Y \mid X, D = 0]] , \qquad (2)$$

where the last equality follows because $E[Y^d \mid X, D = d] = E[Y \mid X, D = d]$. In words, the potential outcome for treatment status *d* for those who actually received treatment *d* is identical to the *observed* outcome.

Note that *X* should contain all variables that are determinants of *D* and at least one of the potential outcomes Y^1 and Y^0, but should contain neither variables that are directly linked to the definition of *D* nor variables that are already outcome variables of the contract status. The latter argument implies that we should not control for *job satisfaction* which is likely to be an outcome of the teacher status *D*. We also do not want to control for teacher's *education* and *training* since these are part of the definition of a contract teacher's status. By definition, education and training change when we replace a regular teacher with a contract teacher. In other words, the definition of $E[Y^1 - Y^0]$ reflects the effects of the teacher contract *together* with the effects of concurrent changes in education and training. (This strict relationship may not be exactly true as there could be individuals who received the regular teacher training and education, but entered in the teaching profession only later at a time when only contract teachers were hired. Hence, well trained and educated teachers can also be observed among the contract teachers. In addition, since the rules regulating entry into the teaching profession were not always strictly adhered to, we also observe civil servants with little education and training. We therefore also examine alternative specifications with controls for education and training in Section 5.3.)

On the other hand, we want to control for characteristics such as a teacher's *age* and *job experience* that are determinants of *D*. Contract teachers are on average younger and less experienced. But this relationship is not strict, i.e. we also observe contract teachers with several years of teaching experience and also regular teachers with very little experience. We therefore do *not* want to lump job experience into the definition of a contract teacher. In the data, contract teachers have lower job experience. However, one can well imagine that in a few years time average job experience and its distribution could exhibit a similar pattern, as we currently observe for regular teachers.

We also include in *X* a large number of variables characterising the school (location, equipment and facilities, management and parental involvement), the socioeconomic background of the parents, and the child's proficiency in French and Mathematics at the beginning of the school year as measured by the pre-test scores.

According to the identifying assumption (1), when we compare only individuals with the same characteristics *X*, the allocation to a contract or a regular teacher is not related to the students' potential outcomes. The pre-test scores are important for the plausibility of (1) as they reflect the child's otherwise unobserved performance as well as the impact of previous inputs into the education production process. For example, contract teachers are often hired in remote areas, where civil servant teachers are less willing to move, and where we generally also observe lower educational performance, which would in turn be reflected in the pre-test scores. We also expect that students who are taught by a contract teacher might also have been taught by a contract teacher in the previous year since the school might have many contract teachers. The effects of the contract teacher of the previous year, however, should also be reflected in the pre-test score of French and/or Math. Similarly, it might also be that parents, who are particularly interested in their child's education, place their child in a school or a class with a regular (or a contract) teacher. Again, we would expect that this extra interest of the parents would also be visible in the pre-test scores.

In a conventional linear regression notation

$$Y_i = \alpha + \beta D_i + \gamma X_i + U_i \tag{3}$$

assumption (1) is usually interpreted as $E[U \mid D, X] = 0$. I.e. conditional on *X*, the error term *U* should not be related to *D*. This basic intuition holds here as well, but the nonparametric approach outlined below and based on (1) is less restrictive in several dimensions. First, it does not require linearity as in (3). Second, it permits arbitrary treatment effect heterogeneity, i.e. α and β do not have to be constants but can be stochastic and correlated with *X*. Third, it does not impose monotonicity in *U* as does (3). Monotonicity in U does not permit that $Y_i > Y_j$ when both are taught by a regular teacher while $Y_i < Y_j$ when both are taught by a contract teacher. This corresponds to the rank invariance assumption as discussed before. This implies that, as opposed to our nonparametric approach, a rank reversal is not permitted in the conventional regression framework.

In addition, nonparametric regression allows for endogenous control variables. In other words, assumption (1) interpreted in the model (3) only requires that $E[U \mid D, X] = E[U \mid X]$ but does permit that the latter can be an arbitrary function of *X*, which might be the case for endogenous variables in *X*.

This assumption, however, is generally not sufficient for OLS to be consistent. OLS regression of (3) requires in addition that $E[U \mid X] = 0$ for consistent estimation of β. To give an example, consider *U* as a collection of several unobserved factors, one of them being innate ability. In contrast to OLS, in the nonparametric framework we can thus permit that ability is correlated with the pre-test score in *X* (and also the other variables in *X*). Innate ability has an effect on the post-test *Y* but also on the pre-test *X*, but perhaps in different (nonlinear) ways. For more details see Frölich (2008).

In order to identify the average treatment effect from the right-hand side of (2), we also need the *common support* assumption. More precisely, we need $E[Y \mid X, D]$ to be well defined for every value in the support of *X*. This requires that for every value of *X*, contract teachers and regular teachers are observed. An equivalent way of saying this is to require that $0 < \Pr(D=1 \mid X) < 1$. As aforementioned in Section 3, we observe a wide range in the job experience of civil servant teachers but no or only extremely few contract teachers with more than 10 years of job experience. Hence, the conditional mean would not be well defined for being a contract teacher with large job experience. Therefore, all observations with more than 10 years of experience are dropped, which also greatly mitigates differences in the age distribution.

Before describing the nonparametric estimator in more detail, we first discuss the estimation of the ATE. A matching estimator of the right hand side of (2) is obtained by replacing the outer expectation operator by the empirical distribution function of *X* and by plugging in nonparametric estimators of $m(x,d) = E[Y \mid X = x, D = d]$. Hence,

$$\hat{E}[Y^1 - Y^0] = \frac{1}{N}\sum_i (\hat{m}(X_i,1) - \hat{m}(X_i,0)), \tag{4}$$

where *N* is the sample size and the summation in (4) is over the entire sample, i.e. including observations with D_i=1 and those with D_i=0. The intuition behind (4) is straightforward: We pick one observation *i* with characteristics X_i and estimate $m(X_i,1) = E[Y \mid X = X_i, D = 1]$ and $m(X_i,0) = E[Y \mid X = X_i, D = 0]$. These are the expected outcomes in case of being taught by a contract versus a regular teacher for an individual with characteristics X_i. The nonparametric estimation of $m(X_i,1)$ can be considered as a weighted average of Y_j of all observations *j* with D_j=1 and values of X_j that are close to X_i. More precisely, observations *j* are weighted according to the distance between X_j and X_i in such a way that more distant observations receive smaller weights. Note that for the estimation of $m(X_i,1)$ only the contract teacher observations (i.e. with D_j=1) are used. On the other hand, in the estimation of $m(X_i,0)$ only the regular teacher observations (i.e. with D_j=0) are used.

The values of $m(X_i,1)$ and $m(X_i,0)$ are estimated for every individual in the sample (i.e. including contract and regular teachers) and equation (4) takes the average of the differences $m(X_i,1) - m(X_i,0)$.

We consider two different estimators for $m(x,d)$: *local linear* and *local logit*, where the latter is motivated by the boundedness of the outcome variable *Y* between zero and hundred percent correct answers. The local linear estimator is implemented such that the estimates are capped at 0 and 100. The estimators are local in that a weighted regression model (either linear or logistic) is estimated with Kernel weights that decline with the distance from *x*. The implementation of the local logit estimator follows Frölich (2006), and more details on the estimation process are provided in Bourdon, Frölich and Michaelowa (2007).

The estimation of the QTE follows Frölich (2007), who proposes averaging the conditional distribution functions to obtain the unconditional distribution, which can be inverted to obtain the quantiles. The reason for this is that we first have to condition on *X* to make use of the identifying assumption (1), which cannot be done with the quantiles directly since the average of the conditional quantiles does not give the unconditional quantile.

In Figure 1 we examined the unconditional density functions of the potential outcomes $f(Y^0)$ and $f(Y^1)$. Similarly we can define the unconditional distributions functions $F(Y^0)$ and $F(Y^1)$. More precisely, we define

$$F_{Y^d}(a) = \Pr(Y^d \leq a) = E[1(Y^d \leq a)] = E[E[1(Y^d \leq a) \mid X]],$$

by iterated expectations. $1(\bullet)$ is the indicator function, which takes the value one if the expression in brackets is true, and zero otherwise. Making use of assumption (1) we obtain

$$F_{Y^d}(a) = E[E[1(Y^d \leq a) \mid X, D = d]] = E[E[1(Y \leq a) \mid X, D = d]],$$

which is an expression that contains only observed variables. We note that this expression is very similar to the rightmost expression in (2), with the only difference that *Y* is replaced by 1(*Y* ≤ a) We can therefore estimate $F_{Y^d}(a)$ in analogy to (4) as

$$\hat{F}_{Y^d}(a) = \frac{1}{N}\sum_i \hat{m}_a(X_i, d), \qquad d\text{=0,1} \qquad (5)$$

where *N* is the sample size and $m_a(x,d) = E[1(Y \leq a) \mid X = x, D = d]$. The estimator is basically the same as in (4), simply replacing Y_i with 1(Y_i ≤ a). In a first step, we regress nonparametrically the indicator function 1(Y_i ≤ a) on X_i and on *D=1*, as described after equation (4). The estimated value of $\hat{m}_a(X_i,1)$ is then averaged over all observations *N* to obtain $\hat{F}_{Y^1}(a)$. To obtain $\hat{F}_{Y^0}(a)$ we regress nonparametrically 1(Y_i ≤ a) on X_i and on

D=0 and take the average of the predicted values, again over all observations in the sample.

We estimate $F_{Y^0}(a)$ and $F_{Y^1}(a)$, separately, on a grid of 200 different values for *a* between 0 and 100. By linear interpolation we have thus estimated the entire functions $F_{Y^0}(a)$ and $F_{Y^1}(a)$ for 0 ≤*a*≤100. These two functions are now each inverted to obtain the quantiles, from which we calculate the QTE as $Q^\tau(Y^1) - Q^\tau(Y^0)$.

Inference is based on the bootstrap, where we bootstrap the entire estimation process. First, samples of size *N* are randomly drawn from the original sample. Using this bootstrap sample, $F_{Y^0}(a)$ and $F_{Y^1}(a)$ are estimated by equation (5) and the QTE $Q^\tau(Y^1) - Q^\tau(Y^0)$ is obtained. By repeating this entire process 1000 times, the bootstrap distribution of the QTE, for a particular value of τ, is obtained. Since the treatment variable contract teacher is defined at the classroom level, whereas our estimations are at the pupil level, we need to take into account possible correlation of the unobservables determining the test scores *Y* among the pupils of the same class. This is done by re-sampling entire classes instead of pupils.

5. Empirical results

Before examining the estimation of quantile treatment effects, we first discuss the selection of the control variables and estimate average treatment effects.

5.1 Determinants of contract teacher status

The choice of the set of control variables *X* is crucial for the plausibility of the conditional independence assumption. In Section 4 we discussed which kind of variables should be included in *X* and which should not. We therefore examined three different sets of control variables (Xset 1, Xset 2 and Xset 3) to assess the robustness of the empirical results. The regressor set Xset 2 is based on an extensive collection of variables that have been used as possible predictors of test scores in the literature. These include variables that capture the school environment and classroom equipment, school management, characteristics of teachers and the socioeconomic background of the pupils. In addition, we include French and Mathematics proficiency at the beginning of the school year to capture differences in knowledge and ability. We do not include variables that are likely to have been causally affected by contract teacher status, and would thus represent direct outcomes of it, such as teacher job satisfaction or salary.

Most of the variables in Xset2 were found not to be useful predictors of the treatment variable ‘contract teacher’ in our sample, and hence would not be clearly useful as controls on our criteria. Xset 1 therefore retains only a subset of variables that were significant in class-level logit regressions of the treatment variable on all these regressors in *at least one* of the three countries considered. (See Bourdon, Frölich and Michaelowa 2007 for details). These are the school characteristics (*drinking water in school, school has toilets, male director, school participates in a pilot or exchange programme, distance to next city at least one hour*), the pupil characteristics (*initial test score French, initial test score Mathematics, age, number of schoolbooks*), the teacher variables (*age, job experience*) and class variables (*multi grade class, inspector's visit during the last year*). Finally, we also examined Xset 3, which contains all the variables of Xset 2 but in addition also the variables *educational level of the teacher* and *pedagogical training of the teacher*. As discussed in Section 4, these variables are closely linked to the definition of the contract teacher and one should not control for these. However, since these variables show some variation, indicating that the rules have often not been adhered to in practice, we can nevertheless attempt to examine the effects of these two variables. If effects differ significantly, the difference may be attributed to these two variables.

The weighted kernel functions we use in our estimator depend on fixing variable parameters called bandwidths. This controls the extent of smoothing of values around *X*. A full estimation process will usually incorporate an examination of sensitivity to different values of this bandwidth. In the following subsection, we only show the results for Xset 1 and one set of bandwidth values. (In Section 5.3 we will also examine some variations in the Xset.) Bourdon, Frölich and Michaelowa (2007) also present the results for the enlarged regressor sets Xset 2 and Xset 3 and other bandwidth values. Results are stable across regressor sets and bandwidth choices.

5.2 Average treatment effects

Before estimating quantile treatment effects in Section 6, we first consider average treatment effects. They are presented in Table 5, separately for French and Mathematics achievement. The first two rows show the estimates obtained by nonparametric regression. The results with local logit regression are very similar to the results with local linear regression, where estimates were capped at 0 and 100% correct answers. For Mali, the effects are *positive* throughout, but significant only for 2nd grade Maths. In Niger, the effects are *negative* throughout, but significant only in the 2nd grade. In Togo, the results are more subtle: They are *positive* in the 2nd grade (and significant for French), and *negative* in the 5th grade (and significant for Maths).

In row 3 of Table 5 we consider a simple OLS regression of the outcome variable on a constant, the dummy contract teacher status and the other regressors Xset 1. This is a very restrictive but nevertheless widely used specification in many applied papers, which suppresses all kinds of interaction terms. We term this estimator as the "Naive OLS estimator" and show the coefficient on contract teacher status in the table. However, the OLS estimates are very different from the matching estimates. This therefore cautions against the use of naive OLS without any interaction terms.

Table 5: Average treatment effects of teacher status (Xset 1)

	Mali 2nd		Mali 5th		Niger 2nd		Niger 5th		Togo 2nd		Togo 5th	
	French	Math	French	Math	French	Math	French	Math	French	Math	French	Math
Local linear matching estimator	8.96 (20.84)	**25.18** (11.83)	9.39 (11.46)	11.08 (12.59)	*-12.40* (7.03)	**-12.81** (6.30)	-1.53 (4.44)	-0.51 (4.45)	**11.55** (5.90)	5.21 (5.57)	-5.19 (3.74)	**-7.75** (3.87)
Local logit matching estimator	4.47 (20.17)	**24.39** (12.10)	7.98 (11.78)	10.41 (12.73)	*-11.85* (6.76)	*-11.84* (6.18)	0.52 (4.43)	-0.20 (4.67)	*11.10* (6.02)	4.64 (5.39)	-5.50 (3.54)	**-7.59** (3.70)
Naive OLS estimator	-11.19 (7.80)	-2.81 (8.13)	5.17 (3.17)	4.63 (3.18)	**-8.21** (3.68)	-8.02 (3.74)	-0.81 (2.20)	*-4.40* (2.31)	0.01 (4.15)	0.75 (3.25)	-1.07 (1.75)	*-3.59* (1.85)

Note: Average treatment effects are in the percentage points of test scores. Standard errors are given in parentheses. Inference for OLS is based on asymptotic formula with standard error estimates adjusted for class clustering. The dependent variable is French and Mathematics proficiency, respectively, at the end of school year. Estimates significant at 0.10 level are marked in *italics*, and significant at 0.05 level are marked in **bold**.

A potential concern with treatment effects based on achievement test scores is that they are an imperfect measure of true cognitive development, i.e. contaminated by measurement error. A supplementary appendix, Bourdon, Frölich and Michaelowa (2007), examines the extent and possible consequences of measurement errors. Although measurement error clearly seems to be present, it does not qualitatively change the conclusions of our empirical findings. (Our estimates would tend to be somewhat downward biased in magnitude.)

5.3 Alternative sets of control regressors

In this section we briefly examine the average treatment effects when controlling for alternative sets of regressors. Results are given in Table 6. The first two rows reproduce the nonparametric estimates of Table 5 for Xset 1. In this latter main specification we control for teacher's age and job experience, but do *not* control for the teacher's *educational attainment* nor for his *pedagogical pre-service training*. The latter two variables are strictly tied to the definition of the treatment variable. In principle, by definition, contract teachers receive only very little training and are subject to different educational requirements for admission. Hence, a change in the treatment variable would automatically entail a change in these two variables such that they cannot and should not be used as control variables. In other words, while we would like to compare contract and regular teachers with similar job experience in order not to confound any contract teacher effects with the effects of job experience, they should have different education and training. By not including education and training in *X*, we allow the distributions of these two variables to vary freely according to their occurrence among contract and regular teachers. However, we recognise that part of the treatment effect could be attributable to differences in these distributions.

To shed more light on this issue, Table 6 provides the results for two alternative specifications of the regressor set (for further alternatives see Bourdon, Frölich and Michaelowa 2007). First and notwithstanding our above discussion we include teacher's education and training as additional covariates. Although the previous discussion would imply lack of common support, in practice, as explained in Section 4, the various administrative rules have not always been adhered to. This kind of variation would thus enable us (to some extent) to disentangle the effects of contract teacher status from the effect of education and training. Since contract teachers have on average lower education and training, controlling for these two characteristics should thus increase the effects of teacher status in favour of the contract teachers.

As shown in the second panel in Table 6, for most countries and grades, the results *with* education and training are remarkably similar to the main specification, suggesting that the differences in education and training are not the main driving forces behind the effects of contract teacher status. The only notable change leading to a higher and significant effect can be observed for French in Mali 2nd grade. Given the general tendency to find about equal effects for French and Mathematics (see also the parametric analysis), we consider this as an indication that the previous estimate was too low.

Table 6: ATE of teacher status (with different controls for teacher variables)

	Mali 2nd		Mali 5th		Niger 2nd		Niger 5th		Togo 2nd		Togo 5th	
	French	Math	French	Math	French	Math	French	Math	French	Math	French	Math
Xset 1 (Results identical to Table 5)												
Local linear matching estimator	8.96	**25.18**	9.39	11.08	*-12.40*	**-12.81**	-1.53	-0.51	**11.55**	5.21	-5.19	**-7.75**
Local logit matching estimator	4.47	**24.39**	7.98	10.41	*-11.85*	*-11.84*	0.52	-0.20	*11.10*	4.64	-5.50	**-7.59**
Xset 1 plus teacher's education plus teacher's training												
Local linear matching estimator	20.6	21.0	12.0	9.7	**-13.5**	**-12.2**	3.0	-0.6	6.7	-0.8	*-7.2*	**-8.3**
Local logit matching estimator	*25.2*	**26.0**	*13.0*	10.3	**-12.9**	**-11.4**	5.1	0.3	6.1	-0.7	*-7.7*	**-8.2**
Xset 1 without teacher's age and job experience												
Local linear matching estimator	3.6	7.7	3.7	6.4	**-7.6**	*-7.3*	-2.5	*-3.9*	6.1	4.7	-0.5	-2.2
Local logit matching estimator	-2.5	-0.4	2.7	5.4	**-7.5**	*-7.0*	-2.3	*-3.7*	5.7	4.5	-0.9	-2.5

Note: Average treatment effects in percentage points of test score. The dependent variable is French and Mathematics proficiency, respectively, at the end of the school year. Estimates significant at 0.10 level are marked in *italics*, and significant at 0.05 level are marked in **bold**.

In addition, we observe an unexpected decrease in the effect for 2nd grade French in Togo. Given the very similar average educational attainment of contract and regular teachers in our sample for 2nd grade Togo (see Table 4), this could only be due to differences in pre-service training. An explanation might be that the few contract teachers who effectively received training, received training of strongly improved quality, so that on average this increases this group's overall performance. However, this does not appear very plausible, and given that we observe very similar estimates when we actually drop some regressors (in the third panel of Table 6), this result is probably a statistical artefact. The third panel in Table 6 represents the results when using the regressor set Xset 1 *without* teacher's age and without job experience. Comparing the estimates to the first two rows of Table 6, we observe a substantial drop in the effects for Mali. Given that job experience is substantially lower for contract teachers than for regular teachers in Mali (as we observed in Section 3), *not* controlling for these differences masks the large positive effect of contract teachers in Mali.

For the 2nd grade in Niger and 5th grade in Togo, the effects become less negative when we omit age and job experience. This result may appear unexpected as it implies that job experience has a negative effect. It could be that such a phenomenon may be related to teachers' motivation decreasing over the years. In our context, an alternative explanation might be that the (few) contract teachers with relatively lengthy job experience perform particularly poorly. These individuals were most likely teachers at the time when traditional civil service employment prevailed, and yet did not manage to attain a fixed-term employment situation. However, these results should not be over-interpreted since the differences are not statistically significant.

For Togo 2nd grade, we also find that the effects drop when we omit age and job experience. However, and as previously discussed, we also observed such changes from the second panel which included education and training. These differences should therefore be interpreted with care.

Overall, when comparing the first and third panel of Table 6, a general pattern emerges: In countries and grades where the effects of contract teachers are positive, controlling for the lower job experience (and age) of contract teachers increases the effect, whereas in countries and grades where the effects of contract teachers are negative, taking lower job experience into account makes the effects even more negative. However, the latter is never significant, so we should not over-interpret these comparisons.

6. Nonparametric estimation of quantile treatment effects

In this section we examine the effects of contract teachers on the distribution of the test scores. Table 7 shows the *quantile treatment effects*, estimated separately by country and grade for Xset 1. The results for the 2nd grade in *Niger* are very stable to the set of regressors. The effects are close to zero for the lower quantiles, and decrease almost monotonously to about -20 to -25 percentage points for the high performance students. Hence, whereas weak students do not seem to be affected by teacher status, strong students seem to suffer from being taught by a contract teacher. The effects are very similar for French and for Maths.

For the 5th grade, the estimates also tend to be negative but are small and insignificant. At the lower quantiles the effects remain slightly negative with the larger regressor sets 2 and 3. For the higher quantiles, the estimation results are unclear and sensitive to the choice of regressors included. (For Mathematics, the effects tend to be negative. For French the results are unstable due to local collinearity.)

In *Togo* the impacts tend to be positive for the 2nd grade and negative for the 5th grade. The positive effects for the 2nd grade are a bit larger for French than for Maths. For French, they are significant around the median (about 10 to 14 percentage points) while they remain insignificant for Maths. The estimates with enlarged Xsets are similar with often somewhat larger effects on Mathematics (see Bourdon, Frölich and Michaelowa 2007).

The situation is different for the 5th grade. Here the effects are about zero for the low quantiles, and then decrease almost monotonously until -13 percentage points for Mathematics for the strong students. This pattern is very stable across bandwidth choices and Xsets. Effects are negative for both French and Mathematics but more strongly so for Maths.

In *Mali*, to the contrary, effects tend to be positive for both grades. For the 2nd grade, effects are insignificant for French, but significantly positive for Maths. Judging from the significance levels, this effect seems to be most precise for the lower to medium quantiles. This implies that children at the *lower* end of the performance distribution seem to benefit most clearly from contract teachers. The effects are smaller and less precisely estimated for enlarged Xsets but remain positive for Mathematics, perhaps except for very large quantiles.

For the 5th grade, the effects are also positive for French and Mathematics for the low to medium quantiles, but again significant only for Mathematics and now only at the 10% level. For enlarged Xsets the results are less precise and smaller. Hence, the evidence is less stable for the 5th grade, but overall tends to be positive rather than negative.

All in all, just as the discussion of average treatment effects in the previous sections, the above analysis strongly confirms our expectations that given the different characteristics of the contract teacher programmes in the three countries considered, we should also observe quite different results. As suggested by our initial hypotheses on both the incentive and selection effects, the contract teacher programme in Niger shows the worst results, i.e. either insignificant or clearly negative. To the contrary, the Malian results are consistently positive and significant in Mathematics for both grades. This suggests that the potentially negative role of low salaries (if any) was overcompensated by the positive incentive effect induced notably by parental responsibility and monitoring in the case of Malian community teachers. As expected, Togo occupies an intermediate position which may be related to the fact that parental monitoring responsibility was reduced through the partial integration of contract teachers into the public administration system.

One could also suspect that the relatively bad performance of contract teachers in 5th grade in Togo is related to their lack of pedagogical training. However, the use of an Xset where this variable is controlled for, does not lead to any significant change. There is thus no evidence that lack of training could drive the results. At the same time, as stated earlier, it may not be possible to fully separate the effect of education and training from the treatment itself. As already noted in the discussion of the average treatment effect, this aspect must therefore be interpreted with caution.

The second general observation is that results do indeed differ depending on students' performance. Across all countries, the analysis by quantiles suggests that contract teachers do a relatively better job when teaching weak students. Whenever we observe positive effects, they tend to become significant in low quantiles, and whenever we observe negative effects, they tend to become significant for the strong students. This implies that contract teachers tend to reduce inequalities in the student distribution.

In Mali and Togo, we also observe that contract teachers tend to be relatively more successful with younger students, i.e. they show a better performance in 2nd as compared to 5th grade. However, the opposite is the case in Niger. Overall, the influence of contract teachers is thus more clearly related to students' performance within each grade level than to the difference in students' performance across grades.

A possible explanation for this phenomenon could be that traditional civil servant teachers may be unwilling to move to disadvantaged areas, and unmotivated if they are compelled to do so. Contract teachers, however, are often locally recruited and therefore should not face this problem. Moreover, coming from a similar background, these teachers are more aware of the specific problems of children with learning difficulties; they speak their language, know their parents and thus may be better able to deal with the situation. Contract teachers may thus outperform traditional teachers when teaching weak students, even if traditional teachers do better in a less disadvantaged student environment.

Table 7: Quantile treatment effects of teacher status (Xset 1)

	Mali 2nd		Mali 5th		Niger 2nd		Niger 5th		Togo 2nd		Togo 5th	
	French	Math	French	Math	French	Math	French	Math	French	Math	French	Math
τ=0.1	-2.5	7.8	11.9	5.5	-2.0	0.2	-0.2	-0.5	4.3	2.5	0.5	-2.7
τ=0.2	1.8	**14.7**	15.0	7.2	-1.5	1.3	-3.2	-4.4	*10.9*	8.0	-2.5	-7.6
τ=0.3	3.4	**22.9**	11.9	11.4	0.0	-4.3	-3.5	-2.5	4.4	3.0	-3.4	**-9.1**
τ=0.4	5.1	**28.3**	13.9	10.7	-2.0	-5.6	-2.1	-3.1	7.8	7.9	-3.3	-4.8
τ=0.5	0.1	**33.5**	10.6	*13.2*	-12.9	-11.1	-2.6	-0.2	**14.0**	5.4	-5.5	*-6.8*
τ=0.6	1.6	*21.4*	6.2	12.9	**-19.0**	**-23.1**	-0.1	-2.3	*13.9*	2.2	**-9.5**	-7.0
τ=0.7	3.5	*25.3*	2.9	13.4	**-24.6**	**-30.1**	-0.6	2.7	8.1	2.7	**-9.3**	**-12.7**
τ=0.8	10.0	28.8	2.6	-0.1	**-21.3**	**-25.1**	-1.4	-0.5	0.0	6.4	**-7.8**	**-11.9**
τ=0.9	18.9	13.5	9.1	-0.2	**-11.4**	**-10.1**	-8.7	-0.5	-1.3	1.9	-6.4	**-13.0**

Note: Percentage points treatment effects at the quantiles τ=0.1, 0.2, ..., 0.9. The dependent variable is French and Mathematics proficiency, respectively, at the end of the school year. Estimates significant at 10% are marked in *italics*, significant at 5% in **bold**, significant at 1% in **bold underlined**.

7. Conclusions

In this paper we have analysed the impact of 'contract teacher' reforms on educational quality. Traditionally, in most countries, teachers are hired as *civil servants* after a well regulated initial education and training period and with clear career advancement tracks. Due to financial pressure, high schooling demand and other reasons, many African (as well as Latin American and South Asian) countries have experimented with alternative forms to engage new teachers, usually providing only fixed term contracts with lower remuneration and reduced entry requirements. In this paper we provide an overview of these contract teacher reforms in francophone Africa and nonparametrically estimate *average* and *quantile treatment effects* for Niger, Togo, and Mali. Our empirical analysis is based on data collected by PASEC in several francophone African countries, with the advantage that sampling, data collection, testing, interviewing, data cleaning etc. follow similar procedures, and that, with few exceptions, variables are coded and defined in the same way.

Our estimates point to two major findings. First it seems that contract teachers do relatively better for weak than for strong students. *When positive treatment effects are found, they tend to be more positive at the low to medium quantiles, and when negative effects are observed they tend to be more pronounced at the high quantiles.* This pattern is remarkably stable across countries, and most pronounced for Niger 2nd grade and Togo 5th grade. This suggests that relative to traditional civil servant teachers, contract teachers are in a better position to work in a more difficult learning environment and to react to the needs of students with the most serious learning difficulties. As such, contract teachers tend to reduce existing inequalities in overall student outcomes.

Second, we observe *clear differences in the impact of contract teachers between the three countries*. The treatment effects are highly *negative* for *Niger* in the 2nd grade and tend to be very small, and slightly negative, for the 5th grade. In *Togo* we found clearly positive effects for 2nd grade students. For the 5th grade, the effects were very small and negative for weak students but very large and negative for strong students. Here contract teachers do well in low grades, but clearly fail in the 5th grade for the more advanced students. In contrast, in *Mali*, the treatment effects tend to be positive for both 2nd and 5th grade, particularly in Mathematics.

Hence, when ordering the three countries by their overall effects, we find that effects are *positive in Mali*, somewhat *mixed in Togo* with positive effects in 2nd and negative effects in 5th grade, and *negative in Niger*. This ordering can be related to the manner in which the contract teacher programmes have been implemented. In Mali and Togo, contract teachers have been introduced in a less centralised way and, especially in Mali, the system continues to work predominantly through local communities. This may have led to the closer monitoring and more effective hiring of contract teachers, contributing to the positive outcome. In Niger, the centrally governed approach and the rapid recruitment of large numbers of contract teachers, relative to a rather limited base of qualified secondary graduates, may also have contributed to poor performance.

Hence, a focus on local community and parental involvement in the recruitment and monitoring of contract teachers, as well as a gradual rather than an immediate reform process, seems to be warranted.

Overall, our results indicate that the success or failure of contract teacher programmes, in terms of student performance, depends on careful implementation of the system. Despite lower pay and adverse working conditions, the incentive effect may be positive, especially if teachers are directly engaged by parents or the local community. In this case, teachers will be exposed to recognition for their efforts and they can also be held directly responsible for their work.

These findings are encouraging, especially if we consider that in terms of purely quantitative objectives such as universal primary education, hiring contract teachers appears to be inevitable. According to the most recent UNESCO estimates, until the year 2015, 60 000 new teaching positions must be created in Niger, 55 000 in Mali and 12 000 in Togo (UNESCO-UIS 2006, p. 41). Taking into account the usual rates of turnover and attrition, it could be estimated that overall about 150 000 teachers will have to be recruited in the three countries. This number is twice as high as the stock of teachers currently on the job.

Our results suggest that the involvement and empowerment of parents and local communities can help to overcome this challenge. It should be noted, however, that relying on the cooperation of these groups tends to reinforce existing inequalities. If poor communities pay and monitor their teachers while other well-to-do schools are equipped and managed by public authorities, education policy becomes unacceptably regressive. The new challenge therefore is to achieve a pro-poor distribution of educational expenditure while, at the same time, encouraging local initiative and autonomy.

Acknowledgements: We are grateful to the editor, an associate editor and two anonymous referees for very helpful comments.

References

Bourdon, J., M. Frölich and K. Michaelowa (2007) Teacher Shortages, Teacher Contracts and their Impact on Education in Africa. IZA Discussion paper 2844.

Duthilleul, Y. (2005) Lessons Learned in the Use of Contract Teachers, Paris: IIEP.

Frölich, M. (2006) Nonparametric Regression for Binary Dependent Variables. *Econometrics Journal*, **9**, 511-540.

Frölich, M. (2007) Propensity Score Matching Without Conditional Independence Assumption - with an application to the gender wage gap in the UK. *Econometrics Journal*, **10**, 359-407.

Frölich, M. (2008) Parametric and nonparametric regression in the presence of endogenous control variables. *International Statistical Review*, 76 (2), 214-227.

Glewwe, P. and M. Kremer (2006) Schools, Teachers, and Education Outcomes in Developing Countries. In *Handbook on the Economics of Education* (eds E. Hanushek and F. Welch) vol. 2, pp. 945-1017. New York: Elsevier.

Hanushek, E. A., J. Kain, D. O'Brien and S. Rivkin (2005) The Market for Teacher Quality, NBER Working Paper 11154, http://www.nber.org/papers/w11154 (accessed 4/02/2007).

Michaelowa, K. and A. Wechtler (2006) The Cost-Effectiveness of Inputs in Primary Education: Insights from the Literature and Recent Student Surveys for Sub-Saharan Africa, Study commissioned by the Association for the Development of Education in Africa (ADEA), ADEA Working Document, Paris.

PASEC (2005) Les enseignants contractuels et la qualité de l'enseignement de base I au Niger : Quel bilan?, Dakar: CONFEMEN.

UNESCO-UIS (2006) Teachers and Educational Quality: Monitoring Global Needs for 2015, Montreal: UNESCO Institute for Statistics, http://www.uis.unesco.org/TEMPLATE/pdf/Teachers2006/TeachersReport.pdf (accessed 2/21/2007).

World Bank (2002) Le système éducatif togolais – Eléments pour une revitalisation, Africa Region Human Development Working Paper (Education sector Country Status Report for Togo).

World Bank (2004) La dynamique des scolarisations au Niger: Evaluations pour un développement durable, Africa Region Human Development Working Paper, Development Research Group (Education sector Country Status Report for Niger).

World Bank (2006) Eléments de Diagnostic du Système Educatif Malien: Le besoin d'une Politique Educative Nouvelle pour l'atteinte des objectifs du millénaire et la réduction de la pauvreté, Africa Region Human Development Working Paper, Development Research Group (Education sector Country Status Report for Mali).

World Bank (Africa Region) and Pôle de Dakar (2007) Le système éducatif congolais: Diagnostic pour une revitalisation dans un contexte macroéconomique plus favorable (Education sector Country Status Report for Congo), preliminary version, February, Dakar.

Wößmann, L. (2005) Competition, Decentralization and Accountability: Lessons for Education Policies in Developing Countries from International Achievement Tests. *Nord-Süd Aktuell*, **19**, (2), 142-153.

Teacher Shortages, Teacher Contracts and their Impact on Education in Africa

Supplementary appendix (complementary information, not published in JRSS-A)

A Flexible parametric estimation of average treatment effect

This subsection of the appendix provides supplementary material to Table 5 of the main paper. The first three rows of Table A.1 below are reproduced from Table 5.

The rows below show results based on parametric regressions with alternative specifications and interaction terms. In all of these regressions, the outcome variable is regressed on Xset 1 *separately* in the subpopulation of contract teachers and regular teachers and compared thereafter. This corresponds to a specification where all regressors are interacted with contract teacher status. In addition, the predicted values of the linear regressions (lower part of the table) are capped at 0 and 100% correct answers. In the table, we then report the average difference in the predicted outcomes. In essence, these estimators are identical to the nonparametric estimators above, but with bandwidth values ∞.

In rows 4 and 9, there is no additional change, i.e. no interaction terms are used in the separate regressions. These two rows differ only with respect to the use of a linear versus a logistic regression function. Overall the results of rows 4 and 9 are rather similar. For Mali and Niger, they are also rather similar to the nonparametric results of rows 1 and 2, at least in those cases where the latter are significant. The situation is somewhat different in Togo. Although the signs of the parametric and nonparametric estimates coincide throughout, the parametric estimates are less than half the size of the nonparametric ones.

The rows 5 to 8 and 10 to 13 show the estimation results of more complex parametric specifications with different numbers of interaction terms included in addition to Xset 1. Specification A is the most complex and includes a fourth order polynomial in each of the two pre-test scores, a second order polynomial in each of the other non-dummy variables of Xset 1, and the interaction between each of the two pre-test scores and all the other variables of Xset 1. In Specification B, the fourth-order polynomial is reduced to a second-order polynomial. Specification C is similar to B in that it retains the second-order polynomials but drops all the interaction terms between the two pre-test scores and the other variables of Xset 1. The results of these three different specifications are rather similar. Apart from 2nd grade French in Mali, where estimates become inflated, the estimates stay close to the nonparametric matching estimates. For Togo the difference to the nonparametric estimates is still pronounced, but the more flexible parametric estimates are somewhat closer to the nonparametric ones. Finally, Specification D starts from Specification B but drops the second-order polynomials, i.e. retaining only linear and interaction terms. Now overall results for 2nd grade French are similar to the nonparametric result, but for Togo 5th grade the differences remain pronounced. At the same time, the results with Specification D sometimes change a lot relative to the more flexible Specifications A and B, which is particularly visible in 2nd grade French in Mali. This specification which dropped all quadratic, cubic and higher order terms thus appears to be too restrictive. Linear terms alone do not capture well all the information contained in the pre-test scores. Specification B, on the other hand, appears to be sufficiently flexible in that the higher-order polynomials permitted in Specification A do hardly affect the estimates.

Interestingly, in Specifications A and B, and for each country and grade, the effects for Math and French are almost identical. This also holds relative to the standard deviation of scores as the latter are also quite similar (see Table B.2).

To summarize, the findings of this parametric analysis clearly confirm the nonparametric matching estimates: Positive effects in Mali, negative effects in Niger, and for Togo positive effects in the 2nd grade but negative effects in the 5th grade. Whereas the estimates are very close for Mali and Niger, the parametric estimates are less than half the size of the nonparametric results for Togo. We thus suspect that the global parametric estimation approach did not fit so very well for Togo and led to underestimating the true effects (though still giving the correct sign).

As a final word, one should repeat that the simple but common OLS regression (row 3) of the outcome variable on a constant, contract teacher status and Xset 1, which assumes equal slope parameters for contract and regular teachers, produces highly misleading results (particularly for 2nd grade in Mali).

Table A.1: Average treatment effects of teacher status (Xset 1)

	Mali 2nd		Mali 5th		Niger 2nd		Niger 5th		Togo 2nd		Togo 5th	
	French	Math	French	Math	French	Math	French	Math	French	Math	French	Math
1) Matching (local linear)	8.96	**25.18**	9.39	11.08	*-12.40*	**-12.81**	-1.53	-0.51	**11.55**	5.21	-5.19	**-7.75**
2) Matching (local logit)	4.47	**24.39**	7.98	10.41	*-11.85*	*-11.84*	0.52	-0.20	*11.10*	4.64	-5.50	**-7.59**
3) Naive OLS estimator	-11.19	-2.81	5.17	4.63	**-8.21**	**-8.02**	-0.81	*-4.40*	0.01	0.75	-1.07	*-3.59*
4) Logit	5.46	**22.74**	**12.53**	**15.25**	**-8.37**	**-10.30**	-0.22	**-2.66**	4.95	**6.71**	-0.74	**-2.91**
5) Logit Interaction A	**28.67**	30.04	**11.71**	**7.93**	**-10.38**	**-11.30**	3.16	2.01	6.47	*7.02*	-1.33	**-3.26**
6) Logit Interaction B	**28.68**	30.28	**10.95**	**9.07**	**-10.27**	**-10.55**	3.60	2.28	6.82	6.89	-0.88	**-2.85**
7) Logit Interaction C	**29.01**	**30.86**	**13.58**	**14.80**	**-12.69**	**-11.39**	**4.04**	2.53	4.40	*7.98*	-0.92	**-3.53**
8) Logit Interaction D	1.58	**18.76**	**8.77**	**10.27**	**-6.61**	**-9.56**	-0.24	**-3.16**	**8.49**	**6.51**	-0.70	**-2.65**
9) Linear	6.91	**24.05**	**14.02**	**17.73**	**-8.25**	**-10.29**	-0.11	*-2.48*	5.43	**7.25**	-0.83	**-2.94**
10) Linear Interaction A	**26.40**	27.51	**14.89**	**10.58**	**-11.13**	**-11.85**	2.32	1.45	7.04	*7.77*	-1.24	**-3.19**
11) Linear Interaction B	**26.64**	27.75	**14.69**	**12.01**	**-10.99**	**-11.10**	2.74	1.65	7.41	*7.66*	-0.61	*-2.73*
12) Linear Interaction C	**26.23**	**29.46**	**15.77**	**17.35**	**-12.61**	**-11.43**	**3.20**	2.06	4.90	*8.72*	-0.89	**-3.48**
13) Linear Interaction D	3.29	**19.46**	**12.61**	**14.74**	**-7.19**	**-9.79**	-0.16	**-3.00**	**8.64**	**7.17**	-0.57	*-2.54*
14) Linear only one pre-test	4.63	**23.21**	**14.32**	**17.97**	**-7.06**	**-10.27**	-0.48	*-2.18*	5.38	**9.79**	-0.65	**-4.55**

Note: Average treatment effects in percentage points. Inference is based on the bootstrap, by resampling classes. Bandwidth values are 0.5, 0.5, 0.25 for the nonparametric matching estimates. The dependent variable is French and Math proficiency, respectively, at the end of school year. Estimates significant at 10% are marked in *italics*, significant at 5% are marked in **bold**.

B Measurement error in the cognitive achievement tests

A potential concern with treatment effects based on achievement test scores is that they are an imperfect measure of true cognitive development. Test scores enter in the empirical analysis twice: once as a measure of the outcome variable and once as a control variable (the pre-test scores). Both kinds of test scores are likely to be affected by some measurement error.

We will first examine the extent and impact of measurement errors in a linear model. As a basis for comparison required in this context (and to enable an instrumental variable strategy), row 14 of Table A.1 presents the parametric estimate for the main specification when using only one of the two pre-test scores as regressors. The specification is identical to row 9 (of Table A.1), but the Math pre-test is dropped when estimating the French outcome and vice versa. Overall, the rows 9 and 14 produce fairly similar results.

Given that we have four different tests of ability (French and Math at the beginning and end of the school year), we can attempt to estimate the size of the measurement error in different ways. Suppose the test score *Y* for student *i* at time *t* (0 = beginning of year, 1 = end of year) in subject *s* (F = French, M = Math) is given as

$$Y_{i,t}^{s} = \xi_{i,t}^{s} + \varepsilon_{i,t}^{s} \tag{6}$$

where $\xi_{i,t}^{s}$ is true ability and $\varepsilon_{i,t}^{s}$ is a measurement error. We suppose in the following that the measurement errors are mean zero, and independent of each other and of the true abilities. Let σ^2 be the variance of the measurement error. Suppose for the moment that all the other regressors in Xset 1 are measured without error. Consider first the regression of the French test at *t*=1 on the French test at *t*=0, a constant and the other regressors in Xset1 (separately for contract and regular teachers). This corresponds to the columns for French in row 14 of Table A.1 and is reproduced as the first row in Table B.1. The measurement error in $Y_{i,1}^{F}$ inflates the noise in the regression whereas the measurement error in $Y_{i,0}^{F}$ leads to inconsistent (downward biased) estimates. Assuming that the true French and Math ability at *t*=0 are correlated, we can use the Math test at *t*=0 as an instrument for $Y_{i,0}^{F}$. Vice versa, we use the French test at *t*=0 as an instrument for $Y_{i,0}^{M}$. (This instrument is not entirely convincing since the performance on the French test may have a direct effect on Math since the Math test is administered in French such that better knowledge of French helps a student on the Math test.)

Using this instrumental variable, equation (3) is estimated by linear IV estimation. The vector *X* in (3) consists of the subject performance at *t*=0 and all the other control variables. The vector of instrumental variables *Z* consists of the performance in the *other* subject at *t*=0, in addition to all the other control variables. Define $\mathbf{X}_i = (D_i, X_i^{'})'$ and $\mathbf{Z}_i = (D_i, Z_i^{'})'$. For consistency of the IV estimator we need $E[\mathbf{Z}_i^{'}\mathbf{X}_i]$ to be of full rank and require also that $E[\mathbf{Z}_i^{'}U_i] = 0$. A sufficient condition for the latter assumption is that $E[U \mid D, Z] = E[U \mid Z] = 0$. This is similar to the independence condition in (1) with *X* being replaced by *Z*. In other words, *U* should be unrelated to *D* conditional on all control variables (except the mismeasured regressor). In addition, *U* should be independent of *Z*.

The resulting instrumental variable estimates are shown in the second row of Table B.1.[1] These instrumental variable estimates serve two purposes. First, we see that the IV estimates are not very different from the OLS estimates in the first row (except for the noisy estimate for 2nd grade French in Mali). This shows that measurement error seems to have only little impact on the estimates. Second, we can use the IV estimator to estimate the variance of the measurement error σ^2, as we discuss below.

We also consider an alternative instrumental variables strategy that does not rely on multiple pre-test measures, but exploits higher order moment conditions, as in Dagenais and Dagenais (1997). (An extension of their approach is developed in Lewbel 1997.) If one assumes that the measurement errors are *symmetrically* distributed and also that the error term in the main regression is *symmetrically* distributed, the following regressors are valid instrumental variables

$$x_{i,k}^2, \quad x_{i,k}y_i, \quad y_i^2, \quad x_{i,k}^2 y_i - 2x_{i,k}E[x_k y] - y_i E[x_k^2], \quad x_{i,k}y_i^2 - x_{i,k}E[y^2] - 2y_i E[x_k y],$$

where the subscript *k* refers to the *k*-th regressor and the variables *x* and *y* are de-meaned, i.e. $x_{i,k} = X_{i,k} - E[X_k]$ and $y_i = Y_i - E[Y]$. These instrumental variables conditions are based on higher-order moment calculations and require that the true regressors X^* themselves are *not* symmetrically distributed. The conditions apply to any regressor x_k that is considered to be affected by measurement error. Dagenais and Dagenais note that $x_{i,k}^3 - 3x_{i,k}E[x_k^2]$ and $y_i^3 - 3y_i E[y^2]$ are also valid instruments, but only under the

[1] Below, we also show the estimated bias of OLS as a fraction of the absolute value of the IV estimate, calculated as (estimate$_{\text{OLS}}$ – estimate$_{\text{IV}}$)/ abs(estimate$_{\text{IV}}$).

assumptions of normal measurement errors and normal regression error, which we do not want to impose in our analysis. In their simulations, Dagenais and Dagenais further note that the estimator using only the squared regressor $x_{i,k}^2$ and these cubic terms as instruments performed often better than when using the full set of instrumental variables. In Table B.1, we therefore report only the estimates with $x_{i,k}^2$ as instrumental variable (To be sure, we also computed estimates with the full set of instrumental variables, and they turned out to be very similar.) The third row shows the IV estimate when only one pre-score is used in the regression (French for French and Math for Math) and instrumented with its de-meaned square, whereas row 5 shows the IV estimates with both pre-test scores (and both instrumented by their de-meaned square). For comparison, row 4 reproduces the OLS estimates with both pre-test scores. Comparing row 3 to row 1 and row 5 to row 4, we find again that the OLS and IV estimates are quite similar (except for the noisy estimate for 2nd grade French in Mali). Hence, the presence of measurement error does not seem to invalidate our interpretation of the results.

Table B.1: Average treatment effects with measurement error (Xset 1)

	Mali 2nd		Mali 5th		Niger 2nd		Niger 5th		Togo 2nd		Togo 5th	
	French	Math	French	Math	French	Math	French	Math	French	Math	French	Math
1) OLS using only one pre-test	4.63	**23.21**	**14.32**	**17.97**	**-7.06**	**-10.27**	-0.48	*-2.18*	5.38	**9.79**	-0.65	**-4.55**
2) IV with the other pre-test	**10.46**	**25.50**	**13.77**	**17.28**	**-7.05**	**-10.15**	-1.31	*-2.27*	4.85	**8.22**	1.43	**-3.13**
Bias of OLS as fraction of IV	-56 %	-9 %	4 %	4 %	0 %	-1 %	63 %	4 %	11 %	19 %	-145 %	-45 %
3) IV using one squared pre-test	*9.03*	**23.10**	**13.87**	**17.96**	**-6.83**	**-10.41**	-0.80	-2.11	5.36	**10.49**	-1.91	**-5.38**
Bias of OLS as fraction of IV	-49 %	0 %	3 %	0 %	-3 %	1 %	40 %	-3 %	0 %	-7 %	66 %	15 %
4) OLS with both pre-tests	6.91	**24.05**	**14.02**	**17.73**	**-8.25**	**-10.29**	-0.11	*-2.48*	5.43	**7.25**	-0.83	**-2.94**
5) IV with both squared pretests	**11.84**	**23.79**	**13.58**	**17.58**	**-8.37**	**-10.48**	0.06	-1.95	4.57	***6.21***	-2.55	**-4.12**
Bias of OLS as fraction of IV	-42 %	1 %	3 %	1 %	1 %	2 %	-283 %	-27 %	19 %	17 %	67 %	29 %
Estimation of the measurement error in the *D*=0 sample (= regular teachers)												
$\hat{\sigma}^2$ (2) vs. (1)	0.0105	0.0119	0.0049	0.0090	0.0159	-0.0025	0.0083	0.0075	0.0036	0.0221	0.0060	0.0089
$\hat{\sigma}^2$ (3) vs. (1)	0.0086	-0.0006	0.0035	0.0002	-0.0175	0.0043	-0.0561	-0.0116	0.0027	-0.6195	-0.0021	-0.0077
$\hat{\sigma}^2$ (5) vs. (4)	0.0048	0.0007	0.0006	0.0006	0.0755	0.0153	0.0305	0.0127	0.0172	0.0187	-0.0042	0.0046
Estimation of the measurement error in the *D*=1 sample (= contract teachers)												
$\hat{\sigma}^2$ (2) vs. (1)	0.0093	0.0083	0.0113	-0.0058	0.0084	0.0116	0.0074	0.0053	0.0091	0.0207	0.0074	0.0081
$\hat{\sigma}^2$ (3) vs. (1)	-0.0025	-0.0242	-0.0339	-0.0011	-0.0085	-0.0058	0.0038	-0.0091	-0.0158	0.0276	-0.0147	-0.0195
$\hat{\sigma}^2$ (5) vs. (4)	0.0035	0.0092	0.0497	0.0140	-0.0238	0.0050	-0.0099	0.0216	0.0242	0.0212	-0.4208	-0.0471

Note: Average treatment effects in percentage points. Inference is based on the bootstrap, by resampling classes. The dependent variable is French and Math proficiency, respectively, at the end of school year. Estimates significant at 10% are marked in *italics*, significant at 5% are marked in **bold**. The bias of OLS as a fraction of the IV estimate is calculated as (estimate$_{\mathrm{OLS}}$ – estimate$_{\mathrm{IV}}$)/ abs(estimate$_{\mathrm{IV}}$).

Using the IV and OLS regression results, we now also attempt to estimate the size of the measurement error. Write the regression of $Y_{i,1}^F$ on $Y_{i,0}^F$, a constant, and the other variables in Xset 1 in the standard notation as

$$Y = X^*\beta_0 + U$$

where X^* contains the true regressors (including true ability), β_0 is the true coefficient vector, and *U* is an error term. As before, we estimate the entire vector β_0 separately in the *D*=1 and *D*=0 population. The observed regressors *X* are obtained as X^* plus measurement error where the variance matrix of the measurement errors is assumed to be

$$\Sigma_\varepsilon = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

In other words, only the pre-test (which is the first regressor) is measured with error. Let us now write $\Sigma_\varepsilon = \sigma^2 e_1 e_1^{'}$ where e_1 is a column vector of zeros with first element being one. Moreover, let us define $\Sigma_X = plim \frac{1}{N} \sum_i X_i X_i '$ and $\Sigma_{X^*} = plim \frac{1}{N} \sum_i X_i^* X_i^* '$. Since $\Sigma_X = \Sigma_{X^*} + \Sigma_\varepsilon$, the OLS estimator converges to

$$plim \hat{\beta}_{OLS} = (\Sigma_X)^{-1} \Sigma_{X^*} \beta_0 = (\Sigma_X)^{-1} (\Sigma_X - \sigma^2 e_1 e_1^{'}) \beta_0$$

This expression can be rewritten as

$$\sigma^2 \Sigma_X^{-1} e_1 e_1^{'} \beta_0 + plim \hat{\beta}_{OLS} - \beta_0 = 0 .$$

Obviously, we do not know β_0 but if our previous instrumental variables assumptions are correct, the IV estimates are consistent for β_0. We therefore estimate σ^2 as the value that sets the following condition to zero as close as possible

$$\sigma^2 \Sigma_X^{-1} e_1 e_1^{'} \hat{\beta}_{IV} + \hat{\beta}_{OLS} - \hat{\beta}_{IV} = 0$$

Since this equation is a vector equation with a scalar unknown, we choose σ^2 as the minimizer of the Euclidean norm of the deviations from zero.

The estimates of σ^2 are shown in the lower panel of Table B.1, separately for the *D*=0 and *D*=1 sample. σ^2 is estimated for each of the three different IV strategies: one pre-test instrumented by the other pre-test (row 2), one pre-test instrumented by its de-meaned square (row 3), both pre-tests instrumented by their de-meaned squares (row 5). The estimates of σ^2 are rather noisy and some estimates are negative. If we ignore the negative estimates, the average σ^2 for the 2nd grade is 0.015 for the *D*=0 sample and 0.013 for the *D*=1 sample. For the 5th grade the average σ^2 is 0.007 for the *D*=0 sample and 0.014 for the *D*=1 sample. (If we replace the negative estimates by zeros, we obtain an average σ^2 for the 2nd grade as 0.012 for the *D*=0 sample and 0.009 for the *D*=1 sample, and for the 5th as 0.005 for the *D*=0 sample and 0.007 for the *D*=1 sample.) Before we discuss the magnitude of these measurement errors in more detail, we consider an alternative estimation approach, which leads to very similar, albeit slightly larger results.

Our *alternative* strategy to estimate the extent of measurement error uses variance restrictions as an alternative to instrumental variables approaches.

For each student we observe the four test scores $Y_{i,1}^F$, $Y_{i,0}^F$, $Y_{i,1}^M$ and $Y_{i,0}^M$, which we relate to true ability as in (6)

$$Y_{i,t}^s = \xi_{i,t}^s + \varepsilon_{i,t}^s .$$

The first four rows of Table B.2 show the standard deviations of the test scores $Y_{i,1}^F$, $Y_{i,0}^F$, $Y_{i,1}^M$ and $Y_{i,0}^M$, which are more or less similar in size.

To identify the variance σ^2 of the measurement error, we can write the relationship between the four true abilities $\xi_{i,1}^F$, $\xi_{i,0}^F$, $\xi_{i,1}^M$ and $\xi_{i,0}^M$ in the following way

$$\xi_{i,t}^s = \xi_{i,0}^M + (\xi_{i,1}^M - \xi_{i,0}^M) \cdot t + \mu_{i,0}^F \cdot 1(s = F) + t \cdot 1(s = F) \cdot (\mu_{i,1}^F - \mu_{i,0}^F) , \qquad (7)$$

where $\xi_{i,0}^M$ is the Math ability at baseline, and $\mu_{i,0}^F = \xi_{i,0}^F - \xi_{i,0}^M$ is the differential ability in French at baseline. Here, $1(s = F)$ is the indicator function which is one if *subject* is French. $\xi_{i,1}^M - \xi_{i,0}^M$ is the individual learning gain in Math from the beginning to the end of the school year and $\mu_{i,1}^F - \mu_{i,0}^F$ is the differential learning gain in French. We do not restrict the distributions of $\xi_{i,1}^M, \xi_{i,0}^M, \mu_{i,1}^F, \mu_{i,0}^F$ in any way. Clearly, model (7) is completely general and therefore does not identify σ^2. As a first identifying restriction, we impose that the learning gains are identical for Math and French, which yields the relationship

$$\xi_{i,t}^s = \xi_{i,0}^M + (\xi_{i,1}^M - \xi_{i,0}^M) \cdot t + \mu_{i,0}^F \cdot 1(s = F) . \qquad (8)$$

This model implies that after double differencing only pure measurement error is left, i.e. that

$$Var(\Delta Y_i^F - \Delta Y_i^M) = 4\sigma^2 ,$$

where $\Delta Y_i^s = Y_{i,1}^s - Y_{i,0}^s$. We can thus estimate the variance from the variance of the differential test score gains, which are given in the fifth row of Table B.2. $\hat{\sigma}^2$ is about 0.017 for the 2nd graders and 0.008 for the 5th graders. We note that these results are very similar across countries, and also to the results we obtained before. (The differences between the 2nd and 5th grade cannot be interpreted since the tests are completely different at the different grades.)

Table B.2: Estimates of variance of measurement error

	Mali 2nd	Mali 5th	Niger 2nd	Niger 5th	Togo 2nd	Togo 5th
	Standard deviations of test scores					
Stddev(Y_1^F)	0.24	0.15	0.26	0.15	0.23	0.18
Stddev(Y_1^M)	0.24	0.15	0.28	0.16	0.21	0.17
Stddev(Y_0^F)	0.21	0.19	0.18	0.12	0.26	0.20
Stddev(Y_0^M)	0.25	0.21	0.26	0.16	0.23	0.18
	Model (8) with equal trend in French and Math					
$\hat{\sigma}^2$	0.018	0.008	0.017	0.008	0.016	0.008
	Estimates of model (9)					
$\hat{\sigma}^2$	0.018	0.009	0.017	0.009	0.015	0.008
Var($\xi_{i,1}$)	0.037	0.013	0.062	0.017	0.031	0.020
Var(ξ_i)	0.044	0.034	0.050	0.016	0.039	0.023
Var(μ_i^F)	0.006	0.002	0.014	0.003	0.018	0.004
Cov($\xi_{i,1}$, ξ_i)	0.019	0.015	0.045	0.013	0.021	0.019
Cov($\xi_{i,1}$, μ_i^F)	-0.010	-0.005	-0.028	-0.006	0.009	0.003
Cov(ξ_i , μ_i^F)	-0.010	-0.005	-0.022	-0.006	-0.003	0.002
a	1.256	6.426	3.579	1.678	-0.096	0.688
b	0.305	2.473	1.600	0.461	-0.324	-0.447
Noise fraction	32.9 %	28.3 %	26.8 %	41.6 %	26.8 %	25.7 %

Note: Only the results with the regular teachers are shown here (D=0).
Corresponding results for contract teachers are very similar.

Model (8) is rather strict in that it treats any differences between French and Math individual learning gains as measurement error. We would therefore like to relax model (8), while noting again that the completely general model (7) would not permit us to identify σ^2. The covariance matrix of the observed $Y_{i,1}^F$, $Y_{i,0}^F$, $Y_{i,1}^M$ and $Y_{i,0}^M$ contains 10 different elements, whereas the model (7) contains 10 unknown covariance elements plus the unknown σ^2.

As a middle ground we consider the model

$$\xi_{i,t}^s = \xi_{i,0}^M + (\xi_{i,1}^M - \xi_{i,0}^M)\cdot t + \mu_{i,0}^F \cdot 1(s=F) + t\cdot 1(s=F)\cdot(a\cdot\mu_{i,0}^F + b\cdot(\xi_{i,1}^M - \xi_{i,0}^M)), \qquad (9)$$

where *a* and *b* are unknown constants, i.e. not individual specific. For example, for *a*=0 and *b*=-0.5, the individual learning gains in French are only half the learning gains in Math. Model (9) contains nine unknown elements (six unknown covariance elements and the parameters *a*, *b* and σ^2). Model (9) implies restrictions on the covariance matrix of $Y_{i,1}^F$,

$Y_{i,0}^{F}$, $Y_{i,1}^{M}$ and $Y_{i,0}^{M}$, which has 10 distinct elements, such that we can estimate all model parameters by GMM via the sample covariance matrix of $Y_{i,1}^{F}$, $Y_{i,0}^{F}$, $Y_{i,1}^{M}$ and $Y_{i,0}^{M}$. (Details on the implementation of the estimator are available upon request.)

The estimation results are provided in the lower part of Table B.2. We note that the estimates of $\hat{\sigma}^2$ are very similar to those obtained from model (8), and are also similar but slightly larger than the ones we had obtained in Table B.1. Given that all these various different approaches to estimate σ^2 lead to very similar results, we consider the $\hat{\sigma}^2$ obtained from model (9) as an upper bound. From these estimates we calculate the average noise fraction at the very bottom of Table B.2 as

$$\frac{4\hat{\sigma}^2}{Var(Y_{i,1}^{F})+Var(Y_{i,1}^{M})+Var(Y_{i,0}^{F})+Var(Y_{i,0}^{M})},$$

which is 25-30% of the variation in the test scores (40% in Niger 5th grade).

Given the results from the upper part of Table B.1, this measurement error does not seem to invalidate the basic findings on the treatment effect of the contract teacher status.

C Quantile treatment effects with measurement error

In Section A we explored the impacts of measurement errors in the test scores on the average treatment effects. We estimated the amount of measurement error and also found that various instrumental variables estimators produced quite similar treatment effects as the nonparametric analysis, which ignored measurement error. In any case, the qualitative interpretation of the treatment effects between countries, subjects and grades was stable throughout. Regarding the estimation of quantile treatment effects, one could imagine that the effect of measurement error might be different here as it also affects the variance (and thus the quantiles) of the test scores. Taking account of measurement error in the estimation of the QTE is more complex, since we first have to generate the conditional distribution of *Y* given *X*, integrate it over the distribution of *X* to obtain the unconditional distribution of *Y* given *X*, and finally invert this to obtain the *unconditional* quantiles. This is done separately for contract teachers and regular teachers, before taking the difference between the two.

In principle, one could use parametric (instrumental variable) quantile regression to obtain the *conditional* quantiles, invert them to obtain the conditional distributions, integrate over *X* to obtain the unconditional distributions and invert again to obtain the *unconditional* quantiles. One could extend this idea to localize the estimator to obtain a fully nonparametric approach to incorporating measurement error. This in-depth analysis, however, is beyond the scope of this paper.

To roughly assess the consequences of measurement error in the nonparametric estimation of QTE (as in Section 6), we limit ourselves to some Monte Carlo simulations where we subject the nonparametric estimator to artificial data with measurement error and compare this to the results without measurement error. To this end, we make use of the estimate of the variance of the measurement error $\hat{\sigma}^2$ that we obtained in Tables B.1 and B.2.

We consider a simple setup with four different data generating processes to assess the impact of measurement error. There are 3 independent control variables: the French pre-test $X_1 \sim N(0.5, 0.03)$, the Math pre-test $X_2 \sim N(0.5, 0.03)$ and a dummy variable X_3, which is distributed Bernoulli(0.5). Hence both pre-tests are continuous and have *variance* 0.03. This value is chosen to mimic the results of Table B.2 as discussed below.

The binary treatment variable is defined as $D = 1(X_1 + X_2 + X_3 + U > a)$, where *U* is a standard normal random variable and *a* is chosen such that the mean of *D* is 0.5. Hence, X_1, X_2 and X_3 are strong predictors of *D* but the variance of *U* is sufficiently large to ensure common support.

The Y^0, Y^1 variables are generated according to one of four different data generating processes (DGP), as given below. The parameters of the DGP have been chosen such that the variances of Y^0 and Y^1 are also very close to 0.03, and such that only very few draws have to be capped at zero or at one to be consistent with the range of our outcome variable.

After the generation of the *X*, *D* and Y^0, Y^1 variables, random measurement errors are added to X_1, X_2 and to Y^0, Y^1. These measurement errors are mutually independent, mean-zero and normally distributed with variance 0.0125. (This variance is the average of $\hat{\sigma}^2$ in Table B.2.) Since the variances of X_1, X_2, Y^0 and Y^1 (before adding measurement error) are each 0.03, their variance is 0.0425 after adding measurement error, such that the fraction of noise is 0.0125/0.0425 = 29.4%, which corresponds to the average of the noise fraction of Table B.2.

Finally, *Y* is generated as $Y = Y^1 D + Y^0 (1 - D)$.

The data generating processes are:

DGP 1: $$Y^0 = 0.8(X_1 + X_2)(0.2 + 0.1U^2) - 0.1X_3 + 0.15 + 0.115U$$

$$Y^1 = 0.5(X_1 + X_2) + 0.1X_3 - 0.1 + 0.11U$$

DGP 2: $$Y^0 = 0.6(X_1 + X_2) - 0.1X_3 + 0.08U$$

$$Y^1 = 0.6(X_1 + X_2) - 0.1X_3 - 0.1 + 0.08U$$

DGP 3: $$Y^0 = 0.8(X_1 + X_2)(0.2 + 0.1U^2) - 0.1X_3 + 0.15 + 0.115U$$

$$Y^1 = 0.8 - 0.125(X_1 + X_2)^4 + 0.1X_3 + 0.06U$$

DGP 4 $$Y^0 = 0.5(X_1 + X_2) + 0.1X_3 - 0.1 + 0.11U$$

$$Y^1 = 0.8 - 0.125(X_1 + X_2)^4 + 0.1X_3 + 0.06U$$

where *U* is a standard normal random variable.

The resulting distributions for Y^0 (solid line) and Y^1 (dashed line) are shown in Figure C.1 for the different DGP.

With this artificial data, the same nonparametric estimation process as in Section 6.1 is applied (with and without measurement error) and repeated 1000 times. (Sample size is 1000.) As a first finding, we note that, not surprisingly, the variance of the nonparametric estimates increases with measurement error. With respect to bias, Table C.1 presents the average estimates over these 1000 replications. The results are most clear-cut for the DGP 2, where the treatment effect is constant at all quantiles. Here we observe that measurement error leads to attenuation bias at all quantiles: The true effects are underestimated (in absolute value) by about 20%. For the other DGP with variable treatment effects, the results are more complex. In most cases, we observe an attenuation effect at the quantiles 0.3 to 0.7. In the tails of the distribution (τ=0.1 and τ=0.9), however, the estimated QTE tend to be larger when measurement error is present. In any case, the differences are not very large, though. All in all, the results of our Monte Carlo simulation therefore do not give rise to concern with respect to the interpretation of our quantile treatment effects in the previous subsection.

Figure C.1: Artificial proficiency distributions with measurement error

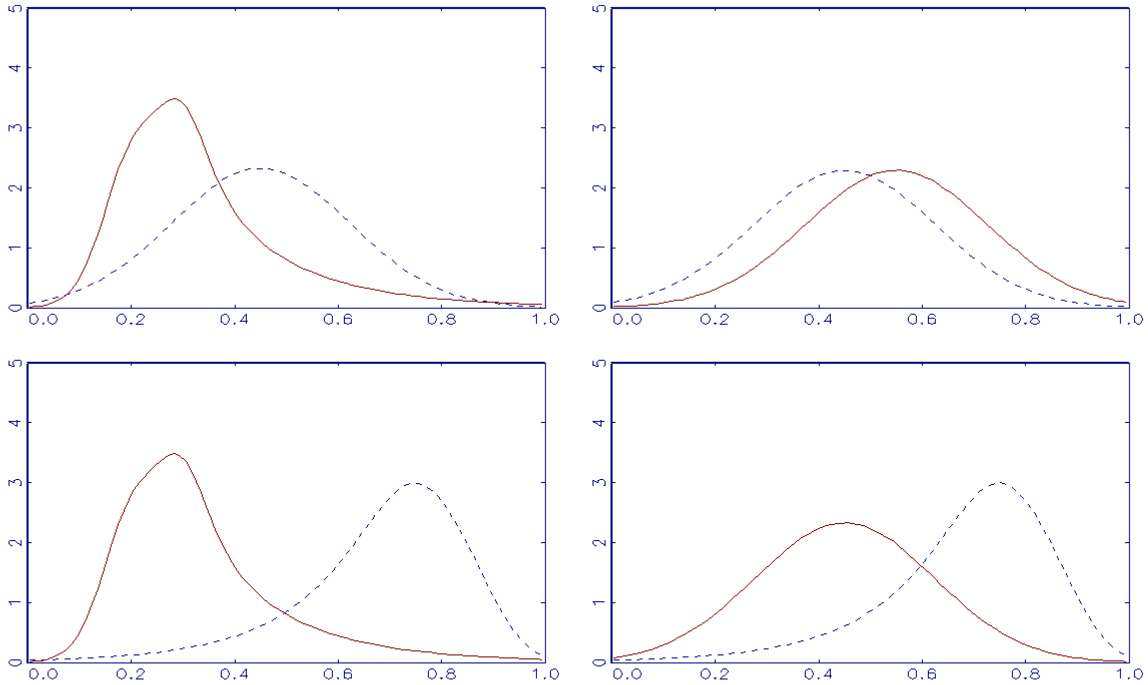


Table C.1: Quantile treatment effects (in percentage points) with / without measurement error

	DGP 1			DGP 2			DGP 3			DGP 4		
	true	no error	with error	true	no error	with error	true	no error	with error	true	no error	with error
ATE local logit	11.2	11.3	12.4	-10.0	-10.2	-8.2	34.2	33.8	33.3	23.0	22.6	22.5
ATE local linear		11.2	12.3		-9.9	-8.1		33.5	33.1		22.3	22.3
QTE τ=0.1	6.0	5.8	9.0	-10.0	-10.3	-8.3	28.8	26.3	28.3	22.8	21.1	21.2
QTE τ=0.2	9.7	9.4	11.3	-10.0	-10.0	-8.2	35.7	34.2	33.7	26.0	25.0	24.1
QTE τ=0.3	11.9	11.8	12.8	-10.0	-9.9	-8.1	38.7	38.0	36.2	26.8	26.1	24.9
QTE τ=0.4	13.5	13.4	13.9	-10.0	-9.8	-8.1	40.2	39.9	37.6	26.7	26.3	25.1
QTE τ=0.5	15.0	14.8	14.8	-10.0	-9.8	-8.1	41.1	40.8	38.3	26.1	25.8	24.9
QTE τ=0.6	16.2	16.0	15.5	-10.0	-9.8	-8.1	41.3	41.1	38.5	25.2	24.9	24.5
QTE τ=0.7	16.7	16.7	15.8	-10.0	-9.9	-8.2	40.6	40.6	38.1	23.9	23.7	23.8
QTE τ=0.8	15.7	16.0	15.4	-10.0	-10.0	-8.3	37.8	38.0	36.7	22.2	22.0	22.8
QTE τ=0.9	11.1	11.5	12.7	-10.0	-10.3	-8.5	30.3	30.5	32.2	19.4	19.5	21.1

Note: For each DGP, the first column shows the true ATE and QTE. The column 'no error' shows the effects estimated by the nonparametric estimator when the DGP is not contaminated with measurement error. (The average estimates over 1000 replications are shown, sample size 1000.) The column 'with error' shows the estimates of the nonparametric estimator when the DGP is contaminated with measurement error.