# Compulsory tutorial programmes and performance in undergraduate microeconomics: A regression discontinuity design

Schöer, Volker and Shepherd, Debra

African Micro-Economic Research Unit (AMERU), School of Economic and Business Sciences, University of the Witwatersrand, Research on Socio-Economic Policy (ReSEP), Department of Economics, Stellenbosch University

September 2013

# Compulsory tutorial programmes and performance in undergraduate microeconomics: A regression discontinuity design

**Volker Schöer[1] and Debra Shepherd[2]**

**September 2013**

## Abstract

As South African universities experience extremely low graduation rates, academic staff implement a range of interventions, such as tutorial programmes, in order to improve student performance. However, relatively little is known about the impact of such tutorial programmes on students' performance. Using data from an introductory microeconomics course, this paper investigates the impact of a compulsory tutorial programme on students' performance in their final examination. Due to the fact that the tutorial programme was only compulsory for students that obtained less than a pass in the first test, while otherwise offered on a voluntary basis, this paper employs a fuzzy RD design to investigate the impact of the tutorial programme on final exam performance. Findings indicate that assignment to the compulsory programme positively affects students' performance. However, this result is mainly driven by students who already seem to have the ability to perform but, for whatever reason, underperformed in the first test. Thus, while assignment to the tutorial programme itself leads to an improvement in performance, the mechanism is unclear.

**Keywords**: Peer tutoring, compulsory tutorial programme, introductory microeconomics, regression discontinuity design, South Africa

**JEL:** A2, A20, A22

[1] African Micro-Economic Research Unit (AMERU), School of Economic and Business Sciences, University of the Witwatersrand, South Africa, email: volker.schoer@wits.ac.za

[2] Research on Socio-Economic Policy (ReSEP), Department of Economics, Stellenbosch University, South Africa, email: debrashepherd@sun.ac.za

## INTRODUCTION

At 15 percent, South Africa has one of the lowest university graduation rates in the world (Lekseka and Maile, 2008). Drop-out rates amongst first-year students has also been reported to be as high as 35 percent at some universities during recent year. These worrying trends in higher education come at high financial and social costs. At the same time, university departments are taking strain as enrolment numbers continue to rise and resources are becoming even more limited. As a result of these factors, university departments have the dual requirement of improving the quality of teaching while improving cost effectiveness ("doing more with less"). A further concern within the current teaching and learning environment of universities is that the traditional approaches to curricula and assessment have promoted a surface approach to learning rather than a deep or strategic approach which may bring disproportionate gains to minority student groups (Entwistle et al, 2001). There are therefore both practical and ethical reasons for the move towards adopting peer tutoring as part of the learning support structure in higher education. The increase in use of peer tutoring in higher education courses clearly raises important questions of assessment, acceptance and the eventual success of such a programme, as poor design can be damaging to the positive features of what could be an important component of teaching and learning (Boud et al, 1999).

The specific microeconomics course used for purposes of this study initiated its own tutorial programme in 2009 parallel to formal lecture sessions.[3] Attendance of these tutorials was made mandatory for poor performing students (obtained below 50 percent) identified through early assessment testing. Students who achieved at least 50 percent in the first test were still permitted to attend tutorials on a voluntary basis. The 2010 class cohort is used for analysis purposes given the stricter enforcement of the policy. The specific design of this policy has presented an opportunity to directly assess the impact of tutorial attendance on academic performance through the use of regression discontinuity design. Specifically, a fuzzy regression discontinuity design is employed to estimate a local average treatment effect of the tutorial programme within a bandwidth of the policy cut off. Estimates using both parametric and non-parametric models are presented.

This paper begins with an overview of the literature that empirically investigates the effectiveness of peer tutoring on undergraduate performance in economics. The following section describes the data and policy design of the programme, followed by a discussion of the methodology. The next two sections present the empirical results and robustness checks, while the final section concludes.

## OVERVIEW OF THE LITERATURE

The body of research on peer tutoring saw tremendous growth in recent decades as illustrated by the many reviews and surveys (c.f. Goldschmid and Goldschmid, 1976; Whitman, 1988; Lee, 1988; Maxwell, 1990; Topping, 1996). The literature spans a range of elements of the peer tutoring process from practice to design and organisation (c.f. Schmidt and Moust, 1995), as well as assesses the relative advantages of peer tutoring for both tutees and tutors *inter alia* cognitive

[3] The tutorial programme existed prior to 2009, although a full year undergraduate economics course was presented; that is, the first-semester microeconomics course was combined with the second-semester macroeconomics course. There is therefore limited comparability prior to and post 2009.

processes and emotional support as well as the impact on various outcomes such as performance, retention and drop-out. In determining the effectiveness of peer tutoring, one should be cognisant that programmes tend to be diverse and therefore may have very little in common. For example, tutors may be staff or students; the tutor and tutees may meet in individual or group settings; frequency of meetings may range from several times a week to once a week to once a month; tutors may receive special training or may be unsupervised; tutors may receive some form of remuneration or may volunteer to participate; tutors and tutees may have some or no choice in their pairings; and so forth. Additionally, tutoring programmes may differ in their aims and objectives, be it improved achievement, reduced attrition or increased interest in the subject. Three methods of peer tutoring have been widely used in higher education and have demonstrated to be quite effective (Topping, 1996). These are: cross-year small-group tutoring, where upper year undergraduates or postgraduates function as tutors to a small group of lower undergraduate students; the personalised system of instruction (PSI), where students are able to progress through the study material at their own pace and the role of peer tutors are largely to check, test and record the advancement of tutees; and supplemental instruction (SI).

The evaluation of peer tutoring programs in higher education has traditionally tended to use weak programme designs, with much of the empirical work relying on cross-sections of subjective outcome measures that are largely retrospective in nature (Jacobi, 1991). Often the data are reported without adequate evidence of reliability and validity. However, recent research has become more empirically rigorous, with greater use of experimental and randomly controlled programme designs that attempt to correct for potential selection biases. While student-to-student tutoring has been used with some success in several disciplines, there have been relatively few evaluations of its impact on student learning in economics (see for example Kelley and Swartz, 1976; Johnston et al, 2000; Munley et al, 2010). Research is even more limited in a South African context (see Jansen and Horn, 2009).

The few empirical studies that have been published tend to be fraught with methodological weaknesses that seriously limit both internal and external validity of the results. For example, research of tutorial programmes that are based on systematic selection rather than random assignment need to make adequate attempts to control for sampling and self-selection biases, although there should be recognition that the corrections are likely to be imperfect or incomplete (Cook and Campbell, 1979). A further concern problem with peer tutoring research is the potentially low levels of external validity. Most research is based on data collected within a single department within a specific university. The scope for generalizing these findings based on these studies to other tertiary institutions and other students is limited.

A study of a peer tutoring program at Duke University by Kelley and Swart (1976) made use of weekly computer based tests to differentiate between good and poor performing students after which top performers were given the option to tutor weaker students in exchange for exemption from a forthcoming examination. The performance of students who accepted an invitation to attend the tutorial sessions was compared to the group of students who declined. A significant positive impact of 0.67 standard deviations (4.2 percentage points) on the final course score was estimated. However, it is posited that these results may understate the true impact of tutorials as it excludes the performance of the tutors themselves. The authors correctly recognise that the group of tutees are a self-selected group and that their results are likely to be inflated by selection on unobservables, most notable motivation, despite the two groups being very similar on observables.

In a South African context, Jansen and Horn (2009) make use of ordinary least squares regression to model the impact of various factors, including tutorial attendance, on the course mark in an undergraduate economics course at a South African university. Student attendance of these tutorials was voluntary, although students who performed poorly in the first test were encouraged to attend. The group of students who attended regularly were found to have better school-leaving grades and a better average performance in economics. This therefore raises concerns that the coefficient on tutorial attendance may be biased due to sample selection. Class attendance was included as a proxy for motivation, which may serve as a control against the voluntary attendance. A significant positive effect of tutorial attendance on performance was found, with a larger effect for first-time registered students than for repeat students.

More recently a number of studies have attempted to estimate the impact of peer tutor programs through experimental design so as to correct for selection bias. Johnston et al (2000) evaluate the impact of a collaborative problem solving (CPS) approach to tutorials in a second-year macroeconomics course. Treatment and control groups were generated where one group was exposed to the CPS approach whilst the other attended tutorials that continued to use the traditional approach. Programme evaluation was based on qualitative measures such as student attitude and teaching-evaluation questionnaires, as well as quantitative information regarding tutorial attendance and examination performance. Students attending CPS were found to both value their tutors' performance and enjoy their tutorials more. They also spent significantly more time preparing for the tutorial sessions. No consistent gain was observed for the control versus treatment groups, except in the case of foreign students. The researchers posit, though not convincingly, that the non-significant change in performance and learning may be due to spill over effects or inappropriate selection of the control and treatment groups. Munley et al (2010) evaluate the effect of participating in a tutoring programme across several courses (including undergraduate economics) using two methodological approaches. First they model the exogenous effect of participation or level of participation on the final grade; and second, given voluntary participation, they adopt a treatment model per Greene (2008) where participation and performance are modelled jointly using selection and outcome equations. They use two policies regarding intercollegiate athletics as an exclusion restriction. Under the first model treating participation as exogenous, they find a *negative* and statistically significant coefficient on the binary choice to participate in tutorials, which they put down to participation likely being higher amongst weaker students. Modelling the choice to participate, the coefficient on tutorial participation turns positive but is statistically insignificant. However, modelling the level of participation rather than the choice to participate yields positive and significant results. Therefore, the amount of participation appears to be more relevant for improving performance, with a sufficient amount of tutorial attendance required in order to see notable gains.

It is clear from the already existing research that the results are mixed, which may in part be due to differences in the underlying programmes and their participants, or the choice of modelling strategy. This study aims to add to the current empirical evidence on the effectiveness of peer tutoring in economics through the use of what the authors believe to be a truly exogenous tutorial programme that addresses the issue of sample selection bias.

## DATA AND POLICY DESIGN

This study uses tutorial attendance data from an undergraduate micro economics course that was run during the first academic semester (February to May) of 2010. The course has one of the largest enrolments amongst undergraduate modules, with 1767 students enrolled in the year analysed.[4] Students were sub-divided by language (English or Afrikaans) into one of seven formal lecture classes. Students were expected to attend 50-minute lectures per week for 14 weeks, as well as one 50-minute tutorial session that began two weeks after the start of the formal academic semester and lasted for the remaining 12 weeks of the semester. The tutorial programme is one of structured academic support where students are able to benefit from a small-class environment (less than 30 students per tutorial). Students are instructed to attempt a tutorial question set that tackles problems related to coursework material covered in the formal lectures in the preceding week. This is provided to all students one week prior to the tutorial. Tutors are expected to cover as many of the answers to these problem sets, time permitting.

Attendance of tutorial classes is voluntary up until a week following the first semester test, after which students with a test score below a passing score of 50 percent were required to attend the tutorial classes on a compulsory basis. Students who did not write the first semester test were also subject to compulsory tutorial attendance. Given that we do not observe their performance, and therefore cannot necessarily include them in the group of "just failers", these students are dropped for analysis purposes. Furthermore, in order to make comparisons from test 1 to test 2, we only consider those students who wrote both tests. Our final sample is therefore comprised of 1653 students (93.5 percent of the original sample). Tutorial attendance remained voluntary for students that scored at least or above 50 percent in the first test. The compulsory tutorial policy was announced in the first week of classes, with further reminders given in the weeks prior to and after the first test. Students that scored below 50 percent were alerted via e-mail that they were required to attend the tutorial classes. Tutorial attendance was recorded by tutors as students arrived for each tutorial class. Any student that left before the end of the tutorial was not marked down as attending.

The first semester test (or early assessment test) was written fairly early into the semester a few weeks after the start of tutorial classes.[5] In addition to this early assessment test, students are required to write at least one of two remaining semester tests, although students are permitted to write all three if they choose. Admission to the examination is contingent on achieving an average of at least 40 percent on the semester tests. Therefore, whilst 1767 students were enrolled for the course at the beginning of the academic year, only 1489 achieved the required semester average to gain entrance to the exam. Those students who did not gain access are likely to be compulsory tutorial students. This could cause concern for our analysis as the sample of students between test 2 and the exam are not the same. However, given that we are only interested in the effect of tutorial attendance for students who perform within a neighbourhood around the cut off, the two samples

[4] 27 students deregistered from the course during the semester, and are therefore dropped from the analysis.

[5] The first semester test comprised of 10 true/false and 10 multiple choice questions (referred to as short answer questions). Subsequent tests and exams consisted of both descriptive and short answer questions. Tests are marked by postgraduate teaching assistants, whilst the course lecturers are involved in the marking of the examinations. In general, markers are unaware of which students are subject to compulsory tutorial attendance.

are unlikely to be that dissimilar. Students were also offered a choice of writing one of two exams, both of which are set to be of the same level of difficulty. Students who wrote the first exam and did not achieve a passing mark (50 percent) but achieved a sub-minimum average of 40 percent for their semester tests and exam were permitted to write the second examination option. Students who chose only to write the second exam therefore only received one exam opportunity. For purposes of this study, we consider the mark obtained by the student in their first exam attempt.[6] As part of the course administration, each student's tutorial attendance, tutoring sessions attended, semester test, class mark and final exam scores, gender, year of enrolment and degree major were recorded. Additional information regarding the student's high school leaving performance, school department, home language, age and test scores in additional undergraduate courses taken in the same semester were also obtained.

## METHODOLOGY: THE REGRESSION DISCONTINUITY DESIGN

We are interested in estimating the effect that participation in the tutorial programme, $T_i$, has on test scores $Y_i$. We assume that $Y_i$ is further related to some vector of observables $W_i$, such that

$$Y_i - \beta_0 + \alpha T_i + W_i \beta_1 + u_i \tag{1}$$

where α represents the effect of $T_i$, assumed to be constant across individuals, and the error term $\varepsilon_i$ is assumed to be uncorrelated with $W_i$. Unless treatment has been randomly assigned conditional on $W_i$, identification of $\alpha$ is hampered by selection bias due to some dependence between $T_i$ and $u_i$. This arises when treatment is related to some unobservable/s not included in $W_i$. The resulting dependence between $T_i$ and $u_i$ will therefore be erroneously attributed to the impact of the programme on the outcome of interest.

We solve for the selection issue using information about the mechanism by which participation in the tutorial programme was assigned. Specifically, compulsory tutorial attendance was determined by performance in the first semester test: students scoring below a given cut off c (50 percent) were required to attend tutorials on a mandatory basis, while students scoring at or more than c were not subject to the compulsory tutorial policy. Therefore, students are assigned to tutorials based on the following deterministic rule:

$$D_i(X_i) = 1\{X_i \geq c\} \tag{2}$$

where $X_i$ is student i's first semester test score, c is the cut off test score and 1{.} is the indicator function.

The above corresponds to the selection rule of a sharp Regression Discontinuity design (Thistlethwaite and Campbell, 1960). The assignment mechanism is clearly not random (there is little

[6] Robustness checks will be performed considering the final exam mark following all attempts, as well as controlling for whether or not the student chose to write the second exam option or not (if we believe that weaker students are more likely to delay sitting the exam). A comparison of means indicates that compulsory students are no less likely to write the second option than non-compulsory students are. However, compulsory tutorial students are more likely to write both exam options. This is to be expected given that they are weaker performing students.

reason to suppose that $X_i$ is unrelated to $Y_i$), therefore a simple comparison of means between the treatment and non-treatment (control) groups would not suffice to provide an unbiased estimate of $\alpha$. However, if we expect that for some arbitrarily small number ε>0 that $E[\alpha_i|X_i = c + \varepsilon] \cong E[\alpha_i|X_i = c - \varepsilon]$ and further assume that both $E[u_i|X]$ and $E[\alpha|X]$ are continuous in X at c (Hahn et al, 2001; van der Klaauw, 2002), then we have

$$\lim_{\varepsilon \downarrow c} E[Y|X] - \lim_{\varepsilon \uparrow c} E[Y|X] = \alpha \tag{3}$$

Therefore, by comparing individuals arbitrarily close to c who did and did not receive treatment, we are able to identify (in the limit) the causal impact of the tutorial programme on performance.

However, given that tutorials were not denied to the group of students scoring at or above the cut off, the rate of tutorial attendance as a function of semester test 1 performance is now a discontinuous function in $x_i$ at c. This represents the discontinuity "fuzzy" or stochastic RD design. Under the same two continuity assumptions listed above and the additional assumptions of local "monotonicity"[7] and "excludability"[8] (Hahn et al, 2001; Imbens and Angrist, 1994) gives

$$\frac{\lim_{\varepsilon \downarrow c} E[Y|X] \quad \lim_{\varepsilon \uparrow c} E[Y|X]}{\lim_{\varepsilon \downarrow c} E[T|X] - \lim_{\varepsilon \uparrow c} E[T|X]} = \alpha_F \tag{4}$$

where the subscript F represents the fuzzy treatment estimator. Taking the limit of both sides of (4) as $\varepsilon \rightarrow c$ would identify the "local Wald" estimator, α, per Hahn et al (2001) as:

$$\alpha_F = \frac{Y^+ - Y^-}{T^+ - T^-} \tag{5}$$

## Estimation

### PARAMETRIC: IV estimator

In a context such as this where treatment, T, is continuous and there is a randomized binary instrument, D, an instrumental variable approach is an obvious way of obtaining an estimate of the impact of T on Y. The treatment effect, $\alpha_{IV}$, is calculated as the reduced form impact D on Y divided by the first-stage impact of D on T, and uses the entire sample of observations. The model set-up is the same as in (1) except with an added second equation that allows for imperfect compliance and observables and unobservables to impact the rate of tutorial attendance:

$$Y_i = \beta_0 + \alpha T_i + W_i \beta_1 + \varepsilon_i$$

$$T_i = \theta_0 + D_i \pi + W_i \alpha + \nu_i$$

$$D_i = 1.\{X \geq c\}$$

$$X = \beta_2 W_i + \xi_i \tag{6}$$

[7] X crossing c cannot simultaneously cause some units to take up and others to reject.
[8] X crossing c cannot impact Y except through impacting receipt of the treatment.

where we make no assumptions about the correlations between W, ε, ν and ξ. It is simple to show that

$$\lim_{\varepsilon \downarrow c} E[Y|X = c + \varepsilon] - \lim_{\varepsilon \uparrow c} E[Y|X = c + \varepsilon] = \left\{\lim_{\varepsilon \downarrow c} E[T|X = c + \varepsilon] - \lim_{\varepsilon \uparrow c} [E[T|X = c + \varepsilon]\right\} \alpha]$$

(7)

where the left-hand side represents the reduced-form discontinuity in the relation between Y and X, and the term in front of $\alpha$ is the "first-stage" discontinuity in the relation between T and X. The ratio of the two discontinuities yields the treatment estimator α.

There is no particular reason to believe that the true model is linear, and the consequences of incorrect functional form are more serious in the case of RD design as misspecification generates bias in the estimator of interest, α. Allowing for non-linearities in the underlying function of X can be important, especially in cases where we suspect X and Y to be non-linearly related, for example, when we have reason to expect this relationship to change as a result of the program. One way of circumventing this is to augment the outcome equation with a regression function f(X), known as the control function approach (Heckman and Robb, 1985). We can generalise this function by allowing the $X_i$ terms to differ on each side of the cut off by including the $X_i$ terms individually and interacted with $\mathbf{D_i}$ (van der Klaauw, 2002; McCrary, 2008; Lee and Lemieux, 2010). The reduced-form outcome function is now

$$Y_i = \beta_0 + D_i\alpha\pi + \delta_{01}\widetilde{X}_i + \delta_{02}\widetilde{X}_i^2 + \cdots + \delta_{0p}\widetilde{X}_i^p + \delta_1 D_i\widetilde{X}_i + \delta_1 D_i\widetilde{X}_i^2 + \cdots + \delta_p D_i\widetilde{X}_i^p + W_i\beta_1 + \varepsilon_i$$

(8)

We can also allow for a control-function g(X) in the first-stage equation

$$T_i = \theta_0 + D_i\pi + \gamma_{01}\widetilde{X}_i + \gamma_{02}\widetilde{X}_i^2 + \cdots + \gamma_{0p}\widetilde{X}_i^p + \gamma_1 D_i\widetilde{X}_i + \gamma_1 D_i\widetilde{X}_i^2 + \cdots + \gamma_1 D_i\widetilde{X}_i^2 + W_i\alpha + \nu_i$$

(9)

where $\mathbf{\tilde{X}_i = [(X]_i - c)}$. The instrumental variable estimate of treatment is obtained by taking the ratio $\frac{\alpha\pi}{\pi}$. Given that the model is exactly identified, a two-stage estimation procedure per van der Klaauw (2002) will be numerically identical to $\frac{\alpha\pi}{\pi}$. This involves estimating the control function augmented second-stage outcome equation by replacing $\mathbf{T_i}$ with the first stage estimate. With correctly specified control functions *f(X)* and *g(x)*, this two-stage procedure yields a consistent estimate of the treatment effect. If we assume the same functional form for *f(x)* and *g(x)*, then the two-stage estimation procedure described here will be equivalent to a two-stage least squares estimation with $\mathbf{D_i}$ and the terms in *f(x)* serving as instruments.

It should be noted that the instrumental variable estimate may still be biased by omitted variables if the compulsory tutorial policy changes student behaviour with regards to other learning such as studying, effort and class attendance. Student behaviour may be adjusted in a number of ways: first, students who are required to attend Economics tutorials on a mandatory basis may decrease the amount of time they spend studying or attending class, thereby underestimating the impact of the tutorials; secondly, compulsory tutorial students may feel that there is a stigma attached to the programme, and therefore will put in more effort than students who just passed semester test 1, leading to an overestimate in the impact of tutorials. In terms of adjusted effort,

**NON-PARAMETRIC : Wald estimator**

The estimation procedure described above is a parametric one that uses polynomial regression. Parametric estimation typically uses data away from the cut off, therefore providing global rather than local estimates of the regression function. However, in practice one can consider using a narrower window of observations around the cut off. Non-parametric techniques offer more flexible estimates of the regression function, as well as address the "boundary problem" of RD (we are interested in computing an effect at the cut off using only the closest observations). We could consider using kernel regression given that it is well suited from estimating regression functions at a particular point. However, in finite samples, precise estimation requires sufficiently wide bandwidths, and wider bandwidths come at the cost of greater bias. Local linear regressions have been introduced as a means of reducing bias in standard kernel regression methods (Fan and Gijbels, 1995; Hahn, Todd and van der Klaauw, 2001). Estimates under local linear regression are obtained by solving:

$$\min_{a,b} \sum \mathbf{1}(X_i \geq c)(y_i - a - b(X_i - c))^2 K\left(\frac{X_i - c}{h}\right) \tag{10}$$

in the case of $Y^+ = \lim_{\varepsilon \downarrow c} E[Y|X = c + \varepsilon]$, and

$$\min_{a,b} \sum \mathbf{1}(X_i < c)(X_i - a - b(X_i - c))^2 K\left(\frac{X_i - c}{h}\right) \tag{9}$$

in the case of $Y^- = \lim_{\varepsilon \uparrow c} E[Y|X = c + \varepsilon]$, with K(.) a kernel function and *h* a bandwidth that converges to 0 as $n \to \infty$ . Estimates for $T^+$ and $T^-$ are found in a similar way.

Various techniques are available for choosing the kernel function and bandwidths. Less important is the choice of kernel. RD design studies tend to adopt either the rectangular or triangular kernels, with the difference between the two that the latter places more weight on observations close to the cut off. Of more importance is the choice of bandwidth, as different bandwidth choices can produce quite different estimates. For this reason, it is sensible to report at least three estimates as an informal sensitivity test: one using the preferred bandwidth, one using twice the preferred bandwidth and another using half the preferred bandwidth (McCrary, 2008). In general, choosing a bandwidth in non-parametric estimation involves finding an optimal balance between precision and bias. The default bandwidth from Imbens and Kalyanaraman (2010) is designed to minimize MSE (squared bias plus variance) in a sharp RD design. However, the optimal bandwidth will tend to be larger for a fuzzy design due to the additional variance arising from the estimation of the jump in the conditional mean of treatment. Unfortunately, a larger bandwidth also leads to additional bias. According to McCrary (2008), the best method of bandwidth selection is visual inspection guided by an automatic procedure. A simple automatic bandwidth selection procedure uses a rule-of-thumb (ROT) bandwidth as follows:

$$h_{ROT} = \kappa\left[\frac{\tilde{\sigma}^2 R}{\sum_{i=1}^{N}\{\bar{m}''(X_i)\}^2}\right]^{\frac{1}{5}}$$

where $\kappa$ is 2.702 (3.348) in the case of the rectangular (triangular) kernels respectively, $\tilde{\sigma}^2$ is the estimated standard error of a 4th order polynomial regression of Y on X, R is the range of X and $\bar{m}''(X_i)$ is the second derivative implied by the global polynomial model (Fan and Gibels, 1995). Imbens and Lemieux (2008) recommend using the same bandwidth in the treatment and outcome regressions. When we are close to a sharp RD design, *g(X)* is expected to be very flat and the optimal bandwidth to be very wide. In contrast, there is no particular reason to expect the *f(X)* to be flat or linear, which suggests the optimal bandwidth would likely be less than the one for the treatment equation. As a result, Imbens and Lemieux (2008) suggest focusing on the outcome equation for selecting bandwidth, and then using the same bandwidth for the treatment equation.

### Inclusion of covariates

Up to this point estimation has explicitly allowed for the inclusion of baseline observables as covariates in the regression models. The baseline covariates are useful for testing the validity of the RD design by testing that the local continuity assumptions are satisfied. In their capacity as additional controls for parametric and non-parametric estimation, the only possible gain this affords is a reduction in the sampling variability (assuming they have explanatory power). However, estimation error in their covariates could also reduce efficiency. If the RD design is indeed valid, that is, the distribution of W given X is continuous at the threshold, the inclusion of additional covariates should still provide a consistent estimate of the local treatment effect (Imbens and Lemieux, 2008: 626). If including these controls leads to significant changes in estimates, this would suggest that the continuity assumptions may be violated and the treatment estimates are likely to be biased. Lee (2008) proposed a method to test the sensitivity of RD estimates to the inclusion of covariates by first regressing Y on a vector of individual characteristics and then to repeat the RD analysis using the residuals $\left(Y_i - \hat{Y}_i\right)$ as outcome variable. Intuitively, this procedure nets out the portion of the variation in Y we could have predicted using the pre-determined characteristics, making the question whether the treatment variable can explain the remaining residual variation in Y. The important thing to keep in mind is that if the RD design is valid, this procedure provides a consistent estimate of the same RD parameter of interest.

## RESULTS

From Figure A1 of the appendix it is clear that prior to semester test 1 there was variation in the number of tutorials attended by students. Approximately 55 percent of students attended all tutorials, while more than a fifth of all students did not attend any of the voluntary tutorials. Figure A2 shows how weekly tutorial attendance changed over the semester by compulsory and non-compulsory status. Week 0 indicates the week in which semester test 1 was written. It is immediately clear that prior to test 1, attendance amongst the non-compulsory group was higher than that of the compulsory group. Attendance amongst both groups also appeared to drop during the week in which the first semester test was written. Once the mandatory policy was instituted, the attendance of the compulsory group is approximately 40 percent higher than the non-compulsory group. Noting the trend in tutorial attendance amongst the group of non-compulsory students, the mandatory policy appears to have worked to counteract the tendency for tutorial attendance to decline over the semester.

Table 1 compares the average characteristics of the group of compulsory tutorial students with those of the group of non-compulsory students. Observing the entire student sample, students in the compulsory group were significantly less likely to attend tutorials prior to writing test 1. Additionally, members of this group are more likely to be repeat students, registered for degrees other than Actuarial Science, Accounting, Law and Mathematics and have achieved a lower matric[9] maths mark. They are also less likely to have been part of the NSC matriculant cohort. These differences suggest that identification of the impact of tutorial attendance on test and exam performance using OLS regression would very likely suffer from omitted variables bias. In terms of demographics, the only distinguishing feature of the two groups is that compulsory students are less likely to be from the white population group and more likely to be home-language Afrikaans speakers. The mandatory tutorial policy has a significant effect on attendance subsequent to semester test 1, with compulsory students attending 45 percent more tutorials prior to the exam when considering the entire sample. When the sample is narrowed to within 1 and 0.5 standard deviations from the policy threshold, the tutorial attendance gap prior to test 1 turns insignificant. The differences in post policy performance are also reduced. Despite the substantial increase in tutorial attendance of compulsory students relative to non-compulsory students, the latter continue to significantly outperform the former in tests, despite attending fewer tutorials. However, there are no notable differences in exam performance once the window is narrowed to 0.5 standard deviations. This may be due to the fact that, even with the narrower window, we are still capturing students of differing abilities (note a significant difference in matric maths performance for this sample).

The final column of Table 1 displays coefficients on the binary treatment from a regression of each of the characteristics on the quadratic control function from equation (2) without any covariates. These estimates describe how each variable differs between the compulsory and non-compulsory groups at the policy threshold. It is evident that, at the threshold, the post-test 1 attendance rate is significantly higher for the group of compulsory students. The difference in test and exam performance across the policy threshold is negative and statistically significant (at the 5 and 10 percent levels). This indicates that, at least within a window around the cut-off, performance is higher for the group of compulsory students. There is no significant difference in the other outcomes or characteristics.[10]

Figures A3, A4 and A5 of the appendix present scatter plots of the average tutorial attendance prior to and after test 1 in 0.1 standard deviation wide bins of the normalised test 1 score. It is clear that there is no noticeable discontinuity in attendance prior to test 1 at the threshold. However, once the compulsory tutorial policy is instituted, there are clear discontinuities in attendance at the policy threshold prior to the second semester test and the exam, with non-

[9] "Matric" refers to the final examination at the end of secondary schooling in South Africa. The matriculation exams are centrally set and standardized which allows comparisons of pupils' abilities that graduate from different secondary schools.

[10] Except for the Eastern Cape Education Department.

compulsory student behaviour appearing to change very little between test 2 and the end of the semester. The figures further display linear, quadratic and cubic fits to the underlying data. A linear fit of the running variable appears to capture the primary relationship between attendance and test 1 score the best. The primary analysis will therefore employ a linear form of the control function,

**Table 1**: comparison of compulsory and non-compulsory tutorial attendance groups

| | **Whole sample** | | **1 SD from threshold** | | **0.5 SD from threshold** | | **Parametric RD** |
|---|---|---|---|---|---|---|---|
| | Non-comp | Comp | Non-comp | Comp | Non-comp | Comp | |
| percentage tutorials prior to test 1 | 0.7276 | 0.5927 | 0.6948 | 0.5973 | 0.6483 | 0.5851 | 0.0727 |
| | 0.1349*** | | 0.0975*** | | 0.0632 | | |
| percentage tutorials prior to test 2, post test 1 | 0.5420 | 0.7325 | 0.5197 | 0.7368 | 0.4929 | 0.7207 | -0.2305*** |
| | -0.1905*** | | -0.2171*** | | -0.2278*** | | |
| percentage tutorials prior to exam, post test 2 | 0.4317 | 0.8818 | 0.4218 | 0.8838 | 0.4056 | 0.8915 | -0.4348*** |
| | -0.4501*** | | -0.4620*** | | -0.4858*** | | |
| normalised test 2 score | 0.6671 | -0.2869 | 0.3107 | -0.1371 | 0.1361 | -0.0445 | -0.3094*** |
| | 0.9540*** | | 0.4478*** | | 0.1807*** | | |
| normalised exam mark (first attempt) | 0.4400 | -0.3337 | 0.0968 | -0.3029 | -0.0752 | -0.1725 | -0.3608** |
| | 0.7738*** | | 0.3997*** | | 0.0972 | | |
| female | 0.4441 | 0.4446 | 0.4281 | 0.4161 | 0.4310 | 0.3850 | 0.0270 |
| | -0.0005 | | 0.0120 | | 0.0460 | | |
| Degree other | 0.4800 | 0.7563 | 0.5757 | 0.7346 | 0.6158 | 0.7181 | 0.0861 |
| | -0.2763*** | | -0.1588*** | | -0.1023** | | |
| BA (PPE/VPS) | 0.0439 | 0.0434 | 0.0457 | 0.0412 | 0.0508 | 0.0532 | -0.0322 |
| | 0.0005 | | 0.0046 | | -0.0023 | | |
| BAccounting | 0.3149 | 0.1219 | 0.2524 | 0.1350 | 0.2175 | 0.1383 | -0.0420 |
| | 0.1930*** | | 0.1174*** | | 0.0792** | | |
| BComm (Actuarial Science) | 0.0658 | 0.0033 | 0.0315 | 0.0046 | 0.0198 | 0.0053 | 0.0023 |
| | 0.0625*** | | 0.0270*** | | 0.0145 | | |
| BComm (law/maths) | 0.0754 | 0.0618 | 0.0741 | 0.0664 | 0.0791 | 0.0585 | 0.0075 |
| | 0.0136 | | 0.0078 | | 0.0206 | | |
| BComm (Economics) | 0.0200 | 0.0134 | 0.0205 | 0.0183 | 0.0169 | 0.0266 | -0.0218 |
| | 0.0067 | | 0.0022 | | -0.0096 | | |
| repeater | 0.0620 | 0.1269 | 0.0804 | 0.1373 | 0.1073 | 0.1489 | -0.0682 |
| | -0.0649*** | | -0.0569*** | | -0.0416 | | |
| White | 0.8613 | 0.7752 | 0.8307 | 0.7908 | 0.8247 | 0.7647 | 0.0971 |
| | 0.0861*** | | 0.0399 | | 0.0600 | | |
| Afrikaans | 0.4875 | 0.5638 | 0.4649 | 0.5678 | 0.4540 | 0.5775 | 0.0151 |
| | -0.0763*** | | -0.1030*** | | -0.1235** | | |
| English | 0.4123 | 0.3591 | 0.4313 | 0.3563 | 0.4339 | 0.3529 | 0.0136 |
| | 0.0533** | | 0.0750** | | 0.0810* | | |
| Age | 19.3015 | 19.3222 | 19.2939 | 19.3012 | 19.2730 | 19.3369 | -0.0205 |
| | -0.0206 | | -0.0072 | | -0.0639 | | |
| NSC | 0.1183 | 0.1269 | 0.1309 | 0.1190 | 0.1412 | 0.1223 | -0.0270 |
| | -0.0086 | | 0.0119 | | 0.0189 | | |
| normalised matric maths score | 1.9835 | 1.1501 | 1.7193 | 1.2307 | 1.6062 | 1.3338 | -0.0699 |
| | 0.8333*** | | 0.4886*** | | 0.2724*** | | |
| normalised matric maths score * NSC | 0.1126 | 0.0142 | 0.1164 | 0.0175 | 0.1271 | 0.0190 | 0.0844 |
| | 0.0983*** | | 0.0990*** | | 0.1081** | | |
| Gauteng ED | 0.0867 | 0.0807 | 0.0831 | 0.0737 | 0.0948 | 0.0538 | 0.0425 |
| | 0.0060 | | 0.0093 | | 0.0411 | | |
| OEB | 0.1638 | 0.1294 | 0.1629 | 0.1475 | 0.1724 | 0.1290 | 0.0348 |
| | 0.0344 | | 0.0155 | | 0.0434 | | |
| Eastern Cape ED | 0.0530 | 0.0454 | 0.0511 | 0.0507 | 0.0460 | 0.0591 | -0.0956** |
| | 0.0076 | | 0.0004 | | -0.0132 | | |
| Western Cape ED | 0.5588 | 0.6084 | 0.5623 | 0.5876 | 0.5460 | 0.6075 | 0.0120 |
| | -0.0496 | | -0.0253 | | -0.0615 | | |
| Observations | 1048 | 599 | 610 | 451 | 315 | 220 | 1061 |

**Notes:** difference in means in brackets. The final column is the estimated parameter on the non-compulsory indicator from a parametric regression discontinuity specification, only considering students that fall within 1 standard deviations of the compulsory cut-off of 50 percent in test 1.

with results based on alternative functional forms generated as robustness checks. We can similarly investigate whether or not a discontinuity in test and exam performance exists at the policy threshold. Figures A6 and A7 show similar scatter plots of average normalised test 2 and exam performance over the support of the normalised test 1 score. Students who performed just below 50 percent in the first test perform markedly higher in the second test and exam than those students who scored just above 50 percent. It is evident that there is a positive relationship between the performance in test 1 and subsequent performance throughout the semester. However, students who performed above the 50 percent in test 1 tend to perform worse in subsequent tests, excepting those who perform at the top of the distribution. The opposite is true for those students who performed below 50 percent in test 1. There therefore appears to be a degree of mean reversion in test 2 and the exam. As with tutorial attendance, different functional forms of the running variable are overlaid on the data. Inspection of the graphs prompted the use of a quadratic control function in the final model.

We now employ the parametric regression discontinuity specification from equations (8) and (9) to estimate the effect of the compulsory tutorial policy on tutorial attendance and exam performance. The samples under consideration are the group of students who score within one and half a standard deviation from the policy threshold. This allows for a better fit of the polynomial control function to the attendance rate over the threshold. As mentioned, linear and quadratic control functions are modelled for the first stage attendance and reduced form performance equations respectively. The results of these estimations are shown in table 2. Focusing first on the impact of the compulsory tutorial policy on tutorial attendance prior to the exam, we estimate that attendance for the compulsory student group is 32 percent higher at the threshold prior to the exam. This estimate is statistically significant at the 1 percent level. When the window is narrowed to 0.5 of a standard deviation, the result is largely unchanged. The inclusion of the other covariates in addition to the control function does not have much of an impact on the discontinuity coefficient when observing a window of 1 standard deviation. The instrumental variable results are presented in the final column of table 2. A two-stage regression approach yields an estimated coefficient on tutorial attendance of 1.05, which roughly translates to a 1.5 percentage point increase in exam performance for a 10 percent increase in tutorial attendance.

As stated, local polynomial regressions are used to estimate the local treatment effect. Estimates are generated using a triangular kernel function, as well as several choice of bandwidth, namely, the Imbens and Kalyanaraman (from now on referred to as the IK bandwidth) (2009) and the McCrary (2008) ROT bandwidths. Half and twice the IK and McCrary bandwidths are used for comparison. The results are presented in table 3 below. The optimal IK bandwidth is slightly smaller than the ROT bandwidth. However, the results yielded by the two bandwidth choices are quite similar. The ROT bandwidth predicts a significant increase in exam performance of 7.9 percent of a standard deviation for each additional tutorial attended, whilst the IK bandwidth yields an estimate of 10.3 percent of a standard deviation increase. Both are statistically significant at least at the 5 percent level. It is worth noting the difference in the two bandwidths upon which these estimates are based, as the narrower optimal IK bandwidth yields a larger estimated effect that is very similar to that obtained using parametric regression. This translates to a 1-1.5 percentage point increase in exam score for each additional tutorial attended. It is worthwhile comparing the magnitudes, statistical significance and standard errors on the estimate local treatment effect obtained under the different choices of bandwidths. It is immediately clear that the larger the bandwidth, the smaller is the estimated impact and the smaller the standard error. The contrary is true for smaller bandwidths. This is to be expected, given that a choice of larger bandwidth comes with greater precision. However, it also comes at the cost of greater bias in the estimates. Therefore, the estimates generated using the IK and ROT bandwidths may be downward biased. The following section tests the robustness of our results.

**Table 2:** Regression results for tutorial attendance and performance

| | Within 1 standard deviation | | | | |
|---|---|---|---|---|---|
| | First stage | | Reduced form | | IV |
| $D_i$ | -0.3172*** | -0.3179*** | -0.3531** | -0.3334** | |
| | (0.033) | (0.032) | (0.156) | (0.144) | |
| Attendance | | | | | 1.0503** |
| | | | | | (0.456) |
| $X_i$ | -0.0014 | -0.0005 | 0.0753** | 0.0566** | 0.0556** |
| | (0.002) | (0.002) | (0.030) | (0.028) | (0.027) |
| $X_i*D_i$ | 0.0050 | 0.0035 | -0.0377 | -0.0270 | -0.0230 |
| | (0.003) | (0.003) | (0.038) | (0.035) | (0.037) |
| $X_i^2$ | | | 0.0024 | 0.0018 | 0.0018 |
| | | | (0.002) | (0.001) | (0.001) |
| $X_i^2*D_i$ | | | -0.0026 | -0.0020 | -0.0020 |
| | | | (0.002) | (0.002) | (0.002) |
| Other controls | No | Yes | No | Yes | Yes |
| Observations | 947 | 937 | 947 | 937 | 937 |
| Adjusted $R^2$ | 0.191 | 0.315 | 0.094 | 0.214 | 0.214 |
| | Within 0.5 standard deviations | | | | |
| | First stage | | Reduced form | | IV |
| $D_i$ | -0.2927*** | -0.3182*** | -0.5135** | -0.3265 | |
| | (0.047) | (0.044) | (0.235) | (0.217) | |
| Attendance | | | | | 1.0258 |
| | | | | | (0.728) |
| $X_i$ | -0.0066 | -0.0027 | 0.1592 | 0.0718 | 0.0745 |
| | (0.005) | (0.005) | (0.103) | (0.098) | (0.099) |
| $X_i*D_i$ | 0.0102 | 0.0100 | -0.0878 | -0.0365 | -0.0468 |
| | (0.008) | (0.008) | (0.119) | (0.113) | (0.110) |
| $X_i^2$ | | | 0.0113 | 0.0037 | 0.0037 |
| | | | (0.010) | (0.010) | (0.010) |
| $X_i^2*D_i$ | | | -0.0164 | -0.0067 | -0.0067 |
| | | | (0.012) | (0.012) | (0.012) |
| Other controls | No | Yes | No | Yes | Yes |
| Observations | 491 | 484 | 491 | 484 | 484 |
| Adjusted $R^2$ | 0.204 | 0.338 | 0.013 | 0.145 | 0.145 |

*Notes*: *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors in parentheses. Standard errors of IV estimates generated from 500 bootstraps

**Table 3:** Non-parametric results

| | | IK bandwidth | ROT bandwidth | 0.5*IK bandwidth | 2*IK bandwidth | 0.5*ROT bandwidth | 2*ROT bandwidth |
|---|---|---|---|---|---|---|---|
| **Exam** | *bandwidth* | *0.879* | *1.679* | *0.440* | *1.758* | *0.840* | *3.360* |
| 1 | $E[T^+] – E[T^-]$ | -0.3149*** (0.041) | -0.3172*** (0.030) | -0.3200*** (0.052) | -0.3169*** (0.029) | -0.3141*** (0.041) | -0.3159*** (0.026) |
| 2 | $E[Y^+] – E[Y^-]$ | -0.3253** (0.134) | -0.2496*** (0.093) | -0.4298*** (0.165) | -0.2456*** (0.093) | -0.3313*** (0.118) | -0.2272*** (0.083) |
| 2/1 | LATE (Implied IV) | 1.0330** (0.434) | 0.7868*** (0.294) | 1.3430** (0.528) | 0.7751*** (0.293) | 1.0549*** (0.390) | 0.7190*** (0.261) |

*Notes*: *** p<0.01, ** p<0.05, * p<0.10. Bootstrapped standard errors generated from 500 bootstraps shown in parentheses.

## ROBUSTNESS CHECKS

One concern regarding identification of the treatment effect is that the compulsory policy may induce behavioural changes in effort. We may suspect that students who are subject to the policy are “labelled” as weak students. This may motivate compulsory tutorial students to exert more effort to better their performance relative to students who just passed and do not suffer the stigma of being a weak student. Therefore, the estimated impact of tutorial attendance may overstate the actual impact of the tutorials. One way of testing this assertion might be to analyse the behaviour of students in a subject which does not offer tutorial support. Unfortunately, such information was not readily available for this study. Alternatively, we propose to use the second test as a potential “treatment” by comparing the average exam outcomes of “just failers” and “just passers” in the second test amongst the group of compulsory students. If we find a negative estimate, this will indicate that compulsory students who scored below 50 percent in the second test performed better in the exam than compulsory students who scored above 50 percent. Due to the fact that both groups are required to attend tutorials on a compulsory basis, and therefore receive the same “treatment”, any divergence in exam performance may be ascribed to behavioural responses to “just failing” or “just passing”. We used local polynomial regression to compare the average exam performance of compulsory tutorial students who scored below 50 percent in the second semester test to the average exam performance of compulsory students who scored at least 50 percent or higher. Using the optimal IK bandwidth, we estimate a LATE of -0.231. This translates to approximately an exam performance that is 2 percentage points higher for the group of compulsory students who just failed test 2, indicating that there is potentially a stronger motivation for “just failers” to pass subsequent testing relative to “just passers”. However, this effect is not statistically significant.

Another issue is the potential bias in the LATE that could derive from a discontinuity in the covariates over the threshold. As mentioned, this can be tested by repeating the analysis using the residuals from a regression of the covariates (other than the control function) on performance. Alternatively, we can control for the covariates in estimation of the LATE. Both methods are employed here. We use the same optimal IK bandwidths for the regression corrected estimations as in table 3, and the results are shown in columns 2 and 3 of table 4 below. Correction for covariates reduces the estimated effect of tutorial attendance. This may suggest violation of the continuity assumption of one or more of the covariates. A visual inspection of local linear regression graphs for each covariate indicates no significant discontinuity in the covariates at the threshold (see figures A8-A26 of the appendix), except in the case of "White race group" where we find a significantly (10 percent level) higher density of non-compulsory students than compulsory students close to the cut off. However, this discontinuity disappears with a smaller bandwidth. The reduced LATE could also suggest a discontinuity in one ore more of the unobservables that may be related to the observable characteristics, such as ability and effort. Comparisons of the estimates from table 3 with the regression corrected estimates in table 4 indicate that the results are not statistically significant; therefore we can conclude that inclusion of the covariates in the non-parametric model results in consistent estimates of the LATE and has improved precision as evidenced by smaller standard errors.

Students were permitted to leave the compulsory tutorial programme if they were able to score at least 65 percent in the second semester test. As a result, 56 of the 599 compulsory students were no longer compelled to attend the tutorials. The results may be biased to the inclusion of this group of students as their behaviour may have been altered before test 2 (more motivated to leave the programme) and before the exam (refrained from attending tutorials on a regular basis). However, students were only made aware of their performance in test 2 in the $9^{th}$ week of tutorials, therefore only leaving 3 of the 5 remaining compulsory tutorials optional for this group of students. As a result, only 9 of these 56 students did not attend at least 4 of the 5 compulsory tutorials between test 2 and the exam. The LATE on the exam was re-estimated for two sub-samples of students: sample excluding all compulsory students who scored at least 65 percent in test 2; and a sample excluding only those compulsory students who scored at least 65 percent in test 2 and "left" the programme. The results of these estimations based on the same IK bandwidth from table 3 are shown in columns 4 and 5 of table 4. Excluding all students who achieved at least 65 percent in test 2 dramatically reduces the local treatment effect to 0.46. The effect is also non-significant. Excluding only those students who "left" the programme reduces the estimate slightly. This result may be of concern, as it suggests that the tutorials only had impact (so to speak) for a relatively small group of students; that is, those students who failed test 1, but performed well in test 2. The question then becomes: is this group of students different to the other compulsory students? Comparison of average observables indicates that this group of students tend to have a significantly higher proportion of students enrolled in accounting and actuarial science, as well as higher average performance in matric mathematics. This suggests that this group of students are most likely more able than the other compulsory students, and may have been more motivated to pass in future tests. The positive impact of tutorial attendance may therefore mask a change in behaviour that is policy driven. However, without other information with which we could test how student effort changes in response to this policy, it is difficult to say how much of the positive effect is due to motivational factors and that which is due to the tutorials. On the other hand, the fact that only 9 of the 56 students decided to exit the compulsory tutorial programme suggests that even these higher performing students attached value to being exposed to the compulsory tutorial programme.

Finally, the result may be sensitive to the choice of exam score used. The dependent variable includes exam scores from both the first and second exam papers. Compulsory students were no more likely to opt to write the second exam option than non-compulsory students. However, we may be concerned that the two papers were of different quality. Furthermore, students who wrote the second exam may have had access to the first exam paper, which may have benefited them. We therefore re-estimate the LATE excluding those students who only wrote the second exam option. The results are shown in the final column of table 4. Exclusion of this group of students has no significant effect on the predicted LATE.

**Table 4:** Sensitivity checks

| | | Compulsory students: test 2 as treatment | Regression corrected (residual) | Regression corrected (inclusion of covariates) | Excluding compulsory students who scored >=65% in test 2 | Excluding compulsory students who scored >=65% in test 2 & left programme | Excluding students who only wrote exam 2 |
|---|---|---|---|---|---|---|---|
| **Exam** | | | | | | | |
| 1 | $E[T^+] - E[T^-]$ | - | -0.3042*** | -0.3184*** | -0.3356*** | -0.3308*** | -0.3045*** |
| | | | (0.038) | (0.034) | (0.040) | (0.038) | (0.040) |
| 2 | $E[Y^+] - E[Y^-]$ | -0.2312 | -0.2657** | -0.3013** | -0.1538 | -0.2702** | -0.2885** |
| | | (0.211) | (0.121) | (0.116) | (0.126) | (0.119) | (0.135) |
| 2/1 | LATE (Implied IV) | - | 0.8735** | 0.9463** | 0.4585 | 0.8166** | 0.9464** |
| | | | (0.416) | (0.372) | (0.372) | (0.352) | (0.456) |

*Notes*: *** p<0.01, ** p<0.05, * p<0.10. Bootstrapped standard errors generated from 500 bootstraps shown in parentheses.

## CONCLUSION

The poor academic performance and retention of undergraduate students has prompted the adoption of alternative methods of learning and teaching that not only provide the necessary support to students and enhance their learning approaches, but are also cost-effective. The literature has provided mixed results regarding the impact of peer tutoring on the academic performance of undergraduate students (Topping, 1996). Although much of the existing research includes a cross-sectional component that typically compares the performance of students who have had tutoring versus those who have not, efforts have been made towards the adoption of quasi-experimental and random control designs that include both cross-sectional and longitudinal components that can control for potentially confounding factors or eliminate the sample specific biases that explain the observed effects.

This study aimed to contribute to the literature through the use of a fuzzy regression discontinuity design that potentially corrects for the issue of selection on unobservables that may bias the point estimates of tutorial attendance. The local average treatment effect is estimated using a bandwidth of observations around the policy threshold of 50 percent in the first semester test. It is clear that the policy significantly increases the tutorial attendance amongst compulsory tutorial students following the first semester test. IV regression results indicate a positive impact of tutorial attendance on test and exam performance. A 10 percent increase in tutorial attendance results in approximately a 10 percent standard deviation increase in test and exam performance. However, this result is only statistically significant in the case of the latter. Quantitatively similar impacts are found using local linear polynomial regression, although the results are sensitive to choice of bandwidth and specification of the control function. Robustness checks indicate that the results are fairly insensitive to the inclusion of the other covariates. However, the exclusion of the best performing compulsory students who were permitted to leave the programme decreases the treatment effect. This raises the concern that the result may be biased by unobservable factors such as motivation and effort that are not exogenous to the tutorial policy. Nevertheless, the fact that only 9 of 56 students took advantage of this exit option indicates that the students themselves attach value of attending these tutorials.

In conclusion, being assigned to the compulsory tutorial programme does affect performance but only for students that seem to have the ability to perform anyway. Unfortunately, this study is not able to unpack the mechanism through which assignment to the compulsory tutorial programme impacts on these students. The analysis would have benefited greatly through the inclusion of additional information, unavailable to the authors at the time of this study, regarding the performance of students in other coursework besides microeconomics where such interventions are not currently in place, as well as attitudinal and behavioural changes towards class attendance and time spent in studying. The longitudinal aspect of these types of programmes needs to also be considered, as the benefits of peer tutoring may only emerge at a later stage, or may even be short-lived. Differences between tutored and untutored students may either decline or increase over time depending on the adaption strategies of individual students (Jacobi, 1991).

## REFERENCES


BOUD, D., COHEN, R. and SAMPSON, J. (1999). Peer Learning and Assessment. *Assessment & Evaluation in Higher Education*, 24(4): 413-426.

ENTWISTLE, N.J., THOMPSON, S. and TAIT, H. (1992). *Guidelines for Promoting Effective Learning in Higher Education*. University of Edinburgh: Centre for Research on Learning and Instruction.

ENTWISTLE, N., McCUNE, V. and WALKER, P. (2001) Conceptions, styles and approaches within higher education: analytic abstractions and everyday experience, in: R. J. Sternberg & L. F. Zhang (Eds) *Perspectives on thinking, learning, and cognitive styles* (Mahwah, NJ, Lawrence Erlbaum Associates), 103–136.

FAN, J. and GIJBELS, I. (1995). Adaptive order polynomial fitting: bandwidth robustification and bias reduction. Journal of Computational and Graphical Statistics, 4: 213-227.

GOLDSCHMID, B. and GOLDSCHMID, M.L. (1976). Peer teaching in higher education: a review. *Higher Education,* 5: 9-33.

GREENE, W.H. (2008). *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

HAHN, J., TODD, P. and VAN DER KLAAUW, W. (2001). Identication and Estimation of Treatment Effects with a Regression-Discontinuity Design. Econometrica, 69: 201-209

HECKMAN, J. and ROBB, R. (1985). Alternative Methods for Evaluating the Impact of Interventions: An Overview. Journal of Econometrics, 30(1-2): 239-67.

IMBENS, G.W. and ANGRIST, J.D. (1994). Identification and Estimation of Local Average Treatment Effects. Econometrica, 62(2): 467-75.

IMBENS, G and KALYANARAMAN, K. (2010). Optimal bandwidth choice for the regression discontinuity estimator. CeMMAP working papers CWP05/10, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

IMBENS, G. and LEMIEUX, T. (2008). Regression discontinuity designs: a guide to practice. Journal of Econometrics, 142(2): 615-635.

JACOBI, M. (1991). Mentoring and undergraduate academic success: A literature review. *Review of Educational Research*, 61(4): 505-532.

JANSEN, A.I. and HORN, P.M. (2009). Tutorial classes – why bother? An investigation into the impact of tutorials on the performance of economics students. *South African Journal of Economics*, 77(1): 179-189.

JOHNSON, W.B. (2007). Student–faculty mentorship outcomes, pp. 189-210 in *Blackwell handbook of mentoring*, ed: T.D. Allen and L.T. Eby. Blackwell, Oxford.

KELLEY, A.C. and SWARTZ, C. (1976). Student-to-student tutoring in economics. *The Journal of Economic Education*, 8(1): 52-55.

LEE, R.E. (1988). Assessing retention program holding power effectiveness across smaller community colleges. *Journal of College Student Development,* 29(3): 255-262.

LEE, D.S. (2008). Randomized experiments from non-random selection in U.S. House elections. Journal of Econometrics, 142(2): 675-697.

LEE, D.S. and LEMIEUX, T. (2010). Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48(2): 281–355.

LEKSEKA, M and MAILE, S. (2008). High University dropout rates: a threat to South Africa's future.

Human Science Research Council, 1-7.
MCCRARY, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics, 142(2): 698-714.
MAXWELL, M. (1990). Does tutoring help? A look at the literature. *Review of Research in Developmental Education* 7(4): 3-7.
MUNLEY, V.G., GARVEY, E. and MCCONNELL, M.J. (2010). The effectiveness of peer tutoring on student achievement at the university level. *American Economic Review, Papers and Proceedings* 100: 277-282.
SCHMIDT, H.G. and MOUST, J.H.C. (1995). What makes a tutor effective? A structural-equations modelling approach to learning in problem-based curricula. *Academic Medicine*, 70(8): 708-714.
THISTLETHWAITE, D. and CAMPBELL, D. (1960). Regression-Discontinuity Analysis: An alternative to the ex post facto experiment. Journal of Educational Psychology , 51: 309-317.
TOPPING, K.J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education* 32: 321-345.
VAN DER KLAAUW, W. (2002). Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach. International Economic Review, 43(4): 1249-1287
WHITMAN, N.A. (1988). Peer Teaching: To Teach Is To Learn Twice. ASHE-ERIC Higher Education Report. Washington DC: ERIC Clearinghouse on Higher Education.

## APPENDIX

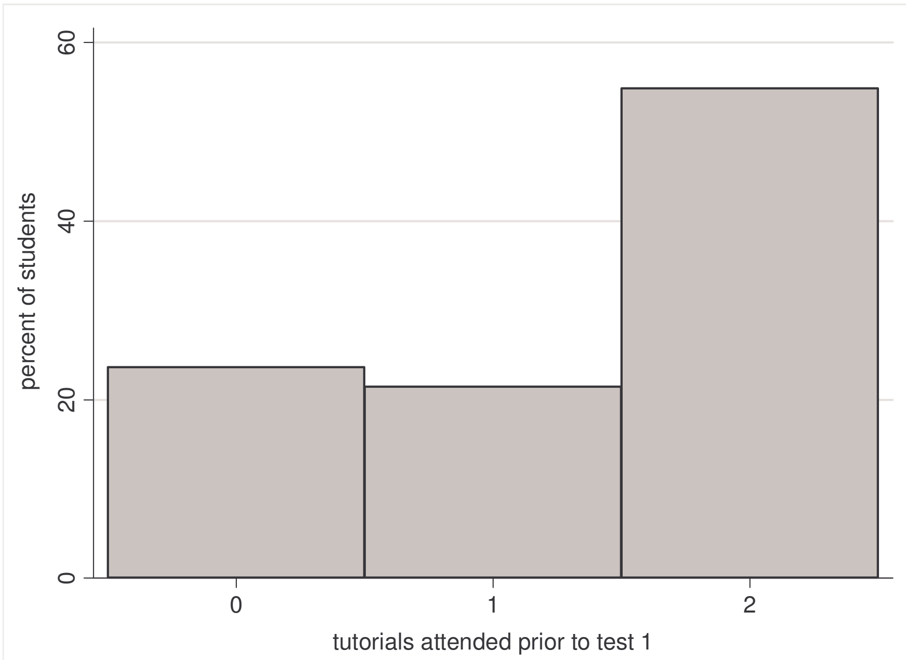



**Figure A1**: student attendance prior to test 1

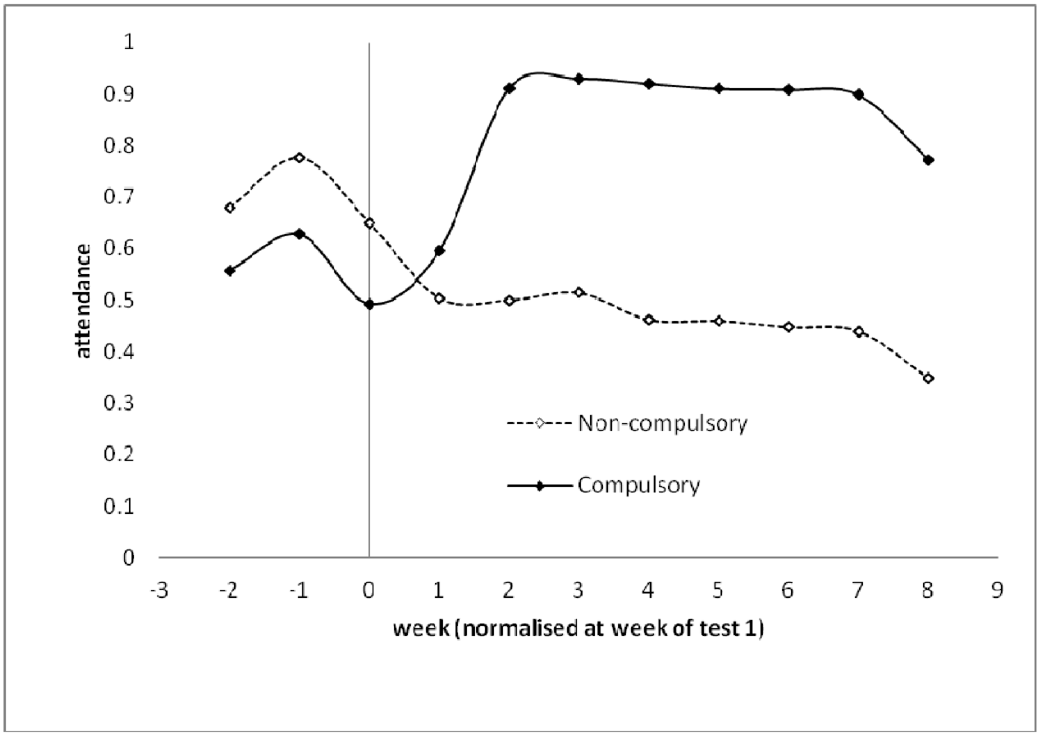



**Figure A2**: student attendance by treatment

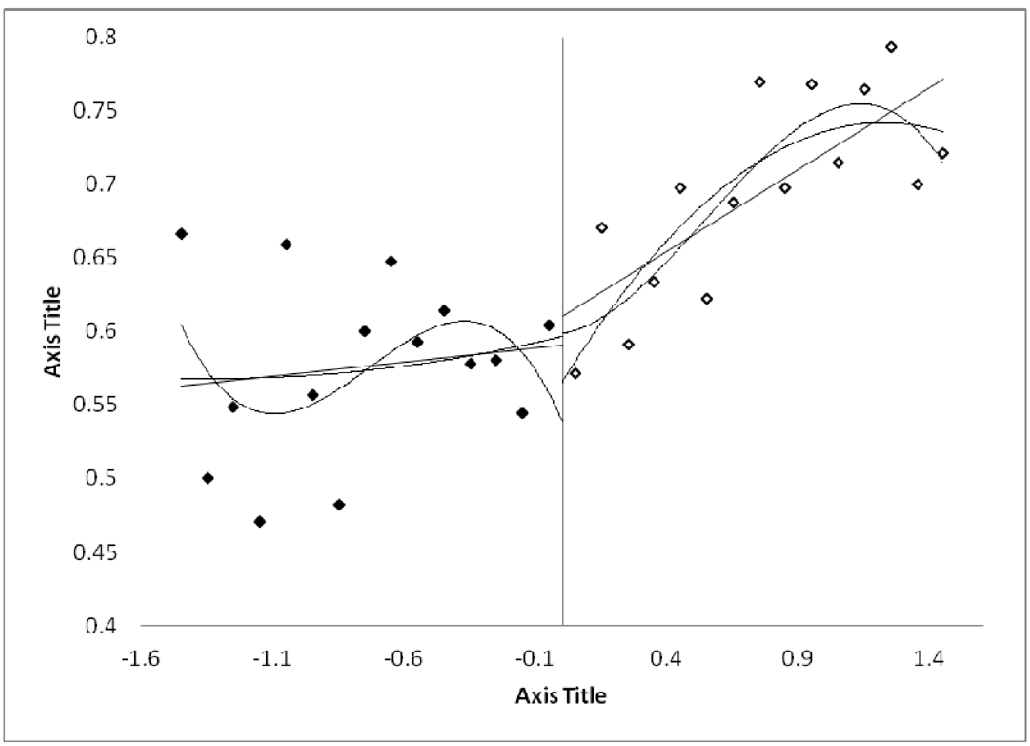



**Figure A3:** student attendance prior to test 1

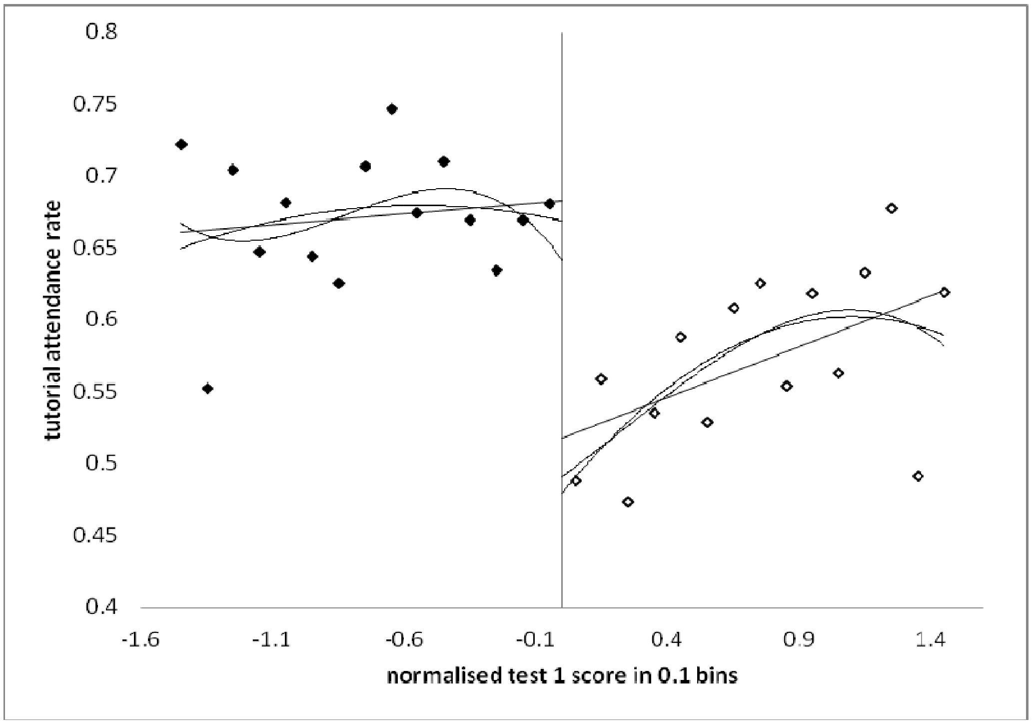



**Figure A4:** student attendance prior to test 2

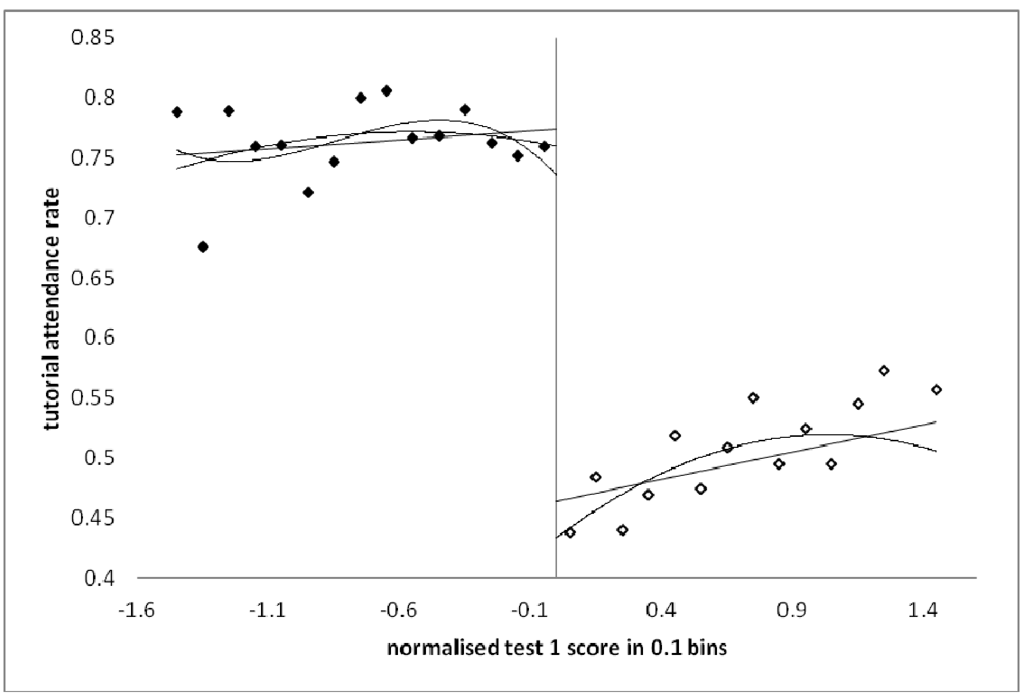



**Figure A5:** student attendance prior to exam

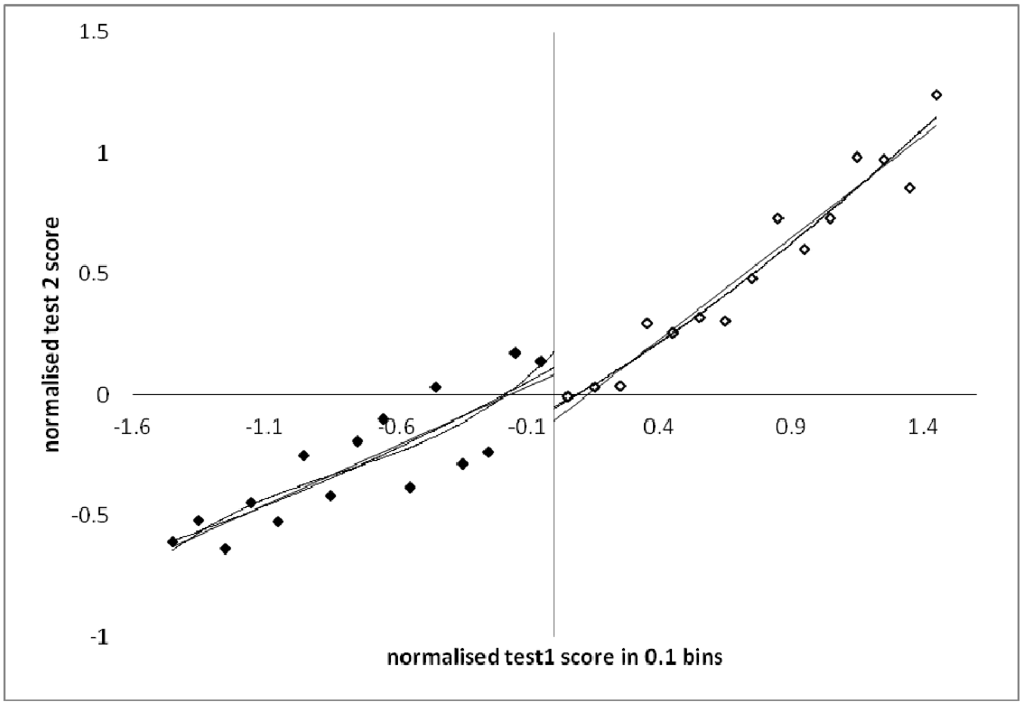



**Figure A6:** student performance in test 2

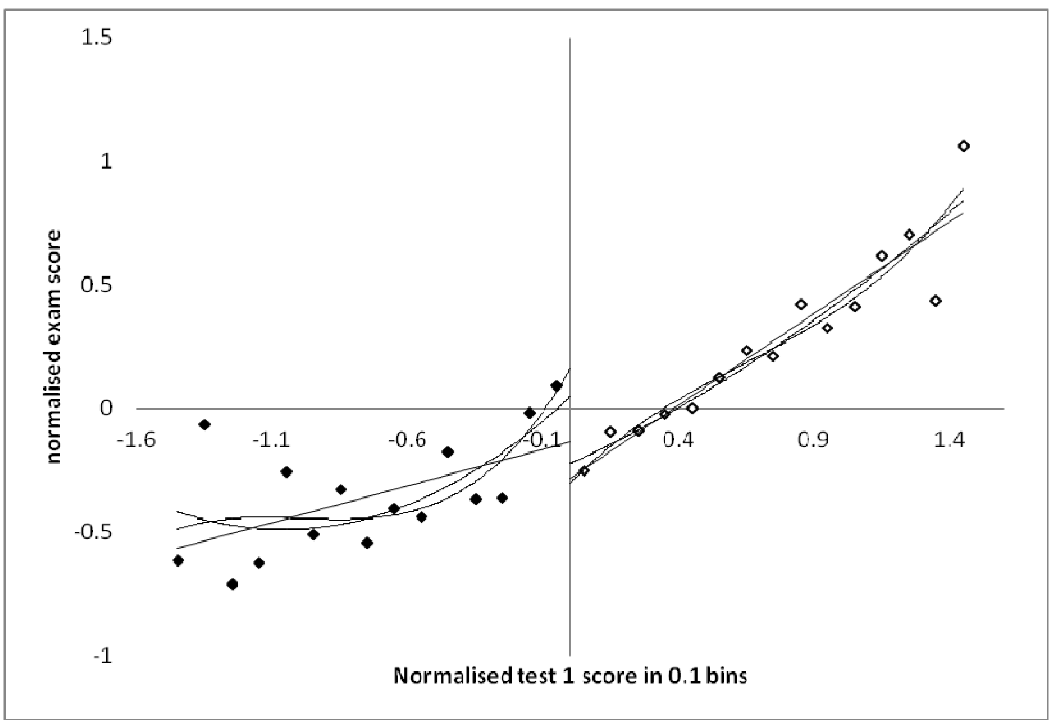



**Figure A7:** student performance in exam

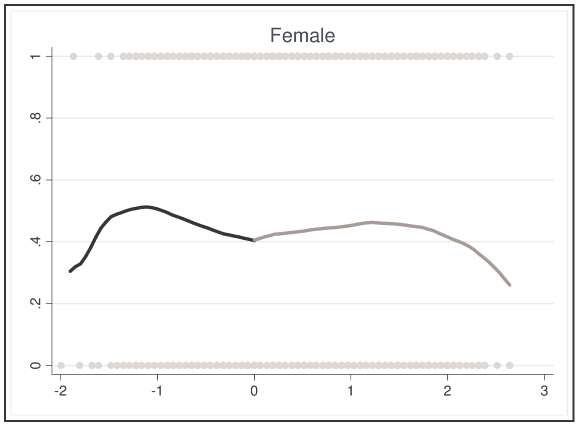



**Figure A8:** student performance in exam

**Figure A9:** student performance in exam

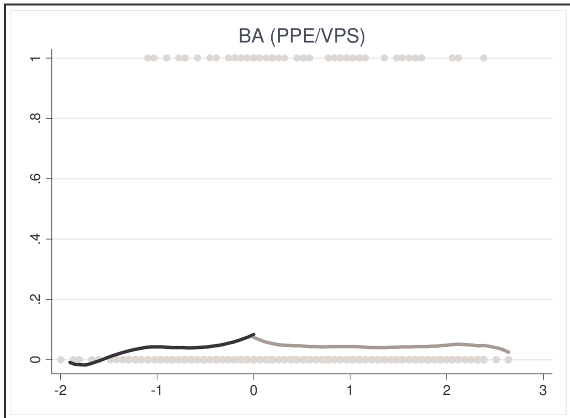



**Figure A10:** student performance in exam

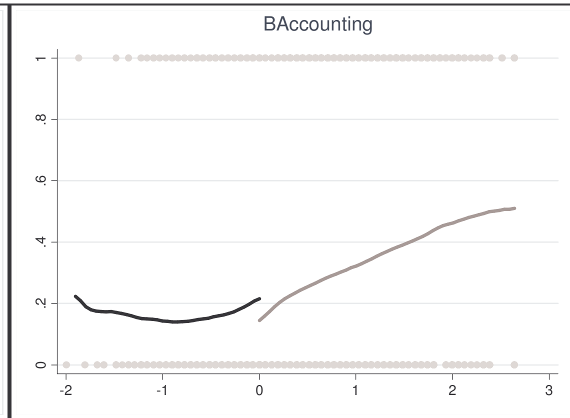
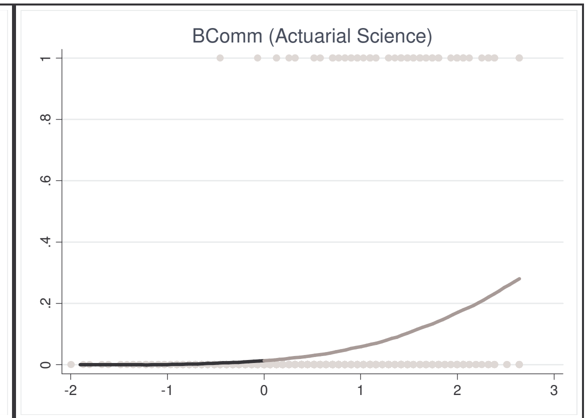



**Figure A11:** student performance in exam



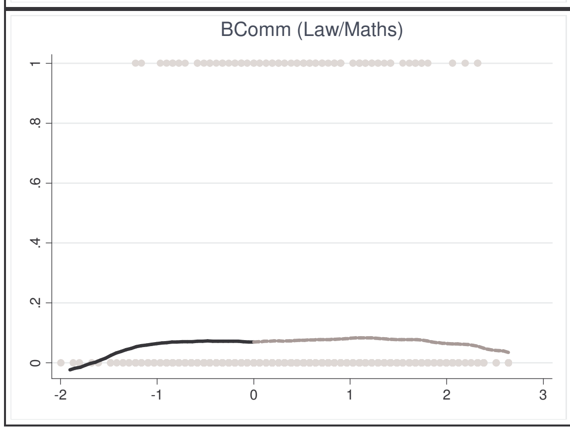




**Figure A12:** student performance in exam

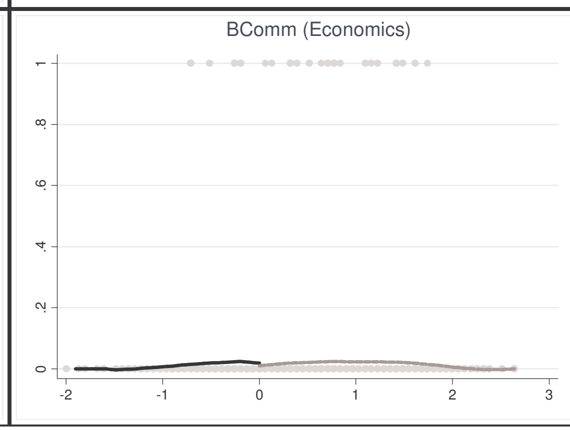



**Figure A13:** student performance in exam

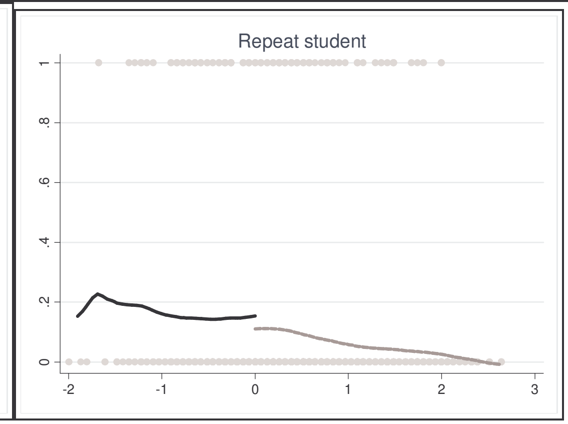
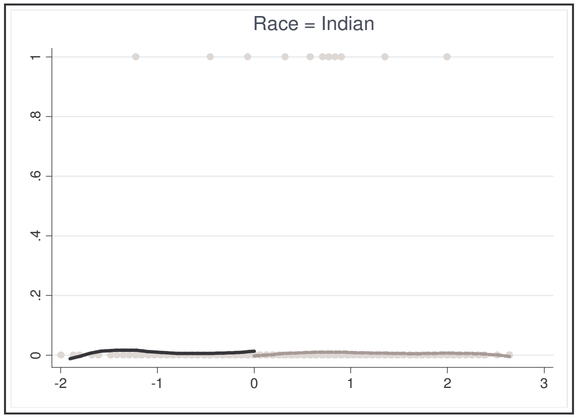


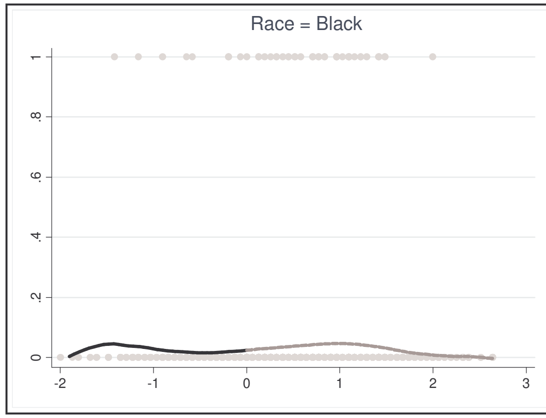

**Figure A14:** student performance in exam

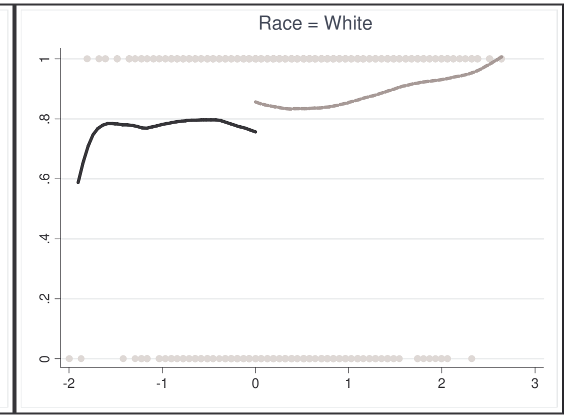
**Figure A15:** student performance in exam

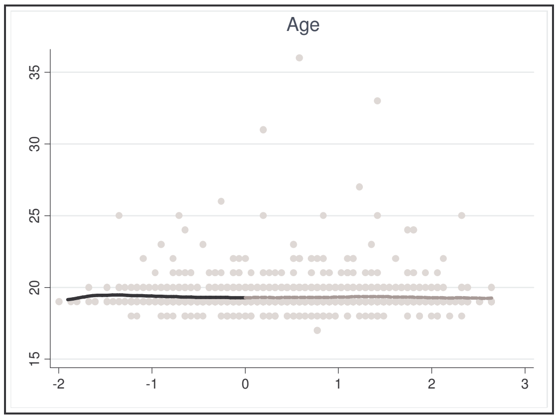
**Figure A16:** student performance in exam

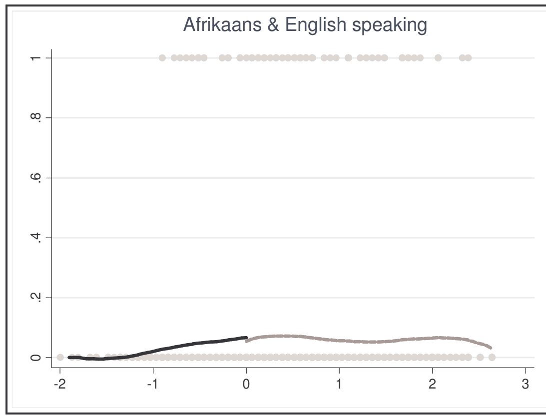
**Figure A17:** student performance in exam



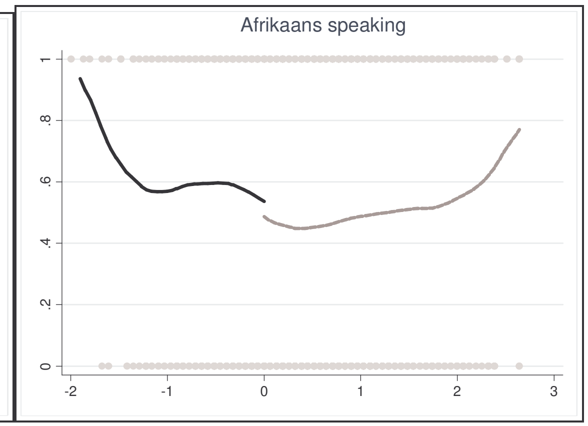




**Figure A18:** student performance in exam

**Figure A19:** student performance in exam

**Figure A20:** student performance in exam

**Figure A21:** student performance in exam

**Figure A22:** student performance in exam

**Figure A23:** student performance in exam

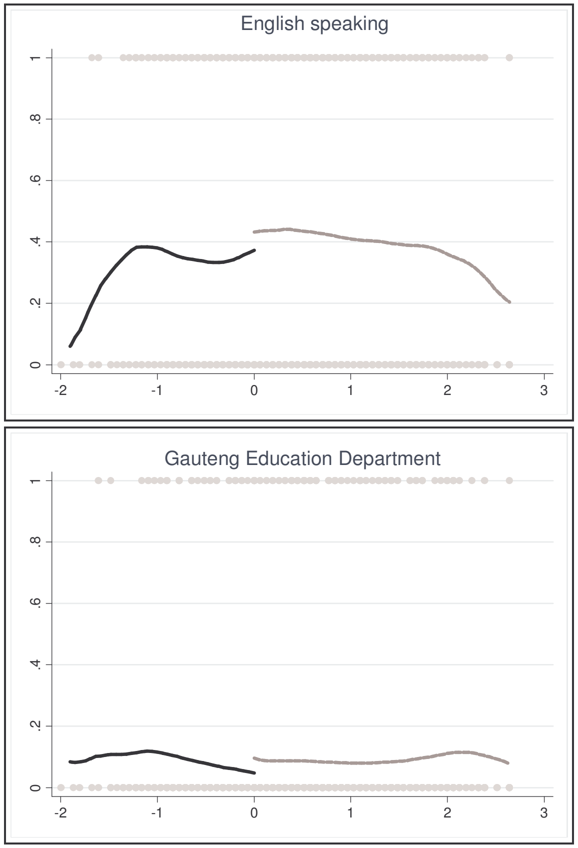



**Figure A24:** student performance in exam

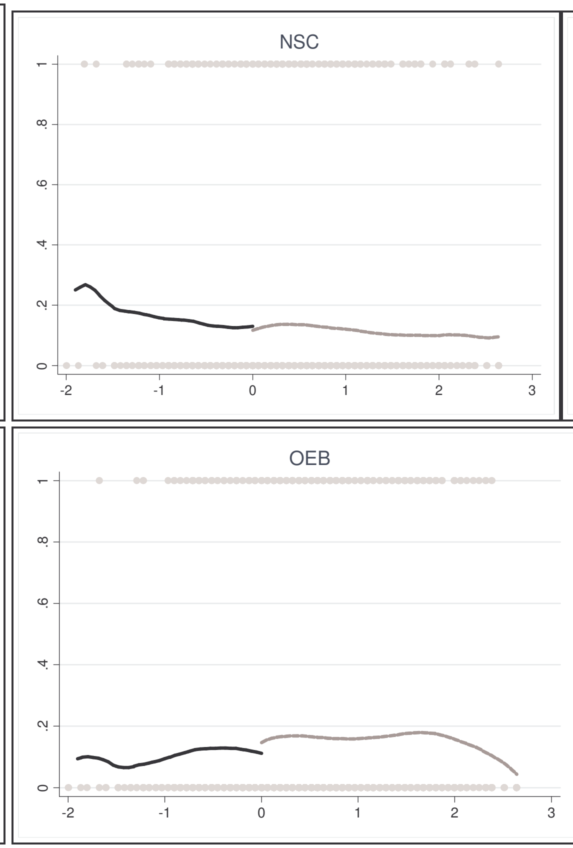



**Figure A25:** student performance in exam

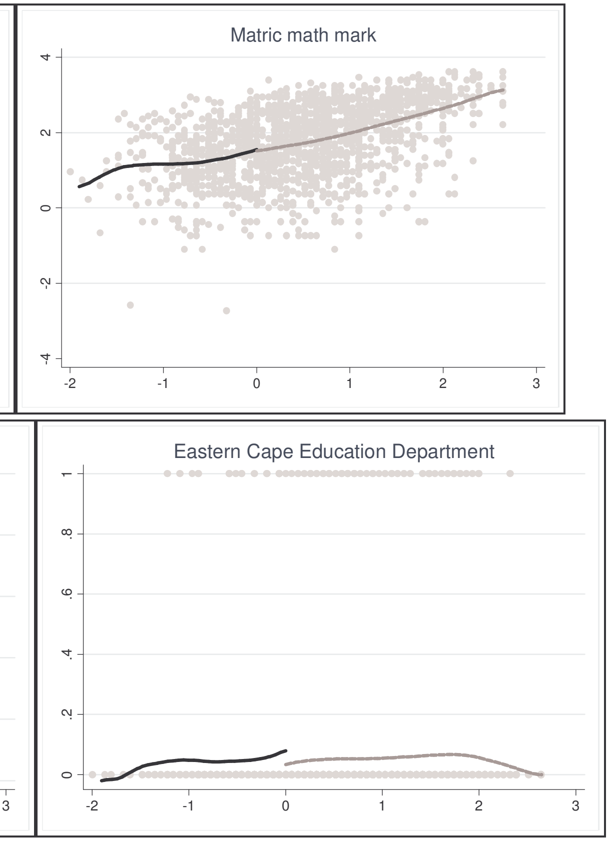



**Figure A26:** student performance in exam

**Table A1**: alternative specifications of the control function, test 2

| | First stage: tutorial attendance | | | | Reduced form: Test 2 performance | | | |
|---|---|---|---|---|---|---|---|---|
| | (1 SD) | | (0.5 SD) | | (1 SD) | | (0.5 SD) | |
| **Linear control function** | | | | | | | | |
| $D_i$ | -0.1673*** | -0.1789*** | -0.1328** | -0.1832*** | -0.1269 | -0.0912 | -0.2379* | -0.1627 |
| | (0.040) | (0.037) | (0.057) | (0.052) | (0.096) | (0.090) | (0.138) | (0.127) |
| $X_i$ | -0.0000 | 0.0010 | -0.0063 | -0.0004 | 0.0234*** | 0.0112 | 0.0451** | 0.0256 |
| | (0.003) | (0.003) | (0.007) | (0.007) | (0.007) | (0.007) | (0.020) | (0.019) |
| $D_i.X_i$ | 0.0061* | 0.0040 | 0.01247 | 0.0097 | 0.0146 | 0.0139* | -0.0059 | -0.0062 |
| | (0.004) | (0.003) | (0.010) | (0.010) | (0.009) | (0.008) | (0.025) | (0.023) |
| $R^2$ | 0.034 | 0.204 | 0.041 | 0.232 | 0.122 | 0.266 | 0.032 | 0.222 |
| **Quadratic control function** | | | | | | | | |
| $D_i$ | -0.1213** | -0.1491*** | -0.2054** | -0.2306*** | -0.2691* | -0.1887 | -0.4692** | -0.2511 |
| | (0.059) | (0.054) | (0.087) | (0.080) | (0.145) | (0.134) | (0.219) | (0.203) |
| $X_i$ | -0.0126 | -0.0010 | 0.0363 | 0.0248 | 0.0721** | 0.0499* | 0.2013** | 0.1002 |
| | (0.011) | (0.010) | (0.034) | (0.033) | (0.029) | (0.027) | (0.093) | (0.086) |
| $X_i^2$ | -0.0007 | -0.0006 | 0.0044 | 0.0026 | 0.0026* | 0.0021 | 0.0161* | 0.0077 |
| | (0.001) | (0.001) | (0.003) | (0.003) | (0.001) | (0.001) | (0.009) | (0.009) |
| $D_i.X_i$ | 0.0171 | 0.0171 | -0.0303 | -0.0115 | -0.0400 | -0.0346 | -0.1906* | -0.1118 |
| | (0.014) | (0.013) | (0.042) | (0.040) | (0.035) | (0.033) | (0.106) | (0.099) |
| $D_i.X_i^2$ | 0.0008 | 0.0005 | -0.0044 | -0.0030 | -0.0023 | -0.0015 | -0.0128 | -0.0041 |
| | (0.001) | (0.001) | (0.004) | (0.004) | (0.002) | (0.002) | (0.011) | (0.010) |
| $R^2$ | 0.035 | 0.205 | 0.043 | 0.233 | 0.125 | 0.268 | 0.038 | 0.224 |
| **Cubic control function** | | | | | | | | |
| $D_i$ | -0.1685** | -0.2446*** | -0.2008 | -0.3410** | -0.4114** | -0.2175 | 0.3447 | 0.3498 |
| | (0.080) | (0.072) | (0.148) | (0.137) | (0.197) | (0.181) | (0.392) | (0.359) |
| $X_i$ | 0.0156 | 0.0394 | 0.0115 | 0.1186 | 0.1697** | 0.0893 | -0.5087 | -0.4409 |
| | (0.028) | (0.026) | (0.119) | (0.114) | (0.074) | (0.069) | (0.318) | (0.293) |
| $X_i^2$ | 0.0029 | 0.0057* | -0.0015 | 0.0250 | 0.0150* | 0.0071 | -0.1531** | -0.1214* |
| | (0.003) | (0.003) | (0.027) | (0.026) | (0.009) | (0.008) | (0.072) | (0.066) |
| $X_i^3$ | 0.0001 | 0.0002** | -0.0004 | 0.0015 | 0.0004 | 0.0002 | -0.0113** | -0.0087** |
| | (0.000) | (0.000) | (0.002) | (0.002) | (0.000) | (0.000) | (0.005) | (0.004) |
| $D_i.X_i$ | -0.0124 | -0.0243 | 0.0414 | -0.0746 | -0.1582* | -0.1047 | 0.3387 | 0.3344 |
| | (0.036) | (0.033) | (0.132) | (0.127) | (0.085) | (0.079) | (0.339) | (0.311) |
| $D_i.X_i^2$ | -0.0027 | -0.0069* | -0.0128 | -0.0348 | -0.0118 | -0.0022 | 0.2116*** | 0.1539** |
| | (0.004) | (0.004) | (0.032) | (0.030) | (0.010) | (0.010) | (0.079) | (0.073) |
| $D_i. X_i^3$ | -0.0001 | -0.0002 | 0.0015 | -0.0008 | -0.0005 | -0.0003 | 0.0071 | 0.0065 |
| | (0.000) | (0.000) | (0.002) | (0.002) | (0.000) | (0.000) | (0.005) | (0.005) |
| $R^2$ | 0.029 | 0.188 | 0.032 | 0.196 | 0.126 | 0.268 | 0.053 | 0.193 |
| Other covariates | No | Yes | No | Yes | No | Yes | No | Yes |
| Observations | 1061 | 1050 | 535 | 528 | 1061 | 1050 | 535 | 528 |

*Notes*: *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors generated from 500 bootstraps shown in parentheses.

**Table A2**: alternative specifications of the control function, exam

| | First stage: tutorial attendance | | | | Reduced form: Test 2 performance | | | |
|---|---|---|---|---|---|---|---|---|
| | (1 SD) | | (0.5 SD) | | (1 SD) | | (0.5 SD) | |
| | Linear control function | | | | | | | |
| $D_i$ | -0.3172*** | -0.3179*** | -0.2732*** | -0.3123*** | -0.2095** | -0.2202** | -0.2769* | -0.2335* |
| | (0.033) | (0.032) | (0.048) | (0.045) | (0.106) | (0.099) | (0.148) | (0.139) |
| $X_i$ | -0.001418 | -0.0005169 | -0.007800 | -0.002723 | 0.03208*** | 0.02465*** | 0.05063** | 0.03619* |
| | (0.002) | (0.002) | (0.005) | (0.005) | (0.008) | (0.007) | (0.020) | (0.020) |
| $D_{i.}.X_i$ | 0.005016 | 0.003502 | 0.01211 | 0.01065 | 0.001894 | -0.0002600 | -0.02328 | -0.02849 |
| | (0.003) | (0.003) | (0.009) | (0.008) | (0.010) | (0.009) | (0.027) | (0.026) |
| $R^2$ | 0.193 | 0.330 | 0.209 | 0.367 | 0.097 | 0.233 | 0.019 | 0.185 |
| | Quadratic control function | | | | | | | |
| $D_i$ | -0.2923*** | -0.3050*** | -0.2771*** | -0.2922*** | -0.3531** | -0.3334** | -0.5135** | -0.3265 |
| | (0.049) | (0.046) | (0.069) | (0.066) | (0.156) | (0.144) | (0.235) | (0.217) |
| $X_i$ | -0.0099 | -0.0086 | -0.0031 | -0.0128 | 0.0753** | 0.0566** | 0.1592 | 0.0718 |
| | (0.007) | (0.007) | (0.023) | (0.024) | (0.030) | (0.028) | (0.103) | (0.098) |
| $X_i^2$ | -0.0005 | -0.0004 | 0.0005 | -0.0010 | 0.0024 | 0.0018 | 0.0113 | 0.0037 |
| | (0.000) | (0.000) | (0.002) | (0.002) | (0.002) | (0.001) | (0.010) | (0.010) |
| $D_{i.}.X_i$ | 0.0138 | 0.0157 | 0.0041 | 0.0184 | -0.0377 | -0.0270 | -0.0878 | -0.0365 |
| | (0.012) | (0.011) | (0.034) | (0.033) | (0.038) | (0.035) | (0.119) | (0.113) |
| $D_i.X_i^2$ | 0.0004 | 0.0002 | -0.0001 | 0.0013 | -0.0026 | -0.0020 | -0.0164 | -0.0069 |
| | (0.001) | (0.001) | (0.004) | (0.003) | (0.002) | (0.002) | (0.012) | (0.012) |
| $R^2$ | 0.194 | 0.331 | 0.209 | 0.367 | 0.099 | 0.234 | 0.022 | 0.186 |
| | Cubic control function | | | | | | | |
| $D_i$ | -0.2821*** | -0.3370*** | -0.3390*** | -0.4397*** | -0.4069* | -0.2777 | -0.1529 | -0.1855 |
| | (0.064) | (0.059) | (0.110) | (0.106) | (0.211) | (0.195) | (0.422) | (0.398) |
| $X_i$ | -0.0132 | 0.0096 | 0.0395 | 0.1175 | 0.1189 | 0.0613 | -0.1970 | -0.0947 |
| | (0.020) | (0.020) | (0.085) | (0.085) | (0.081) | (0.077) | (0.351) | (0.340) |
| $X_i^2$ | -0.0009 | 0.0019 | 0.0106 | 0.0300 | 0.0081 | 0.0024 | -0.0744 | -0.0363 |
| | (0.002) | (0.002) | (0.020) | (0.020) | (0.010) | (0.009) | (0.080) | (0.079) |
| $X_i^3$ | -0.0000 | 0.0001 | 0.0007 | 0.0021 | 0.0002 | 0.0000 | -0.0058 | -0.0027 |
| | (0.000) | (0.000) | (0.001) | (0.001) | (0.000) | (0.000) | (0.005) | (0.005) |
| $D_{i.}.X_i$ | 0.0136 | -0.0014 | 0.0013 | -0.0828 | -0.0965 | -0.0817 | 0.2770 | 0.2002 |
| | (0.029) | (0.028) | (0.102) | (0.100) | (0.096) | (0.090) | (0.375) | (0.360) |
| $D_i.X_i^2$ | 0.0014 | -0.0023 | -0.0224 | -0.0387 | -0.0061 | 0.0043 | 0.0667 | 0.0116 |
| | (0.004) | (0.003) | (0.025) | (0.024) | (0.012) | (0.011) | (0.088) | (0.087) |
| $D_i.\ X_i^3$ | -0.0000 | -0.0001 | 0.0002 | -0.0014 | -0.0003 | -0.0003 | 0.0060 | 0.0043 |
| | (0.000) | (0.000) | (0.002) | (0.002) | (0.000) | (0.000) | (0.006) | (0.006) |
| $R^2$ | 0.194 | 0.331 | 0.210 | 0.368 | 0.100 | 0.235 | 0.096 | 0.187 |
| Other covariates | No | Yes | No | Yes | No | Yes | No | Yes |
| Observations | 947 | 937 | 491 | 484 | 947 | 937 | 491 | 484 |

*Notes*: *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors generated from 500 bootstraps shown in parentheses.