The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers*

Natalie Bau[†] Jishnu Das[‡]

May 17, 2016

Abstract

We use a unique dataset of 1,533 teachers from 574 public schools to present among the first estimates of teacher value added (TVA) and its correlates in a low income country, as well as estimates of the link between TVA and teacher wages. There are three main findings. First, teacher quality matters as much or more for student outcomes in these contexts as in OECD countries: moving a student from a teacher in the fifth percentile to the ninety-fifth percentile leads to a 0.64 standard deviation increase in test scores, relative to a 0.39-0.55 increase in the United States. Second, observed teacher characteristics are closely linked to teacher compensation but explain no more than 5 percent of the variation in TVA. Finally, there is no correlation between TVA and wages in the public sector, and a policy change that shifted hiring from permanent to temporary contracts, reducing wages by 35%, had no adverse impact on TVA, either immediately or after 4 years. This suggests that teacher quality is inelastic to wage reductions at current wages. The study confirms the importance of teachers in low income countries, extends previous experimental results on teacher contracts to a large-scale policy change, and provides striking evidence of significant misallocation between pay and productivity in the public sector.

*Natalie Bau gratefully acknowledges the support the National Science Foundation Graduate Research Fellowship and the Harvard Inequality and Social Policy Fellowship. We are also grateful to Christopher Avery, Deon Filmer, Asim Khwaja, Michael Kremer, Nathan Nunn, Roland Fryer, Owen Ozier, Faisal Bari and seminar participants at the World Bank, CERP, NEUDC and the University of Delaware for helpful comments. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments they represent.

[†]University of Toronto. (email: `natalie.bau@utoronto.ca`)

[‡]World Bank. (email: `jdas1@worldbank.org`)

1 Introduction

How to recruit and reward teachers is one of the most contentious topics in education today. Some policymakers believe that the best way to improve the quality of teachers is to hire the brightest college graduates by offering high salaries.[1] Others argue that public school teachers are overpaid, and with increasing fiscal stress in many countries, teacher salaries are a natural target for retrenchment (Biggs and Richwine, 2011). Understanding the characteristics that make an individual a good teacher and whether the same characteristics are also highly rewarded by the outside labor market is key to this debate. If, for instance, the brightest college graduates can earn high salaries in other professions but are not better teachers, they may be the wrong population to target for recruitment.

We examine both the question of what makes a good teacher and the link between wages and productivity using a unique dataset that we collected between 2003 and 2007 from the province of Punjab, Pakistan as part of the Learning and Educational Achievement in Pakistan Schools or LEAPS project. These data contain test-score information on matched teacher-child pairs, permitting teacher value-added (TVA) estimation for 1,533 public school teachers from 574 public schools.[2] We are able to combine these data with a regime change at the beginning of the data collection period, which led all new hires to receive lower wages and tempoary contracts. By contrasting the TVA of teachers under the old and the new regime, we assess whether the change adversely affected TVA, either for those hired immediately after the change or those hired several years later.

We first construct TVA estimates and examine how TVA is correlated with a rich set of observed teacher characteristics, including teacher test scores, which we collected, in math, Urdu (the vernacular), and English. Teachers have large effects on students' outcomes. A 1 standard deviation increase in TVA leads to a 0.16 standard deviation in student test scores, but observed teacher characteristics explain no more than 5% of the variation in subject-specific TVA (math, Urdu, and English). The only characteristic that affects TVA systematically (and positively) is the first two years of teaching experience. Apart from

[1] An influential report by McKinsey discusses the importance of closing the talent gap in teaching and how wages can play a part in closing the gap (Auguste et al., 2010). The authors write, "The research presented here suggests the need to pursue 'bold, persistent experimentation' to attract and retain top graduates to the teaching profession" (p. 5). They go on to add, "Given the real and perceived gaps between teachers' compensation and that of other careers open to top students, drawing the majority of new teachers from among top-third students would require substantial increases in compensation" (p. 7).

[2] Only 1,383 teachers from 471 schools appear in most of our analysis since questions about certain teacher characteristics, such as the year a teacher started teaching, were only asked in the fourth round of the data collection.

the first two years of experience, the effect of any single teacher characteristic on TVA is small and not stable to the inclusion of different fixed effects. We do find that higher content knowledge is associated with slightly higher TVA for English but not for math and Urdu. For a smaller sample of teachers, for whom we have repeat test-scores, we account for measurement error using an instrumental variables strategy and again find no strong, systematic relationship between content knowledge and TVA.

These results almost precisely replicate those found in the U.S. (see Rockoff, 2004; Chetty et al., 2014a; and Rivkin et al., 2005). This is surprising because one potential explanation for the low explanatory power of observed characteristics in the U.S. is that stringent eligibility criteria standardize the teacher pool and limit underlying variation in teacher characteristics. In our data, there is enormous variation: 49% of public school teachers do not have a bachelor's degree, and (self-reported) mean days absent per month range from 0.5 at the 10th percentile to 5 at the 90th percentile of the absentee distribution.[3] Despite greater variation in the teacher characteristics variables and additional data on content knowledge, more than 95% of the variation in TVA continues to be driven by characteristics that we do not observe.

We show that our observed TVA estimates pass tests designed to assess bias due to the systematic sorting of children to teachers. In our most novel test, we estimate the correlation between the TVAs of a child's current and future teacher among children who switch schools. We first show that the child's future teacher's TVA after she switches schools does not predict her current teacher's TVA, suggesting that children's unobserved characteristics are not systematically correlated with teacher quality. This is consistent with "as good as random" matching of children to teachers. We then show that the gain in test-scores for a child who switches schools is precisely predicted by the TVA of the teacher that she is matched to, suggesting that TVAs are a meaningful measure of teacher quality.

To understand whether more productive teachers are rewarded with higher wages, we study the link between TVA and wages.[4] We first show that there is zero correlation between wages and TVA in the public sector, with public sector wages rewarding seniority and

[3]For comparison, the absence rate of an average teacher in the U.S. is 5%, and the rate is only 3.5 percentage points higher for schools whose proportion of African American students is in the 90th percentile (Miller, 2012).

[4]Here, we assume that TVA is a useful measure of productivity for teachers. We recognize that teachers have other functions that we do not capture, and how these are rewarded is an important agenda in its own right. However, examining the link between TVA and pay is key to understanding how teacher compensation may affect child test scores, which are a key component of educational performance in any system. Moreover, it has been shown that children's test scores are a strong predictor of a variety of long-term outcomes (for example, see Chetty et al. (2014b)).

education – both of which have small effects on TVA. Although this "zero-gradient" result is widely believed to be true, to our knowledge this is the first direct empirical test in a low-income country. Using similar data from 380 *private schools*, we compare the gradient in the "market" with that in the public sector. When we do so, we find that the rewards to seniority are one-fifth as high and more strikingly, that a one standard-deviation increase in TVA increases wages by 11%. Therefore, even in the absence of a formal testing regime, TVA is somewhat observable and can be rewarded, but the public sector does not have a mechanism to do so.

Turning from the association between TVA and wages, we next assess whether teachers are "paid too much" in the sense that baseline wages are higher than is necessary to attract high quality teachers. Determing whether wages are "too high" is typically difficult since researchers must determine how teachers are compensated relative to their outside options. To do this, they must adjust for different schedules (summer vacation), education levels, cognitive ability and the type of teacher, and different adjustments lead to different conclusions about the relative size of teacher compensation.[5]

Using a fuzzy regression discontinuity in month hired, we approach this question in a different way, by directly comparing the TVA of teachers hired just before and after a hiring regime change led to a large wage decline. If teacher quality is positively correlated with a teacher's outside option, higher quality teachers will be less likely to enter the public sector when wages fall. If there is no such correlation or if high quality teachers' outside options are sufficiently small relative to public sector wages, the outside option will not bind and the quality of new entrants will not change. While we do not observe the precise experiment of a wage decline without any other changes, our regime change is a close approximation.

In the mid-nineties, the government of Punjab (Pakistan's largest province with 60% of the country's GDP) started to explore the use of temporary contracts in the health sector to address the dual problems of poor accountability and high wage and pension costs of public employees (Cyan, 2009). In 1998, after Pakistan conducted two wholly unanticipated nuclear tests, foreign direct investments declined and government budgets came under stress. Figure 1 shows the precipitous decline in dollar deposits in Pakisan following the tests.[6] The decline in FDI provided the political impetus for expanding the use of employees on temporary contracts to the education sector. The government first decreased teacher hiring and then, soon after, replaced the hiring of permanent teachers with contract teachers. In

[5] *The Atlantic* summarizes these different studies and critiques of each of them (Weissman, 2011).
[6] This figure is taken with permission from Khwaja and Mian (2008).

our data, 93% of new hires in 1997 were permanent teachers and by 2002, 89% were contract teachers. Contract teachers differed in the permanence of their tenure (by definition) and their salaries. Average wage payments were *35% lower*, not counting further cost savings in terms of long-term benefits such as pensions.[7]

We first show that contract and regular teachers taught students with similar learning trajectories, although contract teachers were typically assigned to smaller schools with fewer facilities like electricity and libraries. Nonetheless, being assigned a contract teacher in the future does not predict trends in test scores in a school, and test-score gains for students do not predict whether they will be assigned to a contract teacher. Moreover, in our main results, we compare contract teachers and permanent teachers *within* the same school. Combined with extensive controls for selection of students to teachers, including prior test-scores, in the estimation of TVA measures (in line with the methodology of Chetty et al. (2014a)), we believe that any bias in our estimates arising from the systematic allocation of contract teachers to specific schools is likely to be small.

When we combine the policy change with our TVA estimates, we find no evidence that the lower wages led to a corresponding decline in TVA. The precision of our estimates varies according to the estimation procedure, but typically we find a positive impact of contract status on TVA. Adjusting appropriately for the gain in TVA over the first year of teaching suggests that contract teachers had even *larger* positive effects. We also do not find evidence that the pool of new teachers worsened over time with similar TVA estimates for later and earlier hires. More remarkably, even the average *education levels* of the new hires did not change significantly after the regime shift. This suggests that the absence of a decline in the TVA was not only because there is no correlation between education and TVA, as we find in our analysis, but also because wages remained sufficiently high that outside options for the more educated were still lower than what they could earn as a contract teacher.

To see this more clearly, note that in 2003 (the first year of our data), teacher salaries in the public sector were 500 percent higher than those in the private sector (Andrabi et al., 2008) and by 2011, despite very high inflation between 2007 and 2011 that could have been used to erode real wages, they were 8 times as high. When public sector wages are an order of magnitude higher than private sector wages, decreasing the public wage by one-third appears not to affect the pool of job-seekers. The outside option is *never* more attractive than the public sector, even after the wage decline.

[7] We arrive at 35% by regressing log teacher salaries on teacher characteristics, including seniority, as well as an indicator variable for contract status. Thus, we report the difference between contract teacher and non-contract teacher wages after accounting for any differences in observable teacher characteristics.

These results add to a recent literature that highlights the misallocation in public resources between teachers wages and other inputs in low-income countries (Pritchett and Filmer, 1999). In related work in Kenya, Duflo et al. (2011) and Duflo et al. (2014) show that contract teachers cost less, but the test-score gains of children randomly matched to contract teachers are higher. They also show that contract teachers with higher performance were more likely to be rewarded with tenure in later years, and thus career concerns provide additional incentives to exert effort. In India, Muralidharan and Sundararaman (2013) allocate a contract teacher to randomly chosen schools. They show that schools where contract teachers were assigned gained more in test-scores. Using observational data, they suggest that there is an independent contract-teacher effect, beyond the reduction in student teacher ratios caused by the additional teacher. Finally, Bold et al. (2013) repeat the experiment in Duflo et al. (2014) with a NGO and the government. They are able to replicate Duflo et al.'s (2014) results when the NGO implements the policy, but not when the government implements. Bold et al. (2013) suggest that learning from experiments is limited by the non-randomness of the implementation partner as discussed by Allcott and Mullainathan (2012).

Our first contribution it to demonstrate the robustness of the experimental results to (a) short and medium-term recruitment, and (b) comparisons of contractual status for marginal hires assigned to different contract status instead of comparisons of marginal contract hires with average permanent hires.[8] Our results suggest that experimental designs underestimate the long-term effectiveness of contract teachers since they do not adjust for differences in experience between contract teachers and permanent hires. Thus, we can confirm the findings reported in Duflo et al. (2014) and Muralidharan and Sundararaman (2013). Given that we see no signs of selective sorting of students to teachers and no association between contract status and teacher characteristics, our study approximates an experiment in which teachers are assigned their contract status randomly after they have been accepted into teaching, and children are randomly allocated across teachers.[9]

[8]Comparisons based on the performance of the average permanent and marginal contract hire could conflate differences due to experience or cohort effects with differences in contract status (for instance, an average permanent hire is older and more experienced in our data).

[9]Our results also demonstrate the predictive power of experimental results for full-scale policy implementations. Bold et al. (2013) have recently argued in the context of contract teachers that results from experiments cannot be extrapolated to large government programs. Their experiment initiated a new program that likely required a transitional adjustment period. For instance, when the government implemented the program, many contract teacher slots went unfilled over the course of the experiment (Bold et al., 2013). In contrast, our study took place within the context of the regular apparatus of the state for teacher hiring, which exploits a change in the regime through a pen-stroke reform. These reforms retain the "business

Our second contribution is to the debate on wage differentials between the public and the private sector. A large literature from the OECD typically finds public sector premia of 5 to 15 percent, with some portion of the gap explained by differential motivation, sector-specific productivity and the selection of workers (Disney and Gosling, 1998, Dustmann and Van Soest, 1998 and Lucifora and Meurs, 2006). In contrast, the wage differential in our study is 500% and within the public sector, we show that a decline in wages of (at least) 35% has no negative impact on productivity as measured by TVA.[10] A related experiment in Indonesia doubled teachers salaries but again found no increase in learning; this is a conceptually separate experiment from ours since it focuses on the link between effort and wages for existing teachers rather than the link between wages and recruitment (De Ree et al., 2014). Nevertheless, the overall message is consistent across all these studies: there are large and significant misallocations in the pay and productivity of public sector teachers in low-income countries.

We should caution that although we get close to the natural experiment of a wholesale reduction in wages, this policy change combined changes in remuneration levels with changes in the returns to effort through career concerns. Here, we are unable to separate the two effects; doing so would require a separate experiment. The results therefore show that temporary contracts induce a combination of teacher effort and quality that can yield the same learning at half the cost.

The remainder of our paper is as follows. Section 2 describes the setting and context, and section 3 discusses the data. Section 4 discusses TVA estimation, the results of regressions of TVA on teacher characteristics, and the robustness of the TVA measures. Section 5 describes a simple model of teacher selection into the public and private sector as a function of wages. Section 6 presents the empirical strategy and the results for our study of the effect of the regime change, focusing on the difference between contract and permanent teachers' value-addeds. Section 7 concludes.

as usual" aspects of the public program, thus offering a window into how such a reform would operate in practice. However, such a regime change is harder to evaluate experimentally, since the legal and regulatory requirement of uniformity in hiring would have to be suspended. The approach used here – a fuzzy regression discontinuity in month hired – is a natural analogue to differences-in-difference and event studies that exploit regime changes across states in a country.

[10]Since we do not include future liabilities such as pensions in this accounting, the wage difference is a lower bound in our study.

2 Setting and Context

Our study uses data from rural areas of Punjab, Pakistan, the largest province in the country with a population of 70 million. The majority of children in the province can choose to attend free public schools, or they can pay to attend private school, and at the primary level, one-third of enrolled children choose to do so.[11] Although funding for public schools has traditionally been small, in recent years, the government of Punjab has ratcheted up education budgets from 468 million dollars in 2001-2002 to 1.680 billion dollars in 2010-2011 (Ishtiaq, 2013). Much of this expenditure is on recurring budget items, and, similar to other low-income settings, teachers' salaries account for 80 percent of spending (UNESCO Islamabad, 2013).[12] Given the high share of teacher salaries in overall education budgets, the link between pay and productivity is critical for education policy.

Whether public sector teachers wages in Pakistan are 'adequate' depends on the comparison. Comparisons across countries, in this case with Indian states, show that both Pakistani and Indian teachers earn, on average, 5-7 times GDP per-capita (Siniscalco, 2004 and Aslam, 2013). Comparisons to other "comparable" professions suggests that teacher remuneration is comparable to salaries for similar professionals. Each of these comparisons has obvious problems. Comparisons across countries require that teachers are efficiently compensated in the "benchmark" country. Comparisons across professions are subject both to selection concerns and differences in the job profiles across occupations.

Teacher salaries in the private sector provide an alternative benchmark. Andrabi et al. (2008) show that teachers' wages in private schools were one fifth of teacher salaries in public schools in 2003-2004, and public school salaries have only grown relative to private school salaries since then (figure 2). Similar wage gaps have been documented in Colombia, the Dominican Republic, the Philippines, Tanzania, Thailand, and India (see Jimenez et al.; Kremer and Muralidharan). These large wage premiums may reflect a lack of accountability and the strength of teachers' unions rather than greater productivity. Absenteeism is high in the public sector and firings are rare since teachers are protected by permanent contracts (Chaudhury et al., 2006). In our own sample, public school teachers self-reported absences of 2.6 days per month compared to 1.9 days per month for private school teachers. Recent research accounting for selection bias in both Pakistan (Andrabi et al., 2010) and India (Mu-

[11] Religious schools, or madrassas, account for 1-1.5% of primary enrollment shares, and their market share has remained constant over the last two decades (Andrabi et al., 2006).

[12] Bruns and Rakotomalala (2003) show, in a study of 55 low-income countries, that teacher salaries account for 74 percent of recurring spending by the government on education.

ralidharan and Sundararaman, forthcoming) shows that attending private schools, despite a lower per-student cost, improves student outcomes.

Unfortunately, a direct public-private comparison of the wage gap is also confounded by the vast differences in observed teacher characteristics between the two sectors. Appendix table A1 shows the large differences in training (90% versus 22% in the public sector relative to the private sector), education (51% hold a bachelor's degree compared to 26% in the private sector), gender (45% female versus 77%), and local residence (27% local versus 54%). Private school teachers also report 11 years less teaching experience on average. Using an Oaxaca-decomposition exercise, Andrabi et al. (2008) argue that controlling for observed characteristics explains little of the wage gap between public and private school teachers, but there is currently little direct evidence on the link between pay and productivity in the public sector.

2.1 Natural Experiment

The second half of this paper exploits a natural experiment, combined with extensive data on matched student-teacher pairs, to assess how a large wage reduction affects the quality of those becoming public teachers. As we discussed previously, the government of Punjab started exploring changes in hiring practices in the mid-nineties, responding to both reports of low accountability and performance and concerns about the budgetary implications of high wages and benefits for public sector employees. Unanticipated nuclear tests in 1998 led to international sanctions and a worsening of the budgetary position of the province, providing the final impetus for changes in public sector hiring practices and leading to a much wider use of contract teachers in public schools. Thus, using a fuzzy regression discontinuity design in month hired, we can estimate the effect of offering temporary contracts with lower wages on teacher characteristics and productivity.

Figure 3 shows the distribution of years hired for the teachers we observe in our sample. While the number of teachers hired each year varies, corresponding to the practice of "batch" hiring in the province, the period following the sanctions (1998-2001) is a uniquely long period of low hiring. After normal hiring resumed in 2002, almost all teachers in the province were hired on temporary contracts.[13] The contract teachers were not tenured and were paid 35% less than observationally similar permanent teachers.

Cyan (2009) notes that the institution of contract hiring was supported by a more cen-

[13]Contract arrangements in Punjab became more common from 2000-2001 on (Hameed et al., 2014) and in 2004, the Government of Punjab announced its Contract Appointment Policy (Cyan, 2009).

tralized hiring process that relied on a point system based on employee qualifications, as well as interview performance. The policy also dictated that contract employees would undergo increased performance evaluation, though Cyan (2009) suggests that, in reality, this often was not the case. Nonetheless, in surveys, 45% of contract teachers said that performance evaluations were linked to their contract renewal (Cyan, 2009). Performance evaluation may have increased teacher effort: 74% of surveyed contract teachers said that they were made to work more than regular teachers, and Cyan (2009) reports that absenteeism and disciplinary infractions appeared to be lower among contract teachers. Cyan (2009) also notes that, as of 2009, there was no formal process for regularizing the contract teachers who are typically employed on 3-5 year contracts. Consistent with this, 71% of teachers said that they did not think their jobs offered them an opportunity for "professional growth," and 95% of teachers reported working on a temporary contract for more than three years. Therefore, in 2009, it seems that most contract teachers did not expect to be regularized in the future. However, following a victory in a court case in 2012, most contract teachers did become permanent teachers and received higher wages thereafter.

This natural experiment allows us to conduct a simple but important exercise to understand the effects of changes to teacher hiring policies. By examining what happens to teacher quality when the government decreases salaries by more than one-third for *all* incoming teachers, we can directly assess how large-scale contract teacher policies affect teacher labor supply and student outcomes.

3 Data

We use data collected across four rounds (2003 to 2007) of the Learning and Educational Achievement in Punjab Schools Survey (LEAPS). The original sample includes 823 schools (496 public) in 112 villages of 3 districts in the province of Punjab, with an additional 111 public schools entering the sample over the next four years.[14] The project was designed as part of a study of the rise of private schooling and, as a result, all the villages included in the study had at least one private school when the study began in 2003.[15]

For our purposes, three parts of the data collection are key. First, a teacher roster was

[14]The three districts were chosen on the basis of an accepted stratification of the province into the better performing north and central regions and the poorly performing south.

[15]Thus, sample villages are generally wealthier, larger, and more educated than average rural villages. At the beginning of the study, one-third of all villages in the province reported having a private school, but 50 percent of the province's population lived in such villages (Andrabi et al., 2008).

completed for all teachers within the school in each year of the survey. This roster included socio-demographic data on teachers (gender, age, educational attainment) and in the fourth round, month-level data on when the teacher began teaching in public schools. We use variables from the teacher roster to look at the difference between contract and permanent teachers in demographic characteristics, salaries, and subject knowledge. Appendix table A1 provides summary statistics on these characteristics for public school and private school teachers across the four rounds of the survey.[16]

Second, to assess learning outcomes, LEAPS tested children in the survey schools. English, Urdu, and mathematics tests were administered to children in grade 3, grade 4, and grade 5 between 2004 and 2007. The tests were low-stakes and designed by researchers to maximize precision over a range of abilities in each grade. To avoid the possibility of cheating, project staff, with clear instructions not to interfere, administered the test directly to students. Test booklets were retrieved after class, so there was no missing testing material. Tests were scored and equated across the four rounds using Item Response Theory, yielding scores in each subject with a mean of 0 and a standard deviation of 1 (Das and Zajonc, 2010). Item response theory weights questions differently according to their difficulty and allows us to equate tests over years so that a standard deviation gain in year 1 is equivalent to a standard deviation gain in year 4. The tests could be equated because we included linking questions across any two years and for some questions, across multiple years. Appendix table A2 shows average test score gains by year over the four rounds of testing in a balanced panel of public school students. Appendix table A3 provides information on the type of questions that were asked and how students in different years of the balanced panel performed on these questions. These tests form the basis for the estimation of TVA in this paper. Appendix table A1 reports summary statistics for yearly changes in student test scores in math, English, and Urdu. Third, the teacher roster was supplemented with more detailed information on the teachers of tested children, including the results of tests im math, Urdu, and English given to teachers.

Teacher quality is identified following the TVA literature (for example, Rockoff, 2004; Chetty et al., 2014a; and Kane and Staiger, 2008) by regressing student test scores on a function of their lagged test scores, round, grade, and teacher fixed effects. Teacher value-

[16]At times, we wish to compare teachers in terms of measures that were collected in different survey rounds (or collected over multiple survey rounds) such as school facilities or teacher absences. To normalize these measures, we regress them on year fixed effects and teacher or school fixed effects, depending on the level at which the characteristic is observed. We then use the teacher or school fixed effect as the teacher-level measure. This process is analogous to how we combine test score data from multiple years to calculate teacher value-added measures.

added is the estimated teacher fixed effect. The panel structure of the data, where both students and teachers are observed multiple times, is important for identification: to be included in the value-added calculations, students must be observed at least twice across consecutive years, since they require a lagged test score to control for selection. To separate correlation in student outcomes within years from TVA, at least some teachers must also be observed across years.

Figures 4 and 5 document the sources of variation in our data. Figure 4 shows that 914 public school teachers were observed across multiple rounds, allowing us to separately identify the portion of student test score gains due to teacher assignment from the portion explained by testing year. While this variation is necessary to separately identify testing round fixed effects, we are still able to estimate TVA for teachers who appear in only one round of the data. Figure 5 shows that 16,386 students are observed across multiple rounds, allowing us to compute lagged test scores and therefore, better account for the selection of students to teachers. In the end, of the 1,756 teachers observed with tested students, we are able to estimate TVAs for 1,533 teachers.[17]

To account for unobservable variables that may bias teacher quality estimates or for unobservable school quality apart from teacher quality, we can also de-mean TVA estimates at the school level. However, we cannot separately identify pure school effects (as opposed to a school simply having better teachers on average) since we do not observe teachers in more than one school. Therefore, the demeaned TVAs should be interpreted as a within school ranking of teacher quality. Demeaning at the school level requires that more than one teacher was observed in the school over the course of the study. TVAs for teachers in the 158 public schools where only one teacher was ever observed with tested students are left out of the within-school TVA sample. These teachers account for 2,357 child-year observations (1,771 unique children).

Appendix table A4 provides more information on the sources of variation for the TVA calculations. In year one, since only 3rd graders were tested, very few students were observed in schools where more than one classroom was tested. In future years, some students were held back, others were promoted, and another sample of 3rd graders was added in year 3, allowing students in a larger number of classrooms to be tested. Columns 1 and 2 describe the sample used to calculate the cross-school TVA estimates. Columns 3 and 4 describe the variation used to calculate the within school TVA measures. Finally, while hiring was largely

[17] The sample size falls because we are unable to estimate TVAs for teachers that we only observe in year 1 since there are no lagged test scores available for their students.

frozen from 1998 to 2001, a small group of 113 teachers in our sample were hired during this four year period. When we test for the effect of temporary contracts on student outcomes, we drop this group, since they are likely to be highly selected. Otherwise, they would be the main source of identifying variation in our fuzzy regression discontinuity specifications, which rely on a sharp discontinuity in contract status following 1998. However, we retain these teachers and their students when we estimate TVAs.[18]

4 Teacher Value-Added

4.1 Estimating Teacher Value-Added

To compute TVA, we estimate the following regression, including all child-year test score observations:

$$y_{it} = \beta_0 + \sum_a \beta_a y_{i,t-1} I(grade = a) + \gamma_j + \alpha_t + \mu_g + \epsilon_{it},$$

where y_{it} is student i's test score in year t, γ_j is the teacher fixed effect, α_t is the round fixed effect, and μ_g is the grade fixed effect. Then, γ_j is the TVA, equivalent to the underlying unexplained variance in test score gains associated with students having the same teacher. This formulation, conventional in the TVA literature, is similar to those used by Kane and Staiger (2008), Harris and Sass (2006), Chetty et al. (2014a), Chetty et al. (2014b), and Araujo et al. (2014).

Like Chetty et al. (2014a) and Kane and Staiger (2008), we do not include child fixed-effects to account for additional unobservable selection of students to teachers, which may bias TVA estimates (Harris and Sass, 2006). Identification with child fixed-effects would be based on a smaller sample of 16,512 children (51% of the sample) who are observed with multiple teachers over time.[19] More worryingly, measurement error in how teacher codes are entered into the data will lead to false switchers – students who appear to be switching teachers but actually are not. Even with a small number of false switchers, this could lead to large biases in the estimation of TVA by inducing spurious correlations between the TVA of teachers with similar ID numbers.

To see this, suppose that 1 percent of teacher IDs are randomly entered incorrectly. This

[18] 113 teachers over a four year period is in contrast to 212 teachers hired in 1997, directly before the hiring freeze, and 110 teachers hired in 2002 alone.

[19] In Pakistan, teachers teach multiple grades, and students change teachers less frequently.

will have little impact on TVA estimates that utilize the full sample. Now suppose that 10 percent of students change teachers each year. When identifying variation comes only from the test scores of students who change teachers, these incorrect entries account for 9 percent of the variation.[20] In other words, when we restrict the sample to students who change teachers, we always include incorrect ID entries, but we shrink the number of correct ID entries, increasing the percentage of the variation that is driven by students with incorrectly entered IDs.[21]

We also do not use empirical Bayesian methods to estimate TVA. The empirical Bayes approach proposed by Kane and Staiger (2008) relies on the assumption that TVA is time invariant. Since teacher experience, particularly in the first two years of teaching, increases teacher effectiveness (Rockoff, 2004; Chetty et al., 2011), controlling for teacher experience is necessary for this assumption to be valid. Experience is collinear with year hired, which in our setting, is highly correlated with contract status due to the sharp change in the hiring regime. Virtually all teachers with 0-5 years of experience are contract teachers, and virtually all teachers with more than 5 years of experience are permanent teachers. Since we cannot flexibly control for experience without subsuming the temporary contract effect, our estimates of γ_j utilize the full sample of students and teachers, averaging over teacher effectiveness at different experience levels.

Nevertheless, while we cannot fully non-parametrically control for experience effects (a necessary step in the standard empirical Bayes estimation procedure), in future analysis, we control for 0-1 years of experience to account for most of the experience effect and separate experience effects from the effects of other teacher characteristics on the TVA estimates.

[20] To arrive at this number, note that there are three cases where a student-year observation will be included in the sample: (1) the teacher ID was incorrectly entered, but no switch actually occurred (probability = $0.01 \times 0.9 = 0.009$), (2) the teacher ID was correctly entered and a switch occurred (probability = $0.99 \times 0.1 = .099$), and (3) the ID was incorrectly entered and a switch occurred (probability = $0.1 \times 0.01 = 0.001$). Then the probability that the teacher ID is mis-attributed in an observation included in the sample is $\frac{0.01}{(0.009+0.099+0.001)} = 0.09$.

[21] More formally, consider a case where students are identical and TVA is randomly distributed, so there is no correlation between a student's future TVA and his current TVA. Now, also assume that a student has a probability p of changing teachers each year, and an ID has a probability e of being incorrectly entered. Then, when the TVA of teacher is calculated for teacher j, it will be a weighted mean of the teacher's true TVA and the TVAs of teachers of any students with mis-attributed IDs. Therefore,

$$E(\widehat{TVA_j}) = \frac{p}{e(1-p)+p(1-e)+ep} TVA_j + \frac{e}{e(1-p)+p(1-e)+ep} \overline{TVA_j},$$

where $\overline{TVA_j}$ is the mean TVA in the teacher population and $\widehat{TVA_j}$ is the estimate of the TVA for teacher j. This expression formalizes the intuition that the bias decreases in the true probability of switching p and increases in the error rate e.

This methods exploit non-linearity in the experience effect, as the literature shows that after the first two years of teaching, experience's effect on student outcomes plateaus (Rockoff, 2004). Using within-teacher observations of student test-scores in different years, we also verify that this is the case in Pakistan in the following section.

One shortcoming of our TVA estimates is that they do no capture teachers' heterogeneous effects on different students. In reality, such heterogeneity may be important for students' outcomes. For instance, Bau (2015) shows that schools in Pakistan can have different effects on the outcomes of more and less advantaged students. Relatedly, Aucejo (2011) shows that teachers responded to the incentive structure of No Child Left Behind in the United States by increasing the outcomes of their lower ability students at the expense of higher ability students. In another example of match mattering for teacher effectiveness, Muralidharan and Sheth (2013), Antecol et al. (2015), Dee (2007), and Hoffmann and Oreopoulos (2009) measure the effects of the teacher-student gender match. While understanding these heterogeneous effects is important, we do not attempt to capture them. If teachers have heterogeneous effects on students, our TVA measures can be thought of as capturing the effect of a teacher on the *average* student.

In the next sections, we first estimate the TVA and examine TVA's correlations with teacher characteristics. We then use the TVA estimates to assess the link between productivity and wages, both in terms of the gradient (do higher TVA teachers earn more?) and the intercept (does lowering wages reduce average TVA?).

4.2 Teacher Value-Added Results

Using our TVA estimates, we first estimate the association between TVA and student performance and the link between TVA and observed teacher characteristics. Specifically, we estimate

$$TVA_j = \beta_0 + \Gamma X_j + \alpha_d + \varepsilon_j,$$

where TVA_j is a teacher j's average value-added over math, Urdu, and English; X_j consists of teacher characteristics, including an indicator variable for some training, an indicator variable for having a bachelor's degree or greater, an indicator variable for having 3 or more years of experience in 2007, an indicator variable for female, an indicator variable for whether a teacher is local, an indicator variable for whether a teacher has a temporary contract, controls for age and age squared, and in some specifications, controls for a teacher's average test scores in math, Urdu, and English; and α_d is a district fixed effect. In some specifications, we also include a school fixed effect. Table 1 presents the results from this specification. The first two columns report regression results without controlling for teachers' own test scores. When these are included, in columns 3 and 4, the sample size drops since not all teachers took the tests. Much like in the United States, most teacher characteristics are not consistently associated with greater mean teacher value-added. Across specifications, we are never able to explain more than 5% of the variation in mean TVA.

Content knowledge does appear to increase TVA, but only for English. Since pedagogues frequently point to low content knowledge as a severe problem in low-income countries, this is surprising. In appendix table A5, we further investigate this result by regressing subject-specific TVAs on teacher characteristics and content knowledge. The results in appendix table A5 confirm those in table 1. The mean English test score is highly correlated with English TVA in the district fixed effects specification – a standard deviation increase on the English test score leads to a 0.06 standard deviation higher TVA measure. However, Math and Urdu content knowledge are not correlated with a teacher's math and Urdu TVAs.[22] Presaging our discussion on wages and productivity in later sections of the paper, we find no correlation between TVA and two key characteristics–education (measured as whether a teacher has a Bachelor's degree) and whether the teacher has some training (with a negative point estimate, significant when we also include content knowledge).

Columns 1-4 treat each teacher as a single observation, but unlike the other characteristics in these regressions, experience is not time invariant. Since we observe teachers multiple times at different levels of experience, we can use our panel data to better identify the experience effect. In column 5, an observation is a student-year, and we regress mean test scores on teacher experience and lagged student test scores, controlling for teacher fixed effects, which capture any time invariant teacher characteristics, using the following specification:

$$mean\,test\,score_{ijt} = \beta_0 + \beta_1 I(exp_j \leq 1) + \beta_2 mean\,test\,score_{ij,t-1} + \gamma_j + \varepsilon_{ijt},$$

[22]One concern is that the effect size of content knowledge is attenuated because of measurement error in the test-scores. For a smaller sample of 622 teachers, we have two tests scores, and we estimate the specifications in appendix table A5 using instrumental variables regression (instrumenting one test-score with another). These estimates are imprecise and insignificant, with a coefficient of 0.202 for English, 0.279 for Urdu and -0.325 for math in the mean TVA specification (the equivalent of column 7 in appendix table A5). Thus, even when we account for measurement error, we see no systematic positive relationship between teacher knowledge and student test scores.

where i denotes a student, j denotes a teacher, and t denotes a year. $I(exp_j \leq 1)$ is an indicator variable equal to 1 if a teacher j has 0 or 1 years or experience in time t, and γ_j is a teacher fixed effect. Then, we report β_1, our coefficient of interest. The results in column 5 suggest that the experience effect is large: the outcomes of students of teachers with only 0 or 1 years of experience are 0.3 standard deviations worse than those of other students. In appendix table A6, we expand this specification to include controls for 0-1, 2, 3, and 4 years of experience. Our results confirm that the most important gains from experience occur in the first 0-1 years of teaching. We find very large experience effects in the first year: students of a teacher with 0-1 years of experience have test scores between 0.54 standard deviations (in English) and 0.75 standard deviations. (in math) lower than students of teachers with 5 or more years of experience. In the second year, the penalty is lower, ranging from -0.48 for Urdu to (a statistically insignificant) -0.25 for English, and by the third year, there are no further significant experience effects. Note that because we are comparing the outcomes of the students of the same teacher over time, these effects are more likely to have a causal interpretation relative to other attributes that are fixed over the observation period.

The results from the regressions of TVA on observed characteristics mirror broader results from the literaturein high income countries (for a review, see Hanushek and Rivkin (2006)). In public schools, few conventional teacher observable characteristics are consistently correlated with student outcomes besides teacher experience, which has a large, non-linear effect in the first few years of teaching. This is not to suggest that teacher quality itself does not matter. In fact, like in the United States (Rockoff, 2004 and Chetty et al., 2014a) and Ecuador (Araujo et al., 2014), we find that teacher quality is important. To estimate a comparable measure of the variance of teacher effects to the estimates from other countries, which control for experience, we re-estimate our TVAs controlling for year-hired fixed effects. Estimating our TVAs with year-hired fixed effects means that the variance of the TVAs captures variation in teacher quality that is *not* related to contract type or experience. Since the variance of these estimated TVAs will be biased upward due to sampling bias, we solve for the sampling bias and subtract it from the variance of our estimates (see the sampling bias estimation appendix). Araujo et al. (2014) make similar corrections, and like Araujo et al. (2014), we account for the fact our TVA estimates are de-meaned within schools.

After implementing this sampling bias correction, we find that a one standard deviation increase in TVA increases student test scores by 0.16 standard deviations (averaging across subjects) using the within school estimates. In contrast, using within-school estimates, Rockoff (2004) finds that a 1 standard deviation better teacher will lead to a – still meaningful

– test score gain of 0.1 standard deviations in the United States, and Araujo et al. (2014) find that a 1 standard deviation better teacher will lead to a 0.07-0.13 standard deviation gain in student test scores. Chetty et al. (2014a), who account for drift in their TVA estimates, find that a 1 standard deviation incrase in teacher quality will increase test scores by 0.1-0.14 standard deviations. Our somewhat larger effect sizes are consistent with the idea that variation in teacher quality may be higher in areas where variation in underlying teacher characteristics is also higher, but we caution against drawing strong conclusions from the diffence. A standard deviation in test scores in the United States or Ecuador may not be comparable to a standard deviation in Pakistan, and the results are sufficiently similar to those found in other countries that they could merely reflect variation across different samples.

4.3 TVA Robustness

In this section, we consider possible threats to the validity of our TVA measures. When we estimate our TVAs, we control for a rich function of lagged student test scores to account for the non-random assignment of students to different teachers. However, if these lagged tests scores do not sufficiently account for selection, our TVA estimates may be biased. Therefore, we first develop a test for selection that examines how sensitive our results are to the inclusion of additional controls. We recalculate out TVA estimates including demographic, socioeconomic, and school inputs controls. In the new estimates, we control for the student's household assets index,[23] gender, and parental schooling. We also include controls for two indices of school facilities that vary over time and time-varying school-level student-teacher ratios. Appendix table A7 reports the correlations between the new TVA estimates and the main estimates that used only functions of lagged test scores as controls. Even though the new TVAs use substantially more information about students' socioeconomic status and school-level inputs, they are highly correlated with the old TVAs (correlations of .977 in English, .969 in math, and .965 in Urdu). Therefore, our TVA estimates are not particularly sensitive to the inclusion of demographic controls, again suggesting that these estimates are not greatly biased by the selection of students to different teachers.

Next, we test whether our TVA estimates predict actual student test score gains. To do so, we will use an "out-of-sample prediction" test following Chetty et al. (2014a). We focus

[23]Following Filmer and Pritchett (2001), we create an asset index by predicting the first factor of a principal components analysis of indicator variables for ownership of different assets. These indicator variables are for owning beds, a radio, a television, a refrigerator, a bicycle, a plow, agricultural tools, tables, fans, a tractor, cattle, goats, chicken, watches, a motor rickshaw, a scooter, a car, a telephone, and a tubewell.

on school switchers and test whether the TVA of a school switcher's new teacher predicts his test scores after switching to that school, controlling for his lagged test score. If TVA is a meaningful measure of teacher quality, a teacher's TVA should predict her student's learning gains. The specific regression specification is:

$$test\ score_{ijt} = \beta_0 + \beta_1 TVA_j + \beta_2 test\ score_{ij,t-1} + \alpha_s + \varepsilon_{ijt}, \tag{1}$$

where $test\ score_{ijt}$ is the test score of a student i with a teacher j in year t, which can be in math, Urdu, or English or the the average across all three, TVA_j is the value-added of a students teacher in the relevant subject, and α_s is a school fixed effect. The sample consists of students who are in a new public school in period t. Because we limit the sample to school-switchers, β_1 will not be influnced by common shocks at the school-level that are correlated over time.

Before conducting this test, we must ensure that β_1 is not biased by selection between students and teachers. If students who learn quickly are more likely to select to certain teachers, then these teachers will appear to have a higher TVA and these high TVAs will also be related to students' outcomes. To test whether this is the case, inspired by Rothstein (2010), we first test for selection among the population of school-changers by testing whether the TVA of a student's future teacher after a school change predicts the TVA of her current teacher. Focusing on school-changers ensures that our test will not find spurious correlations between future and current TVAs due to the fact that school-grade level shocks to the current teachers' students' outcomes will effect the lagged test scores used to calculate the future teachers' TVAs as described by Chetty et al. (2015). If TVA estimates are biased because students with different unobservable characteristics select different teachers, then the TVAs of two teachers with the same or similar students will be correlated. Table 2 presents the results of this test. Across all subjects, there is no evidence of a correlation between current and future TVAs. These test results suggest that selection of students to different teachers on unobservable characteristics is not a major source of bias among the school-changers.

Table 3 reports the results of the regression of students' outcomes on the TVA of their new teachers when they change schools (equation 1). TVA in a subject is highly predictive of the students' outcomes. For average test scores, a one standard deviation increase in TVA increases student test score gains by 0.852 standard deviations and this coefficient is significant at the 1% level. In fact, we cannot reject that the true coefficient is 1, as would be the case if we had estimated the "true" TVA. Given the potential for measurement error and attenuation bias in our small sample, these results suggest that our TVA estimates are very predictive of real student gains from teacher quality.

5 Framework for Teacher Selection

In subsequent sections of this paper, we will test whether teacher productivity, as measured by TVA, is linked to wages. Before describing our methodology for estimating the effects of the regime change that reduced teacher salaries by 35% on teacher productivity, it is useful to outline a simple framework to help interpret the results. Suppose there are N teachers and $M < N$ positions in the public sector. Each teacher has a productivity θ_j drawn from a bounded distribution F, with a maximum of θ_{max} and a minimum of θ_{min}. In the public sector, teachers receive a wage w_{pub} set exogenously by the government, and the public sector hires randomly from its applicants. In the next section, we test whether the assumption of an exogeneous wage, unrelated to productivity, in the public sector is reasonable. In the private sector, due to free entry, private schools make zero profits and a teacher receives her productivity, θ_j. Note that this implies a link between productivity and wages in the private sector, and we return to this in the results below. For simplicity, teachers can costlessly apply to a public sector job or go directly to the private sector. If she applies, a teacher will get a public sector job with probability $p = \frac{T(w_{pub})}{M}$, where T is the endogeneous number of teachers applying to public positions. If she does not get the public sector job, she enters the private sector and receives the private sector wage. Since applying to the public sector is costless, a teacher will always apply for a public sector job if $\theta_j < w_{pub}$.

Then, for a given w_{pub}, there are three possible outcomes:

1. $w_{pub} < \theta_{min}$: In this case, even the least productive teacher makes more in the private sector than they would in the public sector, so no teachers enter the public sector, and there is a shortage in the public sector.

2. $w_{pub} > \theta_{max}$: In this case, even the most productive teachers would make more in the public sector, so all teachers apply to the public sector, and the average productivity in the public sector is $\int_{\theta_{min}}^{\theta_{max}} \theta_j f(\theta_j) \partial\theta_j$. Therefore, there is no shortage since there are $N > M$ applicants.

3. $\theta_{min} < w_{pub} < \theta_{max}$: Then, there $\exists\ \theta^* = w_{pub}$ such that all teachers with productivity greater than θ^* do not apply to the public sector and all teachers with productivity less than θ^* do. Thus, the average productivity of the public sector is $\int_{\theta_{min}}^{w_{pub}} \theta_j f(\theta_j) \partial\theta_j <$

$\int_{\theta_{min}}^{\theta_{max}} \theta_j f(\theta_j)\partial\theta_j$. In this case, it is ambiguous whether there is a shortage of teachers since $T = N \times F(\theta^*)$ may be less than or greater than M.

The first corner case isn't relevant for our empirical context, since we observe that there are teachers entering the public sector before and after the wage change. Therefore, when we study the effect of a decline in w_{pub} to $w'_{pub} < w_{pub}$, we focus on the second two possible equilibria under w_{pub} and w'_{pub}, in which at least some teachers always enter the public sector. First, consider the case where $w_{pub} > \theta_{max}$. Then, there are two possibilities once wages decline to w'_{pub}:

1. $w'_{pub} > \theta_{max}$, and the average productivity in the public sector is still $\int_{\theta_{min}}^{\theta_{max}} \theta_j f(\theta_j)\partial\theta_j$. In this case, there is no shortage ($T > M$), since all teachers apply for public positions.

2. $w'_{pub} < \theta_{max}$, and $\exists\ \theta^{**} = w_{pub}$ such that all teachers with productivity greater than θ^{**} do not apply to the private sector and all teachers with productivity less than θ^{**} do. Then, the new average productivity of the public sector will be $\int_{\theta_{min}}^{w'_{pub}} \theta_j f(\theta_j)\partial\theta_j < \int_{\theta_{min}}^{\theta_{max}} \theta_j f(\theta_j)\partial\theta_j$. Therefore, under w'_{pub}, average productivity in the public sector declines. As before, it is ambiguous whether there is a shortage.

Now consider the case where $\theta_{min} < w_{pub} < \theta_{max}$. Then, under $\theta_{min} < w'_{pub} < w_{pub} < \theta_{max}$, there will be a new $\theta^{**} < \theta^*$, where all teachers with $\theta_j > \theta^{**}$ do not apply to the public sector. In this case, average productivity in the public sector declines from $\int_{\theta_{min}}^{w_{pub}} \theta_j f(\theta_j)\partial\theta_j$ to $\int_{\theta_{min}}^{w'_{pub}} \theta_j f(\theta_j)\partial\theta_j$. Again, it is ambiguous whether there is also a shortage.

Thus, when we study the effect of the wage decline in our data, there are two empirically relevant possibilities. The first is that the average productivity of entering public school teachers remains the same after the wage decline, suggesting that both w_{pub} and w'_{pub} are greater than θ_{max}, and there is no shortage under either wage. The second possibility is that average productivity of public school teachers declines, suggesting that $w'_{pub} < \theta_{max}$, while w_{pub} may be less than or greater than θ_{max}, and a shortage may occur. Figures 6 and 7 graph these two cases. Figure 6 shows the case where even after the salary reduction public school salaries are greater than private school salaries for all teachers, so all teachers continue to apply to public sector positions. Figure 7 shows the case where the salary reduction leads private salaries to be greater than public salaries for a subset of teachers. In this case, the most productive teachers no longer apply to the public sector.

Since our empirical strategy allows us to identify the average productivity of teachers hired under higher and lower wages, we can directly test which of the figures is more applicable to the Pakistani educational system. If the results are in line with the predictions of figure 6, it suggests that the government can reduce permanent teachers' salaries without reducing student learning and without fear of causing shortages. However, we caution that the regime change we evaluate is more complex the model presented here, since it may have also increased the returns to effort through career concerns.

6 Teacher Productivity and Teacher Wages

To understand whether our assumption that wages are uncorrelated with θ_j is reasonable, it is useful to estimate the association betweeen wages and TVA. In table 4, we first regress log salaries on teacher characteristics (column 1) and then include mean TVA in the regression (columns 2-4). The specification is:

$$log(salary_j) = \beta_0 + \Gamma X_j + \alpha_d + \varepsilon_j,$$

where $log(salary_j)$ is the log of the mean salary of teacher j, and X_j consists of the same teacher characteristics as in equation 1. As before,α_d is a district fixed effect, and some specifications (columns 3 and 4) also include school fixed effects.

Receiving some training is associated with a 52% increase in teacher salaries and having a bachelor's degree is associated with a 26% increase.[24] In addition, seniority is heavily rewarded in the public sector, with every additional year of age resulting in a 5.8% (no school fixed-effects) to 6.3% increase in wages (with school fixed effects). Recall that the first two years of teaching experience have a large effect on TVA. While we cannot include both experience and age non-parametrically in this Micerian regression, we can include an indicator variable for whether the teacher had more than 3 years of experience. Again, there is no additional effect of experience in the first three years on teacher wages beyond the seniority effect. Similarly, teacher content knowledge (column 4) has small and insignificant effects on teacher salaries. Unsurprisingly, teachers with temporary contracts make 35% less than teachers with permanent contracts. Strikingly, every attribute that the public sector appears to reward has no significant effect on TVA. When we add mean TVA to the regressions (olumns 2-4), the coefficient is small, insignificant, and negative. Moreover, adding mean TVA has no effect on the adjusted R^2, suggesting that mean TVA does not explain any of the variation in salaries. We infer that higher quality teachers do not appear to be rewarded with higher salaries in the public sector, consistent with our theoretical

[24]Almost all public school teachers have at least some training. Therefore, the large association between training and salaries relies on 44 individuals (3% of the sample) who have no training.

framework.

Perhaps TVA cannot be rewarded because it is difficult to observe or verify. Using our data on private schools, we replicate the specification in column 3 for private school teachers in column 5. The differences in compensation schemes are striking. As has been noted before (Andrabi et al., 2008), the private sector pays teachers according to their outside option, penalizing women and teachers who are locally resident. The private sector also rewards training and education (in similar ways for education, but less so for training). However, the premium on seniority is much lower (1.6%) and critically, TVA is highly correlated with salaries. A one standard deviation increase in TVA is associated with a 11% increase in wages, and this coefficient is statistically significant at the 5% level.

Our results suggest that teacher compensation in the public sector *does not* reward more productive teachers. This is not because productivity is impossible to observe. In the private sector, more productive teachers earn substantially more than less productive teachers. Moreover, these results are consistent with one of the key assumptions of our simple theoretical framework: public sector wages are not increasing in teacher productivity, θ_j. Our theoretical framework points to the next natural question: would a decline in public sector wages lower the average quality of public school teachers? To answer this question, we now estimate the effects of the contract teacher policy, which lowered wages by 35%, on the characteristics of individuals entering the teaching profession and on TVAs.

6.1 Methodology

While our TVA measures do not appear to be biased, we cannot simply regress TVA or other teacher characteristics estimates on a teacher's contract status to estimate the effect of a contract teacher policy since contract status is not randomly assigned. The 2% of teachers hired and retained on temporary contracts prior to 1998, as well as the 17% hired on permanent contracts after 1998, are likely to be highly selected. The hiring regime change in 1998 following the nuclear tests allows us to instrument for contract status using the budgetary shock. Moreover, because the shock changed contract status for much of the labor pool, our natural experiment allows us to understand the effect of a large-scale contract teacher policy on teacher labor supply.

To estimate the effect of the contract regime on what types of individuals become teachers, on which schools and students those individuals are assigned, and on teacher productivity, we first estimate the ordinary least squares regression:

$$y_j = \beta_0 + \beta_1 TempContract_j + \beta_2 month\_hired_j + \beta_3 month\_hired_j \times Post_j + \alpha_d + \epsilon_j, \quad (2)$$

where y_j are the characteristics of teacher j, including her TVA, her students, and the school to which she is assigned. $TempContract_j$ is an indicator variable equal to 1 if a teacher has a temporary contract and 0 otherwise, $month\_hired_j$ is the month a teacher was hired, $Post_j$ is an indicator variable equal to 1 if a teacher was hired in or after 1998, and α_d is a district fixed effect. We include time trends in teacher quality to account for the fact that most of the variation in contract status is driven by whether teachers were hired before or after the budgetary shock. Even so, the estimates of β_1 from this ordinary least squares regression are likely to be biased for several reasons. First, as we discussed before, contract status is likely correlated with other teacher characteristics. Second, we typically observe teachers hired on a temporary contract with fewer years of experience in our data since these teachers are hired later. Since the effects of experience on student learning are highly non-linear, linear time trend controls that span the entire sample are unlikely to fully account for these experience effects.

To account for these potential sources of bias, we use a fuzzy regression discontinuity design comparing teachers hired right before and after the budgetary shock. This approach is analogous to an instrumental variables regression that incorporates time trends and includes a subset of the sample around the budgetary shock. Therefore, to estimate β_1 without selection on contract status, we instrument for $TempContract_j$ with the indicator variable $Post_j$. The first stage of this two stage least squares strategy is then:

$$TempContract_j = \delta_0 + \delta_1 Post_j + \delta_2 month\_hired_j + \delta_3 month\_hired_j \times Post_j + \alpha_d + \mu_j. \quad (3)$$

Following Lee and Lemieux (2010), who discuss regression discontinuities with discrete data, such as time, we cluster our standard errors at the month hired level.

We first report these regressions with teacher (female, local, bachelor's degree, and receipt of training), student (mother and father education and household wealth), and school characteristics (quality of facilities, number of teachers, and student-teacher ratios) as our outcome variables. This allows us to test whether the contract policy affected the observable characteristics of individuals entering the teaching profession, the allocation of students to teachers, and the allocation of teachers to schools. We then estimate the effect of temporary contracts on TVA to examine the effect of this policy change on teacher productivity.

6.2 Results

6.2.1 *Existence of a First Stage*

Figure 8 shows the discontinuous effect of being hired after 1998 on contract status. Being hired after 1998 is associated with an 80 percentage point increase in the probability that a teacher is hired on a temporary contract. Figure 9 shows the similar discontinuity in salaries, with regression equivalents in table 5. Each coefficient in the table is the result from separate regressions of the form specified either in equation 2 (OLS) or equation 3 (fuzzy RD with a four-year bandwidth such that the sample includes 1994-1997 and 2002-2005). In the OLS regression (row 1, column 1), temporary contract status is associated with a 1,760 Pakistan rupee decline in a teacher's salary (24% relative to average salaries for teachers hired in 1997), which is somewhat less dramatic than the 2SLS estimate of 2,207 or 30% (row 1, column 3). These effect sizes are also consistent with the effect of temporary contracts in the Mincerian regression (35%), which accounts for teacher observable characteristics.

6.2.2 *Effect of the Policy on Teacher Characteristics*

First, we test whether the change in the hiring regime resulted in a change in teachers' characteristics. Interestingly, and consistent with the idea that salaries may be higher than is necessary to incentivize high quality teachers to enter the teaching profession, we find no evidence of a change in the teacher pool. Figure 10 plots trends in teacher characteristics from 1970 to the present. There are broad trends over time toward greater feminization, higher education and a greater proportion of younger teachers, but despite yearly variation in teacher characteristics, there is little evidence of a large trend break following the policy change.

The remaining rows in table 5 formally compare the characteristics of contract and permanent teachers. OLS specifications containing the full sample appear to reflect the general but non-linear trend that teachers hired later are more educated; having a temporary contract increases the probability of having a bachelor's degree by 32 percentage points. However, once endogeneous selection into contract status is appropriately accounted for in the regression discontinuity specifications, and we no longer assume linear trends in month hired across the full sample of teachers, the effect of having a temporary contract on bachelor's degree is no longer significant. In fact, in the fuzzy regression discontinuity specification, there are no significant differences between the characteristics of teachers hired on permanent and temporary contracts, including in test scores for math, Urdu, and English. Notably, the change in regime did not lead to a decline in the fraction of teachers with a bachelor's degree; this suggests that the outside options for these teachers remain below the considerably lower contract teacher wages. Interestingly, the fuzzy regression discontinuity estimates suggest that the fraction of female teachers increased following the establishment of the policy (and continued to increase in figure 10), although the coefficient is insignificant and imprecisely estimated. This observation is consistent with the idea that there is a large surplus labor pool of educated women, as discussed by Andrabi et al. (2013).

6.2.3 *Effect of the Policy on Allocation of Teachers to Schools*

In contrast to the negligible effects of the contract policy on teacher characteristics, it does appear that the schools that contract teachers were assigned to were signfiicantly different. Figure 11 plots the teacher fixed effect estimates of the school extra and basic facilities indexes against the year a teacher started teaching, and figure 12 plots schools' student teacher ratios and the number of teachers in a school. The regression equivalents are presented in table 6. Both the figures and the regression results show that contract teachers were assigned to smaller schools with fewer teachers and extra facilities. Rows 3 to 9 present the coefficients for the components of the extra facilities index separately, and the effect of contract status on the extra facilities index is driven by schools with fewer libraries, who are less likely to have computers or electricity. In addition, there were 5.5 fewer teachers in the schools that contract teachers were assigned to (the average teacher works at a school with 9.7 teachers) and student-teacher ratios were smaller, although the latter is not signficant either in the OLS or the RD.

We can also test whether the parental backgrounds of children taught by contract teachers differed, and figures 13 and 14, as well as the final three rows of table 6, suggest that they did. Although statistical significance differs according to the specification, we find evidence that parental education (particularly father's education) was lower for children assigned to contract teachers. The index of assets appears to be somewhat higher, but this is statistically insignificant. These results are consistent in both across- and within-school regressions.

6.2.4 *Were Contract Teachers Assigned to Lower Ability Children?*

The fact that contract teachers were assigned to smaller schools with fewer extra facilities and worse parental education could suggest that they were teaching children whose learning was systematically lower. If this is indeed the case, our TVA estimates for contract teachers may be negatively biased. Fortunately, the panel structure of our data set allows us to directly test whether higher or lower ability students were selectively matched to contract teachers, which is ultimately the main plausible source of bias in the TVA estimates. We therefore test directly whether contract teachers were assigned to schools with higher ability students by testing whether student test score trends predict a school being assigned a contract teacher. In table 7, we first test whether time trends are the same for schools that never received contract teachers and schools that eventually received them. We estimate

$$y_{it} = \beta_0 + \beta_1 year_{it} + \beta_2 I(Received\,Contract\,Teacher)_s + \beta_3 I(Received\,Contract\,Teacher)_s \times year_{it} + \Gamma X_{it} + \epsilon_{it}$$

where y_{it} is the outcome variable, mean student test scores, $I(Received\,Contract\,Teacher)_s$ is an indicator variable equal to 1 if a school ever hired a contract teacher and 0 otherwise, $year_{it}$ is the survey year, and X_{it} is a vector of controls consisting of district fixed effects and lagged student test scores. The sample does not include any student-year observations from schools that have contract teachers in the survey year (or received one in a past year). As column 1 shows, test scores are not different on average between schools that did and did not receive contract teachers. Moreover, there is little difference in pre-trends in test scores between public schools that do and do not receive contract teachers. If anything, the pre-trends for schools that later received contract teachers are negative. Next, we assess whether test-scores gains predict receiving a contract teacher at the student level instead of the school level. Column 2 shows that, within schools, there is no significant difference between the test scores or test score trends of students who will and will not eventually receive contract teachers. Finally, column 3 tests whether yearly test score gains predict having a contract teacher. It shows that, across schools, a student's average test score gains before receiving a contract teacher are not predictive of whether he or she later receives a contract teacher. In summary, despite the fact that contract teachers were assigned to children with less educated fathers and schools that were smaller, there is no evidence to suggest that learning among children who received a contract teachers was different.

6.2.5 *Effect of Contract Status on Student Test Scores*

Table 8 now presents the results of the OLS and RD specifications of mean TVA on contract teacher status, and figures 15 and 16 are the graphical representations of the reduced form of the regression discontinuity specifications. Since TVA, which approximates productivity, is our main outcome of interest, we report the instrumented results for the full sample and for bandwidthes of 2, 3, and 4 years around the policy change. The effect of teacher contract status on mean TVA both across and within schools in the ordinary least squares regressions is small and the sign varies (-0.004 and 0.024). However, this estimate may be downwardly biased by selection of teachers hired before 1998 into contract teacher status and the relative inexperience of contract teachers during the years students were tested. To account for these effects, we estimate the fuzzy RD in rows 2 to 5 of table 8. The smaller bandwidthes better accomodate non-linear experience effects since they do not assume that the effects of month hired are linear throughout the entire sample. Consistent with this, estimated effect sizes for contract teachers are larger when the bandwidth is 2, 3, or 4 years. The effect sizes are similar within and across schools, with positive effect sizes around 0.2 standard deviations. While the mean TVA estimates across schools are typically imprecise and insignificant, the within-school mean TVA estimates are more precise and frequently significant at the 5% or 10% level.

The analytical standard errors in this procedure do not account for estimation error in the TVA estimates. To account for this, we also estimate the regression discontinuity coefficients with a clustered bootstrap procedure (see appendix table A8). The within-school mean TVA estimates are still often significant at the 5% or 10% level. Another potential issue is that contract teachers appear more effective because they have smaller class sizes (even though student-teacher ratios are not significantly different in the RD). In appendix table A9, we repeat the fuzzy regression discontinuity analysis after re-estimating the TVAs controlling for average school-year student-teacher ratios. The results are qualitatively and quantitatively similar. Overall, we find no evidence of a decline in TVA following the regime change and some evidence that the TVA of contract teachers was higher – potentially due to higher effort – than the permanent teachers hired prior to the regime change.

6.2.6 *Quality of the Teaching Pool Over Time*

While applicants hired right after the budgetary shock appear to be similar to applicants hired previously in terms of characteristics, over time, the quality of applicants may still have declined. Understanding whether this is the case is critical for understanding the equilibrium effects of the policy change. In figure 10, there does seem to be a reduction in teacher training and an increase in workforce feminization after 2002 (although both may continue pre-existing trends and neither characteristic is strongly associated with TVA).

To test whether the quality of new teachers is decreasing over time, we would like to compare the test scores of the students of contract teachers hired earlier to those hired later. This poses several problems. First, on average, we observe more recently hired teachers with fewer years of experience. Thus, we will only compare the outcomes of the students of inexperienced contract teachers (teachers with 0 or 1 year of experience) to mitigate the effects of different levels of teacher experience. A second challenge is that later hires are only observed with students in later testing rounds. If student test scores are improving over time for unrelated reasons and if we do not control for year of testing, the effect of being a later hire will be upwardly biased. Therefore, we also include a control group of permanent teachers hired before 1998 in our regression sample, so that we can include testing round fixed effects. We estimate the regression:

$$y_{it} = \beta_0 + \beta_1 month\_hired_j + \beta_2 Post_j + \beta_3 Post_j \times month\_hired_j + \sum_g \beta_g y_{i,t-1} I(grade = g) + \alpha_t + \epsilon_{it},$$

where y_{it} is the test score of a student i in year t, $month\_hired_j$ is the month a teacher j is hired, $Post_j$ is an indicator variable equal to 1 if a teacher is hired after 1998 and 0 otherwise, $y_{i,t-1}$ is a student's lagged test score, g is her grade, and α_t is a round fixed effect. β_3 then captures the effect of the month a teacher was hired on student outcomes for teachers hired after 2002. The coefficient β_2 does not have a clear interpretation. Because the sample is limited to inexperienced contract teachers, β_2 here is not analogous to the reduced form contract teacher effect in the fuzzy regression discontinuity. Instead, it captures a combination of the contract teacher effect and the inexperience effect.

Table 9 reports the results of this regression for mean student test scores. The estimates of β_3 are small and statistically insignificant. Accordingly, there is little reason to believe that over time teacher quality decreased in response to decreased teacher salaries and teacher tenure. This result may seem surprising, since the contract teacher policy had a large negative impact on teacher salaries. However, even after the contract teacher policy, there was a large (and growing) gap between public sector and private sector teachers' salaries (figure 2). This suggests that high salaries for teachers in the public sector were a rent that did not translate into higher teacher quality. While lowering salaries decreased the rents from teaching in the public sector, it was still more attractive than private sector opportunities.

6.3 Natural Experiment Robustness

In this section, we test whether our results could be driven by student selection to teachers, the selective attrition of the students of contract teachers, or the selective attrition of poorly performing contract teachers themselves.

Student Selection. The robustness tests in the TVA section indicate that our TVA estimates are not biased by selection of students to teachers. Therefore, it is unlikely that differences in student quality between contract and non-contract teachers are driving our results. Nonetheless, we can more formally test whether either observed or unobserved student quality drives the differences in student outcomes between contract and non-contract teachers. In our first additional robustness test, we control for mean student household assets and lagged mean test scores at the classroom-level when we estimate the TVAs. Altonji and Mansfield (2014) argue that the classroom level means of observable student characteristics, which are used to assign students into classrooms, can proxy for unobservable characteristics related to student outcomes as well. In our new TVA estimates, we control for the classroom-level means of two of the most likely determinants of student-to-classroom and student-to-school sorting – ability and wealth. Table 10 presents the results of the fuzzy regression discontinuity regressions, using the new TVA estimates as the outcomes. The point estimates are qualitatively and quantitatively similar to those in table 8.

Selective Attrition: Assessing Student and Teacher Attrition. Our estimates of the contract teacher effect may also be biased if the students of contract teachers are differentially more likely to exit the sample or lower quality contract teachers are more likely to leave schools. Table 11 tests for both these types of attrition. In column 1, we show that conditional on the year a student first appeared in the panel, the percent of times a student had a contract teacher has no significant effect on whether she appears in the fourth round of the panel. Column 2 shows that the percent of rounds a student was observed with a contract teacher does not predict the total number of years she is observed. In the remaining columns, we test for teacher attrition. Column 3 shows that the mean TVA of a contract teacher does not significantly predict whether she was present in the fourth round of the panel, and column 4 shows that it does not predict the number of years a teacher was observed (conditional on the year she started teaching). Overall, we find no evidence of either differential attrition of the students of contract teachers or differential attrition of contract teachers by quality.

Permanent vs. Temporary Income. While the majority of contract teachers did not expect to be normalized in 2009 (Cyan, 2009), contract teachers did win a court case in 2012 which led many teachers to be tenured and receive salaries comensurate with permanent teachers. Therefore, teachers may have entered contract teacher teaching with the expectation that the salary reductions were temporary. If this is the case, to interpret our results, we must determine how much intial salary reductions reduced *permanent* incomes for contract teachers. This exercise requires several additional assumptions. We assume teachers had rational expectations and that the discount rate is a conservative 3%. Furthermore, we assume that a teacher expects to work for 40 years. For teachers hired in 2002, temporary contracts reduced their salaries by 35% for 10 of those 40 years. Even with this very low discount rate, the contract policy reduces permanent wages for teachers hired in 1998 by 15%, suggesting that there is still substantial room to lower wages without negatively affecting teacher quality.

External Validity. In Pakistan, the contract teacher policy was instituted in response to an economic crisis, which may have also negatively affected teachers' outside options. Thus, we should be cautious in applying these results to other contexts where teachers' outside options are unchanged. However, there is reason to believe that lowering teacher salaries, even in the absence of an economic crisis, would not result in a decline in productivity. We observe teachers who are hired as late as 2007, 9 years after the nuclear tests. The results in table 10 indicate that these teachers are no worse than those hired in 2002. The recession in 1998 did not last 10 years. In fact, according to the World Bank, Pakistan experienced a period of relatively high per capita GDP growth from 2003-2007 (2.7-5.5% per year). Similarly, in 1997, the unemployment rate according to the World Bank was 5.8% and by 2007, it was 5.1%. Moreover, in our own data, we do not find that salaries fell for private school teachers hired after 1998. Taken together, these facts suggest that even if teachers' outside options fell after the nuclear tests, they likely recovered by 2007.

7 Conclusion

This paper makes two important contributions to our understanding of the educational production function in low income countries. First, we provide among the first correlations between teacher observable characteristics and teacher quality from a low income country. Like in medium and high-income countries, we still find that, apart from the first two years of experience, other observable characteristics are poor predictors of teacher quality. This result is surprising since the teacher labor market in Pakistan (and many low income countries) differs substantially from that in the United States. For example, while virtually all teachers in high income countries have at least a bachelor's degree, in our public school sample, only 51% of teachers have bachelor's degrees or greater. We also find that teacher quality is an important determinant of students' learning. Having a teacher with a standard deviation greater within-school TVA leads to a 0.16 standard deviation higher test score. This effect size is on the higher side of what is found in the United States (Rockoff, 2004; Chetty et al., 2014a).

Second, this paper builds on work by Duflo et al. (2014) and Muralidharan and Sundararaman (2013) on the quality of contract teachers. Like these papers, which provide clean experimental estimates of the contract teacher effect, we find that contract teachers have as great as and perhaps moderately higher TVAs than permanent teachers. However, our setting differs from those of Duflo et al. (2014) and Muralidharan and Sundararaman (2013) in important ways. We study the impact of a large-scale policy change which caused virtually all new teachers to be hired as contract teachers on much lower salaries. This allows us to assess the equilibrium effects of a large-scale contract teacher policy on the supply of teachers. One major concern is that hiring all teachers as lower-salary contract teachers would lead the quality of new teachers to decline. Instead, we find little evidence that teacher quality declined either directly after the new policy was put into place or in the subsequent years. In fact, given the large experience effects in the first two years of teaching, it is likely that the previous papers actually underestimated the positive effects of contract teachers on TVA.

Our results suggest that, at least in low income countries, policies that increase wage levels to attract higher-skilled teachers, like those advocated by Auguste et al. (2010), would be costly and ineffective. Since higher levels of education are not correlated with TVA, the best teachers are not those with the greatest education, and the public sector wastes money by paying these teachers their outside options in low-income countries. More remarkably, wages are already so high that even a 35% decline has no impact on the education levels of new recruits. In other work, we have argued that this may reflect the existence of a large pool of educated women with few job options (Andrabi et al., 2013). This paper suggests that public sector compensation could be signficantly redesigned to better account for the realities of low-income countries. Combining lower salaries (or salaries that are more strongly tied to teacher productivity) with greater investment in other school characteristics or student incentives could allow low-income countries to greatly improve their educational performance.

References

Allcott, H. and S. Mullainathan (2012). External validity and partner selection bias. *NBER Working Paper*.

Altonji, J. G. and R. K. Mansfield (2014). Group-average observables as controls for sorting on unobservables when estimating group treatment effects: the case of school and neighborhood effects. *NBER Working Paper*.

Andrabi, T., N. Bau, J. Das, and A. I. Khwaja (2010). Are bad public schools public "bads?" Test scores and civic values in public and private schools. *Working Paper*.

Andrabi, T., J. Das, and A. I. Khwaja (2008). A dime a day: The possibilities and limits of private schooling in Pakistan. *Comparative Education Review 52*(3), 329–355.

Andrabi, T., J. Das, and A. I. Khwaja (2013). Students today, teachers tomorrow: Identifying constraints on the provision of education. *Journal of public Economics 100*, 1–14.

Andrabi, T., J. Das, A. I. Khwaja, and T. Zajonc (2006). Religious school enrollment in Pakistan: A look at the data. *Comparative Education Review 50*(3), 446–477.

Antecol, H., O. Eren, and S. Ozbeklik (2015). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics 33*(1), 63–89.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady (2014). A helping hand? Teacher quality and learning outcomes in kindergarten. *Working Paper*.

Aslam, M. (2013). Focusing on teacher quality in Pakistan: Urgency for reform. *Right to Education*.

Aucejo, E. (2011). Assessing the role of teacher-student interactions. *Working Paper*.

Auguste, B. G., P. Kihn, and M. Miller (2010). Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching: An international and market research-based perspective.

Bau, N. (2015). School competition and product differentiation. *Working Paper*.

Biggs, A. G. and J. Richwine (2011). Assessing the compensation of public-school teachers. *The Heritage Foundation*.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur (2013). Scaling up what works: Experimental evidence on external validity in kenyan education. *Center for Global Development Working Paper* (321).

Bruns, B. and R. Rakotomalala (2003). *Achieving universal primary education by 2015: A chance for every child*, Volume 1. World Bank Publications.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers (2006). Missing in action: teacher and health worker absence in developing countries. *Journal of Economic Perspectives 20*(1), 91–116.

Chetty, R., J. Friedman, and J. Rockoff (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *Quarterly Journal of Economics 126*(4).

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–2679.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2015). Response to Rothstein (2014) 'revisiting the impacts of teachers'. *CEPR Discussion Paper 10768*.

Cyan, M. (2009). Contract employment policy review. *Punjab Government Efficiency Improvement Program*.

Das, J. and T. Zajonc (2010). India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics 92*(2), 175–187.

De Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2014). Double for nothing? The effects of unconditional teacher salary increases on student performance. *Working Paper*.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources 42*(3), 528–554.

Disney, R. and A. Gosling (1998). Does it pay to work in the public sector? *Fiscal Studies 19*(4), 347–374.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review 101*(5), 1739–74.

Duflo, E., P. Dupas, and M. Kremer (2014). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics 123*, 92–110.

Dustmann, C. and A. Van Soest (1998). Public and private sector wages of male workers in Germany. *European Economic Review 42*(8), 1417–1441.

Filmer, D. and L. Pritchett (2001). Estimating wealth effects without expenditure data or tears: An application to educational enrollments in states of india. *Demography 38*(1), 115–132.

Hameed, Y., R. Dilshad, M. Malik, and H. Batool (2014). Comparison of academic performance of regular and contract teachers at elementary schools. *Asian Journal of Management Sciences & Education 3*(1), 89–95.

Hanushek, E. A. and S. G. Rivkin (2006). Teacher quality. *Handbook of the Economics of Education 2*, 1051–1078.

Harris, D. N. and T. R. Sass (2006). Value-added models and the measurement of teacher quality. *Working Paper*.

Hoffmann, F. and P. Oreopoulos (2009). A professor like me: the influence of instructor gender on college achievement. *Journal of Human Resources 44*(2), 479–494.

Ishtiaq, N. (2013). Understanding Punjab education budget 2012-2013: A brief for standing committee on education, provincial assembly of the Punjab.

Jimenez, E., M. E. Lockheed, and V. Paqueo (1991). The relative efficiency of private and public schools in developing countries. *The World Bank Research Observer 6*(2), 205–218.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper*.

Khwaja, A. I. and A. Mian (2008). Tracing the impact of bank liquidity shocks: Evidence from an emerging market. *The American Economic Review 98*(4), 1413–1442.

Kremer, M. and K. Muralidharan (2008). Public and private schools in rural India. *School Choice International*.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*, 281–355.

Lucifora, C. and D. Meurs (2006). The public sector pay gap in France, Great Britain and Italy. *Review of Income and Wealth 52*(1), 43–59.

Miller, R. (2012). Teacher absence as a leading indicator of student achievement: New national data offer opportunity to examine cost of teacher absence relative to learning loss. *Center for American Progress*.

Muralidharan, K. and K. Sheth (2013). Bridging education gender gaps in developing countries: The role of female teachers. *NBER Working Paper*.

Muralidharan, K. and V. Sundararaman (2013). Contract teachers: Experimental evidence from India. *NBER Working Paper*.

Muralidharan, K. and V. Sundararaman (forthcoming). The aggregate effect of school choice: Evidence from a two-stage experiment in India. *Quarterly Journal of Economics*.

Pritchett, L. and D. Filmer (1999). What education production functions really show: a positive theory of education expenditures. *Economics of Education review 18*(2), 223–239.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, schools, and academic achievement. *Econometrica*, 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review P & P*, 247–252.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics 125*(1).

Siniscalco, M. T. (2004). Teachers' salaries. *Education for All Global Monitoring Report*.

UNESCO Islamabad (2013). Education budgets: A study of selected districts of Pakistan.

Weissman, J. (2011). Are teachers paid too much: How 4 studies answered 1 big question. *The Atlantic*.

8 Tables

Table 1: Relationship Between Teacher Characteristics and Mean Teacher Value-Added for Public School Teachers

	(1) Mean TVA	(2) Mean TVA	(3) Mean TVA	(4) Mean TVA	(5) Mean TVA
Female	0.070***	-0.036	0.080***	0.207	
	(0.026)	(0.134)	(0.026)	(0.225)	
Local	0.025	0.008	0.024	-0.004	
	(0.025)	(0.031)	(0.028)	(0.049)	
Some Teacher Training	-0.023	-0.101	-0.093	-0.213*	
	(0.055)	(0.072)	(0.075)	(0.126)	
Has BA or Better	0.054**	0.043	0.012	0.010	
	(0.025)	(0.031)	(0.033)	(0.059)	
Had > 3 *Years of Exp in* 2007	0.060	0.076	0.037	0.163*	
	(0.038)	(0.052)	(0.047)	(0.097)	
Temporary Contract	-0.003	0.049	-0.020	0.051	
	(0.036)	(0.048)	(0.043)	(0.083)	
Mean English Test Score			0.032**	0.015	
			(0.015)	(0.022)	
Mean Urdu Test Score			0.034	0.013	
			(0.023)	(0.037)	
Mean Math Test Score			0.023	-0.013	
			(0.022)	(0.034)	
Have 0 *or* 1 *Years Exp.*					-0.305**
					(0.135)
Lagged Mean Score					0.717***
					(0.013)
Fixed Effects	District	School	District	School	Teacher
Number of Observations	1,383	1,383	919	919	27,089
Adjusted R Squared	0.224	0.450	0.228	0.415	0.721
Clusters	471	471	469	469	583
F	2.031	1.194	2.533	0.602	

This table reports estimates of the association between TVA and teacher characteristics. For columns 1-4, observations are at the teacher level and characteristics are time invariant. Column 5, identifies within teacher experience effects, controlling for teacher fixed effects, and regressing student outcomes on whether a teacher had 0 or 1 year of experience. Observations for this column are at the student-year level. Standard errors are clustered at the school level.

Table 2: Does Future Teacher Value Added Predict Current Teacher Value Added When Students Change Schools?

	(1) Coefficient (se)	(2) N
Forward Lag of English	0.051	3,231
	(0.049)	
Forward Lag of English (Within School)	-0.023	1,976
	(0.056)	
Forward Lag of Math	-0.076	3,231
	(0.061)	
Forward Lag of Math (Within School)	-0.017	1,976
	(0.036)	
Forward Lag of Urdu	-0.081	3,231
	(0.072)	
Forward Lag of Urdu (Within School)	0.017	1,976
	(0.040)	
Forward Lag of Mean Score	-0.033	3,231
	(0.067)	
Forward Lag of Mean Score (Within School)	0.002	1,976
	(0.046)	

This table tests for bias in the teacher value-added calculations. The current teacher value-added of students who change schools in the next period is regressed on the value-added of their future teacher. Observations are at the child level and standard errors are clustered at the teacher level.

Table 3: Out-of-Sample Validation of TVAs

	(1) Math Test Score	(2) English Test Score	(3) Urdu Test Score	(4) Mean Test Score
$Math\,TVA$	0.781***			
	(0.065)			
$English\,TVA$		0.857***		
		(0.068)		
$Urdu\,TVA$			0.845***	
			(0.077)	
$Mean\,TVA$				0.852***
				(0.078)
Lagged Score Control	Y	Y	Y	Y
Number of Observations	3,822	3,822	3,822	3,822
Adjusted R Squared	0.557	0.542	0.590	0.636
Clusters	1,090	1,090	1,090	1,090

This table tests if TVAs predict the test score gains of school changers who are allocated to the new teachers. If the TVA estimates perfectly predict the "true" teacher value-added, these coefficients should be 1. Standard errors are clustered at the teacher level.

Table 4: Relationship Between Mean TVA and Log Salary for Public and Private School Teachers

	(1) Log Salary Public	(2) Log Salary Public	(3) Log Salary Public	(4) Log Salary Public	(5) Log Salary Private
$Mean\, TVA$		-0.007	-0.028	-0.044	0.111**
		(0.014)	(0.025)	(0.036)	(0.046)
$Female$	-0.036***	-0.035***	0.154**	0.054	-0.413***
	(0.013)	(0.013)	(0.070)	(0.094)	(0.043)
$Local$	-0.052***	-0.051***	-0.049	-0.019	-0.178***
	(0.019)	(0.019)	(0.032)	(0.043)	(0.029)
$Some\, Teacher\, Training$	0.518***	0.518***	0.392***	0.837***	0.165***
	(0.141)	(0.141)	(0.140)	(0.316)	(0.045)
$Has\, BA\, or\, Better$	0.255***	0.255***	0.263***	0.211***	0.334***
	(0.019)	(0.019)	(0.028)	(0.042)	(0.045)
$Had\ > 3\, Years\, of\, Exp\, in$ 2007	0.063	0.064	0.120*	0.122	0.020
	(0.042)	(0.042)	(0.064)	(0.101)	(0.029)
$Temporary\, Contract$	-0.354***	-0.355***	-0.327***	-0.308***	
	(0.032)	(0.032)	(0.059)	(0.092)	
Age	0.058***	0.058***	0.063***	0.039	0.016**
	(0.015)	(0.015)	(0.020)	(0.029)	(0.007)
Age^2	-0.000***	-0.000***	-0.001**	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$Mean\, English\, Score$				0.016	
				(0.017)	
$Mean\, Urdu\, Score$				-0.006	
				(0.029)	
$Mean\, Math\, Score$				0.020	
				(0.025)	
Fixed Effects	District	District	School	School	District
Adjusted R Squared	0.616	0.615	0.662	0.707	0.459
Number of observations	1,383	1,383	1,383	919	807
F	108.304	96.471	35.025	12.496	38.522
Clusters	471	471	471	469	294

This table reports estimates from regressions of log mean teacher salaries in public (columns 1-4) and private (column 5) schools on teacher characteristics, including mean TVA (columns 2-5) and average teacher test scores in English, Urdu, and math (column 4). All regressions include either district or school fixed effects, and standard errors are clustered at the school level.

Table 5: Effect of the Discontinuity on Teacher Characteristics

	(1) OLS	(2) SE	(3) RD (4 Year)	(4) SE
Salary	-1,759.525***	305.833	-2,206.965*	1,261.971
Bachelor's	0.318***	0.032	0.109	0.140
Some Training	0.003	0.031	0.010	0.096
Local	-0.017	0.037	-0.066	0.134
Age Started	0.072***	0.024	0.943	1.116
Single	0.148***	0.032	0.053	0.136
Female	-0.005	0.044	0.273	0.190
Mean Teacher English Score	0.326***	0.080	0.319	0.248
Mean Teacher Urdu Score	0.076	0.067	0.217	0.336
Mean Teacher Math Score	-0.013	0.080	-0.502	0.375

This table presents OLS and fuzzy regression discontinuity results for the effect of temporary contracts on teacher characteristics. The IV includes teachers hired 4 years before 1998 and 4 years after 2002. Standard errors are clustered at the month hired level for the regression discontinuity results and the school level for the OLS results. Salaries, which were observed multiple times over several years, were normalized by calculating the teacher fixed effect, controlling for year fixed effects. Each cell is a coefficient estimate (or standard error estimate) for the temporary contract teacher effect.

Table 6: Effect of the Discontinuity on Student Characteristics and School Characteristics

	(1) OLS	(2) SE	(3) RD (4 Year)	(4) SE	(5) RD (Within School)	(6) SE
Basic School Facilities	0.076	0.065	0.337	0.240		
Extra School Facilities	-0.480***	0.115	-0.922**	0.434		
Library	-0.118***	0.031	-0.295**	0.125		
Computer	-0.040**	0.017	-0.170**	0.067		
Sports	-0.117***	0.026	-0.200	0.153		
Hall	-0.030	0.019	-0.087	0.076		
Wall	-0.029	0.029	0.134	0.088		
Fans	-0.085**	0.034	-0.131	0.101		
Electricity	-0.099***	0.035	-0.211**	0.087		
Number Teachers	-1.660***	0.496	-5.568**	2.386		
Student Teacher Ratio	-3.945	3.789	-5.471	13.078		
Student Household Assets	0.044	0.091	0.453	0.372	0.236	0.263
Student Mother Education	-0.077***	0.023	-0.001	0.116	-0.121	0.073
Student Father Education	-0.042	0.026	-0.134	0.081	-0.166**	0.071

This table presents OLS and fuzzy regression discontinuity results for the effect of temporary contracts on student and school characteristics. The IV includes teachers hired 4 years before 1998 and 4 years after 2002. Standard errors are clustered at the month hired level for the regression discontinuity results and the school level for the OLS results. Characteristics observed multiple times over several years were normalized by calculating the teacher or school fixed effect (depending on the level at which the characteristic is observed), controlling for year fixed effects. Each cell is a coefficient estimate (or standard error estimate) for the temporary contract teacher effect.

Table 7: Do Student Test Score Trends Predict Being Taught by a Contract Teacher?

	(1) Mean Test Scores	(2) Mean Test Scores	(3) Had a Contract Teacher
$Year$	0.134***	0.145***	
	(0.013)	(0.013)	
$I(Received\,Contract\,Teacher)$	0.048	0.069	
	(0.078)	(0.083)	
$Year \times I(Received\,Contract\,Teacher)$	-0.015	-0.011	
	(0.023)	(0.024)	
$Mean\,Test\,Score\,Gain$			-0.014
			(0.016)
District FE	Y	Y	Y
School FE	N	Y	N
Grade by Lagged Test Score Interactions	Y	Y	N
Number of Observations	25,296	25,296	15,956
Adjusted R Squared	0.637	0.677	0.037
Clusters	478	478	497

This table tests whether better students are allocated to contract teachers. The first column compares trends in student test scores before the receipt of a contract teacher in schools that did and did not receive contract teachers. The next column compares the test score gains of students within schools who did or did not receive contract teachers before the receipt of the contract teacher. The final regression regresses an indicator for whether a student ever had a contract teacher on their mean test score gains (residualized by testing round and grade) in the years prior to receiving a contract teacher. In this sample, each student is observed once. Standard errors are clustered at the school level.

Table 8: The Effect of Teacher Contract Status on TVA

	(1) Mean TVA	(2) SE	(3) N	(4) Within School Mean TVA	(5) SE	(6) N
OLS (Full Sample)	-0.004*	0.042	1,337.000	0.024*	0.026	1,278
RD (Full Sample)	-0.004	0.052	1,337.000	0.056	0.041	1,278
RD (2 Year)	0.840	0.550	227.000	0.360	0.322	201
RD (3 Year)	0.219	0.241	376.000	0.254**	0.123	336
RD (4 Year)	0.350	0.234	393.000	0.193*	0.097	350

This table regresses mean TVAs on whether a teacher has a temporary contract in all public schools using the ordinary least squares and 4 fuzzy regression discontinuity specifications. In the instrumental variables specifications, contract status is instrumented for with an indicator variable for whether a teacher was hired after 1998. All regressions contain linear time trends which are allowed to differ before and after the budgetary shock and district fixed effects. The table presents instrumental variables (RD) specifications for the full sample and for bandwidths of 2, 3, and 4 years before and after the budgetary shock. Observations are at the teacher level, and standard errors are clustered at the school level in the OLS specification and the month hired level in the regression discontinuity specifications.

Table 9: Trends in the Quality of the Teacher Pool Over Time

	(1) Mean Test Scores
Month Hired	0.002**
	(0.001)
$Month\ Hired \times I(Year\ Hired > 2001)$	-0.007
	(0.024)
$I(Year\ Hired > 2001)$	Y
Round FE	Y
District FE	Y
Grade by Lagged Test Score Interactions	Y
Number of Observations	21,788
Adjusted R Squared	0.660
Clusters	450

This table documents trends in the outcomes of the students of contract teachers by month hired. This sample includes the students of inexperienced contract teachers (0 or 1 years of experience) hired between 2002 and 2007 and permanent teachers hired before 1998 (used as a control group to control for year of testing fixed effects). Standard errors are clustered at the school level.

Table 10: Regression Discontinuity Results Controlling for Classroom Mean Characteristics

(1) Bandwidth (Years)	(2) Mean TVA	(3) Standard Error	(4) Within School, Mean TVA	(5) Standard Error
Full Sample	0.003	0.061	0.076**	0.038
2	1.215*	0.598	0.786*	0.423
3	0.272	0.261	0.263*	0.138
4	0.348	0.237	0.188*	0.108

This table replicates the fuzzy regression discontinuity estimates of the effect of temporary contract status on TVA. TVA estimates include controls for the mean household asset index of a classroom and mean lagged student test scores in the classroom.

Table 11: Tests for Selective Attrition of Students and Teachers

	(1)	(2)	(3)	(4)
	Student Attrition		Teacher Attrition	
	Present In Year 4	Years Observed	Present In Year 4	Years Observed
Percent of Rounds Observed with a Contract Teacher	-0.015	-0.038		
	(0.019)	(0.028)		
Mean TVA			0.026	0.015
			(0.070)	(0.142)
Year Student Entered Panel FE	Y	Y	N	N
District FE	Y	Y	Y	Y
Year Teacher Started FE	N	N	Y	Y
Number of Observations	22,596	22,596	298	298
Adjusted R Squared	0.157	0.503	0.057	0.203
Clusters	512	512	200	200

This table examines whether the students of contract teachers selectively leave the sample and whether lower quality contract teachers selectively leave the sample. The outcome variables are an indicator variable for whether a student was in the sample in round 4, the number of rounds a student was observed, an indicator whether a teacher was in the sample in round 4, and the number of rounds a teacher was observed. In the first two columns, the sample is all tested public school students. In the second two, it is all contract teachers. Standard errors are clustered at the school level.

9 Figures

Figure 1: Prevalence of Foreign Currency Deposit Accounts in Pakistan (Khwaja and Mian, 2008)

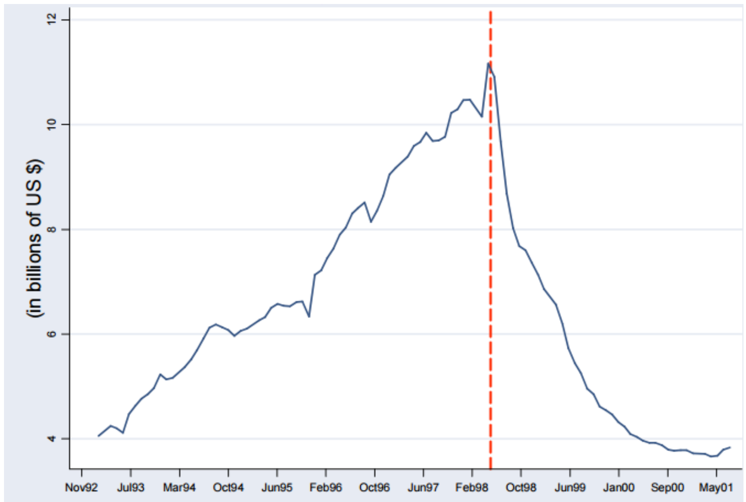


Figure 2: Teacher Salaries in Public and Private Schools

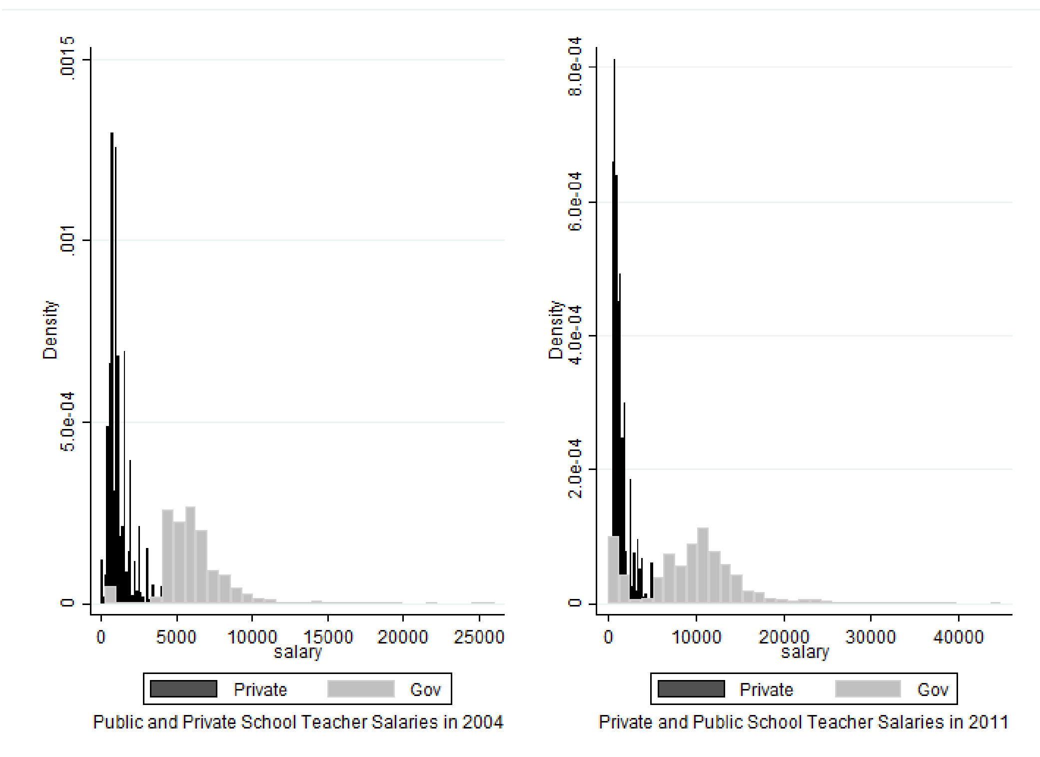


Figure 3: Number of Teachers Hired Per Year in the Sample

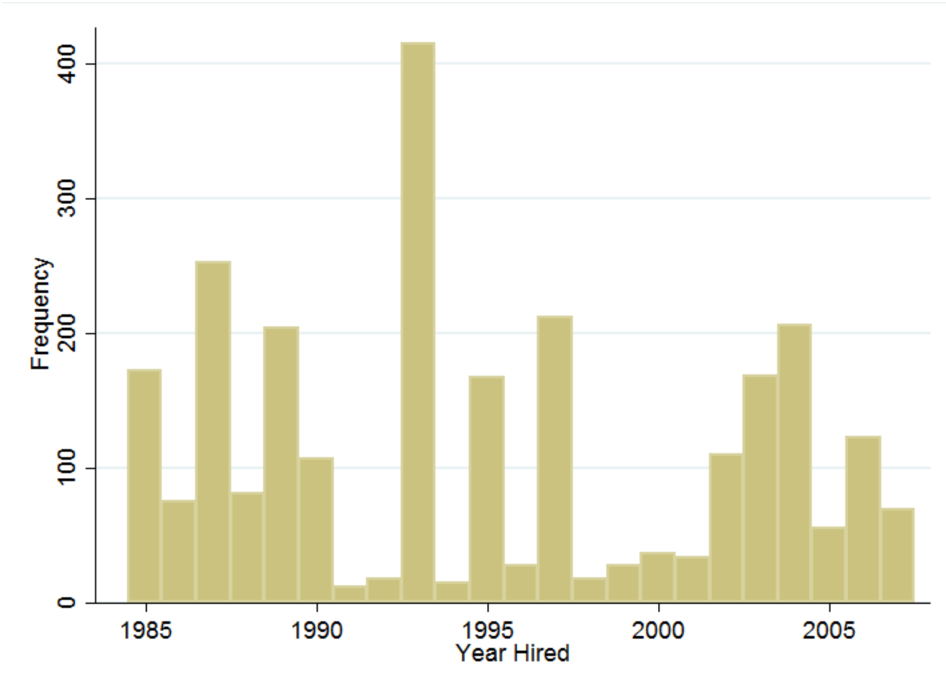


Figure 4: Number of Rounds Public School Teachers are Observed

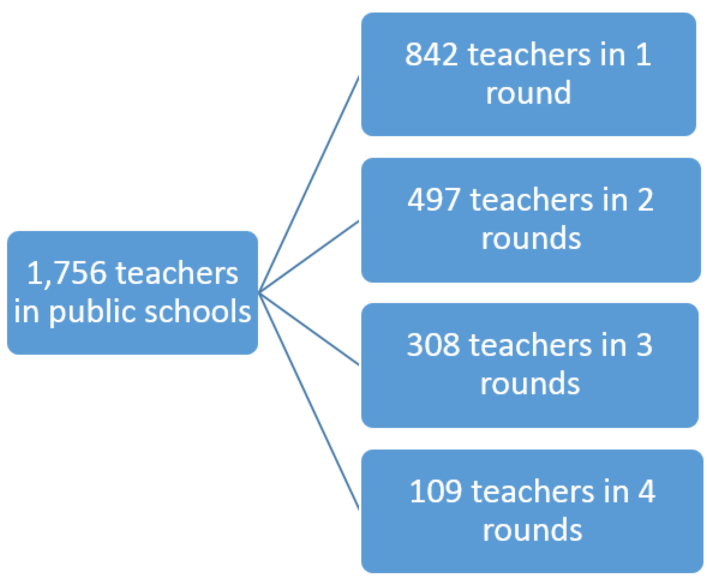


Figure 5: Number of Rounds Public School Students are Observed

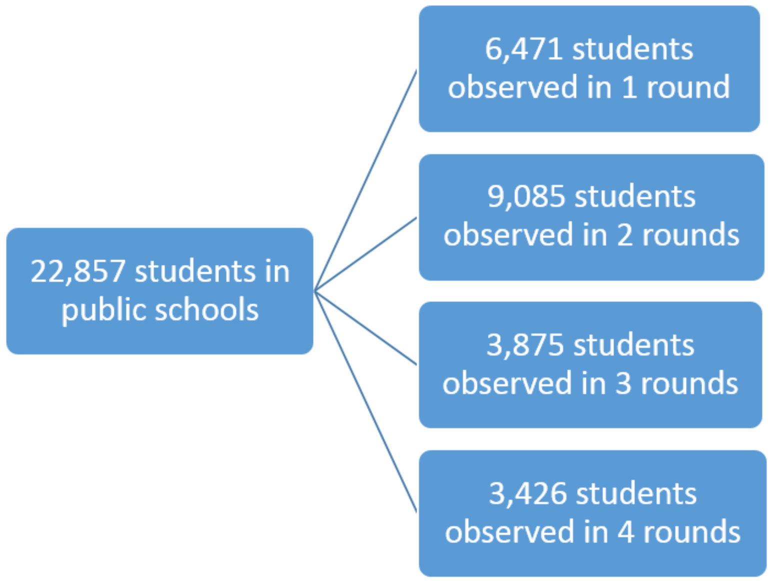


Figure 6: Case I: $w_{pub}, w'_{pub} \geq \theta_{max}$

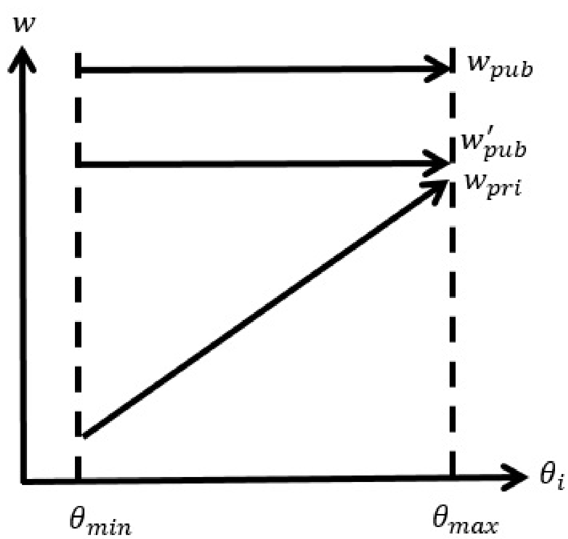


Figure 7: Case II: $w'_{pub} < \theta_{max}$

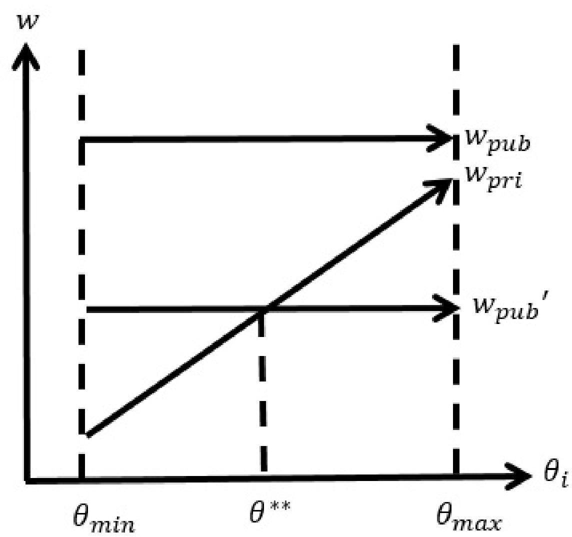


Figure 8: Discontinuity in Contract Status by Month Hired

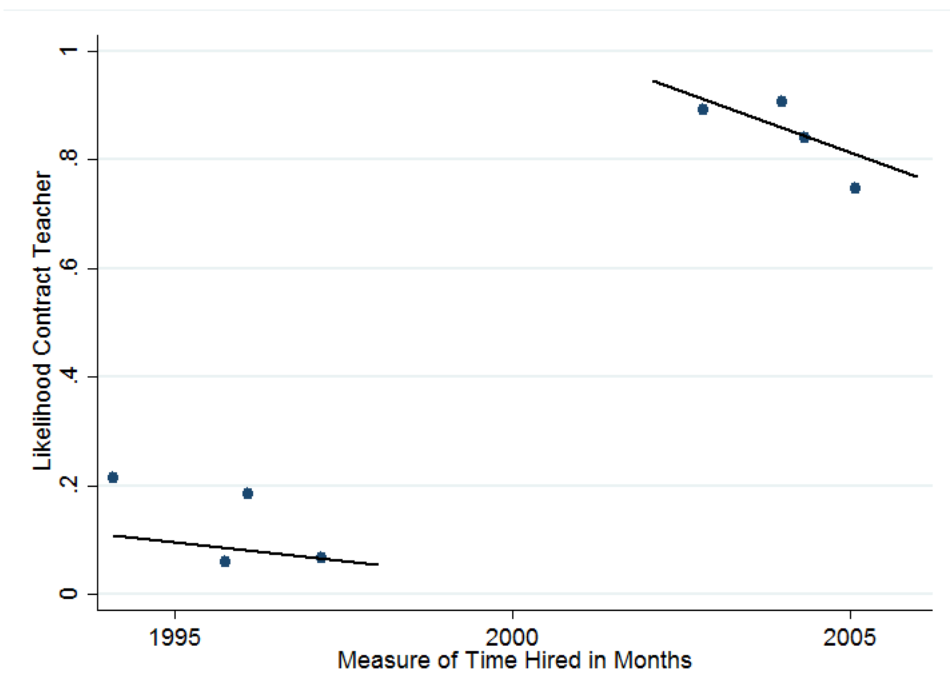


Figure 9: Discontinuity in Teacher Salary by Month Hired

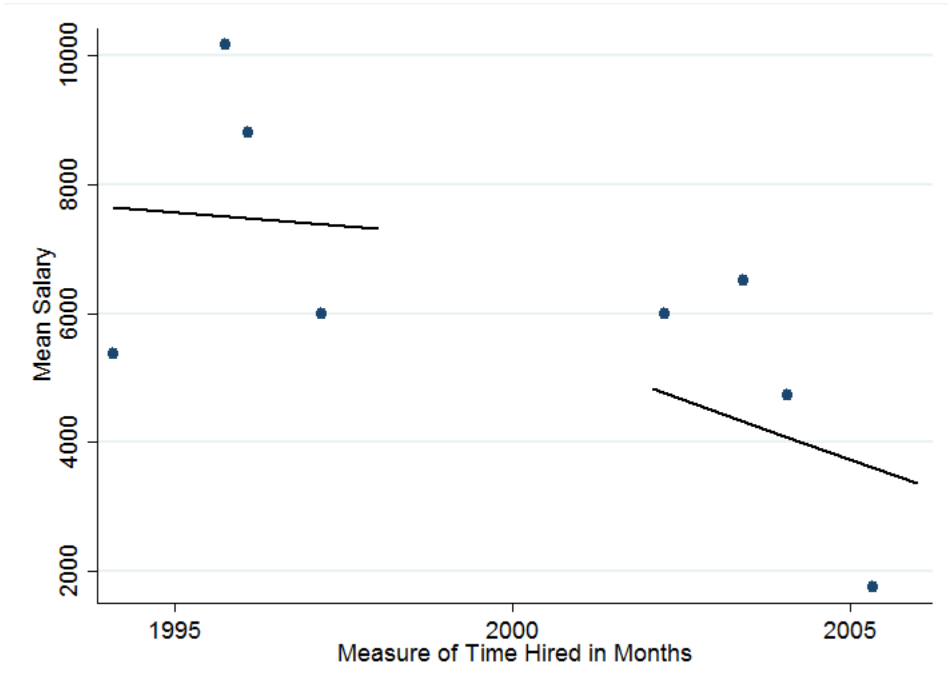


Figure 10: Trends in Teacher Characteristics by Month Hired

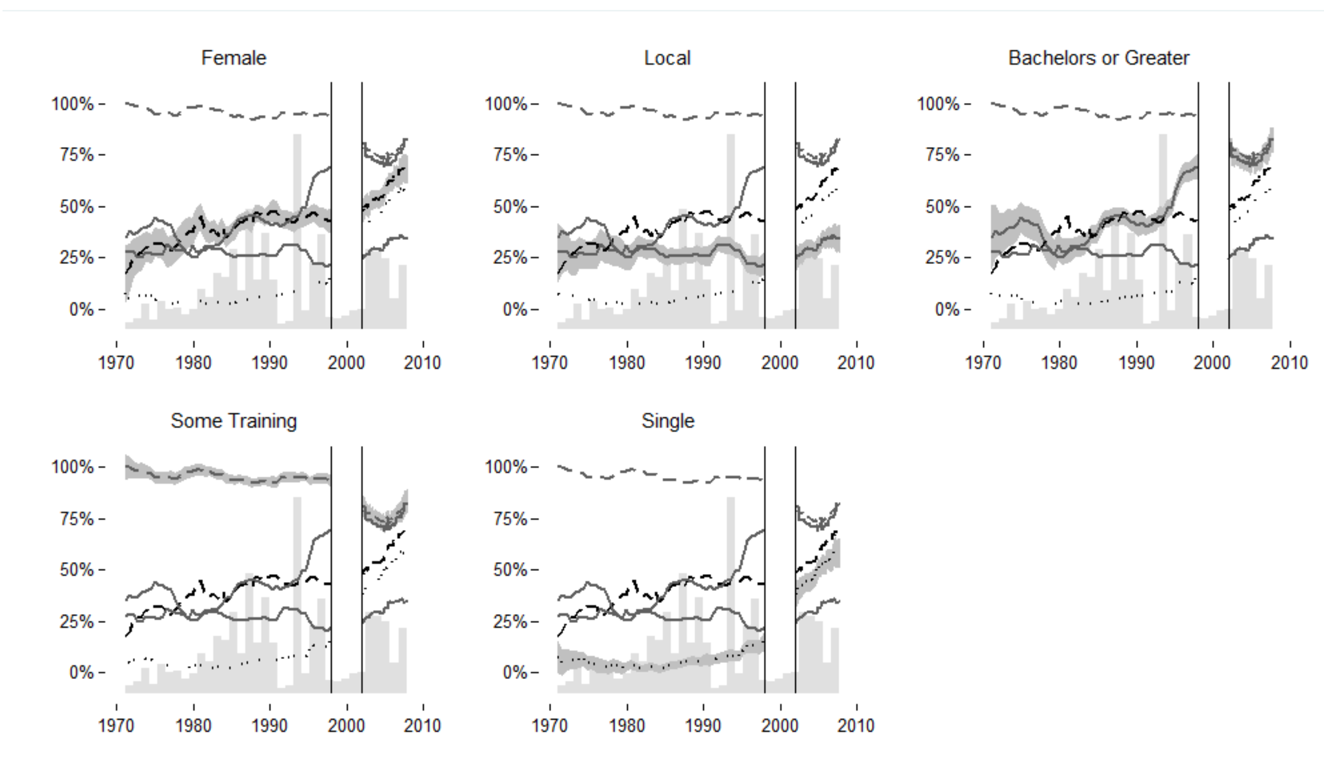


Figure 11: Trends in School Facilities by the Month a Teacher was Hired

Figure 12: Trends in School Characteristics by the Month a Teacher was Hired

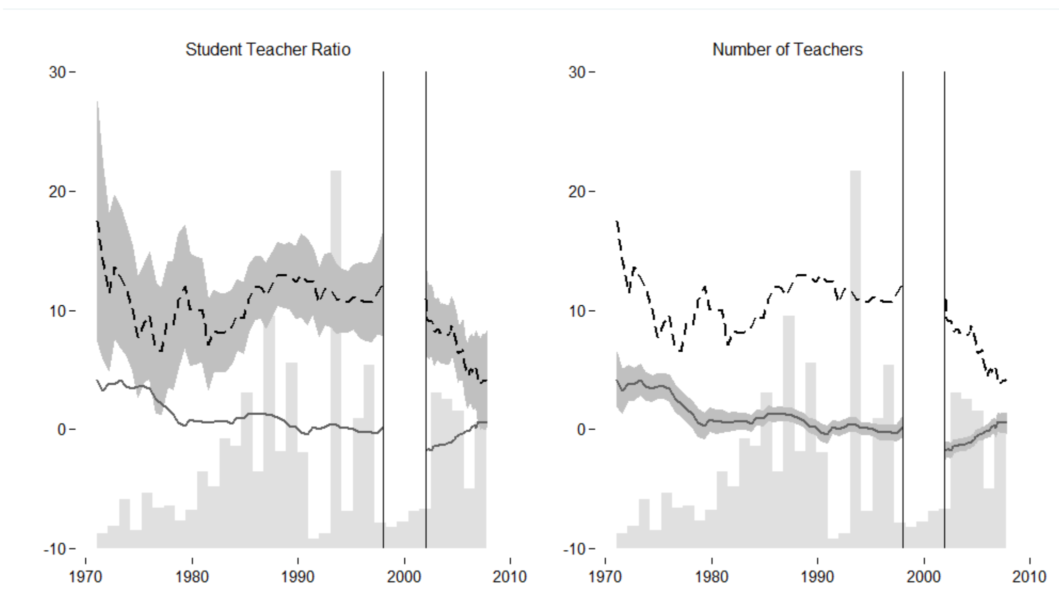


Figure 13: Trends in Student Parent Education by the Month Their Teacher was Hired

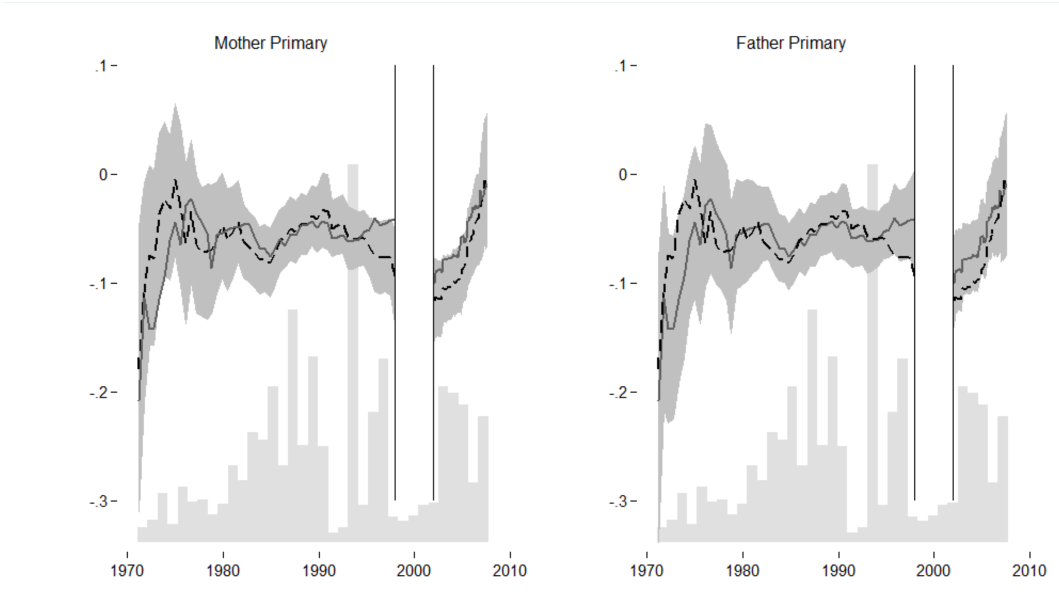


Figure 14: Trends in Student Household Assets by the Month Their Teacher was Hired

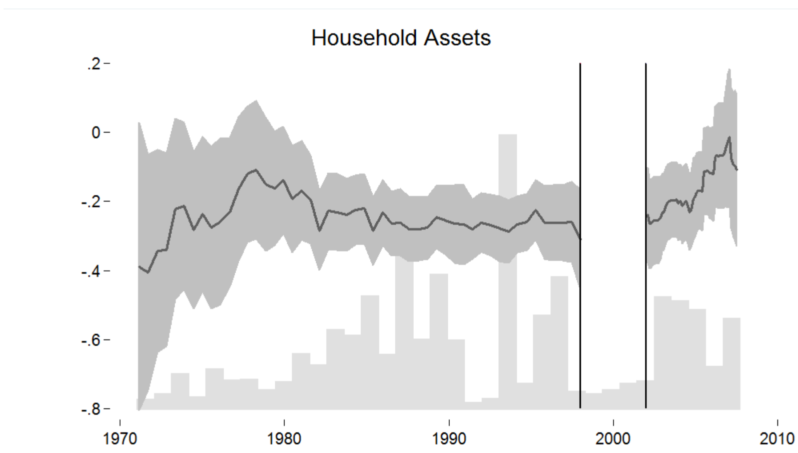


Figure 15: Discontinuity in Mean TVA by Month Hired

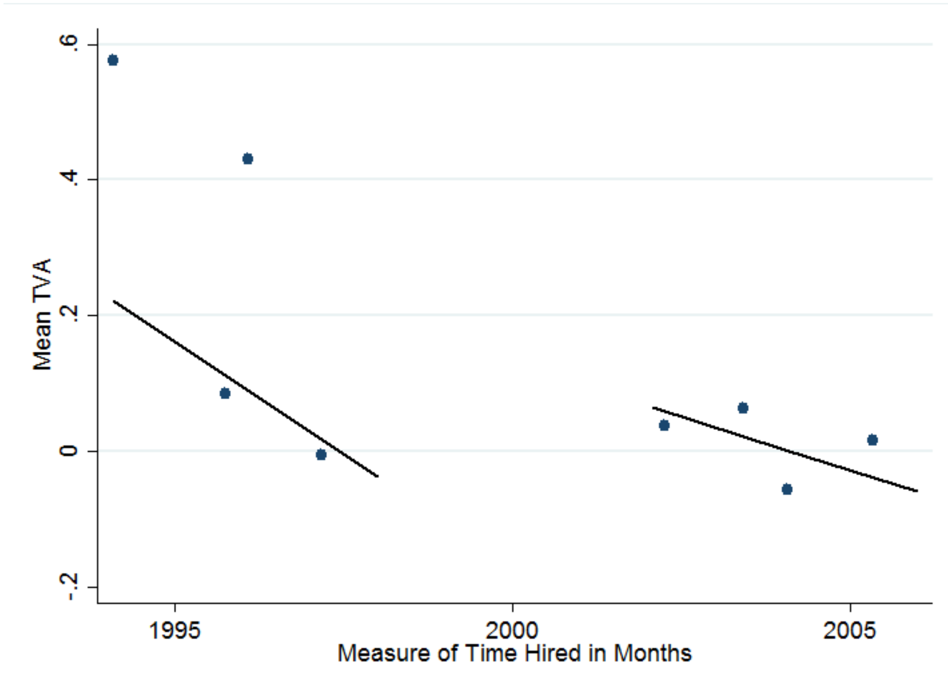
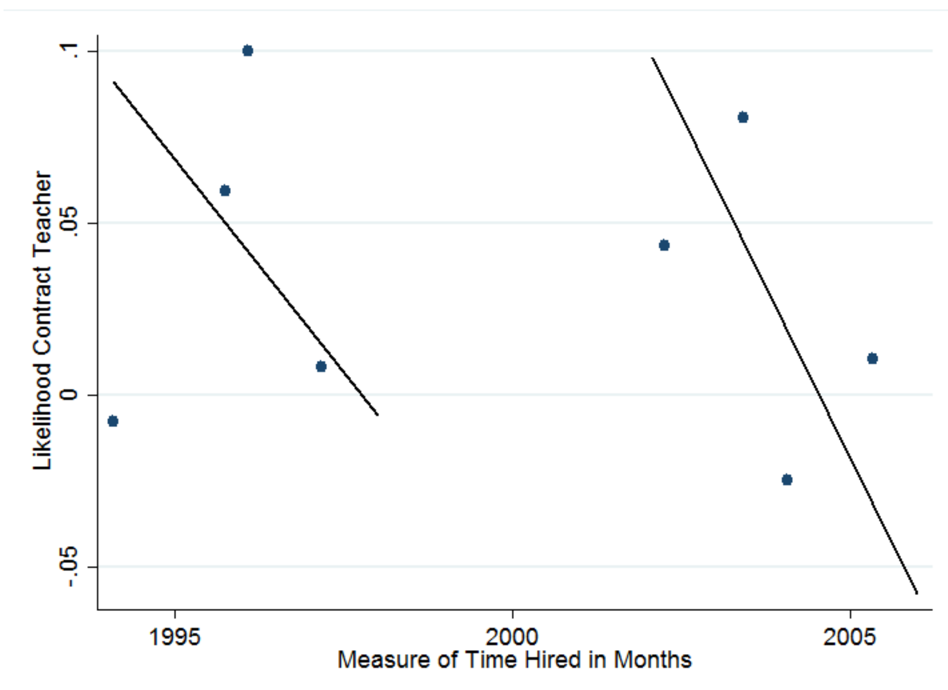


Figure 16: Discontinuity in Mean TVA Within School by Month Hired



10 Sampling Bias Estimation Appendix

Our sampling bias calculation closely follows that of Araujo et al. (2015). The only difference is that we de-mean at the school level by weighting all teachers equally instead of weighting them by the number of observed students. Following, Araujo et al. (2015), let the TVA estimation equation be:

$$y_{ijst} = \delta_{js} + \mathbf{X_{ijst}}\beta + \epsilon_{ijs},$$

where y_{ijst} is the test score of student i of teacher j in school s in year t, δ_{js} is the coefficient on the teacher indicator variable, $\mathbf{X_{ijst}}$ are the controls, and ϵ_{ijs} is a student specific error. We assume any time error is additive and is removed by round fixed effects. Since there is sampling error in our estimates, our estimate for a teacher's fixed effect is $\hat{\delta} = \delta_{js} + \frac{\sum_{i=1}^{N_{js}} \epsilon_{ijst}}{N_{js}}$.

Now define our demeaned teacher effect of interest $\gamma_{js} = \delta_{js} - \frac{\sum_{d=1}^{T_s} \delta_{ds}}{T_s}$, where T_s is the number of teachers in a school s. Accounting for sampling error, our estimate of γ_{js} is then

$$\widehat{\gamma_{js}} = \widehat{\delta_{js}} - \frac{\sum_{d=1}^{T_s} \widehat{\delta_{ds}}}{T_s}.$$

We can rewrite this as:

$$\widehat{\gamma_{js}} = \delta_{js} - \frac{\sum_{d=1}^{T_s} \delta_{ds}}{T_s} + \left(\frac{\sum_{i=1}^{N_{js}} \epsilon_{ijs}}{N_{js}} - \frac{\sum_{d=1}^{T_s} \sum_{i=1}^{N_{ds}} \frac{\epsilon_{ids}}{N_{ds}}}{T_s} \right).$$

Simplifying gives us,

$$\widehat{\gamma_{js}} = \gamma_{js} + \frac{1}{N_{js}} \sum_{i=1}^{N_{js}} \epsilon_{ijs} - \frac{1}{T_s} \sum_{d=1}^{T_s} \sum_{i=1}^{N_{ds}} \frac{\epsilon_{ids}}{N_{ds}}. \tag{4}$$

Assume ϵ_{ijs} are homoskedastic with a variance σ^2 and independent of γ_{js}. At this point, it is also helpful to define some additional notation:

$$\phi = \frac{1}{M} \sum_{js} \gamma_{js}^2,$$

where M is the total number of teachers. ϕ, the variance of true teacher quality, is our object of interest. The variance of the TVA estimates is

$$\widehat{\phi} = \frac{1}{M}\sum_{js}\widehat{\gamma_{js}}^2.$$

Therefore,

$$E(\widehat{\phi}) = \frac{1}{M}\sum E(\widehat{\gamma}^2),$$

and, referring back to equation 4, $E(\widehat{\phi})$ can be rewritten as

$$\begin{aligned}E(\widehat{\phi}) = \frac{1}{M}\sum_{M} E\Bigg(&\gamma_{js}^2 + \left(\frac{1}{N_{js}}\sum_{i=1}^{N_{js}}\epsilon_{ijs}\right)\left(-\frac{1}{T_s}\sum_{d=1}^{T_s}\sum_{i=1}^{N_{ds}}\frac{\epsilon_{ids}}{N_{ds}}\right) + \left(\frac{1}{N_{js}}\sum_{i=1}^{N_{js}}\epsilon_{ijs}\right)\left(\frac{1}{N_{js}}\sum_{i=1}^{N_{js}}\epsilon_{ijs}\right)\\ &+ \left(\frac{1}{T_s}\sum_{d=1}^{N_s}\sum_{i=1}^{N_{ds}}\frac{\epsilon_{ids}}{N_{ds}}\right)\left(\frac{1}{T_s}\sum_{d=1}^{T_s}\sum_{i=1}^{N_{ds}}\frac{\epsilon_{ijs}}{N_{ds}}\right)\Bigg).\end{aligned}$$

Simplifying produces the relationship

$$\begin{aligned}E(\widehat{\phi}) &= \phi + \frac{1}{M}\sum_{js}\left(\frac{\sigma^2}{N_{js}} - \frac{\sigma^2}{N_{js}T_s} + \frac{1}{T_s^2}\sum_{d=1}^{T_s}\sum_{i=1}^{N_{ds}}\frac{\epsilon_{ids}}{N_{ds}}\sum_{d=1}^{T_s}\sum_{i=1}^{N_{ds}}\frac{\epsilon_{ids}}{N_{ds}}\right)\\ &= \phi + \frac{1}{M}\sum_{js}\left(\frac{\sigma^2}{N_{js}}\left(1-\frac{1}{T_s}\right) + \frac{1}{T_s^2}\sum_{d=1}^{T_s}\frac{\sigma^2}{N_{ds}}\right).\end{aligned} \tag{5}$$

We can then rearrange equation 5 to get an estimable equation for the true variance of the TVAs in terms of the variance of the estimated TVAs and the sampling bias:

$$\phi = E(\widehat{\phi}) - \frac{1}{M}\sum_{js}\left(\frac{\sigma^2}{N_{js}}\left(1-\frac{1}{T_s}\right) + \frac{1}{T_s^2}\sum_{d=1}^{T_s}\frac{\sigma^2}{N_{ds}}\right).$$

11 Appendix Tables

Table A1: Summary Statistics

	(1)	(2)	(3)	(4)
	Government		Private	
	Mean	SD	Mean	SD
Female	0.449	0.497	0.768	0.422
Local	0.273	0.445	0.538	0.499
Some Training	0.904	0.294	0.221	0.415
BA Plus	0.514	0.500	0.255	0.436
Mean Salary	7671 ($129)	3746 ($63)	1407 ($24)	997 ($17)
Temporary Contract	0.229	0.420	0.838	0.368
Year Started	1,990.80	10.710	2,002.17	7.749
Mean Days Absent	2.644	3.297	1.936	3.368
Mean School Basic Facilities	-0.519	0.831	0.606	1.353
Mean School Extra Facilities	-0.607	1.401	0.716	1.033
Mean Student Household Assets	-0.236	0.242	0.484	1.022
Mean Student Mother Primary Education	0.298	0.212	0.378	0.276
Mean Student Father Primary Education	0.580	0.245	0.739	0.242
Change in Math Scores	0.393	0.499	0.355	0.488
Change in English Scores	0.393	0.474	0.388	0.461
Change in Urdu Scores	0.444	0.453	0.423	0.434
Change in Mean Scores	0.410	0.413	0.372	0.399

This table presents teacher-level summary statistics across 4 rounds of the LEAPS survey (2004-2007). Changes in test scores are calculated by averaging over the difference between a student's test scores in time t and time $t-1$. Household assets and school basic and extra facilities are predicted from a principal components analysis of indicator variables for the presence of different assets, and school facilities and are normalized by year observed.

Table A2: Learning Gains by Subject and Year

	(1) Math	(2) English	(3) Urdu
Gains Between Years 1 and 2	0.197	0.260	0.278
Gains Between Years 2 and 3	0.479	0.442	0.470
Gains Between Years 3 and 4	0.323	0.396	0.417

This table presents average yearly learning gains in math, English, and Urdu in a balanced panel of 3,426 public school students who appear in all four rounds of data.

Table A3: Proportion Correct on Selected Questions

	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4
English				
English 12: Match picture with word, Banana	0.612	0.715	0.738	0.806
English 18: Fill missing letter for picture, Cat	0.666	0.708	0.741	0.790
English 19: Fill missing letter for picture, Flag	0.283	0.284	0.375	0.443
English 30: Fill missing word in sentence	0.272	0.321	0.340	0.430
English 43: Construct sentence with word "deep"	0.010	0.011	0.022	0.068
English 44: Construct sentence with word "play"	0.024	0.026	0.066	0.145
Math				
Math 1: Count number of moons, write number	0.603	0.650	0.686	0.679
Math 9: Add 3 + 4	0.893	0.896	0.898	0.906
Math 12: Multiply 4 x 5	0.581	0.605	0.618	0.695
Math 24: Add 36 + 61	0.837	0.864	0.850	0.876
Math 25: Add 678 + 923	0.535	0.565	0.583	0.655
Math 27: Subtract 98 - 55	0.685	0.727	0.716	0.778
Math 30: Multiply 32 x 4	0.499	0.530	0.541	0.632
Math 32: Divide 384 / 6	0.186	0.225	0.281	0.375
Math 34: Cost of necklace, simple algebra	0.095	0.138	0.179	0.227
Math 39: Convert 7/3 into mixed fractions	0.017	0.040	0.036	0.087
Urdu				
Urdu 3: Match picture with word, Book	0.715	0.783	0.807	0.853
Urdu 4: Match picture with word, Banana	0.710	0.781	0.798	0.854
Urdu 5: Match picture with word, House	0.517	0.566	0.616	0.683
Urdu 10: Combine letters into word	0.716	0.762	0.770	0.829
Urdu 12: Combine letters into word	0.357	0.415	0.431	0.517
Urdu 19: Antonyms, Chouta	0.424	0.468	0.527	0.622
Urdu 20: Antonyms, Khushk	0.351	0.450	0.482	0.574
Urdu 36: Complete passage for grammar	0.284	0.352	0.381	0.518

This table presents the proportion of correct answers by year tested on a selected set of questions from the math, Urdu, and English exams. The data consists of a balanced panel of 3,426 public school students.

Table A4: Sources of Variation in Teacher Value-Added Calculations

	(1) Number of Teachers	(2) Number of Students	(3) Teachers in Schools With > 1 Teacher With Tested Students	(4) Students in Schools With > 1 Teachers With Tested Students
Public, Rd 1	486	8340	4	131
Private, Rd 1	303	3617	0	0
Public, Rd 2	593	9327	214	3290
Private, Rd 2	336	3340	97	846
Public, Rd 3	1007	16946	884	15320
Private, Rd 3	579	6777	524	6247
Public, Rd 4	1103	15357	812	12610
Private, Rd 4	599	5911	478	5020

This table presents the breakdown of the data used to calculate within and across school TVAs. Within school TVAs require teachers to teach in schools where more than one teacher has tested students (such that the mean school effect is not equal to the sole teacher's TVA). The sample of students driving variation in the within school TVAs are the students who attend schools where more than one teacher has tested students.

Table A5: Relationship Between Subject-Specific Knowledge and Subject-Specific TVA

	(1) Math	(2) Math	(3) English	(4) English	(5) Urdu	(6) Urdu
Female	-0.051*	0.191	0.164***	0.339	0.129***	0.090
	(0.030)	(0.359)	(0.029)	(0.266)	(0.026)	(0.223)
Local	-0.001	-0.020	0.043	0.012	0.029	-0.003
	(0.034)	(0.060)	(0.033)	(0.061)	(0.027)	(0.046)
Some Teacher Training	-0.076	-0.206	-0.111	-0.265	-0.092	-0.167
	(0.088)	(0.128)	(0.097)	(0.221)	(0.078)	(0.126)
Has BA or Better	0.002	0.015	0.016	-0.008	0.018	0.022
	(0.040)	(0.069)	(0.035)	(0.064)	(0.033)	(0.060)
Had > 3 *Years of Teaching Experience in* 2007	0.012	0.172	0.062	0.189	0.036	0.126
	(0.058)	(0.122)	(0.057)	(0.125)	(0.049)	(0.086)
Mean English Test Score	0.017	0.009	0.061***	0.035	0.019	-0.000
	(0.017)	(0.025)	(0.021)	(0.031)	(0.014)	(0.022)
Mean Urdu Test Score	0.048*	0.035	0.037	0.019	0.017	-0.013
	(0.027)	(0.045)	(0.026)	(0.046)	(0.024)	(0.039)
Mean Math Test Score	0.020	-0.018	0.024	-0.022	0.023	0.002
	(0.025)	(0.041)	(0.025)	(0.038)	(0.023)	(0.035)
Temporary Contract	-0.043	0.031	0.015	0.042	-0.032	0.081
	(0.052)	(0.103)	(0.047)	(0.092)	(0.045)	(0.086)
Fixed Effects	District	School	District	School	District	School
Number of observations	919	919	919	919	919	919
Adjusted R Squared	0.161	0.316	0.240	0.406	0.200	0.397
F	1.629	0.464	5.709	0.652	3.349	0.504
Clusters	469	469	469	469	469	469

This table regresses subject-specific TVAs in math, Urdu, and English on teacher characteristics, including teacher test scores in math, Urdu, and English. Standard errors are clustered at the school level.

Table A6: Non-Linearities in Within-Teacher Experience Effects

	(1) Math	(2) Urdu	(3) English	(4) Mean
Has 0 or 1 Years of Exp.	-0.748***	-0.698***	-0.536***	-0.604***
	(0.278)	(0.217)	(0.201)	(0.206)
Has 2 Years of Exp.	-0.434**	-0.481***	-0.250	-0.353**
	(0.196)	(0.160)	(0.154)	(0.143)
Has 3 Years of Exp.	-0.112	-0.115	-0.024	-0.075
	(0.178)	(0.140)	(0.151)	(0.131)
Has 4 Years of Exp.	-0.070	-0.067	-0.030	-0.071
	(0.150)	(0.119)	(0.105)	(0.095)
Teacher FE	Y	Y	Y	Y
Lagged Test Scores	Y	Y	Y	Y
Number of Observations	4,616	4,616	4,616	4,616
Adjusted R Squared	0.623	0.666	0.644	0.722
Clusters	212	212	212	212

This table tests for non-linearities in the effect of teacher experience on student test scores. All regressions control for teacher fixed effects and lagged student test scores. The sample is restricted to temporary contract teachers in government schools. Standard errors are clustered at the school level.

Table A7: Correlation Between TVA Specifications

	(1) Across Schools	(2) Within Schools
English	0.977	0.955
Math	0.969	0.951
Urdu	0.963	0.944

English, math, and Urdu TVAs are calculated with and without individual level controls for gender, age, household assets, basic and extra facilities indices, and student-teacher ratios. Each cell of the table gives the correlation between the TVA estimated with the parsimonious specification and the TVA estimated with the detailed specification.

Table A8: Bootstrapped Regression Discontinuity Results

(1) Bandwidth (Years)	(2) Mean TVA	(3) P-value	(4) Within School, Mean TVA	(5) P-value
1	0.004	0.497	0.045	0.130
2	1.041**	0.026	0.196	0.181
3	0.182	0.244	0.402*	0.053
4	0.398*	0.070	0.217**	0.043

This table replicates the fuzzy regression discontinuity estimates of the effect of temporary contract status on TVA using a bootstrap procedure to account for estimation error. The bootstrap was clustered at the month hired level. The estimates are based on 10,000 random samples of the data. The coefficients reported are the averages of the coefficient estimates over the 10,000 draws.

Table A9: Regression Discontinuity Results Including Student-Teacher Ratio Controls

(1) Bandwidth (Years)	(2) Mean TVA	(3) Se	(4) Within School, Mean TVA	(5) Se
1	-0.007	0.053	0.060	0.040
2	0.683	0.488	0.345	0.306
3	0.263	0.252	0.260**	0.122
4	0.226	0.195	0.198**	0.096

This table replicates the fuzzy regression discontinuity estimates of the effect of temporary contract status including controls for school-year student teacher ratios. Standard errors are clustered at the month hired level.