Economics Working Paper Series

2022/002

Paying Students to Stay in School: Short- and Long-term Effects of a Conditional Cash Transfer in England

Andrew McKendrick

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

Paying Students to Stay in School:

Short- and Long-term Effects of a Conditional Cash Transfer in England

Andrew McKendrick*

Department of Economics, Lancaster University Management School

14 June 2021

Abstract:

I examine the impact of the Education Maintenance Allowance, a conditional cash transfer in England that was available nationally from 2004 to 2011, on a range of short- and long-term outcomes. Average treatment effects are identified, assuming unconfoundedness, using Inverse Probability Weighting Regression Adjustment. Treatment effect heterogeneity is examined using Causal Forests, a new machine learning approach. I find beneficial impacts of EMA on retention, university attendance and, for the first time, insecure work, as measured by the probability of being on a "zero hours" contract. Other outcomes (educational attainment, risky behaviours, and labour market outcomes) are found not to be impacted.

JEL Codes: H52, I12, I28, J22

Keywords: Education Maintenance Allowance, Causal Forest, Heterogeneity, Labour Market Outcomes, Job Security, Risky Behaviours

Acknowledgements: The author was supported by a PhD studentship from the Economic and Social Research Council (ESRC). I am grateful to my supervisors Ian Walker and Maria Navarro Paniagua, participants in seminars at Lancaster University, and to Charlotte Edney, Alexander Farnell, Emma Gorman, and Vincent O'Sullivan who gave helpful comments. I am grateful to the Centre for Longitudinal Studies (CLS) at UCL's Institute of Education, and to the UK Data Service (UKDS) for making available the *Next Steps* and NPD data. However, neither CLS nor the UKDS bear any responsibility for the analysis or interpretation of these data. Access to the data is controlled by the UKDS after registration, application, and training. Contact the author if you require help and guidance with this process, or to access the code used here.

* Email: a.mckendrick@lancaster.ac.uk

1. Introduction

The Education Maintenance Allowance (EMA) was a conditional cash transfer (CCT) that was available in the UK to those in post-compulsory full-time education or training. The level of entitlement depended on family income. EMA had been introduced in 2004 based on favourable evaluation of its pilot studies (Chowdry et al., 2008; Dearden et al., 2009; Middleton et al., 2004). The purpose was to encourage participation in post-compulsory education by pupils from low-income families. It was abolished in England in 2011 on the grounds that most students would have continued their studies after compulsory schooling ended anyway. That is, EMA represented a deadweight loss to the taxpayer (Bolton, 2011).[1]

Where CCTs have been implemented, they have generally been shown to incentivize particular activities. Some of the most notable examples are Brazil's *Bolza Escola* and Mexico's *Progresa* – both of which aimed to improve school attendance for low-income households, and both have been favourably evaluated (see for example Attanasio et al., 2012; Glewwe & Kassouf, 2012; Schultz, 2004)). Indeed, both schemes have developed into more extensive programs with wider objectives beyond education. But examples of CCTs in developed countries are relatively scarce. Given the enthusiasm for CCTs in developing countries, EMA is an important, almost unique, example for a developed economy.[2]

Unlike many CCTs, that often focus on younger students than EMA did, the money was paid directly to the young person, bypassing their parents. It was conditional on a single behaviour – attendance in education or training after age 16.[3] A student in receipt of the full EMA grant – £30 per week – would receive almost £1200 per year; for a family at the threshold of

[1] The compulsory leaving age at the time was 16, it is now 18.

[2] Another example, from the US, is the *Opportunity NYC: Family Rewards* program that was introduced from 2007 (see the MDRC evaluations by Greenberg et al. (2011); Riccio et al. (2010); Riccio & Miller (2016).

[3] Students were required to prove their family income was below the threshold for eligibility and to show that they had their own bank account to be credited.

eligibility for that entitlement (family income of £20,810) this would be an increase of roughly 5 percent in such a household's finances. As such, the potential impact might be expected to be large. Effects were broadly positive in the pilot studies – improving participation at ages 17 and 18. There appears to have been less of an impact on later outcomes (for example university attendance) and the estimated impacts on educational attainment were mixed. In its final full year in England the 643,000 young people who received EMA went to 32% of all 16–18-year-olds (or 47% of those in full-time post-compulsory education) at an annual cost of over £ ½ billion (Bolton, 2011).

The contribution of this paper is that it explores avenues for further exploration that previous EMA work has left undone. Firstly, by the time of its demise, EMA had existed for approaching a decade, with ample opportunity for the nature of effects identified (for example in Dearden et al (2009)) to have changed if, for example, some early success of EMA normalised participation in education after age 16. Secondly, the pilots were in specific areas and, whilst the evaluation analysis was convincingly implemented based on matched control areas, the effect, once the programme was rolled out nationally, may be different for general equilibrium reasons. Moreover, previous analysis looked at only a limited range of short-term outcomes. The CCT literature has increasingly focused on long-term outcomes as more time passes since these schemes were implemented.[4] Finally, previous work predated the widespread adoption of machine learning methods that offer considerably greater flexibility that traditional methods and allow sources of heterogeneous effects to be revealed that would have previously been overlooked.

The *Next Steps* dataset, used in this paper, follows a single cohort of young people in England, some of whom were eligible for EMA, and provides an opportunity to look at the EMA's

[4] See Millán et al. (2019b)

impact when it was no longer a novelty. Moreover, the rich nature of the data enables an analysis of the effect of the CCT on risky behaviours, where there may be potential unintended consequences of giving adolescents relatively large sums of money – for example, in the form of alcohol and cannabis consumption. The data also enables the examination of long-term labour market outcomes at age 25 – fully 8 years after first receipt of EMA.

In the absence of a quasi-experimental strategy, I proceed to identify effects under the assumption of unconfoundedness. In practice, this means that, after estimating linear specifications, I first estimate average effects using Inverse Probability Weighting Regression Adjustment (IPWRA). I then use Causal Forests, a machine learning approach that is new to economic analysis, to examine heterogeneity in treatment in a flexible and systematic way (Athey et al., 2019; Athey & Imbens, 2016; Breiman, 2001). This rapidly developing methodology has already seen use in economics; recently and prominently in Davis & Heller (2020). EMA is a good application for the method, especially as the original research based on the pilots employed fully interacted linear models and examined several dimensions of heterogeneity. Causal Forests can improve upon these traditional methods and in so doing help improve the external validity of programme evaluation. Knowing where the largest treatment effects are to be found can inform the design of future interventions.

Several statistically significant effects are identified. Results of a similar magnitude to the pilot studies (possibly slightly larger at around eight percentage points) are found on retention in full-time education and training. Positive impacts are also found on university attendance by age 25. Attainment and degree subject choice are not impacted. Indeed, neither are any other outcomes other than the probability of being on a zero hours contract at age 25 which is reduced for those on EMA at age 17. In reducing insecure work, EMA likely has a positive impact on welfare; whilst these contracts suit some, they do not suit all, and the opportunities provided by EMA might tip the balance away from the zero hours option.

The paper proceeds as follows. Section 2 covers the relevant literature, section 3 the background to EMA and the data I use, section 4 outlines the empirical strategy, whilst sections 5 and 6 present the results and then discuss and conclude.

2. Related Literature

Analysis of EMA's pilots (see Chowdry et al. (2008); Dearden et al. (2009); Middleton et al. (2004)) suggested that post 16 school participation might rise substantially. The Dearden et al. work examined the pilot schemes that took place in 1999/2000 in mostly more-deprived areas of England. It found improvements of around 4.5 percentage points to participation in the first year of post-16 education (year 12) and 6.7 percentage points in the second year of post-16 education (year 13). Along with EMA participation, the Department for Education report from the pilots also examine outcomes such as attendance at university and employment, though the effects are not found to be significantly different from zero Middleton et al. (2004).

The range of other outcomes examined has not been extensive, although Feinstein & Sabatés (2005) find reductions in crime in the EMA pilot areas when it was introduced. Nor have many studies examined the impacts of EMA outside of the pilot schemes once the scheme was rolled out nationally from 2004 onwards. One of the few examples is Holford (2015) that finds, using the same *Next Steps* dataset used in this paper, that EMA reduced the labour supply of EMA eligible teenagers by around 13 percentage points. Longer term effects have not been investigated.

Britton & Dearden (2015) conducted analysis into the 16-19 Bursary Fund that partially replaced EMA. This scheme began in 2011/12 and was targeted at the lowest income students and was mediated through schools and colleges who had discretion in who received it. Their difference-in-differences analysis uses those who were never eligible for EMA (i.e. had parental income that was too high) as the control group compared to a treatment group of those who would have been eligible but no longer received it. Participation in full-time education post age 16 fell under the new scheme – suggesting that EMA itself had a positive impact on participation.

Analysis of EMA fits into an established literature on the use of CCTs in education; although examples of CCTs in developed nations do exist, it is more common to see them used in developing countries. Developing country examples include Mexico (Attanasio et al., 2012; Schultz, 2004), Nicaragua (Gitter & Barham, 2009), and Brazil (Glewwe & Kassouf, 2012; Peruffo & Ferreira, 2017). Alternatively, cash transfers might be unconditional but are implicitly linked to education as in the Moroccan “Labelled Cash Transfer” in (Benhassine et al., 2015).

In developed nations there have been more limited use of CCTs. A CCT like EMA has been used in Australia and was shown to have increased post-compulsory schooling and attendance at university (Dearden & Heath, 1996). In contrast, a small scheme in New York City yielded modest effects at best, (Greenberg et al., 2011; Riccio et al., 2010; Riccio & Miller, 2016). In the US state of Georgia, student aid conditional on attainment has been used successfully to promote higher education participation (Dynarski, 2003). In contrast, in Denmark Humlum & Vejlin (2013), the results of promoting higher education participation through a CCT proved to have no statistically significant effect despite very precisely estimated parameters.

Generally, studies relating to conditional cash transfers in education have found positive impacts on attendance, but mixed results for attainment (Fiszbein & Schady, 2009). This indicates that CCTs improve what they are conditioned on – but not necessarily on other outcomes. In the case of EMA, which is conditional on school attendance, we may not expect to see improvements in other outcomes. This is the case in the Middleton et al. (2004) work on the EMA pilots where, although full-time education attendance improved, later university attendance did not change, and attainment at A-level did not either (although GCSE attainment did slightly). Although, as attendance has improved there will be those who now have post-compulsory schooling qualifications who would not have had them before. As time passes, the analysis of longer-term effects becomes possible (see Millán et al. (2019a) – a literature that my results also speak to).

Across CCT studies there is a degree of heterogeneity in who the transfer targets (i.e., who is eligible) and who physically receives the payment (the young person being educated or their parents). Often programmes are aimed at young people still in compulsory education and the money goes to their mother, as in PROGRESA in Mexico or, indeed, one arm of the EMA pilots. The money could be paid to the father instead, as in the Moroccan case in Behassine et al (2015). There is less heterogeneity around the conditions for continuing to receive the CCT. Morais de Sa e Silva (2015) finds that 80% of the 43 CCTs surveyed required continued attendance to keep receiving the transfer.

3. Background and Data

EMA was rolled out nationally in 2004 having been piloted from 1999. 55 local authority areas were included in the pilot scheme across two different waves. The choice of pilot areas was not random; they were generally more deprived (Fletcher, 2000). The aim of the policy was to encourage people to stay in education after age 16 – which was then the compulsory schooling age. Cash payments were made directly to pupils (rather than their parents as in the case of many CCTs across the world and one of the EMA pilots) during school term. Amounts varied with household income – students could receive £10, £20, or £30 per week depending on their household income being below £30,810, £25,522, or £20,818 respectively, contingent on remaining in full-time education or training.[5] The relatively high threshold for eligibility means that around 32% of all 16-18-year-olds, and close to half of those 16-18-year-olds in education,

[5] These thresholds were not adjusted over time.

received some amount of EMA. Consequently, at its 2009-10 peak, the annual cost of EMA was £580 million.

Receipt was based on continued attendance in full-time education, with eligible courses being those studied in a further education college or some apprenticeships.[6] Anecdotally, it seems that this was not monitored with closely. It was simple and easy to prove eligibility – all that was needed was a statement of family income, and proof of young person having their own bank account.[7] The young person could keep working part-time and their earnings would not impact entitlement for EMA. Similarly, them receiving EMA would not impact their parents benefit entitlement. Although the guidelines were set out at a national level, the administration was performed by schools and colleges.[8] When EMA ended in England in 2011 it continued in Scotland, Wales, and Northern Ireland.

The *Next Steps* Dataset

The data employed are from *Next Steps* (UCL, 2021b), a cohort study from England, also known as the Longitudinal Study of Young People in England (LSYPE1). The study began in 2004, randomly sampling 650 schools before randomly sampling around 30 Year 9 (age 13-14) pupils from each. These individuals were then resurveyed each year across seven waves until they were age 20. A further, eighth wave, was undertaken in 2015 when respondents were 25. The study is similar in character to the well-known US National Longitudinal Survey of

[6] By "some" I mean – Learning Skills Council (LSC) funded Entry to Employment (E2E) courses, which are a work-based learning route that some might do prior to an apprenticeship, or Programme-led apprenticeships (PLAs) that are largely classroom based – 15 hours week was needed in the classroom for these apprenticeships to qualify.

[7] The EMA website (now archived) is accessible here.

[8] This is different to the new 16-19 Bursary, the less generous successor to EMA (analysed by Brittan and Dearden (2015)), the distribution of which is essentially at the discretion of colleges with minimal national guidance.

Youth (NLSY), albeit for a single cohort, in that it contains a very detailed array of characteristics.

These data are linked to the National Pupil Database (NPD) (UCL, 2021a) – the administrative dataset for education in England. This means that prior attainment, at primary school (known as Key Stage 2 or KS2 at age 11) and secondary school (GCSE point score at age 16), can be controlled for and attainment effects of EMA (on A-level achievement, at age 16) can be estimated. The NPD also contains information on the lower super output area (LSOA) that the individual lived in at age 14 – this means post-code (zip-code) level data on deprivation can be matched into the data. Moreover, the precise local authority that the individual lived in at age 14 is known, too, which means those who lived in an EMA pilot area can be identified. It is plausible that pilot areas may have become more adept at distributing EMA as they have more experience of it, so it is important to control for this.

My sample ultimately contains only those who were eligible for EMA. This mirrors the previous work done on the pilot studies. Eligibility is determined by household income the year prior to attending college as that is likely the income statement used to prove eligibility. This yields a maximum possible sample of 4,859, of whom 66.1 percent receive EMA.

Individuals leave school at age 16. This coincides with Wave 3 of *Next Steps* (in 2006). Wave 4, when individuals are 17, is therefore the first year that individuals can receive EMA. Wave 4 is the first wave from which outcomes appear in this paper. These are risky behaviours: frequency of drinking alcohol and whether the individual has ever tried cannabis. Wave 5 provides information on the main activity of the young person – the outcome that is constructed takes value 1 if the individual is in full-time education or training.[9] This is the measure of retention. Wave 5 gives the attainment measure – UCAS score. This is the grades achieved in

[9] Including some apprenticeships as detailed in footnote 6.

the top 3 A-levels converted into a point score that is used by universities when judging applications. Waves 6 to 8 yield information on university attendance and degree subject choice, where we are interested to see if the latter is impacted if a group of young people from disadvantaged backgrounds choose potentially more lucrative degrees in terms of future earnings. Finally Wave 8 allows examination of labour market outcomes – these are earnings, hours worked, whether the individual has ever been employed (a measure of long-term unemployment), and whether they are currently working in insecure work, defined as working on a zero hours contract – a type of employment contract in the UK that does not guarantee a minimum number of hours of work in any given week. These outcomes are measured when the individual is age 25 and are shown in Table 1, where they are broken down by receipt of EMA.

Controls come from Wave 4 and earlier. The full list of controls is given in Table 2, again broken down by EMA receipt. The second column shows the age at which the responses were collected. These covariates span across personal characteristics such as gender and ethnicity, to household characteristics like parental employment status, and highest educational attainment. A number of these variables are included specifically to account for selection into EMA. Prior knowledge of EMA is important – being able to apply for something is helped along substantially by knowing about it; it may also account for more driven individuals who have been planning to go to college for some time prior to attending. Local authorities that were part of the EMA pilots may be more experienced in advertising and administrating the grant and so an individual living in those areas may have higher likelihood of take up. Similarly, EMA enrolment in your area will account for peer effects that make individuals who have friends who are applying more likely to apply. Other control variables like parental employment and education will also impact (likely positively) selection into EMA. The variables are largely binary (such as free school meal status) or have relatively small numbers of responses (such as general health or English region of residence). This is for purposes of overlap – using continuous variables is not conducive to good overlap.

In both Tables 1 and 2 it is evident that differences exist between those who are on EMA and those who are not. For the outcomes, in Table 1, part of this may be impacts of the CCT – this would indicate large potential impacts of EMA, such as in the case of University Attendance. In the control variables, all of which are pre-treatment, these reflect differences in the type of person who applies and ultimately receives EMA. Naturally, these differences will likely extend to unobserved characteristics as well. Insofar as these unobserved characteristics matter for selection into treatment and for outcomes their impact will be lessened if overlap is good, assuming that unobserved variables follow a similar distribution to those which are included.

As discussed below, good overlap is an important assumption for methods based on unconfoundedness, but in selecting the list of controls that yields good overlap concerns may arise of choosing controls that give the best overlap. I combat this by beginning the analysis by judging which covariates deliver the best overlap whilst using a randomly generated outcome variable. This enables me to judge overlap without knowing what the effect will be. Figures 1 and 2 show that good overlap and covariate balance are achieved. Figure 2 is generated from a propensity score matching estimation using the Stata command teffects, but the modelling of the first stage is analogous to IPWRA. The chart is intended to show how effective matching methods appear to be in this application – and this is shown in Figure 2 where the right and left graphs are visually identical. This same list of covariates is then used in the full OLS specification, too. Balance tables are not shown but are available on request. Population weights are used throughout.

Table 1 Outcomes by EMA Receipt

Variable	Age	Does Not Receive EMA N	Mean	Std Dev	Receives EMA N	Mean	Std Dev
Retention	18	1,408	0.551	0.498	2,916	0.700	0.458
UCAS Score	18	643	-0.029	1.036	2,098	0.029	0.990
University Attendance	25	1,412	0.331	0.471	2,881	0.519	0.500
STEM Degree Subject	25	435	0.379	0.486	1,382	0.413	0.493
Alcohol	17	1,289	0.081	1.000	2,086	-0.073	0.99
Cannabis (Ever Tried)	17	1,619	0.368	0.482	3,139	0.297	0.457
Earnings	25	600	5.827	0.615	1,293	5.812	0.718
Ever Employed	25	847	0.953	0.212	1,864	0.922	0.268
Hours	25	695	38.193	11.284	1,479	36.686	10.724
Zero Hours	25	698	0.079	0.27	1,476	0.065	0.247

Note: Retention is a dummy variable taking value 1 if the young person is still in full-time education (FTE) in Wave 5. UCAS Score is the score attached to the top 3 grades an individual achieved and is standardised to be mean 0. University is a dummy variable taking value 1 if the individual attends university by age 25. STEM degree subject takes value 1 if an individual is studying a STEM degree, and 0 otherwise. Alcohol is frequency of alcohol consumption, standardised. Cannabis consumption is a dummy variable taking value 1 if the individual has ever tried cannabis. Earnings is log employment earnings from all jobs. Ever employed is a binary variable that is 1 if the individual has had any job by age 25, and so is a measure of long-term unemployment in some sense. Hours worked is hours in one's main job. And finally, Zero Hours is a binary variable that is 1 if the individual is on an insecure labour contract that does not guarantee a minimum number of hours in a given week. These are unweighted figures, though survey weights are used in analysis below.

Table 2 Covariates by EMA Receipt

		Does Not Receive EMA			Receives EMA		
Variable	Age	N	Mean	Std Dev	N	Mean	Std Dev
Gender	14	1,623	0.455	0.498	3,179	0.534	0.499
Non-White	14	1,645	0.249	0.433	3,208	0.405	0.491
Special Educational Need	14	1,629	0.209	0.407	3,166	0.173	0.378
On Free School Meals	14	1,604	0.129	0.335	3,155	0.273	0.446
Quartile of KS2 score	11	1,551	2.526	1.100	3,065	2.586	1.124
Quartile of GCSE Points	16	1,620	2.423	1.087	3,180	2.71	1.093
Pilot Area	14	1,648	0.333	0.471	3,211	0.455	0.498
EMA Enrollment in LA	17	1,648	0.416	0.141	3,211	0.474	0.153
Aware of EMA	15	1,613	0.626	0.484	3,178	0.702	0.457
General Health	17	1,638	3.408	0.677	3,170	3.422	0.659
Main Parent Age	17	1,580	44.87	6.032	3,149	45.215	6.643
Single Parent	17	1,571	0.105	0.307	3,138	0.15	0.357
Step Family	17	1,631	0.128	0.334	3,192	0.089	0.284
Main Parent Employment	17	1,584	1.753	0.816	3,172	2.059	0.859
Main Parent Education	17	1,586	3.788	1.742	3,163	4.294	1.935
No. Dep Children in HH	17	1,631	1.505	0.969	3,195	1.897	0.848
Housing Tenure	17	1,588	1.157	0.641	3,142	1.25	0.717
IMD in Year 11	16	1,648	24.89	16.59	3,208	31.142	18.367
Urban or Rural	16	1,648	3.706	0.739	3,209	3.771	0.669
Region	16	1,648	5.345	2.444	3,208	5.056	2.414

Note: Gender, Non-white, Special Educational Needs, and Free School Meals are all binary variables that are 1 if the individual is female, not white, has special educational needs or is on free school meals, respectively. KS2 score is the test taken at the end of primary (junior) school, GCSEs are high-stakes test taken at the end of secondary (high) school. Pilot Area is a dummy variable that is 1 if the individual lives in an area that was an EMA pilot area when EMA was piloted in the late 1990s. Aware of EMA is a binary variable that takes value 1 if the individual had heard of EMA in the penultimate year of secondary (high) school – that is prior to applying for EMA. General Health is a 4-point scale describing one's health in the last 12 months, ranging from "not at all good" to "very good". Main parent age is the age of the parent who responded to the parental section of the survey' single parent is a dummy that is 1 if the individual lives in a single-parent household; Main Parent Employment and Education are the parent's occupation and highest educational qualification; Number of dependent children in the household is a count of the number of people aged under 19 who live with the young person; Housing Tenure indicates whether the house the young person lives in is owned or rented; IMD is the index of multiple deprivation, a composite of six measures (outlined further in the data section of chapter 1); Urban or Rural is over several levels from small village to large city; and finally, region is the NUTS1 statistical regions of England.

Figure 1 Overlap

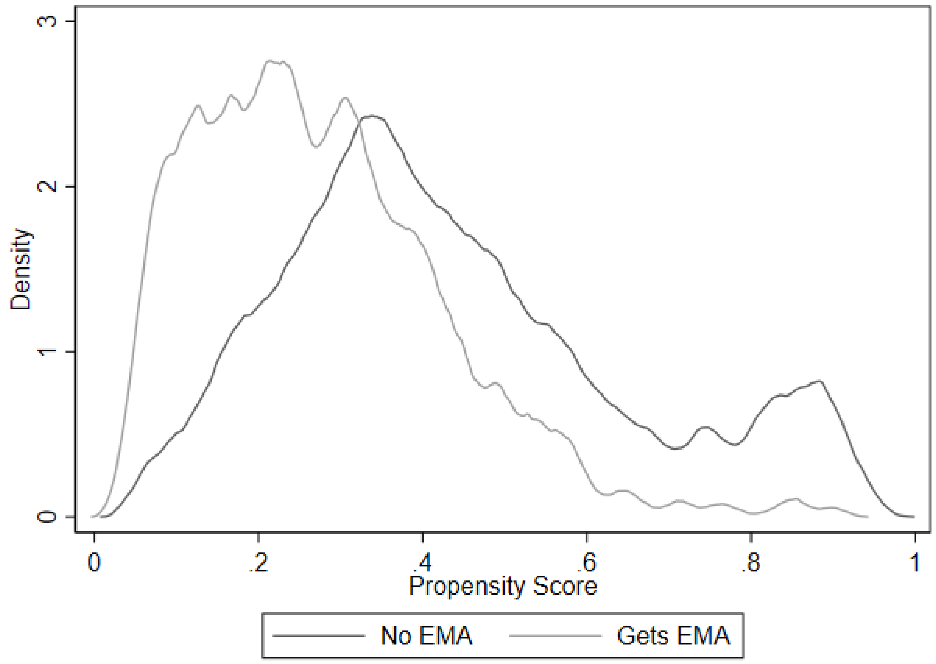
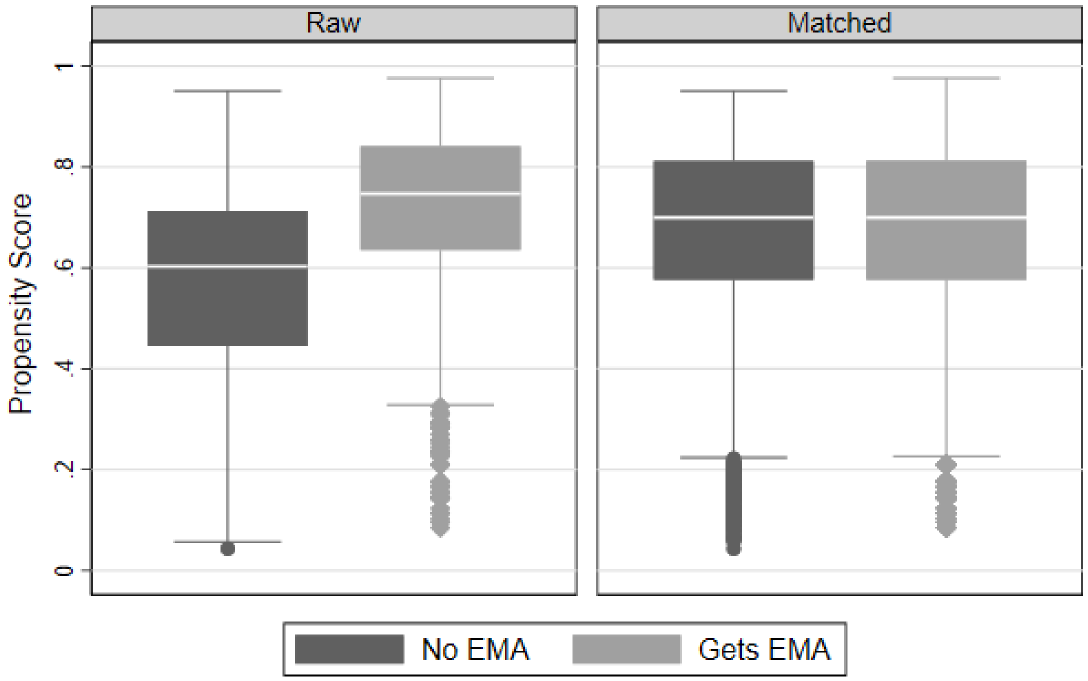


Figure 2 Covariate Balance



4. Empirical Strategy

The treatment effect of interest (τ_i) is that associated with receiving the Education Maintenance Allowance.[10] In a linear regression set up:

$$Y_i = \beta + \tau_i W_i + \boldsymbol{X}_i \gamma_i + \epsilon_i \quad (1)$$

Where Y_i is the outcome of interest (e.g., attendance in full-time education), W_i is the treatment (EMA receipt), $\mathbf{X}_i$ is a vector of control variables, and ϵ_i is a conventional error term.

In the light of the missing counterfactual, the paper proceeds by identifying estimates under the unconfoundedness (or ignorability) assumption – that is, that treatment assignment is unrelated to potential outcomes conditional on observed covariates.[11] Specifically, I estimate average effects using Inverse Probability Weighting Regression Adjustment (IPWRA).[12] Below, W_i is the binary treatment indicator, $Y_i(0)$ is the outcome of individual i in the absence of the treatment, $Y_i(1)$ is the outcome if the individual is treated, and X_i contains the collection of observed characteristics. The treatment is independent of the outcome, conditional on covariates.

$$W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) | X_i \quad (2)$$

Essentially, there is a trade-off between internal validity, due to the strong assumptions that underpin unconfoundedness, and external validity. As suggested by Athey & Imbens (2017), empirical approaches in economics should consider a range of methods rather than simply rely on those with the greatest internal validity (e.g., regression discontinuity, difference-in-differences (DiD), and instrumental variable (IV) methods). A second requirement is the

[10] The i subscript is relevant in for the later Causal Forest analysis, as individual level treatment effects can be estimated using that method.

[11] The thresholds for eligibility might facilitate a regression discontinuity design, but there are a few reasons that we do not proceed in this way. First, the data do not facilitate it – continuous measures of household income are only available two years prior to the time at which EMA is received, the discrete data after that is not appropriate for an RDD. Additionally sample sizes are too small near the cut-offs.

[12] See Imbens & Wooldridge (2009) for a more in-depth description of the method.

overlap assumption – that any given individual has a non-extreme probability (close to zero or one) of being in the treatment or control group.

$$0 < \Pr(W_i = 1 | X_i = x) < 1 \quad (3)$$

IPWRA models both the treatment (EMA receipt) and the outcome in two separate equations. A propensity score is estimated that models the probability of treatment based on included observables. This propensity score is then used to weight the second stage in an attempt to generate better counterfactuals and so strip out the possibility of (observed) selection into treatment from the outcome equation. Based on selection on the unconfoundedness and overlap assumptions, IPWRA is likely to be closer to causal effects than OLS by accounting for two levels of selection – in treatment and outcome. IPWRA exhibits the so-called "double robustness" property that means it produces consistent estimates even if one of the two equations is incorrectly specified. IPWRA also requires good overlap. As outlined above – I generate the overlap "blind" by using a randomly generated "x" as the outcome variable to avoid any temptation of picking the specification that yields a particular outcome.

Beyond average effects, I am interested in heterogeneity, because it is reasonable to assume that treatment will vary by at least some observable characteristics. To examine this, I employ Causal Forests (CFs).[13] CFs are a recent, and still rapidly developing, innovation in the application of machine learning for causal inference (Athey, 2017; Athey et al., 2019; Athey & Imbens, 2016; Mullainathan & Spiess, 2017; Varian, 2014; Wager & Athey, 2018). Recent developments have meant that valid statistical inference can now be made from these methods in the context of estimating causal effects rather than prediction – the more usual setting for machine learning techniques.

[13] Using the R package grf by Tibshirani et al. (2020), available from the CRAN project using the following link: https://CRAN.R-project.org/package=grf.

Causal forests developed from random forests whose trees make splits based on different variables (e.g., male/female, aged under 40/aged over 40) within a dataset to enable the prediction of an individual's outcome (e.g., will they default on a bank loan) based on their characteristics (Breiman, 2001). In essence it is an alternative to nearest neighbour matching in that each decision node (or leaf) defines the set of nearest neighbours for a given observation. In essence, instead of choosing the k closest points to an observation based on distance, close points are defined as those that occupy the same leaf (Wager & Athey, 2018). The resulting forests are improvements on nearest neighbour matching in terms of bias and variance.

When used for prediction, the regression tree algorithm makes splits that optimise performance relative to some metric – commonly, minimising the mean squared error (MSE). A problem arises when it comes to causal inference. In any given causal tree, the MSE cannot be used as one never observes both $Y_i(0)$ and $Y_i(1)$ for every individual. This is different to prediction when the actual Y that the algorithm must predict is known for the training sample – in causal inference we do not have the counterfactual. Instead, developing work by others, Athey & Imbens (2016) minimises the *expected* MSE (EMSE) of predicted treatment and, further, show that this is equivalent to splitting based on the characteristics that yield the biggest differences in treatment effect plus a penalty parameter for within-node (or leaf) variation.

Causal Forests are non-parametric. This is appealing as, unless the underlying data generation process happens to be linear, the linear model may fail to identify the true effect by making assumptions about the true functional form. Finally, Causal Forests are estimated 'honestly'. Honest forests attempt to minimise the risk of spuriously identifying effects by only using any given data sample for either estimating treatment effects or for where to make splits. Not both.

In using Causal Forests, I follow the application outlined in Athey & Wager (2019). I grow initial regression forests of 2000 trees to provide out-of-bag predictions of the propensity score (of treatment) and the main effect. Doing this enables the causal forest to focus on those features identified as most important in these initial forests rather than wasting splits on variables that are unimportant for heterogeneity. These values are inputs into the Causal Forest, which, when grown, has 10,000 trees. The Average Treatment Effect (ATE), Average Treatment Effect on the Treated (ATT), and the Average Treatment Effect on the Control (ATC) can be estimated using the Causal Forest. Conditional Average Treatment Effects (CATE) can also be estimated for each individual.

Arguably the greatest value of Causal Forests is in identifying heterogeneity in treatment effects. As Davis & Heller (2020) point out, testing for heterogeneity generally involves interacting treatment with various covariates where each additional hypothesis test may spuriously identify effects. If several interactions at the same time, or nonlinear functions of covariates are important, then traditional approaches may miss heterogeneity. Causal Forests flexibly and systematically model heterogeneity "based on high-dimensional nonlinear functions of observables" (Ibid., p665). I test for heterogeneity by comparing the average treatment effect for those individuals above the median treatment effect to those below. If this difference is statistically different from zero, there is evidence of heterogeneity. Treatment effect heterogeneity can also be charted.

The previous work on EMA has employed methods that, as in this paper, are not based on quasi-experimental methods. Dearden et al (2009), for example, employ fully interacted ordinary least squares (OLS) and probit techniques alongside PSM (as well as two difference-in-difference based sensitivity checks). In this context "fully interacted" means interactions of their treatment of being in an EMA pilot area with all other control variables. In contrast to this earlier work, I can flexibly and systematically analyse heterogeneity using Causal Forests with a much-reduced possibility of identifying sources of differential effects spuriously.

5. Results

Average Effects

Table 3 presents the first set of results; these relate to educational outcomes. Column (1) gives a simple OLS regression of the outcome on receipt of EMA (referred to as Gets EMA in the tables). Column (2) includes the full list of covariates, column (3) does the same but adds a school fixed effect to judge stability of results to past schooling. Column (4) then gives the average treatment effect (ATE) from an IPWRA specification, and Column (5) gives the average treatment effect on the treated (ATT) from the same specification.

In Table 3 two outcomes display robust and statistically significant effects – retention and university attendance. In the case of retention, once controls are added, this effect is around 8 percentage points. This is a little higher than in the pilot schemes where Dearden et al (2009) report retention effects of around 6.7 percentage points, though in statistical terms the estimate is not different to theirs. For university attendance the impact is similarly large; results appear a little less stable but are statistically the same across columns. There does not seem to be any impact on attainment which is always statistically insignificant. Some of this appears to be precision – as the point estimates are in fact large – in the case of the IPWRA estimates, they are around 8 percent of a standard deviation higher for those in receipt of EMA conditional on attendance. The final outcome in Table 3 is STEM degree subject. The IPWRA estimates differ from the others – in some sense they may be more credible – as IPWRA models both treatment and outcome where OLS models only outcome.

Moving to Table 4, which shows risky behaviours, there are no estimates that are statistically different from zero. This is positive news for EMA in some sense - the fact that giving young people large sums of money does not seem to increase alcohol consumption or the likelihood of them having tried cannabis are encouraging findings. This implies relatively little in the way of negative externalities resulting from EMA.

Table 3 Impact of EMA Receipt on Educational Outcomes

	(1) OLS SIMPLE	(2) OLS FULL	(3) OLS FULL + School FE	(4) IPWRA ATE	(5) IPWRA ATT
Panel A - Retention in FTE/Training					
Gets EMA	0.141***	0.086***	0.088***	0.084***	0.079***
	(0.019)	(0.019)	(0.021)	(0.019)	(0.020)
N	3,722	3,722	3,692	3,722	3,722
R-squared	0.019	0.087	0.257		
Panel B - UCAS Point Score (Standardised)					
Gets EMA	0.021	0.072	0.061	0.081	0.085
	(0.055)	(0.048)	(0.055)	(0.050)	(0.053)
N	2,260	2,260	2,179	2,260	2,260
R-squared	0.000	0.353	0.551		
Panel C - University Attendance					
Gets EMA	0.170***	0.090***	0.077***	0.092***	0.089***
	(0.018)	(0.016)	(0.018)	(0.016)	(0.018)
N	3,714	3,714	3,692	3,714	3,714
R-squared	0.030	0.332	0.463		
Panel D - STEM Degree Subject					
Gets EMA	0.031	0.042	-0.006	0.063**	0.065*
	(0.033)	(0.033)	(0.041)	(0.032)	(0.035)
N	1,591	1,591	1,466	1,591	1,591
R-squared	0.001	0.065	0.408		

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, p<0.1. Column (1) includes no covariates, each other column uses the full list provided in Table 2. IPWRA specifications use the full list of covariates in both stages.

Table 5 shows labour market outcomes at age 25. Broadly, results are not forthcoming. The one exception to this is interesting, however. Whilst earnings, likelihood of ever having been employed, and hours worked are not impacted by getting EMA, the likelihood of being on an insecure "zero hours" employment contract is reduced by a statistically significant 4 percentage points, and this effect is stable across columns. Zero hours contracts are preferred by those whose conventional labour market opportunities are limited. The novel suggestion here is that EMA generates better conventional labour outcomes and thus reduces the probability of choosing zero hours employment.[14]

[14] We also examined mental health, as measured by GHQ score. This is measured at both age 17, when EMA is first received, and at age 25. There were no significant effects on this outcome.

Table 4 Impact of EMA on Risky Behaviours

	(1) OLS SIMPLE	(2) OLS FULL	(3) OLS FULL + School FE	(4) IPWRA ATE	(5) IPWRA ATT
Panel A - Frequency of Drinking Alcohol (Standardised)					
Gets EMA	-0.157***	-0.057	0.004	-0.062	-0.062
	(0.038)	(0.041)	(0.046)	(0.041)	(0.044)
N	2,954	2,954	2,885	2,954	2,954
R-squared	0.006	0.072	0.271		
Panel B - Ever Tried Cannabis					
Gets EMA	-0.054***	0.002	0.012	-0.007	-0.010
	(0.018)	(0.018)	(0.019)	(0.018)	(0.019)
N	4,101	4,101	4,083	4,101	4,101
R-squared	0.003	0.106	0.281		

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Column (1) includes no covariates, each other column uses the full list provided in Table 2. IPWRA specifications use the full list of covariates in both stages.

Table 5 Impact of EMA on Long-Term Labour Market Outcomes

	(1) OLS SIMPLE	(2) OLS FULL	(3) OLS FULL + School FE	(4) IPWRA ATE	(5) IPWRA ATT
Panel A - Log Earnings					
Gets EMA	0.015	0.017	0.056	0.004	-0.013
	(0.040)	(0.035)	(0.047)	(0.033)	(0.034)
N	1,681	1,681	1,565	1,681	1,681
R-squared	0.000	0.166	0.501		
Panel B - Ever Employed					
Gets EMA	-0.042***	-0.019	-0.012	-0.021	-0.024
	(0.012)	(0.013)	(0.015)	(0.014)	(0.016)
N	2,357	2,357	2,294	2,357	2,357
R-squared	0.006	0.121	0.401		
Panel C - Hours Worked					
Gets EMA	-1.393*	-0.977	-0.555	-1.087*	-0.991
	(0.733)	(0.711)	(0.836)	(0.659)	(0.658)
N	1,924	1,924	1,826	1,924	1,924
R-squared	0.003	0.142	0.435		
Panel D - Zero Hours Contract					
Gets EMA	-0.035**	-0.044**	-0.038**	-0.044**	-0.038**
	(0.018)	(0.017)	(0.018)	(0.017)	(0.019)
N	1,926	1,926	1,827	1,926	1,926
R-squared	0.004	0.068	0.387		

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Column (1) includes no covariates, each other column uses the full list provided in Table 2. IPWRA specifications use the full list of covariates in both stages.

Before moving on to systematically examine heterogeneity using causal forests we examine, as Dearden et al (2009) do, heterogeneity by gender. This is shown in Table 6. Significant differences are not generally forthcoming – but with two important exceptions. First, EMA has a large and highly significant effect on STEM subject choice for women but not for men. Second, the effect on hours of work seems to be entirely confined to men which is a remarkably large effect suggesting that EMA may have facilitated quite different employment for men that would have otherwise been the case.

Finally, when faced with a large number of outcomes, as arises here, it is natural to be concerned with the possibility of false discovery – that some results are significant simply by chance. Appendix Table A1 takes the IPWRA specifications from above (as they account for selection in both treatment as control unlike the other two methods) and adjusts the p-values for false discovery – using the method in Benjamini & Hochberg (1995). As is clear, almost every outcome that was significant before is still significant, the exception being hours worked. UCAS Points was marginally significant before (p=0.101) but the adjusted value is not (p=0.202). In fact, the significance level does not change on any of the results – retention, university attendance, STEM degree subject, and working on a zero hours contract remain significant at the 1%, 1%, 10%, and 5% levels.

Causal Forests, Average Effects, and Treatment Heterogeneity

Table 7 reports the average treatment effect (ATE), the average treatment effect on the treated (ATT), and the average treatment effect on the control (ATC) for each outcome, using the full set of controls used in the IPWRA equations in Table 5 above. It shows broad agreement with the previous specifications – retention, university attendance, STEM degree subject, and the probability of being on a zero hours contract continue to show statistically significant effects. The effect sizes are very similar to those in earlier tables.

Table 6 Heterogeneity by Gender (IPWRA ATE Specifications)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A - Educational Outcomes								
	Retention		UCAS Points		University		STEM Subject Choice	
	Male	Female	Male	Female	Male	Female	Male	Female
Gets EMA	0.088***	0.094***	0.153**	0.081	0.100***	0.095***	0.007	0.145***
	(0.027)	(0.027)	(0.061)	(0.064)	(0.023)	(0.022)	(0.049)	(0.036)
N	1,798	1,924	1,019	1,241	1,784	1,930	676	915
Panel B - Risky Behaviours								
	Alcohol Consumption		Cannabis Ever		.	.	.	.
	Male	Female	Male	Female				
Gets EMA	-0.032	-0.086	0.001	-0.010	.	.	.	.
	(0.056)	(0.060)	(0.028)	(0.024)	.	.	.	.
N	1,244	1,254	2,009	2,092	.	.	.	.
Panel C - Labour Market Outcomes (Age 25)								
	Log Earnings		Ever Employed		Hours Worked		Zero Hours Contract	
	Male	Female	Male	Female	Male	Female	Male	Female
Gets EMA	-0.045	0.065	-0.022	-0.012	-3.162***	0.728	-0.028	-0.019
	(0.040)	(0.049)	(0.016)	(0.020)	(0.873)	(0.978)	(0.020)	(0.019)
N	733	948	1,030	1,327	738	876	733	879

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 7 Average Estimates from Causal Forest Estimations

	(1) ATE	(2) ATT	(3) ATC	(4) Heterogeneity	(5) 95% CI
Retention	0.089***	0.086***	0.096***	0.033	+/- 0.087
	(0.019)	(0.019)	(0.022)		
UCAS	0.075	0.076	0.072	0.051	+/-0.199
	(0.051)	(0.054)	(0.048)		
University	0.092***	0.089***	0.098***	-0.039	+/- 0.059
	(0.015)	(0.016)	(0.016)		
STEM	0.064*	0.076**	0.030	0.058	+/- 0.137
	(0.034)	(0.035)	(0.034)		
Alcohol	-0.098**	-0.095**	-0.102**	-1.081	+/- 0.157
	(0.040)	(0.040)	(0.046)		
Cannabis	-0.017	-0.017	-0.016	0.077	+/- 0.074++
	(0.018)	(0.018)	(0.021)		
Earnings	0.022	0.020	0.027	0.072	+/- 0.143
	(0.035)	(0.035)	(0.038)		
Ever Employed	-0.028**	-0.030**	-0.024**	0.058	+/- 0.072
	(0.013)	(0.015)	(0.011)		
Hours	-0.623	-0.496	-0.839	-0.839	+/- 1.252
	(0.706)	(0.711)	(0.761)		
Zero Hours	-0.044**	-0.045**	-0.043**	-0.058	+/-0.073
	(0.019)	(0.020)	(0.017)		

Cluster robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Sample sizes are as in Table 6 for each outcome. Column (5) shows an heuristic test for heterogeneity – whether the difference between the ATE above and below the median is different from zero. + symbols denote significant differences. Included covariates are those in Table 2.

However, two new effects are identified – one on alcohol consumption, and another on the likelihood of ever having been employed. In the case of the latter, it appears that the estimates are similarly precise, but the coefficients are a little larger meaning that effects are significant, where before they were not. In the case of alcohol consumption, however, the effect sizes are themselves larger. Table 7 also includes a measure of heterogeneity – this takes the average individual treatment effect above and below the median treatment effect and tests the difference between them. The confidence interval is provided. By this heuristic, it is only cannabis consumption appears to have substantial amounts of heterogeneity with respect to EMA receipt.

Table 8 reports the variables most important for heterogeneity in each outcome. This measure essentially counts how frequently the causal forest algorithm makes splits using a given variable and weights that by the depth at which these splits are made. Earlier splits (higher up the tree) imply that treatment effect heterogeneity was maximised best by choosing that variable at that point, so shallower splits should be given more weight. The most important dimensions of heterogeneity are identified in a separate regression forest estimated before the causal forest is estimated. An initial regression forest helps to calibrate the causal forest and ensure that splits are not being made that are unimportant for heterogeneity. Athey and Wager (2020)'s application prioritises those variables that are more important than the average variable when it comes to making the splits. Table 8 reports the top five variables judged by this metric. Some variables have fewer than five; some variables appear frequently – the IMD and quartiles of KS2 and GCSE score for example, others appear less often. The latter is true for gender, region of residence, and parental education.

Below, Figures 3 to 5 show the overall distribution of conditional average treatment effects for each outcome, arranged in the same groups as the initial OLS/IPWRA tables above. Essentially, these are frequency plots of individual level treatment effects. Figure 3 gives the educational outcomes. In the case of retention, a small number of individuals experience negative (though close to zero) impacts of EMA, whilst the modal bin is between 0 and 5 percentage points. For some, effects are as large as 30 percentage points. UCAS score displays a stranger distribution with peaks at different points, with the minimum effect being clustered around zero, but the maximum being around 0.1 of a standard deviation. University attendance displays a similar pattern to retention in that most individuals have effects close to zero or around ten percentage points but a small number see much larger impacts. STEM degree subject choice, like retention, most resembles normality.

Table 8 Most Important Variables for Heterogeneity as Identified by Causal Forests

	Rank				
Outcome	Most Important	2nd	3rd	4th	5th
Retention	LA-EMA	GCSE	Dependent Children	Non-white	.
UCAS points	KS2	GCSE	.	.	.
University	GCSE	KS2	Non-white	.	.
STEM	LA-EMA	IMD	Main Parent Age	Region	GCSE
Alcohol	IMD	Gender	Non-white	.	.
Cannabis	GCSE	General Health	.	.	.
Earnings	IMD	GCSE	KS2	Gender	.
Ever Employed	GCSE	IMD	LA-EMA	Main Parent Age	KS2
Hours	IMD	KS2	GCSE	Gender	.
Zero Hours	LA-EMA	IMD3	Parental Education	GCSE	KS2

Note: Table displays the most important dimensions of heterogeneity for each outcome. Variable importance is determined by weighting the number of splits across tress in the Causal Forest by the depth at which those splits occur. A greater number of splits that occur by a particular variable, the more important it is for heterogeneity. Those with a variable importance about the mean are included. LA-EMA is the overall level of EMA enrolment in an LA in Wave 4, GCSE is quartiles of GCSE attainment, KS2 is quartiles of Key Stage 2 attainment, IMD is the Index of Multiple Deprivation.

For risky behaviours (Figure 4), alcohol consumption (Panel A) is fairly concentrated around negative eight percentage points. Cannabis consumption is different – it seems to have few who experience a zero effect of EMA and instead some who experience positive and some who experience negative effects, with the balance tilting towards more individuals experiencing negative effects. In Figure 5 labour market outcomes are displayed. The conditional average treatment effects of EMA on log earnings (Panel A) have a broad distribution. Though similar to STEM degree subject in range (from around -0.05 to 0.10) the frequencies at any given point are less concentrated. The opposite is true of the likelihood of ever being employed (Panel B) which is highly concentrated and seems to present very little heterogeneity, though there is a long tail of larger negative effects. Hours worked looks reasonably heterogeneous spanning a negative impact of three fewer hours worked at age 25 for somebody on EMA to an additional two hours. Finally, zero hours contracts display a similar sort of distribution to cannabis consumption but are exclusively negative.

Figure 3 Distribution of Conditional ATE on Retention, UCAS Score, University Attendance, and STEM Degree Subject Choice

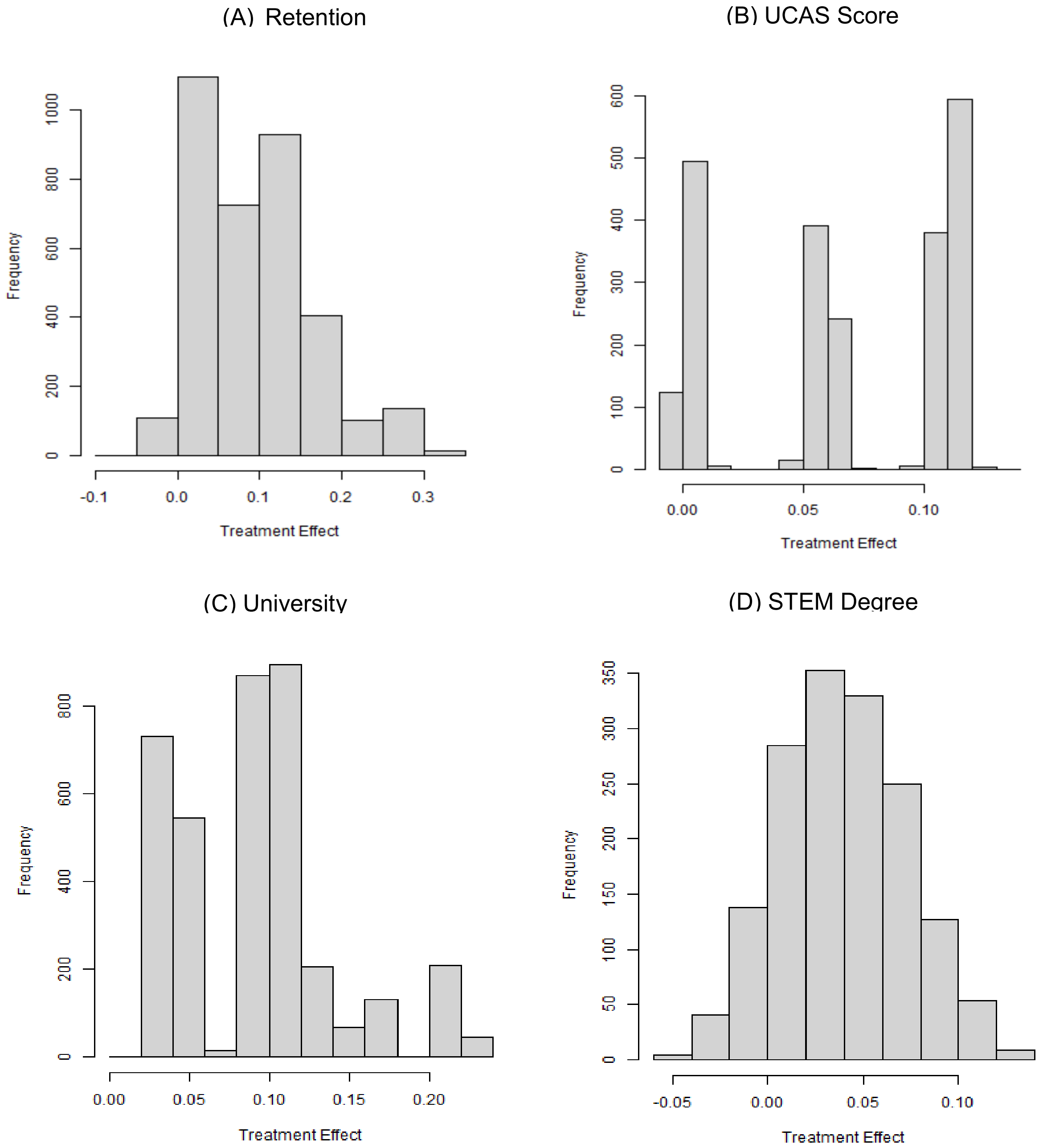


Figure 4 Distribution of Conditional ATE on Alcohol and Cannabis Consumption

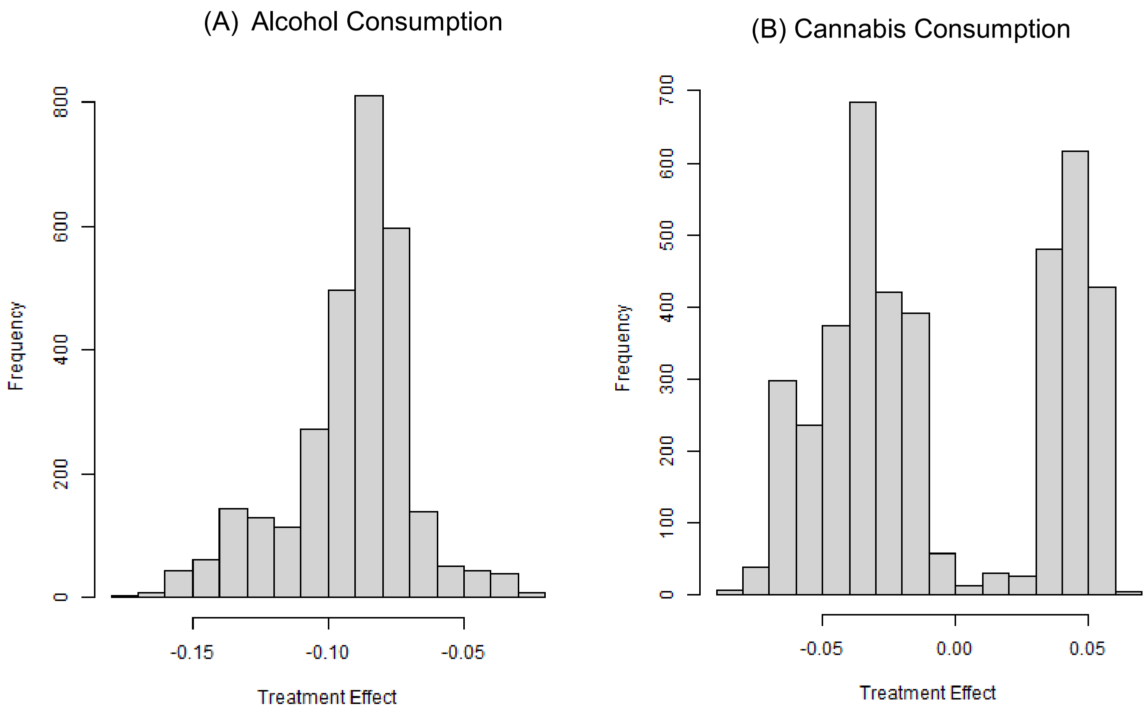
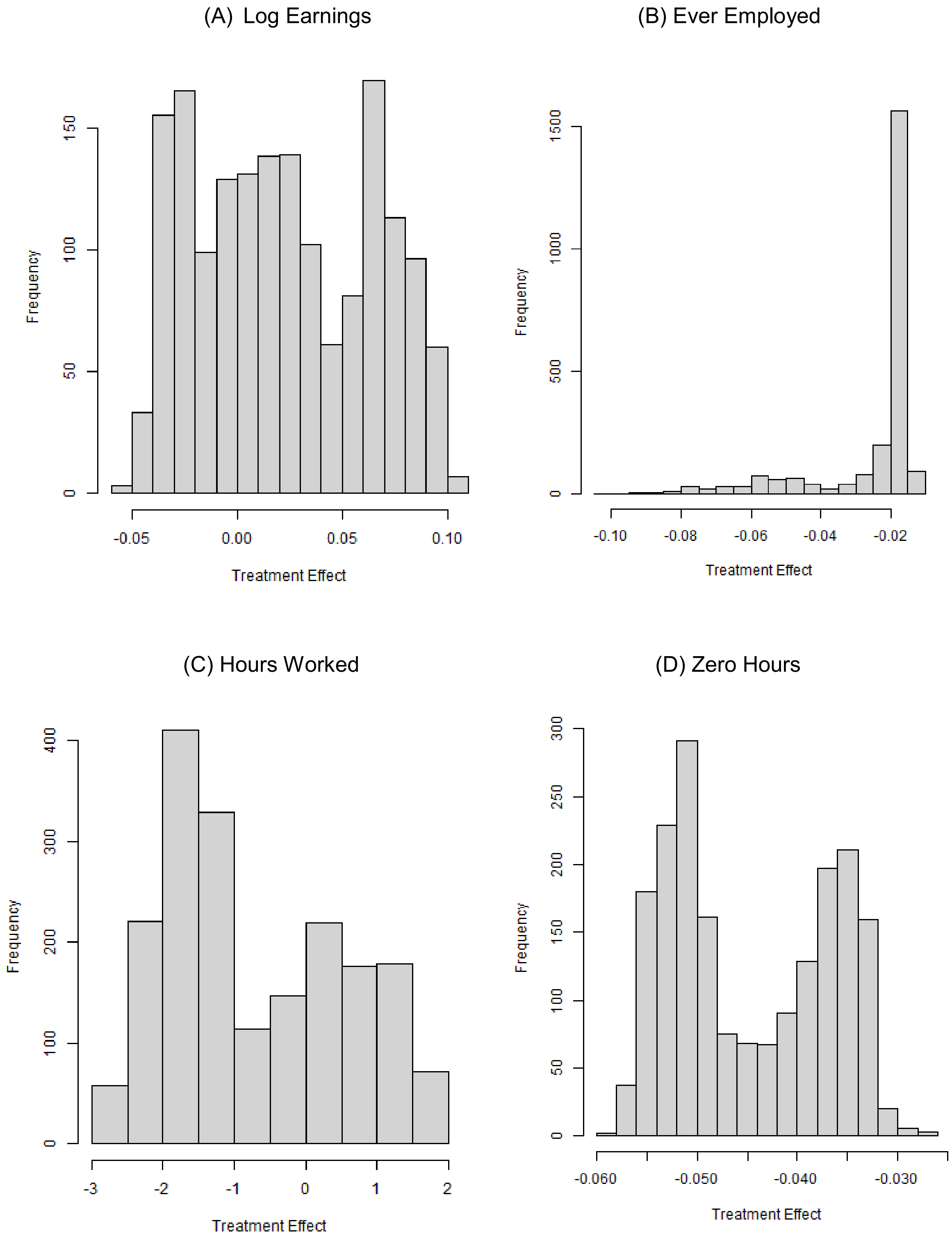


Figure 5 Distribution of Conditional *ATE* on Log Earnings, Employed Probability, Hours Worked, and Zero Hours Contract Probability



Figures 6 to 11 chart the two most important dimensions of heterogeneity for a selection of outcomes; these dimensions come from the 1st and 2nd most important variables listed in Table 8. The equivalent charts for the remaining outcomes (excluded from the main body for ease of presentation) are given in the Appendix. The charts are a mix of scatter and box-and-whisker plots that plot the treatment effect on the y axis and the characteristic by which it is varying on the x axis. There are some interesting results. For retention (Figure 6), EMA's effect seems largest for those in the lowest quartile of prior GCSE attainment (Panel B); higher proportions of people in your LA also being in receipt of EMA is important according to Table 8, but this is less clear when charted. Prior attainment at KS2 and GCSE are important dimensions for heterogeneity in the impact on attainment at A-level (Figure 7). Those who performed better in the past are see smaller improvements as a result of EMA. In Figure 8, university attendance seems to be characterised mostly by greater variance in the effects of EMA by prior attainment rather than seeing the kind of pictures as for retention and A-level attainment. This seems to be true in Figure 9, too, where the effect of STEM degree subject seems to have lower variance as the proportion of those on EMA in your LA and the deprivation of your postcode rise.

In terms of risky behaviours, Appendix Figures 1 and 2 are interesting. For alcohol consumption (Figure 1) it seems that individuals partake less frequently when on EMA if they live in more deprived areas (as measured by IMD). Men and women in the second Panel of Appendix Figure 1 has similar averages, but for females there seems to be a greater chance of greater reductions; in essence, there is greater variability. For cannabis, the outcome where heterogeneity was identified by the heuristic in Table 7, there seems to be substantial variation. Interestingly, those who self-report higher general health seems to be more likely to take cannabis that those with lower self-reported health if they are in receipt of EMA. This is interesting because one may not have thought to examine this dimension of heterogeneity, but the causal forests show that it is important.

Figure 6 Two Most Important Dimensions of Heterogeneity for Retention

(A) EMA Enrolment in LA (B) Quartile of GCSE Score

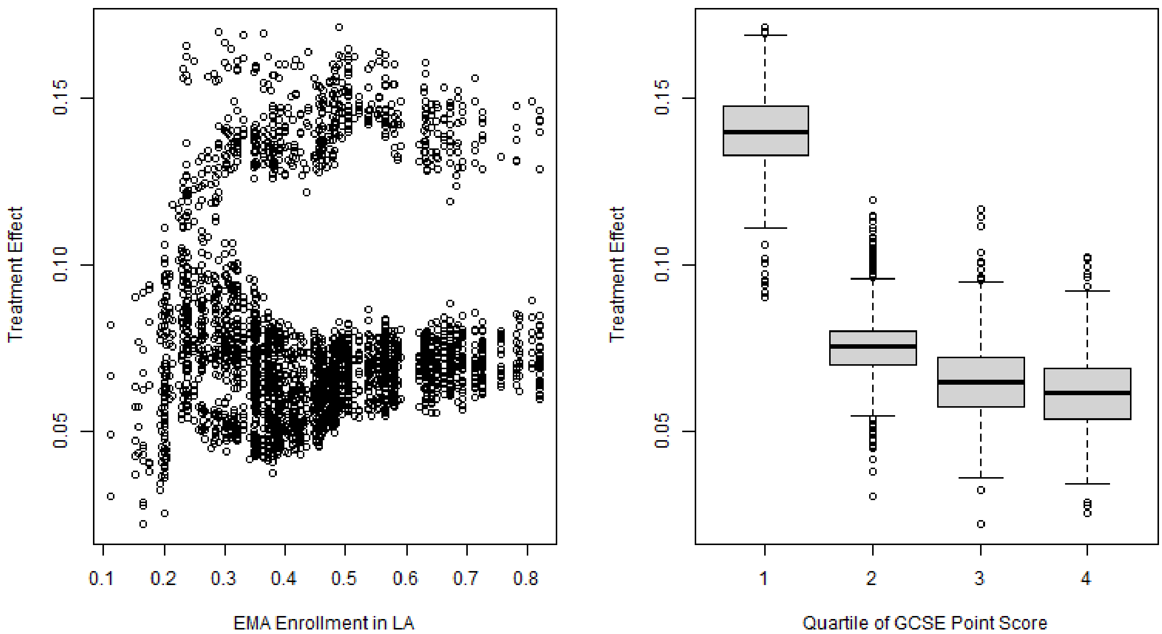


Figure 7 Two Most Important Dimensions of Heterogeneity for UCAS Point Score

(A) Quartile of KS2 Score (B) Quartile of GCSE Score

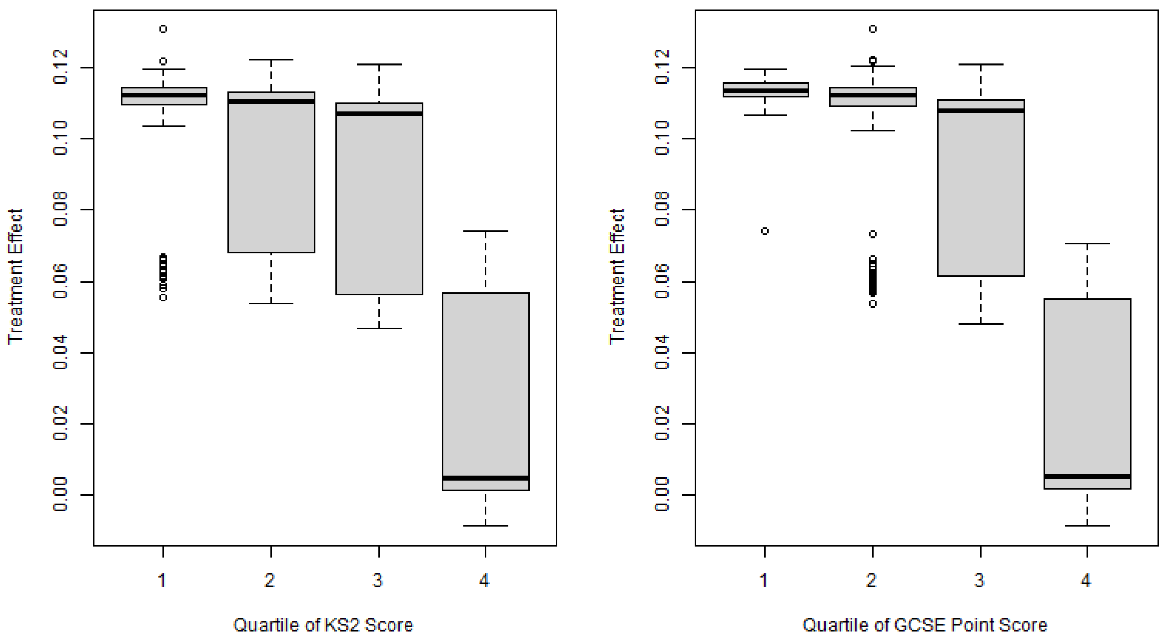


Figure 8 Two Most Important Dimensions of Heterogeneity for University Attendance

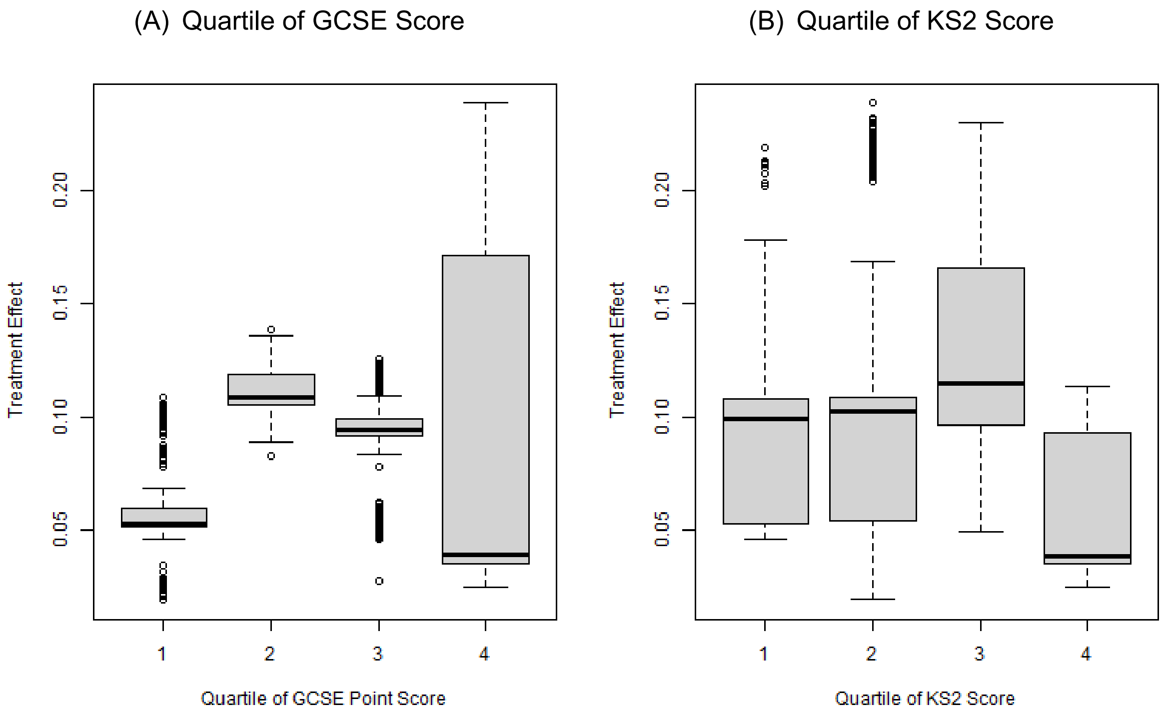


Figure 9 Two Most Important Dimensions of Heterogeneity for Stem Degree Subject

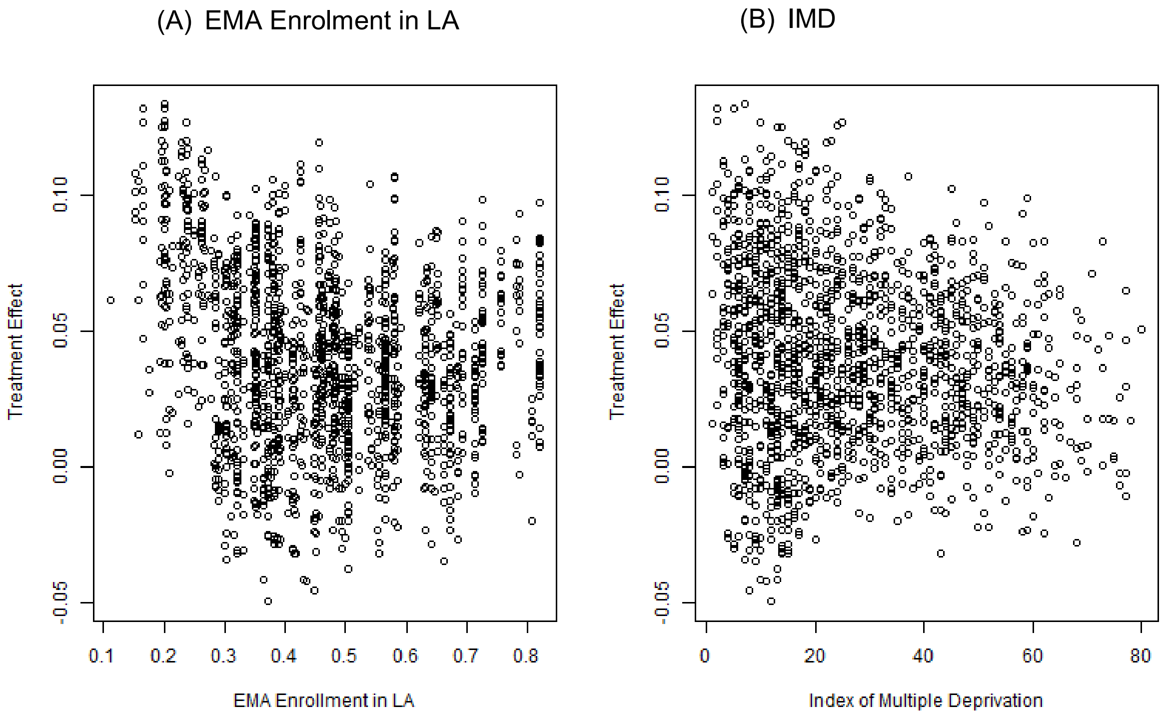


Figure 10 Two Most Important Dimensions of Heterogeneity for Ever Being Employed by Age 25

(A) Quartile of GCSE Score

(B) IMD

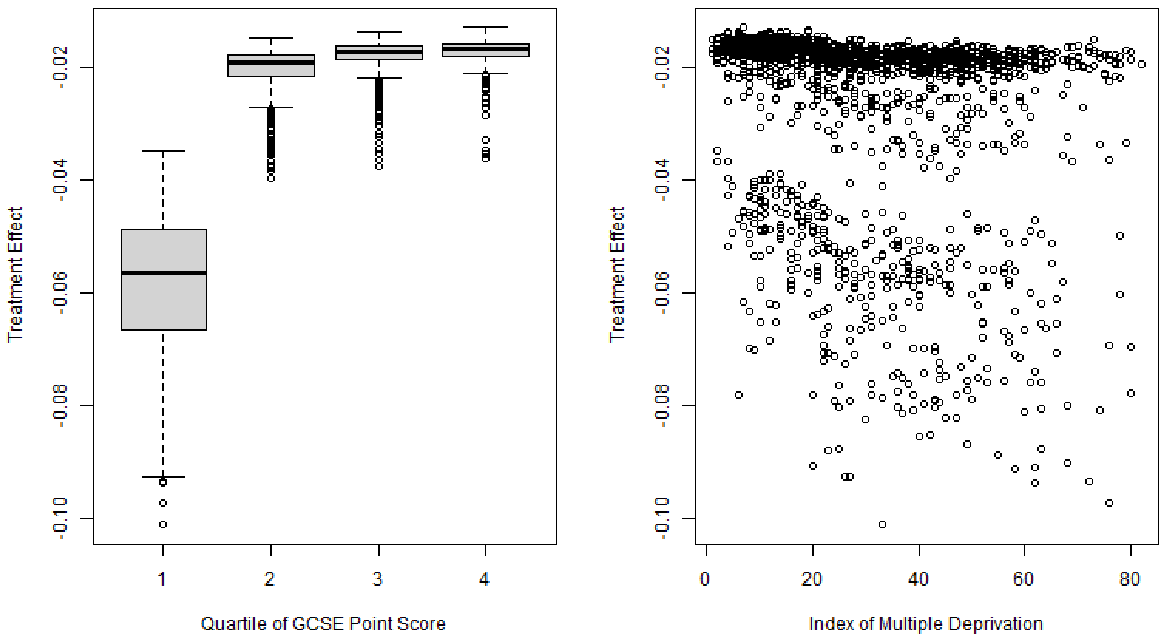
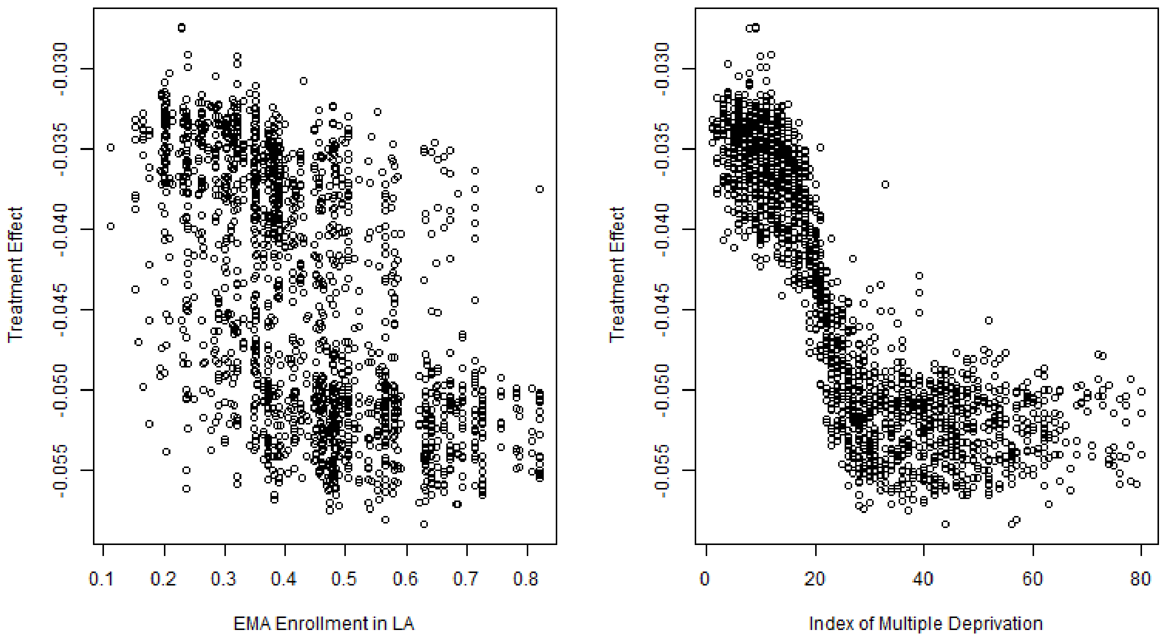


Figure 11 Two Most Important Dimensions of Heterogeneity for Being on a Zero Hours Contract at Age 25

(A) EMA Enrolment in LA

(B) IMD



Finally, in terms of labour market outcomes – there appear to be greater earnings benefits (Appendix Figure 3) for those in more deprived areas if they are on EMA. For some in less deprived areas the effects are actually negative. The likelihood of ever having had a job by age 25 (Figure 10) is roughly zero for the top three quartiles of prior GCSE attainment, but for the lowest EMA seems to mean a negative likelihood of being employed. This is interesting – and perhaps relates to EMA incentivising some people into career paths they were not best suited for. It seems being in a more deprived area means you see slightly longer working hours (Appendix Figure 4) – though this picture is less than clear. The effects seem to most reliably negative for those with higher attainment at KS2. Lastly, having more people in your LA on EMA and living in a more deprived postcode seem to make it more likely EMA has a more negative impact on the probability of working on a zero hours contract at age 25 (Figure 11). This is seen more clearly and easily in the case of the IMD (the right-hand panel of Figure 11).

Taken together, though it was initially indicated that there was not substantial heterogeneity in the effects of EMA by the heuristic in Table 7, it appears that some dimensions exist where there are differences in effects. This may be because in a number of cases it is more the case that some groups see greater variance in the effects rather than different averages. Some dimensions are clear – and few are starker than the large negative effect on the likelihood of working on a zero hours contract at age 25 (Figure 11) for those in more deprived areas.

Amount of EMA received

Up to this point, EMA has been treated as a binary treatment. This is because most recipients receive the full amount – this is shown in Table 9. Now, though, I examine the results by varying amounts of EMA receipt. Figure 12 shows the overlap in the whole sample by amount of EMA. As in Figure 1, the overlap is good.

Table 9 Distribution of Amounts of EMA Received

EMA Payment Amount (£)	Frequency	Percentage
0	1,168	31.38
10	244	6.56
20	298	8.01
30	2,012	54.06
Total	3,722	100

Figure 12 Overlap by EMA Amounts

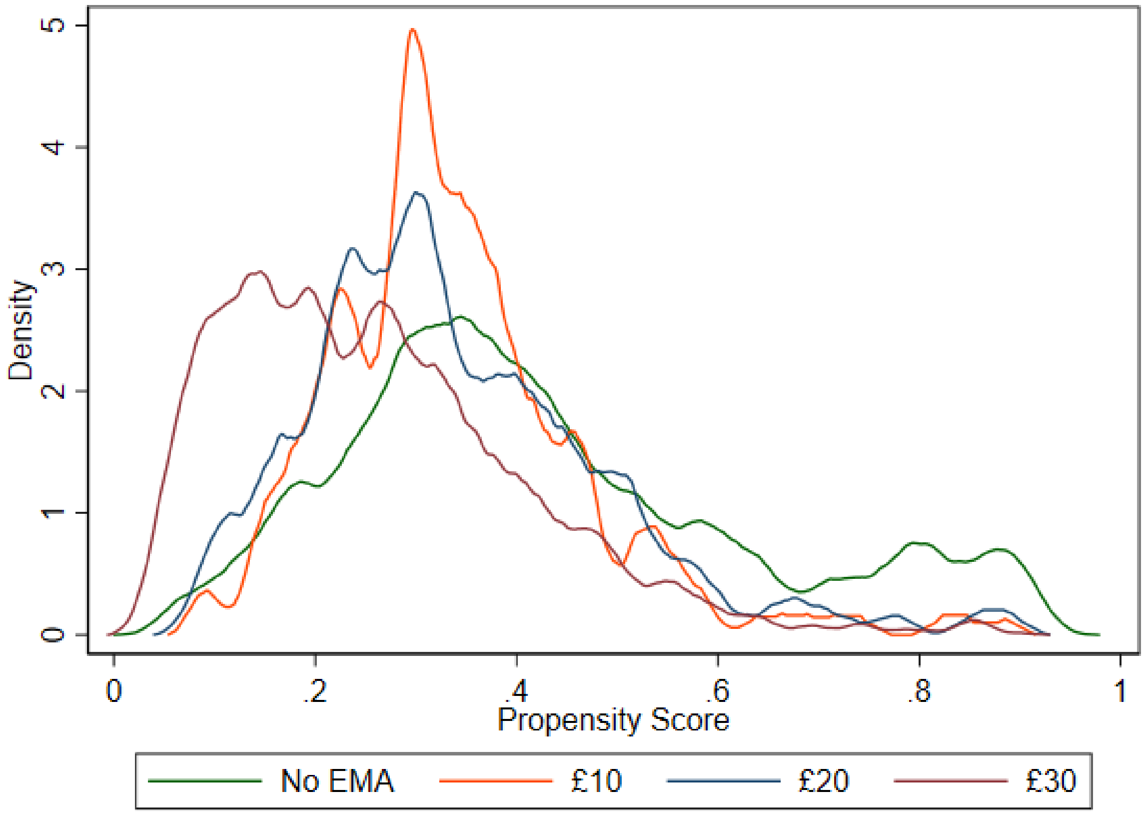


Table 10 takes the IPWRA average treatment effect specifications from above, but the treatment estimates are for the £10, £20, and £30 amounts instead of the binary EMA treatment variable. Two rows are blank – UCAS Points and STEM degree subject. In the case of the former, the overlap was not as good as in Figure 12 (though for all other outcomes, the overlap is very similar to that in Figure 12). For STEM degree subject the IPWRA command yielded no output due to issues of concavity. OLS estimates, of the same specification as the "OLS

Full” specifications above, for these two outcomes (not shown) showed insignificant effects for all EMA amounts for UCAS points whilst only the £10 and £20 brackets were insignificant for STEM degree subject. The £30 bracket was associated with a 6.8 percentage point increase in the likelihood of an individual picking a STEM degree subject at university, significant at the 10 percent level.

In terms of the coefficients in Table 10 the effects vary by outcome. For retention all levels have significant effects. The effect of £10 is larger in magnitude than the others but not statistically different to them. University attendance displays positive and significant effects of the £20 and £30 brackets. Alcohol and cannabis consumption each show negative effects of the £10 and £20 brackets whilst the probability of ever being employed seems only to be significantly impacted for those on the £10 bracket. Finally, hours worked and the likelihood of being on a zero hours contract are only (negatively) impacted for those on the £30 bracket.

This suggests that there is some evidence of heterogeneity in effect by amount received for risky behaviours and later employment. Hours worked and likelihood of being on a zero hours contract are perhaps easiest to understand – the larger amounts of money (given to those with the lowest family incomes) improve outcomes where smaller amounts do not. The other cases – alcohol and cannabis consumption and the likelihood over ever being employed by age 25 – are trickier. In the case of cannabis consumption, it is possible that there is a diminishing desirable (i.e., negative) effect as the amount of money handed out increases – on £10 or £20 the money is used to get the young person to college or to buy resources to learn with. Beyond that some may increase consumption with the additional money. The same reasoning does not apply to alcohol, where, as seen in the heterogeneity charts above, the impacts are either zero or negative, with little evidence of alcohol consumption increasing for any group.

Table 10 - IPWRA ATE Estimates for EMA Amounts

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcomes	£10	Std. Error	£20	Std. Error	£30	Std. Error	Observations
Retention	0.132***	(0.042)	0.097**	(0.039)	0.081***	(0.021)	3,722
UCAS Points	.	.	.	.	.	.	.
University	0.035	(0.034)	0.113***	(0.040)	0.094***	(0.017)	3,712
Stem	.	.	.	.	.	.	.
Alcohol	-0.125*	(0.075)	-0.132*	(0.076)	-0.009	(0.047)	2,497
Cannabis	-0.092***	(0.036)	-0.109***	(0.037)	0.008	(0.020)	4,098
Log Pay	0.052	(0.056)	0.017	(0.041)	0.016	(0.036)	1,416
Ever Employed	-0.204***	(0.030)	0.013	(0.026)	-0.020	(0.014)	2,355
Hours	-0.537	(0.928)	-0.268	(0.796)	-1.814**	(0.745)	1,612
Zero Hours	0.027	(0.026)	0.004	(0.026)	-0.034**	(0.014)	1,610

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: UCAS Points, unlike the other outcomes displayed poor overlap when it came to the amounts and so is omitted. STEM subject choice is omitted as the IPWRA estimation did not produce any output due to issues of concavity. Sample sizes vary when compared to above as the few cases where no overlap existed were omitted, too.

6. Discussion and Conclusion

This paper uses a variety of methods to show robust positive impacts of receiving EMA on retention in full-time education or training and university attendance, and robust negative impacts on the likelihood of being on an insecure (zero hours) employment contract. There is a suggestion, too, that the likelihood of studying a STEM degree subject at university also increases. No other outcomes, including educational attainment, risky behaviours (frequency of alcohol consumption and ever having tried cannabis), and long-term labour market outcomes, are significantly impacted.

The positive impact on retention mirrors that identified in the pilot studies on EMA; indeed, it is slightly larger (at around 8 percentage points instead of 6.7 percentage points). The effect identified in the pilot studies is, therefore, similar once the programme has been rolled out and has been in existence for around 5 years. The university attendance impact is positive too and differs to previous estimates of the impact of EMA. The more novel element of this analysis is in showing that there is no measurable impact on risky behaviours, and that insecure work at age 25 appears to decline. Indeed, the only significant effects on risky behaviours identified (in the causal forest analysis) are negative. It may have been easy to imagine that giving young people large amounts of cash might lead to negative outcomes. That alcohol and cannabis consumption is no higher for those receiving EMA suggests that this drawback is not present, and by implication the associated spillover effects of such behaviours do not increase. In terms of insecure work, it is likely the case that increased time in education (through retention and university attendance) explains the occurrence of this. But it is noteworthy that other labour market outcomes do not improve whilst this one does.

Beyond average effects, I examine heterogeneity systematically using causal forests. Across the range of outcomes, it appears that there are a number of cases where the effect of EMA differs by characteristics. This is true for both retention and attainment as A level (measured by UCAS score) – both are most positive for those in the lowest quartiles of prior attainment; alcohol consumption, which seems most negatively impacted by EMA in more deprived areas; cannabis consumption, which seems reliably to be reduced by EMA for those with lower self-reported health but appears to be more likely to increase for those with high self-reported health (though there is a wide distribution); the likelihood of ever having been employed, which is most negative for those with low GCSE attainment; and the likelihood of being on a zero hours contract at age 25 which varies starkly by deprivation where the individual lives. Of these, it is only the likelihood of ever being employed being worse for lower achievers which could be considered a drawback of EMA and is in need of further investigation.

The debate around EMA is still relevant in policy circles, for example in the United States where many states have compulsory school leaving ages below the UK's 18. But even though the UK now mandates education, where it used to incentivise it through EMA, that does not leave the matter settled. The absence of negative impacts on risky behaviours combined with desirable impacts on university attendance and on the probability of insecure work make a reappraisal of EMA as a policy seem compelling. As attention worldwide turns to recovering lost education during COVID-19 lockdowns, EMA seems less like the deadweight loss government described back in 2010/11 than an idea ripe for potential revival.

1. Reference List

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *485*(February), 483–485.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7353–7360. https://doi.org/10.1073/pnas.1510489113

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, *47*(2), 1179–1203. https://doi.org/10.1214/18-AOS1709

Athey, S., & Wager, S. (2019). Estimating Treatment Effects with Causal Forests: An Application. *Observational Studies*.

Attanasio, O. P., Meghir, C., & Santiago, A. (2012). Education choices in Mexico: Using a structural model and a randomized experiment to evaluate PROGRESA. *Review of Economic Studies*, *79*(1), 37–66. https://doi.org/10.1093/restud/rdr015

Benhassine, N., Devoto, F., Duflo, E., Dupas, P., & Pouliquen, V. (2015). Turning a shove into a nudge?: A "labeled cash transfer" for education. *American Economic Journal: Economic Policy*, *7*(3), 86–125. https://doi.org/10.1257/pol.20130225

Benjamini, Y., & Hochberg, Y. (1995). Controlling for the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Bolton, P. (2011). Education Maintenance Allowance (EMA). In *Parliamentary Briefing Papers*. https://www.studentfinancewales.co.uk/fe/ema.aspx

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1201/9780367816377-11

Britton, J., & Dearden, L. (2015). *The 16 to 19 bursary fund : impact evaluation* (Issue June).

Chowdry, H., Dearden, L., & Emmerson, C. (2008). *Education maintenance allowance evaluation with administrative data: the impact of the EMA pilots on participation and attainment in post-compulsory education*. *November*, 1–33. http://eprints.ucl.ac.uk/18324/

Davis, J. M. V., & Heller, S. B. (2020). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of Economics and Statistics*, *102*(4), 664–677. https://doi.org/10.1162/rest_a_00850

Dearden, L., Emmerson, C., Frayne, C., & Meghir, C. (2009). Conditional cash transfers and school dropout rates. *Journal of Human Resources*, *44*(4), 827–857. https://doi.org/10.3368/jhr.44.4.827

Dearden, L., & Heath, A. (1996). Income Support and Staying in School: What Can We Learn from Australia AUSTUDY Experiment? *Fiscal Studies*, *17*(4), 1–30.

Dynarski, S. M. (2003). American Economic Association Does Aid Matter ? Measuring the Effect of Student Aid on College Attendance and Completion. *The American Economic Review*, *93*(1), 279–288.

Feinstein, L., & Sabatés, R. (2005). Education and Youth Crime: Effects of Introducing the Education Maintenance Allowance Programme. In *Wider Benefits of Learning Report No.14*. https://doi.org/10.2139/ssrn.901421

Fiszbein, A., & Schady, N. R. (2009). Conditional Cash Transfers: Reducing Present and Future Poverty. In *World Bank Report*. https://doi.org/10.1596/978-0-8213-7352-1

Fletcher, M. (2000). *Education Maintenance Allowances: The Impact on Further Education*. http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED441174&site=ehost-live

Gitter, S. R., & Barham, B. L. (2009). Conditional cash transfers, shocks, and school enrolment in Nicaragua. *Journal of Development Studies*, *45*(10), 1747–1767. https://doi.org/10.1080/00220380902935857

Glewwe, P., & Kassouf, A. L. (2012). The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. *Journal of Development Economics*, *97*(2), 505–517. https://doi.org/10.1016/j.jdeveco.2011.05.008

Greenberg, D., Dechausay, N., & Fraker, C. (2011). *Learning Together: How Families Responded to Education Incentives in New York City's Conditional Cash Transfer Program*.

Holford, A. (2015). The labour supply effect of Education Maintenance Allowance and its implications for parental altruism. In *Review of Economics of the Household* (Vol. 13, Issue 3). Springer US. https://doi.org/10.1007/s11150-015-9288-7

Humlum, M. K., & Vejlin, R. M. (2013). The Responses of Youth to a Cath Transfer Conditonal on Schooling: A Quasi-Experimental Study. *Journal of Applied Econometrics*, *28*, 628–649. https://doi.org/10.1002/jae

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Middleton, S., Maguire, S., Ashworth, K., Legge, K., Allen, T., Perrin, K., Battistin, E., Dearden, L., Emmerson, C., Fitzsimons, E., & Megir, C. (2004). *The evaluation of Education Maintenance Allowance Pilots: three years' evidence: a quantitative evaluation*. http://eprints.ucl.ac.uk/18468/

Millán, T. M., Barham, T., Macours, K., Maluccio, J. A., & Stampini, M. (2019a). Long-term impacts of conditional cash transfers: Review of the evidence. *World Bank Research Observer*, *34*(1), 119–159. https://doi.org/10.1093/wbro/lky005

Millán, T. M., Barham, T., Macours, K., Maluccio, J. A., & Stampini, M. (2019b). Long-term impacts of conditional cash transfers: Review of the evidence. In *World Bank Research Observer* (Vol. 34, Issue 1, pp. 119–159). Oxford University Press. https://doi.org/10.1093/wbro/lky005

Morais de Sa e Silva, M. (2015). Conditional cash transfers and improved education quality: A political search for the policy link. *International Journal of Educational Development*, *45*, 169–181. https://doi.org/10.1016/j.ijedudev.2015.09.003

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Peruffo, M., & Ferreira, P. C. (2017). the Long-Term Effects of Conditional Cash Transfers on Child Labor and School Enrollment. *Economic Inquiry*, *55*(4), 2008–2030. https://doi.org/10.1111/ecin.12457

Riccio, J., Dechausay, N., Greenberg, D., Miller, C., Rucks, Z., & Verma, N. (2010). *Toward Reduced Poverty Across Generations: Early Findings from New York City's Conditional Cash Transfer Program*. http://ssrn.com/abstract=1786981

Riccio, J., & Miller, C. (2016). *NEW YORK CITY'S FIRST CONDITIONAL CASH TRANSFER PROGRAM What Worked, What Didn't*. http://ssrn.com/abstract=2821765

Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, *74*(1), 199–250. https://doi.org/10.1016/j.jdeveco.2003.12.009

Tibshirani, J., Athey, S., Friedberg, R., Hadad, V., Hirshberg, D., Miner, L., Sverdrup, E., Wager, S., & Wright, M. (2020). *grf: Generalised Random Forests. R Package* (1.2.0).

University College London, UCL Institute of Education, & Centre for Longitudinal Studies. (2021a). Next Steps: Linked Education Administrative Datasets (National Pupil Database), England, 2005-2009: Secure Access. In *[data collection]. 6th Edition. UK Data Service. SN: 7104*.

University College London, UCL Institute of Education, & Centre for Longitudinal Studies. (2021b). Next Steps: Sweeps 1-8, 2004-2016. . In *[data collection]. UK Data Service. SN: 5545*.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Wager, S., & Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

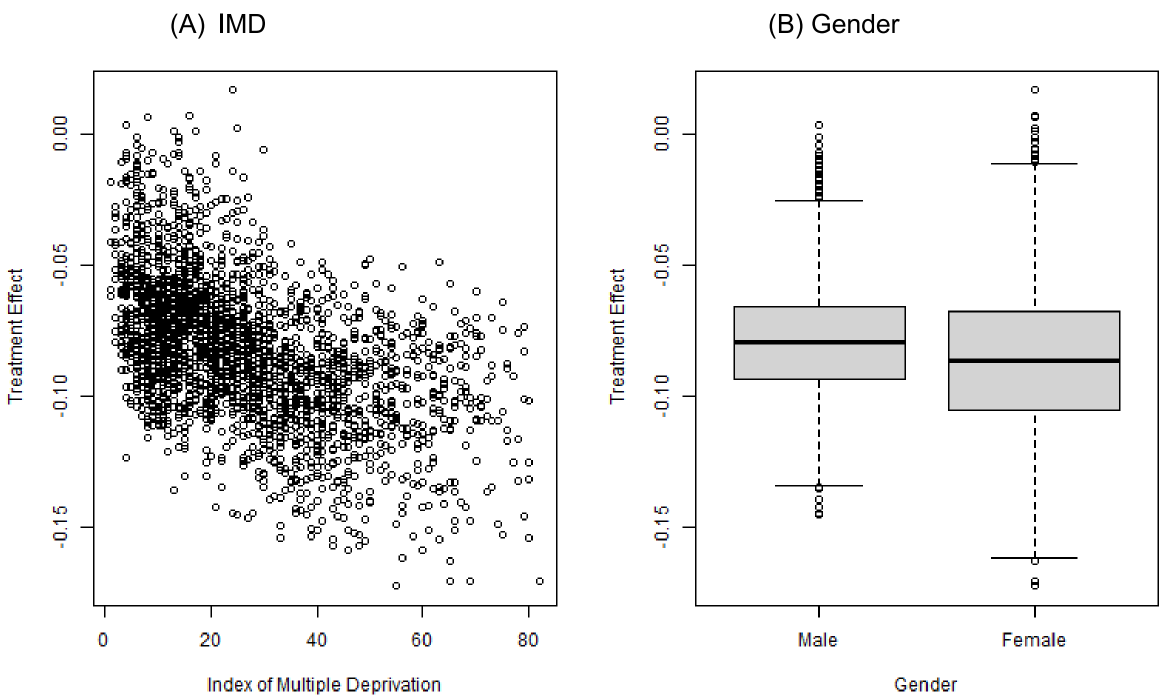
Appendix

Appendix Table 1 IPWRA Specifications with p-values Adjusted for False Discovery

Outcome	Estimate	Std Error	N	p-value	q-value
Educational					
Retention	0.084***	(0.019)	3,722	0.000	0.000
UCAS Points	0.081	(0.050)	2,259	0.101	0.202
University	0.092***	(0.016)	3,714	0.000	0.000
STEM	0.063*	(0.032)	1,591	0.056	0.080
Risky Behaviours					
Alcohol	-0.062	(0.041)	2,954	0.135	0.451
Cannabis	-0.007	(0.018)	4,101	0.689	>0.999
Labour Market					
Earnings	0.004	(0.033)	1,681	0.900	>0.999
Ever Employed	-0.021	(0.014)	2,357	0.116	0.290
Hours	-1.087	(0.659)	1,924	0.099	0.165
Zero Hours	-0.044**	(0.017)	1,926	0.010	0.013

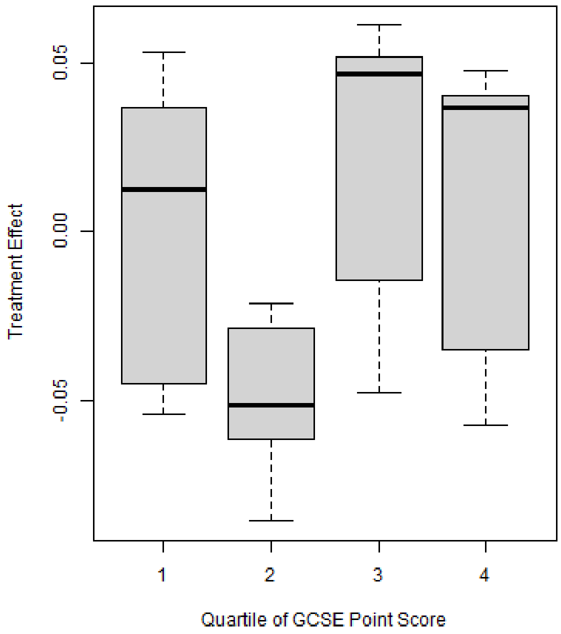
Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Asterisks relate to the q-value. The q-value is the p-value adjusted in line with Benjamini & Hochberg (1995).

Appendix Figure 1 Two Most Important Dimensions of Heterogeneity for Alcohol Consumption

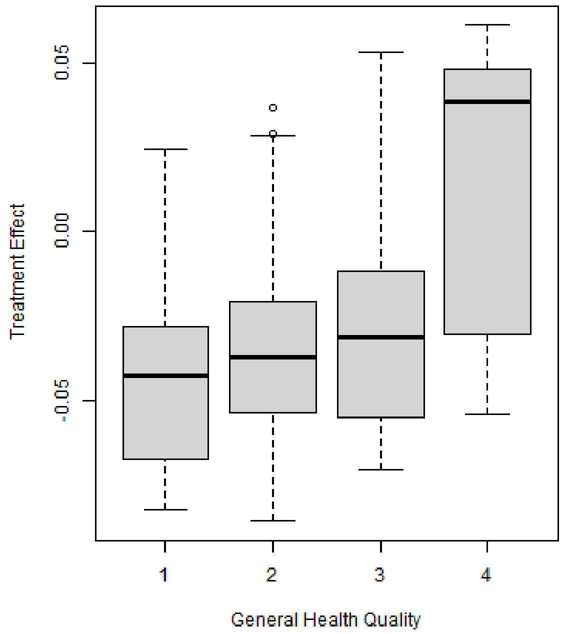


Appendix Figure 2 Two Most Important Dimensions of Heterogeneity for Cannabis Consumption

(A) Quartile of GCSE Score

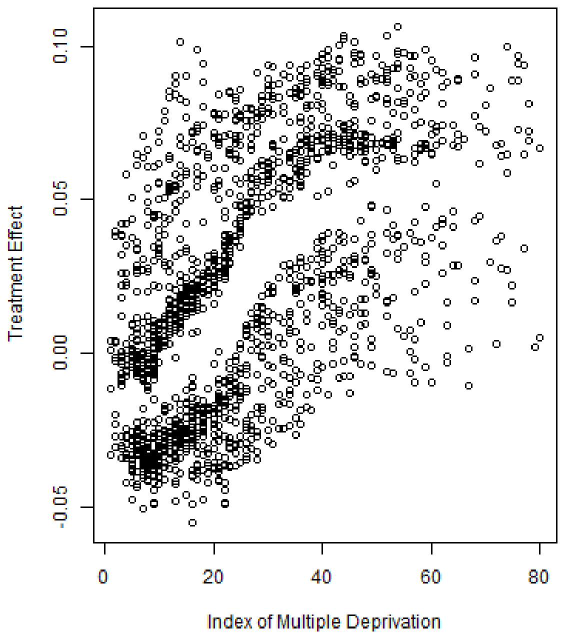


(B) Level of General Health

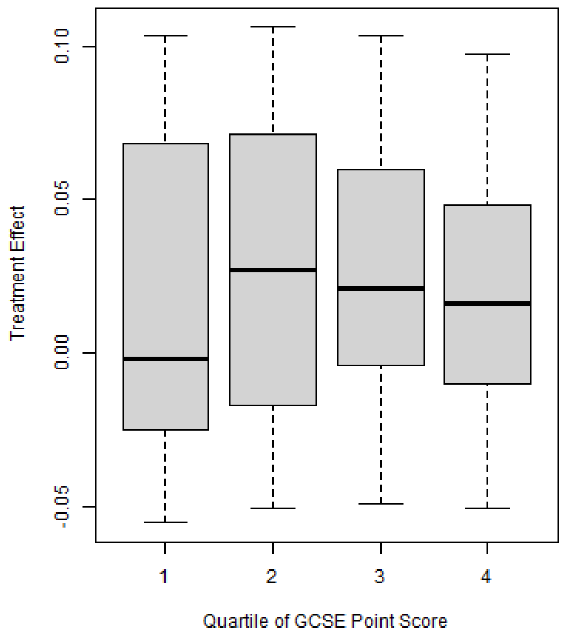


Appendix Figure 3 Two Most Important Dimensions of Heterogeneity for Log Earnings at Age 25

(A) IMD



(B) Quartile of GCSE Score



Appendix Figure 4 Two Most Important Dimensions of Heterogeneity for Hours Worked at Age 25