Department for Work and Pensions

Working Paper No 77

Non-participation in the Employment Retention and Advancement Study: Implications on the experimental first year impact estimates

Supplementary Technical Appendix

Barbara Sianesi

A report of research carried out by Institute for Fiscal Studies on behalf of the Department for Work and Pensions.

This Appendix provides a detailed and formal derivation of the estimation methods considered in the main report, as well as of the conditions for their validity. This Appendix was written as a stand-alone, avoiding the need to constantly refer the reader to the main report. There will thus necessarily be some overlap and repetition with the main report.

First Published 2010.

ISBN 978 1 84712 704 4

Contents

List of figure

1 Analysis framework

1.1 Problem set-up, notation and overview

In setting up the framework and introducing the necessary notation, we reprint Figure 1 from the main report to highlight the structure of the problem we need to address; we also extend summary Box 1 to include some further notation.

The population of interest are those eligible to be offered Employment Retention and Advancement (ERA) services under normal programme operation, i.e. all those becoming unemployed in the six districts over the study intake window. We implicitly condition on this population throughout. The binary variable Q captures the potential selection into the ERA study, with Q=0 denoting individuals who despite being eligible have not been randomly assigned, and Q=1 denoting the ERA study participants. Study participants make up the experimental group which was randomly assigned between a programme group who was offered ERA services (R=1) and a control group who was not (R=0).

The problem we want to address is that because of diversion and of refusal to be randomly assigned, the population under the experimental evaluation (Q=1) does not correspond to the full eligible population, made up by the (Q=1) **and** (Q=0) groups. If selection has taken place into the participating group, the composition of participants will be different from the composition of the eligible population, and impacts estimated on participants will not necessarily be representative of the impacts that the eligibles would have experienced.

Further, let the indicator S denote availability of a survey-based outcome measure conditional on ERA participation. Specifically, S=1 when survey outcomes such as earnings are observed; this happens only for that subsample of participants who (1) were randomly selected to be surveyed, (2) could be contacted, (3) accepted to take the survey and (4) answered the earnings question. For short, we will refer to them as 'respondents'. S=0 by contrast denotes non-surveyed or survey non-respondents or item non-respondents among participants ('non-respondents'). As Figure 1 highlights, it is possible for some selection to have taken place among participants into the responding sample.

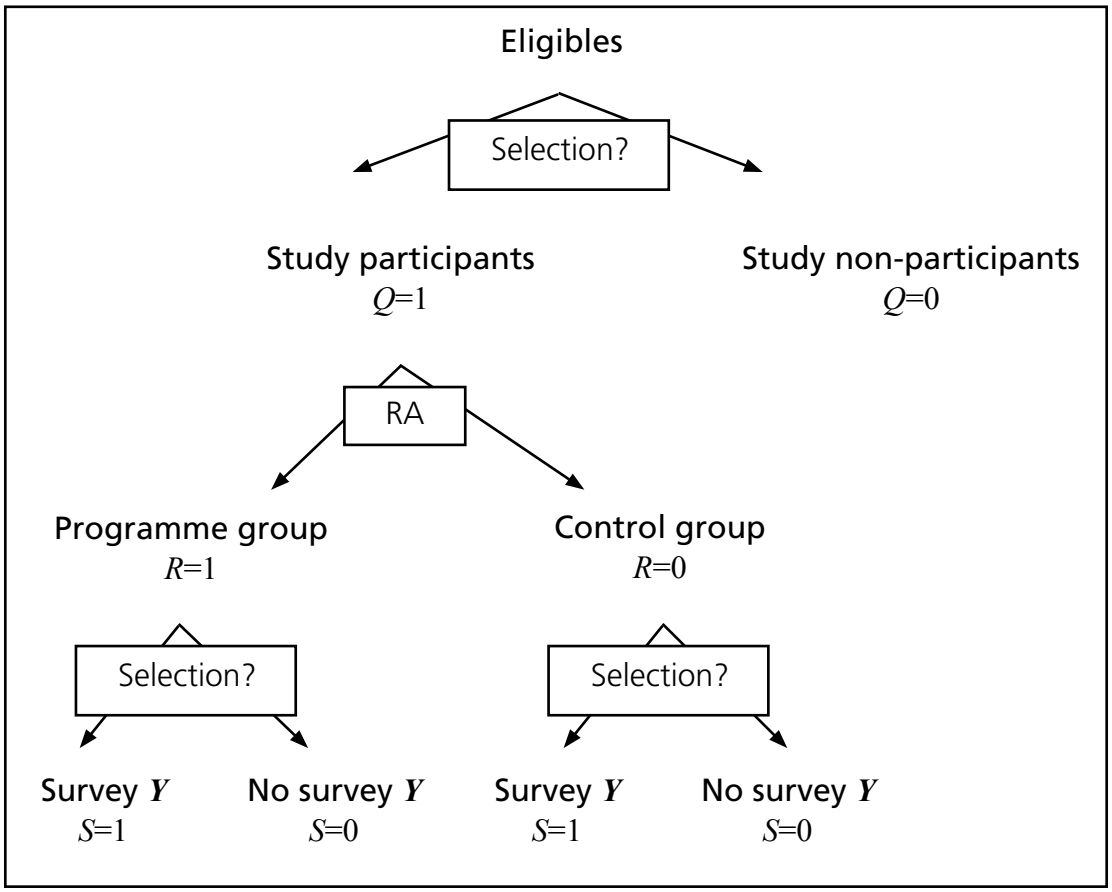
Let $p \equiv P(Q=0)$ be the probability of non-participation among the ERA eligibles. This is directly identified in the data by the proportion of non-participants among the eligibles (see Tables 2.1-3.2 in the main report).

Define the 'propensity score', i.e. the probability that an eligible customer with characteristics $X=x$ does not participate in the ERA study, as:

$$p(x) \equiv \mathrm{P}(Q=0 \mid X=x) = P(Q=0 \mid Q=0 \equiv Q=1, X=x).$$

Turning now to outcomes, we follow the potential outcome framework and let Y_1 be the outcome if the individual were offered ERA services (i.e. the treatment outcome) and Y_0 the outcome if the individual were not offered ERA services (i.e. the no-treatment outcome). The observed outcome is denoted by Y. The individual causal effect of ERA is defined as the difference between the two potential outcomes, $Y_1 - Y_0$.

Figure 1 Simplified structure of the problem



Box 1 Notation

Q=1	ERA study participants (the experimental sample)
Q=0	non-participants
R=1	individuals randomly assigned to the programme group conditional on Q=1
R=0	individuals randomly assigned to the control group conditional on Q=1
S=1	observe survey outcomes conditional on Q=1 ('respondents')
S=0	do not observe survey outcomes conditional on Q=1 ('non-respondents')
X	observed characteristics
p	probability of non-participation among eligibles
$p(x)$	propensity score: $P(Q=0 \mid X=x)$
Y_1	potential outcome if offered ERA services
Y_0	potential outcome if not offered ERA services
Y	observed outcome
ATE	average ERA effect on **all** ERA eligibles (parameter of interest)
ATE_1	average ERA effect on ERA study participants (experimental estimate)
ATE_0	average ERA effect on non-participants
$ATE_{S=1}$	average ERA effect on respondents
$ATE_{S=0}$	average ERA effect on non-respondents
$\Delta_{S=1}$	experimental contrast for respondents

The parameter we are interested in is the average effect of ERA on the **full** ERA eligible population (an Average Treatment Effect):

$$ATE \equiv E(Y_1 - Y_0)$$

What we can, however, directly identify from the available experimental data is the average effect of ERA for participants in the experiment. This is because the experiment provides the average effect of the programme for individuals who have been randomly assigned, which, due to the randomness of R within the Q=1 group, is identified by the difference in the mean outcomes of programme and control groups:

$$\begin{aligned} ATE_1 &\equiv E(Y_1 - Y_0 \mid Q=1) \\ &= E(Y_1 \mid Q=1) - E(Y_0 \mid Q=1) \\ &= E(Y_1 \mid Q=1, R=1) - E(Y_0 \mid Q=1, R=0) \\ &= E(Y \mid R=1) - E(Y \mid R=0) \end{aligned}$$

Denote the average impact of ERA on the excluded eligibles (i.e. on the non-participants) by

$$ATE_0 \equiv E(Y_1 - Y_0 \mid Q{=}0) = E(Y_1 - Y_0 \mid Q{=}0)$$

Using the law of iterated expectations, the parameters ATE and ATE_1 are linked according to:

$$\begin{aligned} ATE &= E(Y_1 - Y_0 \mid Q{=}1)\, P(Q{=}1) + E(Y_1 - Y_0 \mid Q{=}0)\, P(Q{=}0) \\ &\equiv (1{-}p){\cdot}ATE_1 + p{\cdot}ATE_0 \end{aligned} \quad (1)$$

Equation (1) simply states that the parameter of interest, i.e. the average impact of ERA on all the eligibles in the six districts, is given by a weighted average of the parameter we can reliably estimate using random assignment, i.e. the impact on participants, and of the impact on non-participants, with weights given by the relative share of participants and non-participants within the eligible pool.

In the main report we have highlighted two alternative conditions under which the average impact for participants would be the same as the average impact for the full eligible population even in the presence of a non-negligible share of non-participants. The first situation is one of homogeneous ERA impacts, that is $Y_{1i} - Y_{0i} = \beta$ for all eligible individuals i. The second case is one where impacts might be heterogeneous, i.e. $Y_{1i} - Y_{0i} = \beta_i$, but individuals do not base their decisions to participate in the study on the realised individual gain from receiving ERA, β_i. Formally:

$$\begin{gathered} \text{if } Q \perp (Y_1 - Y_0), \text{ i.e. if } P(Q{=}1 \mid Y_1 - Y_0) = P(Q{=}1) \\ \text{then } E(Y_1 - Y_0 \mid Q{=}1) = E(Y_1 - Y_0 \mid Q{=}0) = E(Y_1 - Y_0). \end{gathered}$$

In either of these cases, the ATE_1 based on experimental data would thus still provide unbiased estimates of the ATE of interest.

In the methodological overview in the main report we highlight that for each approach we consider how to deal with non-participants both when follow-up information on their outcomes is available (administrative-data based outcome measures) and when it is not (survey-data based outcomes measures). The implications of these two situations on equation (1) are as follows.

In case of administrative data, equation (1) becomes:

$$ATE = (1{-}p){\cdot}ATE_1 + p{\cdot}\{\boldsymbol{E(Y_1 \mid Q{=}0)} - E(Y \mid Q{=}0)\} \quad (1a)$$

as the observed outcome of the non-participants corresponds to their no-treatment outcome: $E(Y_0 \mid Q{=}0) = E(Y \mid Q{=}0)$.

In case of survey outcomes, both treatment and no-treatment outcomes of the non-participants remain unobserved. Furthermore, in the presence of non-random non-response among ERA study participants, ATE_1 itself will in general remain unobserved:

$$ATE = (1{-}p){\cdot}\boldsymbol{ATE_1} + p{\cdot}\boldsymbol{E(Y_1 - Y_0 \mid Q{=}0)} \quad (1b)$$

1.2 Survey outcomes: Survey and item non-response

Survey outcomes Y, in particular earnings, are only observed for a subsample of participants, i.e. those survey respondents who answered the earnings question.

Define the respondents (S=1) as those ERA study participants (Q=1) with non-missing survey outcome information, as non-respondents those ERA study participants (Q=1, S=0) with missing survey outcome information – whatever the reason (not randomly selected for the survey, not contactable, refused to be interviewed, were interviewed but did not fill in the earnings question). Note thus that in our definition of non-respondents we have lumped survey and item non-respondents, since impact estimates on earnings can only be obtained for our narrower definition of respondents.

In addition to the loss in precision resulting in a reduction of the study's statistical power to detect effects, non-response raises two important validity issues for the evaluation of earnings impacts:

1. **Internal validity**: if the programme and control group experience systematically different non-response, the responding programme and control groups are no longer comparable to one other. In this case the benefits of the original random assignment are lost, and a comparison of the responding programme group members and the responding control group members no longer provides unbiased impact estimates (for the respondents).
2. **External validity**: even if the responding programme and control group members have maintained comparability to one another so that the experimental contrast recovers the average impact for respondents, how do they relate to the original sample? If the responding sample differs substantially from the original one, the results might not generalize to the original target population.

Define $\Delta_{S=1}$ as the experimental contrast calculated on those participants who responded to the earnings question:

$$\Delta_{S=1} \equiv E(Y \mid Q=1, S=1, R=1) - E(Y \mid Q=1, S=1, R=0)$$

$\Delta_{S=1}$ is identified in our data, but we are interested in ATE_1 as one of the two components needed to recover the ATE for the full group of eligibles. The question thus that naturally arises is:

> Under what conditions does the experimental contrast for respondents recover the ATE for the full group of participants, i.e.
>
> $$\Delta_{S=1} = ATE_1$$

Although this condition can indeed be tested on administrative outcomes, which are available for the full group of participants (indeed, for the full group of eligibles), whether it resulted to be met or not would not be easy to interpret. In answering this question it is instead useful to separately consider the following two 'causal-inference' issues related to the internal and external validity issues above.

(a) Internal validity: Under what conditions does $\Delta_{S=1}$ recover the *ATE* for respondents, $ATE_{S=1} \equiv E(Y_1 - Y_0 \mid Q=1, S=1) \equiv E(Y_1 - Y_0 \mid S=1)$?

Since the average ERA impact for respondents is not identified without additional assumptions, to exploit random assignment one has to assume that randomisation keeps holding within the responding sample, i.e. that R is still random (possibly given X) among respondents:

(I-V) $$E(Y_1 \mid S=1, R=1) = E(Y_1 \mid S=1, R=0) = E(Y_1 \mid S=1)$$
$$E(Y_0 \mid S=1, R=1) = E(Y_0 \mid S=1, R=0) = E(Y_0 \mid S=1)$$

Under the internal-validity condition (I-V) that even restricting attention to the subgroup of respondents, randomisation still holds, the ATE for respondents, $E(Y_1 - Y_0 \mid S=1)$, can be estimated using the experimental contrast, $E(Y \mid S=1, R=1) - E(Y \mid S=1, R=0)$:

$$ATE_{S=1} \equiv E(Y_1 - Y_0 \mid S=1) \equiv E(Y_1 \mid S=1) - E(Y_0 \mid S=1)$$
$$= \text{(I-V)} = E(Y_1 \mid S=1, R=1) - E(Y_0 \mid S=1, R=0) = E(Y \mid S=1, R=1) - E(Y \mid S=1, R=0)$$
$$\equiv \Delta_{S=1}$$

Condition (I-V) cannot be directly tested; supporting evidence can however be obtained by assessing whether randomization still holds between the two responding subsamples in terms of their **observed** characteristics.

(b) External validity: Under what conditions can the subsample of respondents be assumed to be a representative subsample of the ERA study participants, in the sense that the *ATE* among respondents is the same as the *ATE* for the full group of participants, i.e. $ATE_{S=1} = ATE_1$?

The average ERA impact is the same for the full sample of participants and for those participants who responded to the survey if participants do not select into responding based on ERA impacts. Formally:

(E-V) $$E(Y_1 - Y_0 \mid Q=1) = (Y_1 - Y_0 \mid Q=1, S=1)$$

Since the impact for respondents is not identified a priori, to 'test' condition (E-V) one has first to assume that condition (I-V) holds. Under (I-V), condition (E-V) can be tested on administrative data as:

$$E(Y_1 \mid Q=1, R=1) - E(Y_0 \mid Q=1, R=0) = E(Y_1 \mid Q=1, R=1, S=1) - E(Y_0 \mid Q=1, R=0, S=1)$$

Note that under (I-V), condition (E-V) is implied by the stronger set of conditions:

(E-V') (a) $E(Y_1 \mid Q=1, R=1) = E(Y_1 \mid Q=1, R=1, S=1) = E(Y_1 \mid Q=1, R=1, S=0)$
(b) $E(Y_0 \mid Q=1, R=0) = E(Y_0 \mid Q=1, R=0, S=1) = E(Y_0 \mid Q=1, R=0, S=0)$

Conditional on random assignment status, non-response is unrelated to potential outcomes, i.e. programme group members do not select into responding based on treatment outcomes, nor do control group members select into responding based on no-treatment outcomes. Put differently, programme and control group members who respond are not selected on outcome-relevant variables. Assumption (E-V') thus rules out selection on outcome-relevant unobservables into responding to the earnings question conditional on random assignment status.

Like assumption (E-V), assumption (E-V') can be tested on administrative outcomes. This is accomplished by testing whether (possibly controlling for observables X), the administrative outcomes of those programme (control) group members who responded to the survey are statistically different from the outcomes of those programme (control) group members for whom we do not observe the survey outcomes.

To conclude, the experimental contrast for respondents, $\Delta_{S=1}$, which is readily obtained from the data, would recover the ATE for the full group of participants, ATE_1, under (I-V) and either (E-V') and/or (E-V). In this case, non-response can be ignored in calculating the average effect on earnings for participants.

It would be hard to believe, though possible, that $\Delta_{S=1}$ just happens to coincide with ATE_1 on administrative outcomes – or that condition (E-V') is met –, even without the need to give a causal interpretation to $\Delta_{S=1}$ (via (I-V)). If there is good support for (I-V), though, the evidence is likely to be more robust.

2 Methodological approaches

2.1 Bounds without assumptions on the selection process

For this type of analysis, outcomes need to be bounded. To fix ideas, suppose in the following that the outcome Y (e.g. employment probability) is bounded between 0 and 1.

a) Follow-up data on the non-participants (administrative outcomes)

From equation (1a), bounds for the parameter of interest can be constructed as $ATE \in [\underline{ATE}, \overline{ATE}]$

where $\underline{ATE} = (1-p)\cdot ATE_1 - p\cdot E(Y \mid Q=0)$
is the worst-case scenario: non-participants would all be non-employed had they received ERA, i.e. $E(Y_1 \mid Q=0) = 0$.

$\overline{ATE} = (1-p)\cdot ATE_1 + p\cdot(1-E(Y \mid Q=0))$
is the best-case scenario: non-participants would all be employed had they received ERA, i.e. $E(Y_1 \mid Q=0) = 1$.

The width of the bounds for the ATE, $\overline{ATE} - \underline{ATE}$, is given by p, the proportion of non-participants among the eligibles.
(If there were none, the upper and lower bounds would trivially collapse on the point estimate $ATE_1 = ATE$).

Sensitivity analysis

We can further explore how sensitive the estimate of the ATE is to assumptions about the selection process into the group of study participants, as reflected by assumptions on the relative magnitude of $E(Y_1 \mid Q=0)$ and $E(Y_1 \mid Q=1)$.

We can thus calculate the ATE as a function of θ, ATE_θ, for various values of θ (θ=0.5, …, 1.50) assuming that:

$$E(Y_1 \mid Q{=}0) = \theta\, E(Y_1 \mid Q{=}1)\ (= \theta\, E(Y \mid Q{=}1, R{=}1))$$

i.e. that the average ERA-treatment outcome that the non-participants would have experienced had they participated in the study is θ times the average treatment outcome experienced by the participants, where the latter is identified by the actual outcomes of the randomised programme group subset of the participants.

By varying the values of θ, we can depict different types of selection processes:

- θ=1 → decisions to participate in the ERA study are unrelated to treatment outcomes;
- θ<1 → negative selection into the non-participants sample (non-participants would have experienced on average lower treatment outcomes than what the participants experience);
- θ>1 → positive selection.

(Note that the θ such that the corresponding ATE_θ coincides to the experimental ATE_1 is given by: $(E(Y \mid Q{=}0) + ATE_1)/E(Y \mid R{=}1)$).

From equation (1a), the ATE as a function of θ, is

$$ATE_\theta = (1{-}p){\cdot}ATE_1 + p{\cdot}\{\theta\, E(Y \mid R{=}1) - E(Y \mid Q{=}0)\}$$

Thus, ATE_θ increases, and linearly, with θ.

The minimum allowable θ for our outcomes is 0, the maximum allowable $\theta = 1/E(Y_1 \mid Q{=}1, X)$ for the binary outcome Y we consider.

b) No follow-up information on the non-participants (survey outcomes)

In this case, we have to construct bounds on ATE based on (1b). It follows that

$$ATE \in [\underline{ATE}, \overline{ATE}]$$

where $\underline{ATE} = (1{-}p){\cdot}ATE_1 - p$

In the worst-case scenario, all non-participants Q=0 would be non-employed if they received ERA, but would be employed in the absence of ERA, i.e. they would have Y_1=0 and Y_0=1, hence a programme effect $Y_1 - Y_0 = -1$, i.e. $E(Y_1 - Y_0 \mid Q{=}0) = -1$.

$\overline{ATE} = (1{-}p){\cdot}ATE_1 + p$

In the best-case scenario, all the non-participants Q=0 would be employed under ERA and non-employed without ERA, i.e. they would have Y_1=1 and Y_0=0, hence a programme effect $Y_1 - Y_0 = 1$, i.e. $E(Y_1 - Y_0 \mid Q{=}0) = 1$.

(Note that the two intermediate cases giving rise to a zero programme effect fall always in between).

The width of the bounds for the ATE is now $2 \cdot p$, double as large as when we did observe the outcomes of the non-participants.

In case non-response cannot be ignored, the bounds will necessarily – and trivially – be the widest possible ones, and unrelated to data content:

$ATE \in [-1, 1]$

since

$$\overline{ATE} = (1-p) \cdot 1 + p \cdot 1 = 1$$

$$\underline{ATE} = (1-p) \cdot (-1) + p \cdot (-1) = -1$$

Inference for partially identified parameters

A final issue concerns the significance of our estimates. A growing body of research in the last years has looked into the problem of constructing confidence intervals for partially identified parameters. In our application, we follow Horowitz and Manski (2000) and derive confidence intervals for bounds that cover the entire identification region with 95 per cent probability. By denoting with $\hat{L}$ and $\hat{U}$ the lower and the upper bounds, we report confidence intervals of the form $[\hat{L} - \zeta, \hat{U} + \zeta]$, where ζ is a positive constant obtained by bootstrapping the distribution of bounds so as to ensure the required probability of coverage. As stated in Imbens and Manski (2004; see Lemma 1), the probability that the interval considered covers the true ATT is at least 95 per cent (thus leading to conservative inference).

2.2 Point estimate under selection on observables

This section focuses on a class of methods that allow point identification of the ATE by relying on the selection-on-observables assumption. While this type of approach differs from the one described in Section A2.3 in terms of the assumption it makes on the selection process into the ERA study, both approaches rely on the assumption that treatment and no-treatment outcomes among the eligibles are not affected by whether an individual is **offered the chance** to participate in the ERA study or not. In other words, participants and non-participants may be drawn from different parts of the distributions of observed and unobserved characteristics, but the mere fact of being offered the chance to participate in the ERA study does not change the relationship between characteristics on the one hand and treatment and no-treatment outcomes on the other. Formally, this requires the potential outcomes of individual i to depend on observables and on the unobservables in a way that is not indexed by Q, i.e.:

$$Y_{1i} = m_1(X_i) + u_{1i}$$
$$Y_{0i} = m_0(X_i) + u_{0i}$$

and not:

$$Y_{1Qi} = m_{1Q}(X_i) + u_{1Qi} \qquad \text{for } Q=0, 1$$
$$Y_{0Qi} = m_{0Q}(X_i) + u_{0Qi}$$

a) Follow-up data on the non-participants (administrative outcomes)

To obtain a point estimate of the ATE, equation (1a) shows that we need to identify $E(Y_1|Q=0)$, the treatment outcome of the non-participants.

This problem is akin to getting the average treatment effect on the non-treated using matching methods, where invoking the 'selection-on-observables' assumption, $E(Y_1 \mid Q=0)$ is estimated based on the (observed) treatment outcome of the participants, $E(Y_1 \mid Q=1) = E(Y \mid Q=1, R=1)$.

In this case, we allow the effect (or treatment outcome) to depend on observable characteristics X in an arbitrary way, as well as for eligible individuals to decide to participate in the experiment based on these Xs.

To clarify the assumptions required, specialise the model as follows (note that additive separability is not required for matching).

$$Y_{1i} = m_1(X_i) + u_i + b_i$$
$$Y_{0i} = m_0(X_i) + u_i$$

where $Y_{1i} - Y_{0i} \equiv \beta_i = [m_1(X_i) - m_0(X_i)] + b_i \equiv b(X_i) + b_i$.

In this set-up, β_i, the individual impact from receiving ERA services, is allowed to be heterogeneous across individuals in both observable and unobservable dimensions: $b(X_i)$ represents the impact for individuals with characteristics X_i and thus captures observable heterogeneity in effects; b_i represents the individual-specific unobserved impact conditional on X_i. The unobserved component u_i represents some unobservable individual trait, such as ability or motivation, that affects the outcome irrespective of treatment receipt.

Assume that for the eligibles, selection into Q is not based on the unobserved, person-specific component of the impact of ERA b, nor on unobserved 'ability' u for given observable characteristics X:

$$Q_i \perp (b_i, u_i) \mid X_i$$

This ensures that the 'selection-on-observables' assumption (A1) is met:

$$\text{(A1)} \qquad E(Y_1 \mid Q=0, X) = E(Y_1 \mid Q=1, X)$$

To give (A1) empirical content, we also need to assume the existence of common support (i.e. overlap in the distribution of observed characteristics X between participants and non-participants:

$$\text{(CS)} \qquad P(Q{=}1 \mid X){>}0 \qquad \text{for all } X \text{ in the support of the eligibles}$$

Specifically, the experimental evaluation cannot provide estimates of the impact of ERA for individuals with observed characteristics $\tilde{X}$ if no participant displays those values. In other words, although there may be **eligibles** with characteristics $\tilde{X}$, if the selection into the ERA experiment is such that nobody with characteristics $\tilde{X}$ is offered ERA or consents to take part so that $P(Q{=}1 \mid \tilde{X}) = 0$, we cannot identify the effect for this subset of eligibles (unless under some arbitrary functional form assumption that allows us to extrapolate).

We can then predict $E(Y_1 \mid Q{=}0)$ as:

$$\begin{aligned} E(Y_1 \mid Q{=}0) &= E_X[E(Y_1 \mid Q{=}0, X) \mid Q{=}0] \\ &= (\text{A1}) = E_X[E(Y_1 \mid Q{=}1, X) \mid Q{=}0] \\ &= (\text{RA}) = E_X[E(Y_1 \mid R{=}1, X) \mid Q{=}0] \\ &= E_X[E(Y \mid R{=}1, X) \mid Q{=}0] \end{aligned}$$

As for implementation, we can match to each non-participant one or more similar programme group member(s) based on the propensity score $p(x) \equiv P(Q{=}0 \mid X) = P(Q{=}0 \mid Q{=}0 \vee Q{=}1, X)$.

To increase matching quality, it might be worth using only the programme group $R{=}1$ (a random hence representative subset of the $Q{=}1$ group) rather than the full $Q{=}1$ group (i.e. both the programme and control groups) to estimate the propensity score; that is, estimate $p(x)$ based on $P(Q{=}0 \mid Q{=}0 \vee R{=}1, X)$.

Sensitivity analysis

As with the bounds, we can explore how sensitive the estimate of the ATE is to straightforward violations of assumption (A1). In particular, replace (A1) by:

$$\text{(A1')} \qquad E(Y_1 \mid Q{=}0, \text{X}) = \theta\, E(Y_1 \mid Q{=}1, X)$$

i.e. thus allowing participants and non-participants with the same observed characteristics X to differ in terms of some unobservable, which translates into a proportional difference of θ.

Under (A1'), $E(Y_1 \mid Q{=}0) = \theta\, E[E(Y \mid R{=}1, X) \mid Q{=}0]$, which simply involves rescaling the matched outcome by θ.

Again, ATE_θ increases (linearly) with θ.

The sensitivity analysis can be easily expanded by allowing θ to depend on X via the propensity score $p(X) \equiv P(Q{=}0 \mid X)$:

$$\text{(A1'')} \qquad E(Y_1 \mid Q{=}0, X{=}x) = \theta(x)\, E(Y_1 \mid Q{=}1, X{=}x) \text{ where } \theta(x) = \theta(p(x))$$

Among customers with the same *a priori* study participation probability p, those who do not participate would have experienced an average treatment outcome which is a fraction $\theta(p)$ of the one of the participants.

Under (A1''), $E(Y_1 \mid Q{=}0) = \theta\, E_p[\theta(p)\, E(Y \mid R{=}1, p) \mid Q{=}0]$

This is most easily performed by stratification matching.

One could consider cases where the selectivity of the non-participants (as captured by $\theta(p){\neq}1$) increases or decreases with the non-participation probability p, or is non-monotonic.[1,2]

In particular:

	p from 0 to 1	Selection into Q=0
Increases	$\theta(p)$ from 1 to <1	negative
	$\theta(p)$ from 1 to >1	positive
Decreases	$\theta(p)$ from <1 to 1	negative
	$\theta(p)$ from >1 to 1	positive
Non-monotonic[1]	$\theta(p)$ from <1 to >1	negative to positive
	$\theta(p)$ from >1 to <1	positive to negative

[1] In the non-monotonic case, one could set $\theta(p)$=1 in correspondence of p equal to the observed non-participation probability.

b) No follow-up information on the non-participants (survey outcomes)

This problem is akin to attrition and involves reweighing the outcomes of the ERA study participants (programme and control groups) on the basis of the characteristics X of the full eligible group (i.e. ERA programme group, ERA control group and ERA non-participants) to make them representative – in terms of X – of the full eligible population.

Assume that, once conditioning on observables X, ERA study participants and non-participants on average experience the same treatment and no-treatment outcomes:[3]

[1] Note that it would not be informative to consider best- and worst-case bounds for all combinations of $\theta(p)$, since the best-case bounds would be obtained for $\theta(p){=}\theta_{\max}$ (=constant that can be derived from the bounded nature of the outcomes considered) for all p and the worst-case bounds for $\theta(p){=}\theta_{\min}$ (=0 for the outcomes considered here) for all p.

[2] The maximum allowable θ is $\mathrm{Max}\{Y_1 \mid Q{=}1, X\}/E(Y_1 \mid Q{=}1, X)$. Specifically, for a binary outcome, it is equal to 1/(rate for participants of type X) and for days in employment or on benefits in the 12 months post inflow, it is equal to 365/(average days for participants of type X).

[3] Note that we cannot test (A2-b), since even though $E(Y_0 \mid Q{=}1, X) = E(Y \mid R{=}0, X)$ and $E(Y_0 \mid Q{=}0, X) = E(Y \mid Q{=}0, X)$, we do not observe the latter outcomes.

(A2) (a) $E(Y_1 \mid Q{=}1, X) = E(Y_1 \mid Q{=}0, X)$ hence = $E(Y_1 \mid X)$ and
(b) $E(Y_0 \mid Q{=}1, X) = E(Y_0 \mid Q{=}0, X)$ hence = $E(Y_0 \mid X)$

In fact, it suffices that (A2) holds in terms of impacts:

(A2) $E(Y_1 - Y_0 \mid Q{=}1, X) = E(Y_1 - Y_0 \mid Q{=}0, X)$ hence = $E(Y_1 - Y_0 \mid X)$

To estimate the *ATE* of interest, write it as:

$$\begin{aligned}\text{ATE} \equiv\ & E(Y_1 - Y_0) = E_X[E(Y_1 - Y_0 \mid X)] \\ & = (\text{A2}) = E_X[E(Y_1 - Y_0 \mid Q{=}1, X)] \\ & = (\text{RA}) = E_X[E(Y_1 \mid R{=}1, X)] - E_X[E(Y_0 \mid R{=}0, X)] \\ & = E_X[E(Y \mid R{=}1, X)] - E_X[E(Y \mid R{=}0, X)] \qquad (2)\end{aligned}$$

The empirical counterpart can be derived in several ways; we consider in particular reweighing and matching estimators, both ignoring and allowing for selective non-response to the survey and/or to the earnings question.

1) Reweighing

1a) Ignoring survey and item non-response

We start by considering the case in which non-response can be safely ignored (see the required conditions in Section A1.2).

As to the first term of equation (2):

$$\begin{aligned}& E_X[E(Y \mid R{=}1, X)] \\ & = \int E(Y \mid R=1, x) f(x) dx \\ & = \int E(Y \mid R=1, x) \frac{f(x)}{f(x \mid R=1)} f(x \mid R=1) dx \\ & \equiv \int E(Y \mid R=1, x) \omega_1(x) f(x \mid R=1) dx \\ & = \int E(\omega_1(x) Y \mid R=1, x) f(x \mid R=1) dx\end{aligned}$$

Using first Bayes' rule (second line), and then the law of iterated expectations (third line):

$$\begin{aligned}P(R{=}1) \quad & = P(R{=}1) \mid Q{=}0) \cdot P(Q{=}0) + P(R{=}1 \mid Q{=}1) \cdot P(Q{=}1) \\ & = P(R{=}1) \mid Q{=}1) \cdot \text{P}(Q{=}1)\end{aligned}$$

noting that $P(R{=}1 \mid Q{=}0) = 0$.

We thus get:

$$\begin{aligned}\omega_1(x) &\equiv \frac{f(x)}{f(x \mid R=1)} \\ &= \frac{f(x)}{\left(\frac{P(R=1 \mid x) f(x)}{P(R=1)}\right)} = \frac{P(R=1)}{P(R=1 \mid x)} \\ &= \frac{P(Q=1) P(R=1 \mid Q=1)}{P(Q=1 \mid x) P(R=1 \mid Q=1, x)} \\ &\equiv \frac{(1-p) p_R}{(1-p(x)) p_R(x)} \overset{RA}{=} \frac{1-p}{1-p(x)}\end{aligned}$$

where

- $p_R \equiv P(R{=}1 \mid Q{=}1)$ is the probability of being randomly assigned to the programme group conditional on participating in the ERA study ($Q{=}1$), and
- $p_R(x) \equiv P(R{=}1 \mid Q{=}1, x)$ is the corresponding conditional probability.

Under randomisation, $p_R{=}p_R(x)$.

The first term:

$E_X[E(Y \mid R{=}1, X)] = E_X[E(\omega_1(x) \cdot Y \mid R{=}1, X) \mid R{=}1]$

can hence be estimated by reweighing the outcomes of the programme group by $\omega_1(x)$ and averaging them:

$$\frac{1}{\#(R=1)} \sum_{i \in \{R=1\}} \omega_1(x_j) y_j = \frac{(1-p) p_R}{\#(R=1)} \sum_{i \in \{R=1\}} \frac{y_i}{\left(1-p(x_i)\right) p_R(x_i)}$$

Under randomisation, $p_R{=}p_R(x)$ so that:

$$\frac{1}{\#(R=1)} \sum_{i \in \{R=1\}} \omega_1(x_j) y_j = \frac{1-p}{\#(R=1)} \sum_{i \in \{R=1\}} \frac{y_i}{1-p(x_i)}$$

Similarly, the second term of (2) can be rewritten as:

$E_X[E(Y \mid R{=}0, X)]$

$= \int E(\omega_0(x) Y \mid R=0, x) f(x \mid R=0) dx = E_X[E(\omega_0(x) \cdot Y \mid R{=}0, X) \mid R{=}0]$

where (noting that due to randomisation, the weight ω is the same):

$$\begin{aligned}\omega_0(x) &\equiv \frac{f(x)}{f(x \mid R=0)} = \frac{P(Q=1)P(R=0 \mid Q=1)}{P(Q=1 \mid x)P(R=0 \mid Q=1,x)} \\ &\equiv \frac{(1-p)(1-p_R)}{(1-p(x))(1-p_R(x))} \overset{RA}{=} \frac{1-p}{1-p(x)} = \omega_1(x)\end{aligned}$$

which can be estimated by reweighing the outcomes of the control group and averaging them as follows:

$$\frac{1}{\#(R=0)} \sum_{i \in \{R=0\}} \omega_0(x_j) y_j = \frac{(1-p)(1-p_R)}{\#(R=0)} \sum_{i \in \{R=0\}} \frac{y_i}{\left(1-p(x_i)\right)\left(1-p_R(x_i)\right)}$$

Under randomisation, p_R=$p_R(x)$, hence:

$$\frac{1}{\#(R=0)} \sum_{i \in \{R=0\}} \omega_0(x_j) y_j = \frac{1-p}{\#(R=0)} \sum_{i \in \{R=0\}} \frac{y_i}{1-p(x_i)}$$

We can thus estimate the ATE in (2) by reweighing and averaging the outcomes of the full group of participants (Q=1):

$$\hat{ATE} = \left[\frac{(1-p)p_R}{\#(R=1)} \sum_{i \in \{R=1\}} \frac{y_i}{\left(1-p(x_i)\right)p_R(x_i)} \right] - \left[\frac{(1-p)(1-p_R)}{\#(R=0)} \sum_{i \in \{R=0\}} \frac{y_i}{\left(1-p(x_i)\right)\left(1-p_R(x_i)\right)} \right]$$

Taking full advantage of the randomisation and noting that #(R=1) = #(R=0) due to the 50-50 random allocation:

$$\hat{ATE} = \frac{1-p}{\#(R=1)} \sum_{i \in \{Q=1\}} \frac{R_i y_i - (1-R_i) y_i}{1-p(x_i)}$$

However, although randomisation worked very well, especially when conditioning on X there might be residual imbalances due to pure chance. Even more crucially, this analysis can only be performed for the survey subgroup, and indeed for that subgroup of survey respondents who responded to the earnings question. For this reason, in implementing this estimator we allow for the more general case, as outlined in the following.

1b) Allowing for survey and item non-response

Assume the selection-on-observables assumption in terms of impacts:

(A2) $\quad E(Y_1 - Y_0 \mid Q{=}1, X) = E(Y_1 - Y_0 \mid Q{=}0, X) \qquad$ hence $= E(Y_1 - Y_0 \mid X)$

Outcomes Y are observed only for a subsample of participants (survey respondents who answered the earnings question). The assumptions discussed in Section A1.2 would allow us to ignore such non-response. Here, by contrast, we want to allow for selective non-response, provided such selection into the responding sample happens only in terms of observable characteristics. Correspondingly, we relax the assumptions from Section A1.2 by invoking them conditional on X:

(E-V'.X) (a) $E(Y_1 \mid R{=}1, S{=}1, X) = E(Y_1 \mid R{=}1, S{=}0, X)$ and
(b) $E(Y_0 \mid R{=}0, S{=}1, X) = E(Y_0 \mid R{=}0, S{=}0, X)$

Then,

$$\begin{aligned} ATE \equiv E(Y_1 - Y_0) &= E_X[E(Y_1 - Y_0 \mid X)] \\ &= (\text{A2}) = E_X[E(Y_1 - Y_0 \mid Q{=}1, X)] \\ &= (\text{RA}) = E_X[E(Y_1 \mid R{=}1, X)] - E_X[E(Y_0 \mid R{=}0, X)] \\ &= (\text{E-V'}.X) = E_X[E(Y_1 \mid R{=}1, S{=}1, X)] - E_X[E(Y_0 \mid R{=}0, S{=}1, X)] \\ &= E_X[E(Y \mid R{=}1, S{=}1, X)] - E_X[E(Y \mid R{=}0, S{=}1, X)] \end{aligned} \quad (3)$$

Under the stated assumptions[4], ATE is thus identified in the data and can be empirically estimated as follows.

As to the first term of expression (3):

$$E_X[E(Y \mid R{=}1, S{=}1, X)]$$

$$= \int E(Y \mid R=1, S=1, x) f(x) dx$$

$$= \int E(Y \mid R=1, S=1, x) \frac{f(x)}{f(x \mid R=1, S=1)} f(x \mid R=1, S=1) dx$$

$$\equiv \int E(Y \mid R=1, S=1, x) \omega_1(x) f(x \mid R=1, S=1) dx$$

$$= \int E(\omega_1(x) Y \mid R=1, S=1, x) f(x \mid R=1, S=1) dx$$

with

$$\begin{aligned} \omega_1(x) &\equiv \frac{f(x)}{f(x \mid R=1, S=1)} \\ &= \frac{P(R=1, S=1)}{P(R=1, S=1 \mid x)} = \frac{P(Q=1) P(R=1, S=1 \mid Q=1)}{P(Q=1 \mid x) P(R=1, S=1 \mid Q=1, x)} \\ &\equiv \frac{(1-p) p_{RS1}}{(1-p(x)) p_{RS1}(x)} \end{aligned}$$

---

[4] An alternative set of assumptions yielding the same expression for the ATE is:
Selection on observables in terms of ERA study participation (impact formulation):
(A2) $E(Y_1 - Y_0 \mid Q{=}1, X) = E(Y_1 - Y_0 \mid Q{=}0, X)$ hence $= E(Y_1 - Y_0 \mid X)$
Selection on observables in terms of non-response (impact formulation):
(E-V.X) $E(Y_1 - Y_0 \mid Q{=}1, X) = E(Y_1 - Y_0 \mid Q{=}1, S{=}1, X)$
Random assignment keeps holding given X within responding sample:
(I-V.X) $E(Y_1 \mid S{=}1, R{=}1, X) = E(Y_1 \mid S{=}1, R{=}0, X) = E(Y_1 \mid S{=}1, X)$
$E(Y_0 \mid S{=}1, R{=}1, X) = E(Y_0 \mid S{=}1, R{=}0, X) = E(Y_0 \mid S{=}1, X)$
Then:

$$\begin{aligned} ATE \equiv E(Y_1 - Y_0) &= E_X[E(Y_1 - Y_0 \mid X)] \\ &= (\text{A2}) = E_X[E(Y_1 - Y_0 \mid Q{=}1, X)] \\ &= (\text{E-V.}X) = E_X[E(Y_1 - Y_0 \mid S{=}1, X)] \\ &= (\text{I-V.}X) = E_X[E(Y \mid R{=}1, S{=}1, X)] - E_X[E(Y \mid R{=}0, S{=}1, X)] \end{aligned}$$

where $p_{RS1} \equiv P(R{=}1, S{=}1 \mid Q{=}1)$ is the probability among ERA study participants of being randomly assigned to the programme group **and** of responding to the survey (indeed to the earnings question), and

$p_{RS1}(x) \equiv P(R{=}1, S{=}1 \mid Q{=}1, x)$ is the corresponding conditional probability.

$E_X[E(Y \mid R{=}1, S{=}1, X)] = E_X[E(\omega_1(x)\cdot \mathrm{Y} \mid R{=}1, S{=}1, X) \mid R{=}1, S{=}1]$

can hence be estimated by reweighing by $\omega_1(x)$ the outcomes of the programme group members who responded to the earnings question and averaging them over this subgroup:

$$\frac{1}{\#(R=1,S=1)} \sum_{i\in\{R=1,S=1\}} \omega_1(x_j) y_j = \frac{(1-p)p_{RS1}}{\#(R=1,S=1)} \sum_{i\in\{R=1\}} \frac{y_i}{\left(1-p(x_i)\right) p_{RS1}(x_i)}$$

Similarly, the second term of expression (3) can be rewritten as:

$E_X[E(Y \mid R{=}0, S{=}1, X)] = E_X[E(\omega_0(\mathrm{x})\cdot Y \mid R{=}0, S{=}1, X) \mid R{=}0, S{=}1]$

with

$$\begin{aligned} \omega_0(x) &\equiv \frac{f(x)}{f(x \mid R=0, S=1)} = \frac{P(Q=1)P(R=0,S=1 \mid Q=1)}{P(Q=1 \mid x)P(R=0,S=1 \mid Q=1,x)} \\ &\equiv \frac{(1-p)p_{RS0}}{(1-p(x))p_{RS0}(x)} \end{aligned}$$

where p_{RS0} is the probability among ERA study participants of being randomly assigned to the control group **and** of responding to the survey (indeed to the earnings question). (Note that p_{RS0} is not equal to $1-p_{RS1}$).

This term can be estimated by reweighing the outcomes of the control group who responded to the earnings question and averaging them over this subgroup:

$$\frac{1}{\#(R=0,S=1)} \sum_{i\in\{R=0,S=1\}} \omega_0(x_j) y_j = \frac{(1-p)p_{RS0}}{\#(R=0,S=1)} \sum_{i\in\{R=0,S=1\}} \frac{y_i}{\left(1-p(x_i)\right) p_{RS0}(x_i)}$$

Hence we can estimate the *ATE* in equation (3) by reweighing and averaging the outcomes of all those participants who responded to the survey ($Q{=}1$ and $S{=}1$):

$$\hat{ATE} = \left[\frac{1}{\#(R=1,S=1)} \sum_{i\in\{R=1,S=1\}} \frac{(1-p)p_{RS1}}{\left(1-p(x_i)\right) p_{RS1}(x_i)} y_i\right] - \left[\frac{1}{\#(R=0,S=1)} \sum_{i\in\{R=0,S=1\}} \frac{(1-p)p_{RS0}}{\left(1-p(x_i)\right) p_{RS0}(x_i)} y_i\right]$$

2) Matching

An alternative to the method of directly weighting the outcomes of the (responding) participant group so as to reflect the distribution of observables in the original eligible population is to construct the weights by performing matching.

The latter offers the advantages that the exact specifications of the propensity score and of the response probabilities are not needed and that one can assess the extent of the actual comparability of groups.

This matching-based idea can be implemented in two ways; either to separately recover the missing ATE_0 and then combining it with the experimental ATE_1 to get the ATE, or to recover the ATE directly.

Again, we consider both a situation where non-response is ignored and one where it is not.

2a) Ignoring survey and item non-response

We start again by assuming that, once conditioning on observables X, ERA study participants and non-participants on average experience the same treatment and no-treatment outcomes:

(A2) (a) $E(Y_1 \mid Q{=}1, X) = E(Y_1 \mid Q{=}0, X)$
(b) $E(Y_0 \mid Q{=}1, X) = E(Y_0 \mid Q{=}0, X)$

Ignoring non-response allows one to treat the responding participants as representative of the full group of participants. We make the following assumptions as to non-response:

(E-V) $E(Y_1 - Y_0 \mid Q{=}1) = E(Y_1 - Y_0 \mid Q{=}1, S{=}1)$

(E-V') (a) $E(Y_1 \mid R{=}1) = E(Y_1 \mid R{=}1, S{=}1) = E(Y_1 \mid R{=}1, S{=}0)$
(b) $E(Y_0 \mid R{=}0) = E(Y_0 \mid R{=}0, S{=}1) = E(Y_0 \mid R{=}0, S{=}0)$

(I-V) $E(Y_1 \mid S{=}1, R{=}1) = E(Y_1 \mid S{=}1, R{=}0) = E(Y_1 \mid S{=}1)$
$E(Y_0 \mid S{=}1, R{=}1) = E(Y_0 \mid S{=}1, R{=}0) = E(Y_0 \mid S{=}1)$

(A) Obtaining the *ATE* after having first obtained the ATE_0

Starting from equation (1b):

$$ATE = (1{-}p){\cdot}ATE_1 + p{\cdot}\boldsymbol{ATE_0} \qquad (1b)$$

- p is observed
- ATE_1 $\equiv E(Y_1 - Y_0 \mid Q{=}1) =$ (E-V) $= E(Y_1 - Y_0 \mid Q{=}1, S{=}1) =$ (I-V)
  $= E(Y \mid S{=}1, R{=}1) - E(Y \mid S{=}1, R{=}0)$.
  Note that we control for X in deriving this estimate.

To recover the ATE_0, we need to estimate $E(Y_1 \mid Q{=}0)$ and, given the absence of survey outcomes for non-participants, $E(Y_0 \mid Q{=}0)$ as well.

- $E(Y_1 \mid Q{=}0) =$ (A2) $= E_X[E(Y_1 \mid Q{=}1, X) \mid Q{=}0] =$ (RA) $= E_X[E(Y_1 \mid R{=}1, X) \mid Q{=}0]$
  $=$ (E-V'.X) $= E_X[E(Y \mid S{=}1, R{=}1, X) \mid Q{=}0]$
  Match to each non-participant in the $Q{=}0$ group one or more 'similar' individuals from the pool of responding programme group members ($S{=}1$, $R{=}1$) and take the latter's reweighted outcomes.
- $E(Y_0 \mid Q{=}0) =$ (A2), (RA), (E-V'.X) $= E_X[E(Y \mid S{=}1, R{=}0, X) \mid Q{=}0]$
  Match to each non-participant in the $Q{=}0$ group one or more 'similar' individuals from the pool of responding control group members ($S{=}1$, $R{=}0$) and take the latter's reweighted outcomes.

With the ATE_0 in hand, we can then use the experimental ATE_1 to get the ATE via (1b).

(B) Obtaining the *ATE* directly

To recover the ATE, we need to estimate $E(Y_1)$ and $E(Y_0)$.

- $E(Y_1)$ $\equiv E(Y_1) \mid Q{=}1 \vee Q{=}0) = (\text{A2}) = E_X(Y_1)[E(Y_1) \mid Q{=}1, X) \mid Q{=}1 \vee Q{=}0]$
  $= (\text{RA}) = E_X[E(Y \mid R{=}1, X) \mid Q{=}1 \vee Q{=}0] = (\text{E-V'}.X) =$
  $= E_X[E(Y \mid R{=}1, S{=}1, X) \mid Q{=}1 \vee Q{=}0]$ or $E_X[E(Y \mid R{=}1, S{=}1, X) \mid (R{=}1, S{=}1) \vee Q{=}0]$

  Match each individual in the group made up by the (Q=0 and Q=1) or the (Q=0 and (R=1, S=1)) groups to individuals in the responding programme group sample (R=1, S=1) and calculate the weight that gets assigned to each individual in the latter group (this weight will be larger than 1). Reweigh the outcomes in this (R=1, S=1) group using these weights and take their average over the (R=1, S=1) group, i.e. use the matched outcome to estimate $E(Y_1)$.
  One can match on the basis of this propensity score $=P(Q{=}0 \mid Q{=}0 \vee (R{=}1, S{=}1), X)$.

- $E(Y_0)$ $\equiv E(Y_0 \mid Q{=}1 \vee Q{=}0) = (\text{A2}) = E_X[E(Y_0 \mid Q{=}1, X) \mid Q{=}1 \vee Q{=}0]$
  $= (\text{RA}) = E_X[E(Y \mid R{=}0, X) \mid Q{=}1 \vee Q{=}0] = (\text{E-V'}.X) =$
  $= E_X[E(Y \mid R{=}0, S{=}1, X) \mid Q{=}1 \vee Q{=}0]$ or $E_X[E(Y \mid R{=}0, S{=}1, X) \mid (R{=}0, S{=}1) \vee Q{=}0]$

  Match each individual in the group made up by the (Q=0 and Q=1) or the (Q=0 and (R=1, S=1)) groups to individuals in the responding control group sample (R=0, S=1) and calculate the weight that gets assigned to each individual in the latter group (this weight will be larger than 1). Reweigh the outcomes in the (R=0, S=1) group using these weights and take their average over the (R=0, S=1) group, i.e. use the matched outcome to estimate $E(Y_0)$.
  One can match on the basis of this propensity score $P(Q{=}0 \mid Q{=}0 \vee (R{=}0, S{=}1), X)$.

Because of random assignment, the two propensity scores above should be the same and should coincide with $p(x)$.

2b) Allowing for survey and item non-response

In this case we weight the outcomes of the respondents among the participants (S=1) so as to reflect the distribution of observables in the full original eligible population.

The first procedure outlined in Subsection 2a) above can correct the ATE_0 for non-response, but would need to be repeated to get a non-response corrected ATE_1 as well:

(A) Obtaining the *ATE* after having first obtained the ATE_0

As in case 2a), to recover the ATE_0, we need to estimate $E(Y_1 \mid Q{=}0)$ and, given the absence of survey outcomes for non-participants, $E(Y_0 \mid Q{=}0)$ as well.

Under (A2) and (E-V'.X):

- $E(Y_1 \mid Q{=}0) = E_X[E(Y \mid R{=}1, S{=}1, X) \mid Q{=}0]$
  This term can be estimated by the matched outcome from matching to each non-participant in the $Q{=}0$ group, one or more 'similar' participants from the $(R{=}1 \text{ and } S{=}1)$ group.
- $E(Y_0 \mid Q{=}0) = E_X[E(Y \mid R{=}0, S{=}1, X) \mid Q{=}0]$
  This term can be estimated by the matched outcome from matching to each non-participant in the $Q{=}0$ group, one or more 'similar' participants from the $(R{=}0 \text{ and } S{=}1)$ group.

However, the experimental contrast obtained as $E(Y \mid R{=}1) - E(Y \mid R{=}0)$ does not take into account non-response.

One could obtain the correct ATE_1 again by reweighing. Under (RA) and (E-V'.X):

- $E(Y_1 \mid Q{=}1) = E_X[E(Y \mid R{=}1, S{=}1, X) \mid Q{=}1]$
  This term can be estimated by the matched outcome from matching to each participant in the full $Q{=}1$ group, one or more 'similar' programme group members from the respondents, i.e. the $(R{=}1 \text{ and } S{=}1)$ group.
- $E(Y_0 \mid Q{=}1) = E_X[E(Y \mid R{=}0, S{=}1, X) \mid Q{=}1]$
  This term can be estimated by the matched outcome from matching to each participant in the full $Q{=}1$ group, one or more 'similar' control group members from the respondents, i.e. the $(R{=}0 \text{ and } S{=}1)$ group.

To allow for non-response it is thus more convenient to follow option (B) and recover the ATE directly:

(B) Obtaining the *ATE* directly

To recover the ATE, we need to estimate $E(Y_1)$ and $E(Y_0)$.

Under (A2), (RA) and (E-V'.X):

- $E(Y_1) \equiv E(Y_1 \mid Q{=}1 \vee Q{=}0) = (\text{A2}) = E_X[E(Y_1 \mid Q{=}1, X) \mid Q{=}1 \vee Q{=}0] = (\text{RA}) = E_X[E(Y \mid R{=}1, X) \mid Q{=}1 \vee Q{=}0] = (\text{E-V'}.X) = E_X[E(Y \mid R{=}1, S{=}1, X) \mid Q{=}1 \vee Q{=}0]$
  Match each individual in the eligible group, i.e. the $Q{=}0$ and $Q{=}1$ groups, to individuals in the subgroup of programme group members who responded to the earnings question $(R{=}1 \text{ and } S{=}1)$ and calculate the weight that gets assigned to each individual in the latter subgroup (this weight will be larger than 1). Reweigh the outcomes in the latter subgroup using these weights and take their average over this subgroup.
  That is, use the matched outcome to estimate $E(Y_1)$.
  One can match on the basis of this propensity score $P(R{=}1 \text{ and } S{=}1 \mid Q{=}0 \vee Q{=}1, X)$.

- $E(Y_0) \equiv E(Y_0 \mid Q{=}1 \vee Q{=}0) = (\text{A2}) = E_X[E(Y_0 \mid Q{=}1, X) \mid Q{=}1 \vee Q{=}0] = (\text{RA}) = E_X[E(Y \mid R{=}0, X) \mid Q{=}1 \vee Q{=}0] = (\text{E-V'}.X) = E_X[E(Y \mid R{=}0, S{=}1, X) \mid Q{=}1 \vee Q{=}0]$

  Match each individual in the eligible group, i.e. the $Q{=}0$ and $Q{=}1$ groups, to individuals in the subgroup of control group members who responded to the earnings question ($R{=}0$ and $S{=}1$) group and calculate the weight that gets assigned to each individual in the latter subgroup (this weight will be larger than 1). Reweigh the outcomes in the latter subgroup using these weights and take their average over this group.

  That is, use the matched outcome to estimate $E(Y_0)$.

  One can match on the basis of this propensity score $P(R{=}0 \text{ and } S{=}1 \mid Q{=}0 \vee Q{=}1, X)$.

Take the difference in the two matched outcomes to obtain the ATE.

3) Analysis of take-up

This analysis aims at answering the two evaluation questions:

1. Are the non-participants individuals who, even if offered ERA services, would not take them up?
2. What kind of involvement would non-participants have had with Jobcentre Plus had they participated in the ERA study and been assigned to the control group?

Let Y be (a measure of) take-up of ERA services.

To answer question (1), we need to estimate $E(Y_1|Q{=}0)$. Under assumption (A2.a):

$$E(Y_1 \mid \text{Q}{=}0) = E_X[E(Y_1 \mid Q{=}0, X) \mid Q{=}0] = (\text{A2.a}) = E_X[E(Y_1 \mid Q{=}1, X) \mid Q{=}0] = (\text{RA}) = E_X[E(Y \mid R{=}1, X) \mid Q{=}0]$$

To implement this estimator, match to each non-participant one or more 'similar' individuals from the pool of programme group members and take the latter's reweighted outcomes.

A similar type of analysis can be performed on the non-participants and the control group to answer question (2) under assumption (A2.b).

2.3 Point estimate under selection on unobservables

This section outlines a set of approaches which allow selection into the group of ERA study participants to depend on outcome-relevant **unobservables**. All of these models fall within the family of 'control function models' and build on the classical sample selection model introduced by Heckman (1979).

As already outlined at the beginning of Section A2.2, this class of models rely on the assumption that ERA and non-ERA outcomes among the eligibles are not affected by whether an individual is **offered the chance** to participate in the ERA study or not.

Note also that due to the lack of a credible instrument, when assessing survey outcomes we rule out selective non-response based on **unobservables**.

As emphasised in the main report, our unique set-up – randomisation coupled with administrative outcomes which are observed for the selected-out sample – allows us to:

(a) test the exclusion restriction of the instrument;

(b) test for the presence of residual selection on unobservables related to no-treatment employment or benefit outcomes;

(c) test how well the various control function models capture the presence and direction of the selection on unobservables we have thus uncovered; and

(d) test how well the various control function models predict the no-treatment outcome for the non-participants.

We start by presenting tests (a) and (b), then move on to describe the various models in some detail, outlining in each case the specific form that tests (c) and (d) take on.

2.3.1 Some initial tests: exclusion restriction and selection on specific unobservables

The following two tests exploit the facts that:

1. the control group is a randomised subset – and hence representative – of the participants
2. the control group, like the non-participants, does not receive ERA, so that for both groups, the observed outcome coincides with the non-ERA outcome Y_0:

   $E(Y \mid R{=}0, X) = E(Y_0 \mid Q{=}1, X)$ and $E(Y \mid Q{=}0, X) = E(Y_0 \mid Q{=}0, X)$
3. for administrative data, the outcomes of the non-participants are observed.

Testing (part of) the exclusion restriction of the instrument

The control function model crucially relies on an exclusion restriction for non-parametric identification. Specifically, we need an observable variable Z satisfying:

(A3) (a) $P(Q{=}1 \mid X, Z)$ is a non-trivial function of Z
(b) $E(Y_0 \mid X, Z) = E(Y_0 \mid X)$
(c) $E(Y_1 \mid X, Z) = E(Y_1 \mid X)$

In other words, the instrument should affect the decision to participate in the ERA study (condition A3.a), but should not otherwise affect potential outcomes directly (conditions A3.b and c).

The power of the instrument in the first stage, i.e. condition (A3.a), can, as usual, be tested. In our case of randomisation coupled with administrative outcome data covering all eligibles, also part the exclusion restriction (condition A3.b) can

be tested when modelling administrative outcomes. The test is implemented by pooling the control group and the non-participants, regressing their observed outcomes on X and Z, and testing the significance of Z. In the following standard OLS model, one would test the null that η is zero:

$$Y_i = Y_{0i} = \beta_0 X_i + \eta Z_i + u_i$$

Testing for selection on specific unobservables

Due to our unique set-up, we are also in a position to test whether there remain differences between participants and non-participants in terms of unobservables related to no-treatment employment or benefit outcomes.

This test is implemented by assessing whether, once controlling for X, the outcomes of the non-participants differ on average from those of control group. Formally, this is a test of:

(A2-b) $$E(Y_0 \mid Q{=}1, X) = E(Y_0 \mid Q{=}0, X)$$

This test can be simply performed by running a regression on the pooled sample of controls and non-participants of observed outcomes Y on the group dummy variable G controlling for X and testing the significance of α:

$$Y = \alpha G + \gamma X + \varepsilon$$

Instead of simple OLS, one could run a Probit or Tobit model whenever the outcome of interest is binary or censored. Also, to minimise all sensitivity to the specification of how the observables should enter the outcome equation or affect differences between the two groups, one can instead perform matching between the two groups (matching to each non-participant one or more similar individual from the ERA control group) and test for the equality of the mean outcomes of the two matched groups. One could also perform the test within propensity score bands (via stratification), to explore whether violations occur at different non-participation probabilities.

The results of this test are not just informative in themselves, but as we show below, form the basis for constructing an important specification check for any given control function model.

2.3.2 Standard control function approach

The non-participation problem can be fruitfully framed as the classical sample selection problem: the treatment outcome is only observed for the ERA study participants (via its representative $R{=}1$ subgroup), but is not observed for the non-participants. In case of survey-based outcomes, it also is the case that the no-treatment outcome is only observed for the participants (via its $R{=}0$ subgroup), but is unobserved for the non-participants.

The model of potential treatment and no-treatment outcomes for the eligible population is the same as the one considered in the previous section, apart than from the additional specification of the distribution of the unobservables:

$$Y_{0i} = \beta_0 X_i + u_i \qquad u_i \sim N(0, \sigma_u^2)$$
$$Y_{1i} = \beta_1 X_i + u_i + b_i \qquad b_i \sim N(0, \sigma_b^2)$$

Treatment outcomes Y_1 are, however, only observed for study participants (as represented by the programme group), not for the non-participants. In case (b), no-treatment outcomes Y_0 are similarly only observed for study participants (as represented by the control group).

Let the observability rule for Y_1 (and Y_0 in case (b)) be:

$$Q_i = 1(\gamma W_i + v_i \geq 0) \qquad W_i = [X_i, Z_i]$$
$$v_i \sim N(0, 1)$$
$$Corr(v_i, u_i) = \rho_{uv}$$
$$Corr(v_i, b_i) = \rho_{bv}$$

The model thus allows for selection into the ERA study based on both unobserved 'ability' (u) and unobserved individual-specific ERA impacts (b).

The crucial set of assumptions implicit in this model is:

(A3) (u_i, b_i, v_i) is a mean zero normal random vector that is statistically independent of W
(note that $Var(v_i)$ is normalised to 1); and $\gamma_Z \neq 0$.

Apart from the parametric choice of the distribution of the unobservables implied by this assumption (in particular, joint normality and homoskedasticity), the control function model crucially relies on an exclusion restriction for non-parametric identification. Specifically, we need an observable variable Z which is contained in W, i.e. which affects the decision to participate in the ERA study (the Q=1 decision), but is not contained in X, i.e. does not affect potential outcomes directly. A way to rewrite assumption (A3) to make these conditions explicit is:

(A3)
(a) $P(Q=1 \mid X, Z)$ is a non-trivial function of Z
(b) $E(Y_0 \mid X, Z) = E(Y_0 \mid X)$
(c) $E(Y_1 \mid X, Z) = E(Y_1 \mid X)$
(d)
$$\begin{pmatrix} u \\ b \\ v \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho_{ub} & \rho_{uv} \\ \rho_{ub} & \sigma_b^2 & \rho_{bv} \\ \rho_{uv} & \rho_{bv} & 1 \end{pmatrix}\right]$$

The power of the instrument in the first stage, i.e. condition (A3.a), can, as usual, be tested. As shown in Section A2.3.1 above, though, in our case, even the exclusion restriction (A3.b) can be tested when modelling administrative outcomes. The parametric assumptions in (A3.d) can be relaxed (and thus tested), as we show in the next subsections.

Going back to our standard selection model, from the joint normality assumption it follows that:

$u_i = \rho_{uv}\,\sigma_u\,v_i + \xi_{ui}$ with $\xi_{ui} \perp v_i$
$b_i = \rho_{bv}\,\sigma_b\,v_i + \xi_{bi}$ with $\xi_{bi} \perp b_i$

For ERA study participants:

$$E(Y_1 \mid Q{=}1, W) = E(Y_1 \mid Q{=}1, R{=}1, \boldsymbol{W}) = \beta_1 X + E(u \mid v > -\gamma W) + E(b \mid v > -\gamma W)$$
$$= \beta_1 X + (\rho_{uv}\sigma_u + \rho_{bv}\sigma_b)\frac{\phi(\gamma W)}{\Phi(\gamma W)} \equiv \beta_1 X + (\lambda_{uv} + \lambda_{bv})H_1(\gamma W)$$
$$E(Y_0 \mid Q{=}1, W) = E(Y_0 \mid Q{=}1, R{=}0, W) = \beta_0 X + E(u \mid v > -\gamma W)$$
$$= \beta_0 X + \rho_{uv}\sigma_u \frac{\phi(\gamma W)}{\Phi(\gamma W)} \equiv \beta_0 X + \lambda_{uv} H_1(\gamma W)$$

Pooling the two subgroups (i.e. programme and control groups) of the participants allows modelling **observed** outcomes Y as:

$$Y = \beta_0 X(1-R) + \beta_1 XR + \lambda_{uv} H_1 + \lambda_{bv} H_1 R + \varepsilon$$

Under the assumptions of the model, the expected unobserved Y_1 for non-participants of characteristics W can be predicted as:

$$E(Y_1 \mid Q{=}0, W) = \beta_1 X + E(u \mid v \le -\gamma W) + E(b \mid v \le -\gamma W)$$
$$= \beta_1 X + (\rho_{uv}\sigma_u + \rho_{bv}\sigma_b)\left(-\frac{\phi(\gamma W)}{1-\Phi(\gamma W)}\right) \equiv \beta_1 X + (\lambda_{uv} + \lambda_{bv})H_0(\gamma W)$$

In case (b), the expected unobserved Y_0 for the non-participants of characteristics W can be predicted as:

$$E(Y_0 \mid Q{=}0, W) = \beta_0 X + E(u \mid v \le -\gamma W)$$
$$= \beta_0 X + \rho_{uv}\sigma_u \left(-\frac{\phi(\gamma W)}{1-\Phi(\gamma W)}\right) \equiv \beta_0 X + \lambda_{uv} H_0(\gamma W)$$

The model can be easily estimated in two steps:

Estimation procedure

1. estimate γ from first-step probit on Q=0 and Q=1 controlling for X and for Z; construct the inverse Mills ratios for each individual, $H_1(\hat{\gamma}W)$ for participants and $H_0(\hat{\gamma}W)$ for non-participants;
2. estimate the β and λ parameters by OLS in the augmented regression on the participants sample (NB: regression includes an intercept and the R indicator).
   a. t-test on λ_{uv} for selection into participation in the ERA study based on unobserved ability;
   b. t-test on λ_{bv} for selection into participation in the ERA study based on unobserved individual impacts;

3. for each individual i (actually, for each individual of type w_i) in the Q=0 group, use the estimates in step 2 to predict $E(Y_1 \mid Q=0, W=w_i)$ and, for case (b), $E(Y_0 \mid Q=0, W=w_i)$ as well;
4. average over all Q=0 individuals to get $E(Y_1 \mid Q=0) = E_W[E(Y_1 \mid Q=0, W) \mid Q=0]$ and, for case (b), $E(Y_0 \mid Q=0) = E_W[E(Y_0 \mid Q=0, W) \mid Q=0]$ as well.

In case (b), in order to identify the ATE, we need to identify both $E(Y_1 \mid Q=0, W)$ and $E(Y_0 \mid Q=0, W)$ and the full model needs to be estimated under the full set of assumptions (A3).

In case (a), i.e. when (administrative) data on the non-participants outcomes is available, to identify the ATE, only $E(Y_1 \mid Q=0, W)$ needs to be identified. In this case, one does not need to make assumption (A3-b) (and one only needs to know that (u_{1i}, v_i) is a mean zero normal random vector, where $u_{1i} \equiv u_i + b_i$). However, without invoking this assumption one cannot separately test for selection based on u as opposed to selection based on b, but just for selection on unobservables. In case (a), using the R=0 group and assuming (A3-b) allows one to estimate the two ρ's separately, i.e.

- selection into the ERA study based on unobserved 'ability' (u);
- selection into the ERA study based on unobserved individual-specific ERA impacts (b).

Since such evidence might be of interest in its own right and since it would be in general hard to argue that (A3-c) holds but not (A3-b), in the following we always invoke the full set of assumptions.

As anticipated, when focusing on administrative outcomes we are in the unusual position of being able to test two features of the performance of the selection model.

Testing how well the control function model captures the actual extent of selection on unobservables

In the following we derive the expression for the standard control function model which is equivalent to the difference in average outcomes for non-participants compared to participants (as represented by the control group) with the same observed characteristics X obtained via OLS. In order to do this, define the binary indicator G=1 if control group member and G=0 if non-participant. The OLS regression described in Section A2.3.1 above would test for selection on unobservables by testing the significance of α in:

$$Y = Y_0 = \alpha G + \gamma X + \varepsilon$$

Note in particular that $\alpha = E(Y_0 \mid X, Q=1) - E(Y_0 \mid X, Q=0)$.

The specification for the no-treatment outcome (for the eligible population) is:

$$Y_0 = \beta_0 X + u$$

Average no-treatment outcomes for the two selected groups of participants and non-participants are thus:

$$E(Y_0 \mid X, Q{=}1) = \beta_0 X + E(u \mid X, Q{=}1)$$
$$E(Y_0 \mid X, Q{=}0) = \beta_0 X + E(u \mid X, Q{=}0)$$

It thus follows that

$$\alpha \equiv E(Y_0 \mid X, Q{=}1) - E(Y_0 \mid X, Q{=}0) = E(u \mid X, Q{=}1) - E(u \mid X, Q{=}0)$$

that is, α is equal to the difference of the two mean unobservables in the selected samples.

Such mean unobservables are constructed by the selection model as the two 'control function' terms:

$$\hat{E}(u \mid W, Q = 1) = \hat{\lambda}_{uv} \bar{H}_{1|Q=1}$$

$$\hat{E}(u \mid W, Q = 0) = \hat{\lambda}_{uv} \bar{H}_{0|Q=0}$$

To recover the necessary parameters, start by considering that participation in the ERA study takes places according to:

$$Q = 1(\mu_0 X + \mu_1 Z + v \geq 0)$$

After having recovered the estimated linear index from a probit model of Q on X and Z (note that we use both the programme and control group in the Q=1 group), we can construct the inverse Mills ratios:

$$H = H_1 \equiv \phi(\text{index})/\Phi(\text{index}) \qquad \text{if } Q{=}1$$

$$H = H_0 \equiv -\phi(\text{index})/(1-\Phi(\text{index}) \qquad \text{if } Q{=}0 \ (\equiv G{=}0)$$

From the regression of Y_0 on X and H_1 for the G=1 group (as represented by the controls), we obtain an estimate of λ_{uv} (and of β_0):

$$E(Y_0 \mid G{=}1) = \lambda_{uv} X + \beta H_1(X, Z)$$

The control function terms are then simply obtained from the estimated λ_{uv} and the mean inverse Mills ratios in the two groups.

Given that the different control function models recover potentially different estimates of such mean unobservables, the difference between α and the two control function terms provides a ready metric to 'order' the performance of these models.

Testing how well the control function model predicts Y_0 for non-participants

We can use the estimated coefficients β and λ and the constructed mills variable H_0 to recover the average predicted Y_0 for the non-participants (G=0):

$$E(Y_0 \mid G{=}0) = \beta_0 X + \lambda_{uv} H_0(X, Z)$$

The average Y_0 which our control function model predicts for non-participants, $E(Y_0 \mid G{=}0)$, is then compared to the average **observed** Y_0 for the non-participants, $E(Y \mid G{=}0)$.

2.3.3 Control function approach relaxing independence

The assumptions (A3) of the standard control function model included independence of the observed characteristics X from the unobservable determinants of treatment and no-treatment outcomes, u and b. While still requesting the X's to be non-correlated with these unobservables, independence can be relaxed to allow for heteroskedasticity of u and b, as well as for the covariances between u and v, and between b and v to depend on X. The latter basically means that the selection process is allowed to be different for customers with different observables.

One way to represent this is to allow the effect of the X's on treatment and no-treatment outcomes to have an unobserved idiosyncratic component η (random effect models):

$$Y_{0i} = (\beta_0 + \eta_{0i})X_i + u_{0i} \qquad e_{0i} \equiv \eta_{0i}X_i + u_{0i}$$
$$Y_{1i} = (\beta_1 + \eta_{1i})X_i + u_{1i} \qquad e_{1i} \equiv \eta_{1i}X_i + u_{1i}$$

It follows that the average no-treatment outcome for the participants can be written out as (note that we are still maintaining the normality assumption):

$$\begin{aligned}
E(Y_0 \mid Q{=}1, W) &= E(Y_0 \mid v > -\gamma W) \\
&= \beta_0 X + E(u_0 \mid v > -\gamma W) + \Sigma_k X_k E(\eta_{0k} \mid v > -\gamma W) \\
&= \beta_0 X + E(u_0 \mid v > -\gamma W) + x_1 E(\eta_{01} \mid v > -\gamma W) + .. + x_K E(\eta_{0k} \mid v > -\gamma W) \\
&= \beta_0 X + \rho_0 \sigma_0 \frac{\phi(\gamma W)}{\Phi(\gamma W)} + \sum_k x_k \rho_{\eta_{0k} v} \sigma_{\eta_{0k}} \frac{\phi(\gamma W)}{\Phi(\gamma W)} \\
&\equiv \beta_0 X + \lambda_0 H_1(\gamma W) + \sum_k x_k \lambda_{0k} H_1(\gamma W) \\
&= \beta_0 X + (\lambda_0 + x_1\lambda_{01} + \ldots + x_K\lambda_{0K}) H_1(\gamma W)
\end{aligned}$$

The unknown parameter vectors β_0 and λ_0 can once again be estimated from a regression of the observed outcomes of the control group ($Q{=}1$, $R{=}0$), in which this time the inverse Mills ratio has been interacted with all of the X's. Note that we can test the null that the error is homoskedastic by performing an F-test on the joint significance of such interaction terms.

One possibility to 'summarise' the extent of selection on unobserved characteristics for the participants is by considering:

$$s_0 \equiv E(\lambda_0(X) \mid Q{=}1) = \lambda_0 + \lambda_{01}\ \overline{x}_{1|Q=1} + \ldots + \lambda_{0K}\ \overline{x}_{K|Q=1}$$

A way to directly obtain an estimate of s_0 as well as its standard error is to reparametrise the augmented outcome equation for $Q{=}1$ group as:

$$Y = Y_0 = \beta_0 X + s_0 H_1 + \lambda_{01}(X_1 - \overline{x}_{1|Q=1})H_1 + \ldots + \lambda_{0K}(X_K - \overline{x}_{K|Q=1})H_1 + e$$

In practice, for the controls we regress the observed (no-treatment) outcome on X, the inverse Mills ratio H_1 and all interaction variables of the latter with the X's in deviation from their means among the participants. We can then directly look at the significance and sign of s_0, the coefficient on H_1.

To compare the control function terms to α, we estimate them as:

$$\hat{E}(u \mid W, Q=1) = \left(\hat{\lambda}_0 + \hat{\lambda}_{01}\overline{x}_{1|Q=1} + \ldots + \hat{\lambda}_{0K}\overline{x}_{K|Q=1}\right)\overline{H}_{1|Q=1}$$

$$\hat{E}(u \mid W, Q=0) = \left(\hat{\lambda}_0 + \hat{\lambda}_{01}\overline{x}_{1|Q=0} + \ldots + \hat{\lambda}_{0K}\overline{x}_{K|Q=0}\right)\overline{H}_{0|Q=0}$$

We predict no-treatment outcomes for the non-participants (which we can then compare to their observed no-treatment outcomes) as:

$$E(Y_0 \mid Q{=}0,\, W) = \beta_0 X + (\lambda_0 + x_1\lambda_{01} + \ldots + x_K\lambda_{0K})\, H_0(\gamma W)$$

Turning now to treatment outcomes, we predict them for the non-participants as:

$$\begin{aligned}
E(Y_1 \mid Q{=}0,\, W) &= \beta_1 X + E(u_1 \mid v \le -\gamma W) + \Sigma_k\, X_k\, E(\eta_{1k} \mid v \le -\gamma W) \\
&= \beta_1 X + E(u_1 \mid v \le -\gamma W) + x_1\, E(\eta_{11} \mid v \le -\gamma W) + \ldots + x_K\, E(\eta_{1K} \mid v \le -\gamma W) \\
&= \beta_1 X + \rho_1\sigma_1\left(-\frac{\phi(\gamma W)}{1-\Phi(\gamma W)}\right) + \sum_k x_k \rho_{\eta_{1k}v}\sigma_{\eta_{1k}}\left(-\frac{\phi(\gamma W)}{1-\Phi(\gamma W)}\right) \\
&\equiv \beta_1 X + \lambda_1\, H_0(\gamma W) + \sum_k x_k \lambda_{1k} H_0(\gamma W) \\
&= \beta_1 X + (\lambda_1 + x_1\lambda_{11} + \ldots + x_K\lambda_{1K})\, H_0(\gamma W)
\end{aligned}$$

where the unknown parameter vectors β_1 and λ_1 have been estimated on the subsample of programme group members:

$$\begin{aligned}
E(Y_1 \mid Q{=}1,\, W) &= E(Y \mid Q{=}1,\, R{=}1,\, W) \\
&= \beta_1 X + (\lambda_1 + x_1\lambda_{11} + \ldots + x_K\lambda_{1K})\, H_1(\gamma W)
\end{aligned}$$

We can summarise selection on unobserved impacts as:

$$s_1 \equiv E(\lambda_1(X) \mid Q{=}1) = \lambda_1 + \lambda_{11}\,\overline{x}_{1|Q=1} + \ldots + \lambda_{1K}\,\overline{x}_{K|Q=1}$$

and again directly obtain it by a suitably reparametrised equation estimated off the programme group.

2.3.4 Control function approach relaxing normality

From the joint normality assumption made among assumptions (A3) it follows that:

$$u_0 = \rho_{01}\, v + \xi$$

(see Section A2.3.2, where we had $\rho_{01} \equiv \rho_{uv}\sigma_u$)

To relax normality, we relax the linearity it implies by adding higher order terms:

$$u_0 = \rho_{01}\, v + \rho_{02}\, v^2 + \rho_{03}\, v^3 + \xi$$

It follows that for participants, the mean of the no-treatment unobservable is:

$$E(u_0 \mid v \geq -\gamma W) = \rho_{01} E(v \mid v \geq -\gamma W) + \rho_{02} E(v^2 \mid v \geq -\gamma W) + \rho_{03} E(v^3 \mid v \geq -\gamma W)$$

For non-participants, the same expression applies, only with the conditioning on $v < -\gamma W$.

One can assume v to be normal without loss of generality; the three terms thus relate to the mean, variance and skewness of a truncated normal. Following Jawitz (2004, Table 1), the expressions for the normalised moments simplify in our case to:

$$E(v \mid v > c) = \frac{\phi(c)}{1-\Phi(c)} \qquad E(v \mid v < c) = -\frac{\phi(c)}{\Phi(c)}$$

$$E(v^2 \mid v > c) = c\frac{\phi(c)}{1-\Phi(c)} + 1 - \Phi(c) \qquad E(v^2 \mid v < c) = -c\frac{\phi(c)}{\Phi(c)} + \Phi(c)$$

$$E(v^3 \mid v > c) = (2+c^2)\frac{\phi(c)}{1-\Phi(c)} \qquad E(v^3 \mid v < c) = -(2+c^2)\frac{\phi(c)}{\Phi(c)}$$

From which it follows that:

$$E(Y_0 \mid Q{=}1, W) = \beta_0 X + \rho_{01} H_{11}(\gamma W) + \rho_{02} H_{12}(\gamma W) + \rho_{03} H_{13}(\gamma W)$$
$$\mathrm{E}(\mathrm{Y}_0 \mid \mathrm{Q}{=}0, \mathrm{W}) = \beta_0 X + \rho_{01} H_{01}(\gamma W) + \rho_{02} H_{02}(\gamma W) + \rho_{03} H_{03}(\gamma W)$$

where

$$H_{11}(\gamma W) \equiv \frac{\phi(\gamma W)}{\Phi(\gamma W)} \qquad H_{01}(\gamma W) \equiv -\frac{\phi(\gamma W)}{1-\Phi(\gamma W)}$$

$$H_{12}(\gamma W) \equiv (-\gamma W)\frac{\phi(\gamma W)}{\Phi(\gamma W)} + \Phi(\gamma W) \qquad H_{02}(\gamma W) \equiv \gamma W \frac{\phi(\gamma W)}{1-\Phi(\gamma W)} + 1 - \Phi(\gamma W)$$

$$H_{13}(\gamma W) \equiv \left(2+(\gamma W)^2\right)\frac{\phi(\gamma W)}{\Phi(\gamma W)} \qquad H_{03}(\gamma W) \equiv \left(2+(\gamma W)^2\right)\left(-\frac{\phi(\gamma W)}{1-\Phi(\gamma W)}\right)$$

To compare the control function terms to α, we estimate them as:

$$\hat{E}(u_0 \mid W, Q=1) = \hat{\rho}_{01}\bar{H}_{11|Q=1} + \hat{\rho}_{02}\bar{H}_{12|Q=1} + \hat{\rho}_{03}\bar{H}_{13|Q=1}$$

$$\hat{E}(u_0 \mid W, Q=0) = \hat{\rho}_{01}\bar{H}_{01|Q=0} + \hat{\rho}_{02}\bar{H}_{02|Q=0} + \hat{\rho}_{03}\bar{H}_{03|Q=0}$$

The model for treatment outcomes is:

$$Y_1 = \beta_1 X + u_1$$
$$E(u_1 \mid v) = \rho_{11}\, v + \rho_{12}\, v^2 + \rho_{13}\, v^3$$

Hence, average treatment outcomes for the non-participants are predicted as:

$$E(Y_1 \mid Q{=}0,\, W) = \beta_1 X + E(u_1 \mid v \leq -\gamma W)$$
$$= \beta_1 X + \rho_{11}\, H_{01}(\gamma W) + \rho_{12}\, H_{02}(\gamma W) + \rho_{13}\, H_{03}(\gamma W)$$

where β_1 and the ρ_1's have been estimated on the subsample of programme group members:

$$E(Y_1 \mid Q{=}1,\, W) = E(Y \mid Q{=}1,\, R{=}1,\, W)$$
$$= \beta_1 X + \rho_{11}\, H_{11}(\gamma W) + \rho_{12}\, H_{12}(\gamma W) + \rho_{13}\, H_{13}(\gamma W)$$

2.3.5 Control function approach relaxing both independence and normality

We can combine the previous two extensions to relax both independence and normality. The corresponding regression functions to predict no-treatment and treatment outcomes for the non-participants are:

$$E(Y_0 \mid Q{=}0,\, W)$$
$$= \beta_0 X$$
$$+ (\rho_{01} + x_1\rho_{01\text{-}1} + \ldots + x_K\rho_{01\text{-}K})\, H_{01}(\gamma W)$$
$$+ (\rho_{02} + x_1\rho_{02\text{-}1} + \ldots + x_K\rho_{02\text{-}K})\, H_{02}(\gamma W)$$
$$+ (\rho_{03} + x_1\rho_{03\text{-}1} + \ldots + x_K\rho_{03\text{-}K})\, H_{03}(\gamma W)$$

$$E(Y_1 \mid Q{=}0,\, W)$$
$$= \beta_1 X$$
$$+ (\rho_{11} + x_1\rho_{11\text{-}1} + \ldots + x_K\rho_{11\text{-}K})\, H_{01}(\gamma W)$$
$$+ (\rho_{12} + x_1\rho_{12\text{-}1} + \ldots + x_K\rho_{12\text{-}K})\, H_{02}(\gamma W)$$
$$+ (\rho_{13} + x_1\rho_{13\text{-}1} + \ldots + x_K\rho_{13\text{-}K})\, H_{03}(\gamma W)$$

where the unknown β and ρ parameter vectors are estimated from the group of controls and of programme group members respectively.

2.3.6 Control function approach allowing for censoring

Full tobit selection model

This extension of the standard model takes into account the censored nature of the outcome variable Y. In particular, the outcome is allowed to be censored (at zero in the case of employment duration or earnings) in both the treatment and no-treatment state:

$$Y_{0i} = \begin{cases} \beta_0 X_i + u_{0i} & if\ \beta_0 X_i + u_{0i} > 0 \\ 0 & if\ \beta_0 X_i + u_{0i} \leq 0 \end{cases}$$

$$Y_{1i} = \begin{cases} \beta_1 X_i + u_{1i} & if\ \beta_1 X_i + u_{1i} > 0 \\ 0 & if\ \beta_1 X_i + u_{1i} \leq 0 \end{cases}$$

with $u_{0i} \sim N(0, \sigma_0^2)$ and $u_{1i} \sim N(0, \sigma_1^2)$

Again, treatment outcomes are, however, only observed for study participants (as represented by the programme group), not for the non-participants. In case (b), no-treatment outcomes are similarly only observed for study participants (as represented by the control group).

The observability rule for Y_1 (and Y_0 in case (b)) is thus still given by:

$$Q_i = 1(\gamma W_i + v_i \geq 0) \qquad W_i = [X_i, Z_i]$$
$$v_i \sim N(0, 1)$$
$$Corr(v, u_0) = \rho_0$$
$$Corr(v, u_1) = \rho_1$$

The unknown parameters β_0, β_1, γ, σ_0, σ_1, ρ_0 and ρ_1 are estimated by Maximum Likelihood:

$$L(\beta_0, \beta_1, \gamma, \sigma_0, \sigma_1, \rho_0, \rho_1) =$$
$$[1 - \Phi(\gamma W)]^{1(Q=0)} \cdot$$
$$[\Phi(\gamma W) \cdot P(\beta_0 X + u_0 \leq 0 \mid X, Q=1, R=0)]^{1(Q=1,\ R=0,\ Y=0)} \cdot$$
$$[\Phi(\gamma W) \cdot f_0(y \mid X, Q=1, R=0)]^{1(Q=1,\ R=0,\ Y>0)} \cdot$$
$$[\Phi(\gamma W) \cdot P(\beta_1 X + u_1 \leq 0 \mid X, Q=1, R=1)]^{1(Q=1,\ R=1,\ Y=0)} \cdot$$
$$[\Phi(\gamma W) \cdot f_1(y \mid X, Q=1, R=1)]^{1(Q=1,\ R=1,\ Y>0)}$$

Hence:

$$\ln L(\beta_0, \beta_1, \gamma, \sigma_0, \sigma_1, \rho_0, \rho_1) =$$
$$1(Q=0) \cdot \ln[1 - \Phi(\gamma W)] +$$
$$1(R=0,\ Y=0) \cdot \ln[1 - \Phi(\beta_0 X/\sigma_0) - \Phi_2(-\beta_0 X/\sigma_0, -\gamma W, \rho_0)] +$$
$$1(R=0,\ Y>0) \cdot \{-\ln\sigma_0 + \ln \phi((Y-\beta_0 X)/\sigma_0) + \ln \Phi((\gamma W + \rho_0/\sigma_0\ (Y-\beta_0 X))/(1-\rho_0^2)^{1/2})\} +$$
$$1(R=1,\ Y=0) \cdot \ln[1 - \Phi(\beta_1 X/\sigma_1) - \Phi_2(-\beta_1 X/\sigma_1, -\gamma W, \rho_1)] +$$
$$1(R=1,\ Y>0) \cdot \{-\ln\sigma_1 + \ln \phi((Y-\beta_1 X)/\sigma_1) + \ln \Phi((\gamma W + \rho_1/\sigma_1\ (Y-\beta_1 X))/(1-\rho_1^2)^{1/2})\}$$

When modelling survey outcomes, participants with missing earnings (because they have not been sampled or because of survey or item non-response) are not dropped from the analysis, but contribute to the likelihood in terms of their participation decision. Average ERA impacts can then be estimated both without including them as well as including them, where in the latter case the average treatment effect is calculated – and separately displayed – for them too. Note also that as was always the case when estimating or bounding treatment effects for survey outcomes, attention is restricted to that subsample of participants and non-participants who would have been eligible to be surveyed.

Once the parameters of the model have been estimated, we can also directly test whether there is selection into participation in the ERA study based on unobserved individual gains from ERA by noting that $\lambda_{bv} \equiv \lambda_1 - \lambda_0$ (since $u_{1i} - u_{0i} \equiv b_i$). A chi-squared test of the null that $(\lambda_1 - \lambda_0) = 0$ is thus informative of the presence and direction of selection into ERA based on such unobserved individual impacts.

Furthermore, with the estimated parameters in hand, we can perform a number of 'tests' on the performance of the extended control function model.

Testing how well the control function model captures the actual extent of selection on unobservables

In order to test how well the control function captures the selection on unobservables, α is calculated as the marginal effect of the binary participation variable on the unconditional expected value of Y_0:

$$\alpha = E(Y_0 \mid X, Q{=}1) - E(Y_0 \mid X, Q{=}0)$$

where

$$E(Y_0 \mid X, Q) = (\eta X + \delta Q)\cdot\Phi[(\eta X + \delta Q)/\sigma] + \sigma\,\phi[(\eta X + \delta\, Q)/\sigma]$$

In estimation, G is used instead of Q.

Note thus that α is calculated viewing Q (or in fact G) just as a regressor; α is simply the difference in average Y between the participants and non-participants holding observed characteristics constant. It is thus a direct indicator of selective differences in terms of unobservables.

Its counterpart is calculated by taking account of selection using the control function and separately considering the selected subsamples defined by $Q{=}1$, thus relying on the control function terms to pick up selection on unobservables.

To calculate $E(Y_0 \mid X, Q{=}1) - E(Y_0 \mid X, Q{=}0)$, we start by noting that because of the censoring:

$$E(Y_0 \mid X, Q{=}j) = E(Y_0 \mid Y_0{>}0, X, Q{=}d)\cdot P(Y_0{>}0 \mid X, Q{=}d) \qquad d{=}0,1$$

To simplify notation, define the following (for later we define the values for the treatment state, state j=1, as well):

$k_i \equiv -\gamma W_i$

$h_{ji} \equiv -\beta_j X_i/\sigma_j \qquad j=0,1$

$\delta_j \equiv -1/(1-\rho_j^2)^{1/2} \qquad j=0,1$

where i denotes the individual and j the state.

The following variable substitutions simplify the calculations:

$q_{ji} \equiv u_{ji}/\sigma_j$

$\eta_i \equiv k_i - v_i$

The four elements needed to calculate the two conditional mean functions for the two groups are calculated as follows:

$$P(Y_0 > 0 \mid X, Q = d) = \frac{P(Y_0 > 0, Q = d \mid X)}{P(Q = d \mid X)}$$

Hence (removing the individual i subscript):

$$P(Y_0>0 \mid X, Q=1) = \frac{1-\Phi(h_0)-\Phi(k)+\Phi_2(h_0,k,\rho_0)}{1-\Phi(k)}$$

$$P(Y_0>0 \mid X, Q=0) = \frac{\Phi(k)-\Phi_2(h_0,k,\rho_0)}{\Phi(k)}$$

where Φ_2 is the joint cumulative distribution of the bivariate normal with correlation ρ.

As to the two expectations:

$$\begin{aligned} E(Y_0 \mid Y_0>0, X, Q=d) &= \beta_0 X + E(u_0 \mid u_0>-\beta_0 X, Q=d) \\ &= \beta_0 X + \sigma_0 E(q_0 \mid q_0>h_0, Q=d) \end{aligned}$$

For Q=1, the formula for the first moment of the truncated bivariate normal distribution provided by Maddala (1983, p.368) can be directly used (leaving out the conditioning on X for ease):

$$\begin{aligned} E(q_0 \mid q_0>h_0, Q=1) &= E(q_0 \mid q_0>h_0, v>k) \\ &= P(q_0>h_0, v>k)^{-1} \cdot \{\varphi(h_0)\Phi(\delta_0 (k-\rho_0 h_0)) + \rho_0 \varphi(k)\Phi(\delta_0(h_0-\rho_0 k))\} \end{aligned}$$

Hence:

$$E(Y_0 \mid X, Q=1) = \frac{1-\Phi(h_0)-\Phi(k)+\Phi_2(h_0,k,\rho_0)}{1-\Phi(k)}\beta_0 X + \frac{\sigma_0}{1-\Phi(k)}\{\phi(h_0)\Phi(\delta_0(k-\rho_0 h_0))+\rho_0\phi(k)\Phi(\delta_0(h_0-\rho_0 k))\}$$

For the group of non-participants, the formula in Maddala can be applied considering that conditioning on $v<k$ corresponds to conditioning on $\eta>0$. After rearranging:

$$E(Y_0 \mid X, Q=0) = \frac{\Phi(k)-\Phi_2(h_0,k,\rho_0)}{\Phi(k)}\beta_0 X + \frac{\sigma_0}{\Phi(k)}\{\phi(h_0)\Phi(-\delta_0\rho_0 h_0)+\rho_0\phi(0)\Phi(\delta_0 h_0))\}$$

To test how well the control function captures the selection on unobservables, α calculated as the marginal effect is compared to the difference of the two terms $E(Y_0 \mid X, Q=1)$ and $E(Y_0 \mid X, Q=0)$, which are estimated according to the two formulae above and averaged over the respective subsample.[5]

Testing how well the control function model predicts observed outcomes

In this extended model, the average observed (no-treatment) outcome of the non-participants is compared to the average predicted outcome from the model, using the formula for $E(Y_0 \mid X, Q=0)$ above with estimated parameters $\hat{\gamma}$ (estimated on the non-participants and the full group of participants), $\hat{\beta}_0$, $\hat{\sigma}_0$ and $\hat{\rho}_0$ (*de facto* estimated on the controls), then averaged over the group of non-participants.

The performance of the model is tested also by predicting outcomes for the participants (treatment outcomes for the programme group and no-treatment outcomes for the control group) and comparing them to the observed outcomes.

Furthermore, we use the model to predict the no-treatment outcomes of the programme group and compare these to the observed no-treatment outcomes of the control group, where the latter provide an unbiased estimate of how well the programme group would have fared had they not participated in ERA.

We also estimate average treatment effects for the participants using the extended model (both the ATT and the ATNT, which should coincide) and compare the estimate to the simple mean difference in outcomes of the programme and control group once controlling for X (i.e. regression model), as well as when taking censoring into account (i.e. tobit model).

---

5 Specifically, we retrieve individual predictions and average over individuals. Note in particular that we include the full participating group $Q=1$ when estimating $E(Y_0 \mid X, Q=1)$.

Estimating treatment effects for the non-participants

Finally, we are in a position to estimate the average effect that non-participants would have experienced had they participated in ERA.

We use the estimated coefficients γ, β_1, σ_1 and ρ_1 to recover the average treatment outcome for the Q=0 group, which we then compare to their observed average (no-treatment) outcome.

The unobserved average treatment outcome for the non-participants is estimated as:

$$E(Y_1 \mid X, Q=0) = \frac{\Phi(k)-\Phi_2(h_1,k,\rho_1)}{\Phi(k)}\beta_1 X + \frac{\sigma_1}{\Phi(k)}\{\phi(h_1)\Phi(-\delta_1\rho_1 h_1)+\rho_1\phi(0)\Phi(\delta_1 h_1))\}$$

(In the case of survey outcome information, we also have to estimate their average no-treatment outcome.)

We finally estimate the ATE using the full model, taking observed treatment outcomes for the programme group and predicted treatment outcomes for the control group and the group of non-participants on the one hand, and predicted no-treatment outcomes of the programme group and observed outcomes of the controls and non-participants on the other.