# Exam Question: QM 2016 – Treatment Effects – Answer Key

Lucas and Mbiti study the effect of school quality on student achievement in Kenya ("*Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya*", American Economic Journal: Applied Economics, 2014). In particular, they are interested in quantifying the effect of attending an elite selective government secondary school (hereafter "elite school") on exam outcomes at the end of secondary school.

They collect data on two exam scores for each student who completes secondary school. These are the Kenyan Certificate of Primary Education (KCPE), and the Kenyan Certificate of Secondary Education (KCSE). Standardised scores for the KCPE are expressed on a 0-500 scale, while the KCSE score varies from 0-12. Finally, they observe a binary measure of whether each student attended an elite school, as well as the school specific cut-off for school entry, which depends upon the score achieved on the KCPE.

In this question we will *only* consider the effect of elite schools on KCSE exam scores, and not on whether they have an effect on rates of highschool graduation.

(a) Let $Y_i$ refer to each student's KCSE score, and $T_i$ be a binary variable which takes 1 if the student attends an elite school, and zero if they attend a non-elite school. Layout what parameter a naïve regression of the form:

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

would identify, and whether this is likely to a useful estimate of the average treatment effect on the treated (ATT). Cast this discussion in terms of the Rubin Causal Model (RCM) and discuss whether you can make any predictions about potential bias. [20%]

**Answer Key** An estimate of $\beta$ from the above regression will contain both the ATT of the impact of attending an elite school on the KCSE score, as well as any residual differences in students owing to selection into an elite school.

From the Rubin Causal Model, we can break this down as:

$$\begin{aligned}\underbrace{E[Y_i|T_i=1]-E[Y_i|T_i=0]}_{\text{Observed difference in average outcomes}} &= \underbrace{E[Y_{1i}|T_i=1]-E[Y_{0i}|T_i=1]}_{\text{average treatment effect on the treated}} \\ &+ \underbrace{E[Y_{0i}|T_i=1]-E[Y_{0i}|T_i=0]}_{\text{selection bias}},\end{aligned}$$

where the term on the left-hand side is what we are estimating in our naïve regression. This notation is based on the "potential outcomes" framework, which asks what would have happened to an individual $i$ if they were to receive treatment ($Y_{1i}$) versus what would happen to the same individual where they not to receive treatment ($Y_{0i}$). The RCM consists of the potential outcome framework, as well as an assignment mechanism, which determines how an individual is classified as a treated or non-treated unit.

From the above, we see that our estimate of $\beta$ will correspond to the ATT only in the case that $E[Y_{0i}|T_i = 1] = E[Y_{0i}|T_i = 0]$. In words, this states that our estimate will equal the ATT only if what would have happened to the treated had they not received treatment is the same (in expectation) as what happened to the untreated when they did not receive treatment.

In this case, it seems highly unlikely that the selection bias term will be equal to 0, and so highly unlikely that a naïve regression will adequately capture the ATT. This is because the assignment mechanism is related to individual performance, and certainly not randomised. In particular, given that the elite school selects students who perform higher on the KCPE, it is likely that, all else equal, these students will score higher on the KCSE whether or not they attended an elite school. In other words, it seems very likely that the selection bias is positive here, and a univariate regression will overestimate the effect of elite schools on KCSE scores.

NOTES: A good answer to the above question will lay out the two components to the estimated coefficient (ATT and selection bias), and may give further details about the RCM. Mentions and definitions of the "potential outcome framework" and "assignment mechanism" involved in the RCM are useful, though not a huge amount of detail is required on this. Responses should go into some depth about why the naïve regression result is likely to be biased. Any relevant explanation of unobservables or uncontrolled variables to explain the bias are acceptable. While it is not strictly necessary that the answer suggests that the selection bias term will be positive, an answer suggesting a negative bias must justify why we would think that people who did comparatively better on the KCPE would do comparatively worse on the KCSE.

(b) Assume that the researchers do not observe the school-specific cut-off for entry, but are able to measure a rich range of additional variables including child and family characteristics, and a range of results on additional pre-secondary tests. How could they use propensity score matching to estimate the ATT? Lay out any identifying assumptions you make, as well as how you estimate the effect using propensity score matching. [20%]

**Answer Key** A propensity score could be used to estimate the likelihood that each individual is treated based on their observable characteristics, and then the impact of the elite school can be calculated by matching each treated person with an untreated counterpart (or group of untreated counterparts) with a similar propensity score. We know from the propensity score theorem (Rosenbaum and Rubin, 1983), that if conditional unconfoundedness holds, then $W_i \perp\!\!\!\perp (Y_{0i}, Y_{1i})|p(X_i)$, and the propensity score is enough to capture all differences between treated and untreated units.

This relies on the following assumptions holding: (1) unconfoundedness: $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp W_i|X_i$, and (2) overlap $0 < \Pr[W_i = 1|X_i] < 1$. If the answer also states that SUTVA is needed, this is fine, though not strictly required.

Estimating with propensity score matching can be done by implementing the following formula:

$$ATT = \frac{1}{N_T} \sum_{i:w_i=1} \left( y_{1,i} - \sum_{j:w_j=0} \phi(i,j) y_{0,j} \right).$$

Thus, the ATT depends on the average difference between each treated unit's outcome $Y_{1,i}$ and a weighted average of matched outcomes $y_{0,j}$. Some discussion should be made of the nature of the weighting function $\phi(i,j)$. This can consist of nearest neighbour matching, in which case $\phi$ is equal to 1 for the nearest neighbour and zero for all others, by kernel matching, using Mahalanobis distance, etc.

NOTES: A good answer should layout what the propensity score does, and the assumptions required, as well as provide details on the estimation of the above formula. They may discuss that the propensity score should be estimated by logit or probit, though this is not required. It may also discuss what the assumptions mean in the context of this question, for example that conditional on family characteristics and past test scores, no other variables correlated with selection into secondary school and the KCSE scores exist, and that overlap requires that there is an untreated propensity score similar to the propensity score of the treated unit. An excellent answer will provide all this, and may include further discussion, such as pointing out that in practice standard errors will not be consistent if estimated by hand, so bootstrapping can be used. It may also point out that if all the individuals who are enrolled in elite schools have higher test scores than non-elite school enrollees, the overlap assumption may not be met.

(c) Briefly, in what circumstances will an estimated ATT and an estimate of the ATE coincide? Why might a policy maker be interested in the difference between the ATT and the ATE in this particular example? [10%]

The ATT: $E[Y_{1i} - Y_{0i}|T_i = 1]$ and the ATE $E[Y_{1i} - Y_{0i}]$ will coincide when those who receive treatment are, on average, representative of those who do not receive treatment. One particular case in which this will hold will be if the effect of treatment is heterogeneous in the population. Another case in which we may expect that ATE=ATT is when an RCT is randomly offered. No formal derivations are needed here, just the intuition, though a student will not lose marks if they derive this via a switching model.

In the case of an elite school program, the differences between an ATT and an ATE are very stark, as those who are treated are very different than those who are untreated. We (or a policy maker) will be interested in the ATT if they are only interested in assessing the effect of the schools on those students who actually attend them (for example, if they want to assess the job that the program is doing currently). On the other hand, an ATE is an estimand of interest for a policy maker interested in expanding access to the whole population of eligible individuals. For example, if a policy maker wanted to make the selective school model available to all individuals, then we are not interested in the ATT, but rather the ATE.

(d) Assume now that the school-specific cut-off score for entry $c_s$ is observed, and for a particular student, they are offered a place in an elite school if $KCPE_i > c_s$, where $KCPE_i$ is the student's score on the primary exam. Note that not all students who are offered a place will necessarily accept the offer. How would you estimate a regression-discontinuity in this case? Be clear to layout (i) how you would define the discontinuity and the running variable (ii) how you would estimate this parameter in practice, and (iii) what parameter of interest you are estimating. [20%]

(i) A discontinuity is based on whether or not the student scores above the eligibility threshold for the school in question: $1\{KCPE_i > c_s\}$. The running variable is then just the difference between the student's score, and the threshold of interest: $r_i = KCPE_i - c_s$.

(ii) Given that assignment is imperfect (not all students who have a score over the threshold necessarily accept the spot), we have a fuzzy regression discontinuity. This requires estimating the regression discontinuity using IV, where the discontinuity is used to instrument the actual decision of whether or not the person attended an elite secondary school. The RD estimate ($\tau$) thus consists of an IV (Wald) estimate of the following form:

$$\tau = \frac{\lim_{x \downarrow c_s} E(Y_i|KCPE_i > c_s) - \lim_{x \uparrow c_s} E(Y_i|KCPE_i < c_s)}{\lim_{x \downarrow c_s} E(T_i|KCPE_i > c_s) - \lim_{x \uparrow c_s} E(T_i|KCPE_i < c_s)}.$$

A very good answer will provide some discussion of estimation practicalities, including how to define the bandwidth, whether to use linear or quadratic polynomials locally in the bandwidth, and potentially flag the existence of the trade-off between increasing sample size and hence efficiency, and increasing the bias as the bandwidth chosen increases. This is a hard question, and if students respond on the mechanics of the regression discontinuity estimator but fail to note that it is fuzzy RD, then partial points should still be awarded.

(iii) The RD estimate gives a paramter $\tau(c_s)$, or an average treatment effect for those individuals local to the cutoff. A good answer will highlight this, while an excellent answer will also note that given that this is an IV estimate, it is a local average treatment effect which holds only for those people who are "compliers" with the cutoff (not never-takers or always-takers).

(e) Lucas and Mbiti report the results in table 1. Interpret these results, and suggest an explanation for why the OLS estimate exceeds the RD estimates. [15%]

Table 1: Estimated Effects: Elite Schools and KCSE Scores

| Dependent Variable: KCSE Score | OLS (1) | RD (2) | RD (3) |
|---|---|---|---|
| Elite School | 0.295*** | 0.014 | -0.015 |
| | (0.026) | (0.052) | (0.080) |
| Window | all pupils | +/- 34 | +/- 17 |
| Observations | 213,988 | 12,704 | 6,150 |
| R-Squared | 0.64 | 0.37 | 0.40 |

OLS refers to an OLS regression of the KCSE score on an indicator of whether the student attends an elite school. RD refers to a regression discontinuity estimate, where the discontinuity is defined based on the school-specific cutoff $c_s$. Window refers to the bandwidth used. The ideal bandwidth is calculated as 34 points on either side of the cutoff.

These parameters suggest a number of interpretations.

- The OLS parameter suggests that when combining both the effect of selection with the ATT (see part a of this question), elite schools have a positive correlation with final grades on the KCSE.
- However, the RD parameter with the ideal bandwidth shows that no significant effects are found, suggesting that the actual effect of the schools over and above selection is minimal. The difference in results seems to be entirely driven by student selection
- The difference between these parameters suggests that the nature of selection is very large. Although they don't seem to improve results, these schools choose students who score nearly 0.3 points higher on a 12 point test.
- A very good answer might point out that as the bandwidth increases, the RD estimate moves in the directin of the biased OLS estimate, though this change is not statistically significant

Various explanations of these results can be accepted, though the most likely seems like "cream skimming". If the elite schools choose the best students, then even if they add no value, a positive selection effect will remain. A good answer might also point out that students who study hard for the KCPE will also study hard for the KCSE, and have various other positive unobservables which are likely to be positively correlated with both selection and KCSE results.

(f) If you were presented with the above regression table, what additional robustness tests or empirical results would you request? Be careful to state what threat(s) to validity each of your requested tests or empirical results addresses. [15%]

Both the McCrary test and placebo tests replacing $Y_i$ in the regression with “exogenous” covariates should be requested. Students may also point out that additional results using alternative measures to control for the running variable (linear, quadratic, non-parametric) would be useful. An excellent answer should suggest that we will be interested in the “first stage”, where we examine whether achieving a score over the cut-off actually does increase the likelihood of going to an elite school. Given that the ideal bandwidth is discussed in table 1, it is not necessary to ask for optimal bandwidth calculations (such as those described by Imbens and Kalyanaraman), though students are not penalised if they do.

When describing these tests, it should be pointed out that the McCrary test and placebo tests are tests of the local unconfoundedness assumption. If these tests are not passed, it may suggest that there is manipulation of the running variable, or some other potential confounder not captured around the discontinuity. If the student requests additional tests using alternative specifications to control for the running variable, it can simply be stated that these are used to adequately capture the effect of the running variable as we move away from the cutoff. Finally, if students have mentioned a first stage test for the instrument, it should be stated that this is a test of instrumental validity.