Picking Winners: Measuring the Effectiveness of Selectively Placed Policy Interventions

Melinda Vigh[1]
Chris Elbers[1]

[1] Vrije Universiteit Amsterdam; Tinbergen Institute, The Netherlands

Picking Winners: Measuring the Effectiveness of Selectively Placed Policy Interventions

Melinda Vigh[*] and Chris Elbers

Vrije Universiteit Amsterdam and Tinbergen Institute

Amsterdam, The Netherlands

October 5, 2017

Abstract

In practice, social and development interventions are often targeted at groups or individuals with the largest expected benefits. In such cases, treatment effects are usually affected by selection on unobservable factors. We show that modeling the process of selective intervention placement allows us to correct for this and identify the Average Treatment Effect using observational panel data. We illustrate the estimation method using simulated data, as well as, data on a large-scale sanitation intervention in Mozambique. Our results provide a useful tool for the assessment of targeted policy interventions, and inform decisions on their continuation or replication.

Keywords: Program evaluation; Selection effect; Correlated random coefficients

JEL: D04, C23

1 Introduction

Social and development programs are rarely implemented following an intended or accidental randomization design. Many social programs are tied to some form of measurable conditionality (e.g. conditional cash transfers), while other programs are allocated based on subjective judgement of the implementers. This paper is interested in the latter type of programs, like community-based development programs that rely on the active participation of community groups. The effects of such programs are usually heterogeneous over the population, and the implementers may target the program on groups most in need, or where they expect the highest benefits. The recent interest of governments in "Pay for Success" financing or social impact bonds is a case in point. Quantifying the effectiveness of such programs is not straightforward because the potential program effects will be different between the treatment and control groups. However, information about the program's impact can help policymakers to evaluate the program's efficiency and to decide whether to continue or replicate the program.

We focus on non-randomized policy interventions where the beneficiaries with the highest expected treatment effects or gains in outcomes are more likely to receive the program due to selective intervention placement or targeting. Heckman et al. (2006) called this phenomena "essential heterogeneity" or "selection on gains".

Providing performance- or outcome-based incentives for the implementers is a good motivator for selective intervention placement. We do not consider incentives provided for the beneficiaries. However, Chassang et al. (2012) discuss incentivized selective (randomized controlled) trials where outcome-based incentives are provided for the beneficiaries in order to disentangle the effects of unobserved heterogeneous effort expenditures by the beneficiaries and the effect of treatment. In this paper, we do not investigate the source of treatment effect heterogeneity. Instead, we rely on the incentive scheme to recover information about the heterogeneity of the potential treatment effect that was observed by the program implementers in the field.

We formalize the program implementation environment in a simple model where the program implementers observe a noisy measure of the heterogeneous potential treatment effect across the eligible population, which is unobserved by the researchers. The implementers use this information to select the location of the intervention based on the highest expected impacts. These characteristics are often a realistic scenario in

participatory interventions.

In this setting, we are interested in estimating the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT). The ATE predicts the impact of the program if it was implemented in the whole population or in a randomly selected sample. This requires stronger identifying assumptions compared to the ATT, which measures the actual impacts of the program on the beneficiary population already reached inclusive of the effects of selective program placement. The ATT is of interest to the policymakers, for example, for cost-benefit analysis purposes.[1]

The key insight for the identification of the ATE is that the timing and placement of the intervention delivery contain useful information about the expected treatment effect even when the treatment effect heterogeneity is driven by unobserved factors correlated with the intervention. We can exploit this information summarized in the time-mean of the intervention variable: beneficiaries with a higher expected treatment effect receive the intervention earlier, which results in a higher value for the time-mean of the intervention variable for this group. This identification strategy requires that we have at least three observations over time, that the intervention is gradually rolled out at different beneficiaries between each of the follow-up surveys, and that the beneficiary-specific treatment effect is independent of time. Under these conditions, the correlated random slopes (CRS) model of Wooldridge (2005, 2010) is ideally suited to estimate the ATE in a reduced-form framework.

In order to investigate the performance of the treatment effect estimators in a selection on gains setting, we carry out Monte Carlo simulations using the proposed model of selective intervention placement with binary treatment and normally distributed disturbances. Our findings show that the correlated random slopes regression provides a good linear approximation of the ATE when the correlation between the heterogeneous treatment effect and the binary intervention variable stems from selection on the expected gains of the beneficiaries and the intervention is rolled out over at least 3 time periods.

We illustrate the applicability of the proposed sample selection model in a real-life scenario using a simplified analysis of the effectiveness of the One Million Initiative

---

[1] Note, however, that if selection on gains occurs then the size of ATT changes as the intervention is rolled out. Hence, the estimate of ATT is specific to the time period and the sample that has been reached.

in Mozambique. The sanitation intervention of the One Million Initiative was implemented by local NGOs, who were allowed to select the location of the interventions. They were also rewarded for working with communities where the intervention was successful (all households are using a latrine). Community participation was also a crucial element of the interventions. Therefore, we expect heterogeneity in the treatment effects, and that the implementing NGOs took this heterogeneity into account when deciding on the locations of the interventions. As a consequence, the sanitation interventions were likely to be subject to selection on gains.

We indeed find evidence of selective placement in the sanitation intervention in terms of the incentivized outcome variable (latrine ownership) but not in terms of the non-incentivized outcomes (hand-washing practices). The estimate of the Average Treatment Effect on the Treated is positive and significant for both variables. However, after controlling for selection on gains, we find that the estimate of the Average Treatment Effect of the sanitation intervention on latrine ownership decreases and is not significantly different from zero. This stems from a decrease in the treatment effect for beneficiaries reached at later stages of the program roll-out, pointing to the effectiveness of selective intervention placement.

Assuming time-invariant beneficiary-specific treatment effects, this finding suggests that, there is an optimal stopping point for rolling out the sanitation intervention focused on increasing latrine ownership in the target program area. However, it may be possible to replicate the positive ATT in other similar populations relying on the incentive scheme for selective intervention placement.

Our findings contribute to the literature on estimating the impacts of selectively placed interventions by placing the correlated random slopes estimator proposed by Wooldridge (2005, 2010) into the selection on gains context. This method allows us to control for selection on unobservables in the treatment assignment, and identify the ATE when treatment effect heterogeneity is time-invariant.

In the following, we first introduce a simple model for selective intervention placement. This is followed by the discussion of the identification strategy for the ATE and ATT. The use of these methods are first illustrated using simulated data in section 4, and then using an empirical example in section 5. Finally, section 6 concludes.

2 The selection model

In order to discuss the estimation of treatment effects of programs targeting beneficiaries with the largest expected treatment effect, we first discuss and formalize the key characteristics of the implementation environment of such programs.

The starting point for selective intervention placement is that the program has heterogeneous impacts in the population. Motivated by performance-based rewards or for other reasons, the program implementers take this into account, and assess the expected impact of the program in the population before deciding where to carry out the program. Based on their noisy observation, they implement the program first where they expect the highest treatment effect.

For example, in section 5 we discuss a sanitation program which provided additional financial rewards to implementing NGOs if all the households in the program communities were found to use toilets during a yearly evaluation campaign. Before starting with the implementation of the sanitation intervention, the NGOs carried out a sensitization campaign of the program in order to collect information about the communities and to elicit the communities' interest in taking part in the program. Hence, the NGOs had information about the expected treatment effect of the program in communities, which is unobservable for the researchers (e.g. community cohesion, leadership and enthusiasm about the program).

We are interested in settings, where the timing of the program delivery contains useful information about the expected value of the treatment effect that would otherwise be unobservable for researchers.[2] We formalize the described implementation environment in a simple threshold-switching model for the intervention placement. We refer to the unit of observation at the individual level. In practice, the intervention placement is often at the community or village level, which we address in section 5.4. The data generating process (DGP) for the outcome of interest (Y_{it}), the individual-specific potential treatment effect (β_i) and the intervention placement (D_{it}) is described by:[3]

---

[2] Hence, matching estimators would not work well in this setting because the heterogeneity is driven by unobservables from the perspective of the researchers.

[3] For simplicity, we assumed a linear model for the outcome equation (1). However, it is possible to use a more general specification, for example for binary outcomes.

$$Y_{it} = \alpha_t + D_{it}\beta_i + X_{it}\theta + \eta_i + \varepsilon_{it} \tag{1}$$

$$\beta_i = \beta + \gamma Z_i + \nu_i \tag{2}$$

$$\tilde{\beta}_i = \beta_i + n_i \tag{3}$$

$$D_i = \begin{cases} (0,1,1) & \text{if } \tilde{\beta}_i > B_2 \\ (0,0,1) & \text{if } B_3 < \tilde{\beta}_i \leq B_2 \quad \text{for } t = (1,2,3) \\ (0,0,0) & \text{if } \tilde{\beta}_i \leq B_3 \end{cases} \tag{4}$$

In the outcome equation (1), α_t denotes the time-varying but common trend for all individuals, X_{it} stands for individual characteristics with constant coefficients (θ), η_i is a heterogeneous intercept or fixed effect and ε_{it} stands for idiosyncratic shocks with variance σ_ε^2.

The heterogeneous potential treatment effect (β_i) is randomly determined in (2). Time-invariant observable individual characteristics (Z_i) may also affect the size of the treatment effect. Note that Z_i may contain the baseline value or time-mean of variables in X_{it}.

The program implementers observe a noisy measure of the potential treatment effect ($\tilde{\beta}_i$) in equation (3) with IID noise n_i that has a variance of σ_n^2. They decide based on the value of $\tilde{\beta}_i$ where to implement the intervention D_{it}, which measures whether individual i has received the intervention before time t. The model assumes that the first observation at $t = 1$ was collected as the baseline prior to the implementation of the program ($D_{i1} = 0$). Following (4), the intervention is first implemented at those individuals where it is expected to have the highest impact based on the noisy observation. The thresholds B_2 and B_3 determine the share of potential beneficiaries receiving the intervention before period $t = 2$ and before period $t = 3$, respectively.[4]

The model represented in (1)-(4) contains the most important features of the setting we described above. The noise-to-signal ratio in (3) (σ_n/σ_ν) and thresholds B_2 and B_3 in (4) determine the correlation (ρ) between the treatment effect β_i and the intervention placement D_i.

In the following, we first discuss the treatment effects of interest and their identification. Before providing an empirical example in section 5, in section 4 we return to

---

[4] Here, we consider 3 time periods, however, the model can easily be extended to include more periods.

the model in (1)-(4) in order to demonstrate the sensitivity of the estimation method to changes in the model settings.

3   Identification strategy

In the context of the selective intervention placement model described above, we are interested in the Average Treatment Effect and the Average Treatment Effect on the Treated. For simplicity, we discuss the case of a binary intervention. However, the model can be generalized for multiple treatment arms or multi-valued interventions as well.

Denoting the outcome variable of interest with the intervention as Y_i^1 and without the intervention as Y_i^0, and the intervention variable as D_i for each beneficiary i, we can write the above treatment effects as

$$ATE = E(Y_i^1 - Y_i^0) \tag{5}$$

$$ATT = E(Y_i^1 - Y_i^0 | D_i = 1) \tag{6}$$

where the expectation is taken over the eligible population for the program. For the program beneficiaries, we do not observe the outcome without the intervention (counterfactual). When the interventions are randomly assigned, Y_i^0 can be estimated by the outcomes of the population that did not receive the treatment assuming that both groups have the same distribution. Then, $ATE = E(Y_i^1 | D_i = 1) - E(Y_i^0 | D_i = 0)$. However, using observational data (without random assignment), more assumptions are required in the form of an assumed data generating process or regression model.

3.1   Regression model

Consider the random-coefficient model with a non-specified beneficiary-specific treatment effect (β_i) following (1):

$$Y_{it} = \alpha_t + D_{it}\beta_i + X_{it}\theta + \eta_i + \varepsilon_{it} \quad \text{for } t = 1, ..., T \tag{7}$$

where α_t, η_i and X_{it} are as in (1). Using a difference-in-difference specification, the first observation of the intervention variable, D_{it}, is standardized as $D_{i1} = 0$. It denotes

that the first observation was measured before the intervention was rolled out. In later periods, $D_{it} = 0$ for the comparison group and non-zero for the beneficiaries that have already been reached by the intervention before time t. The outcome variable, Y_{it}, can be binary, discrete or continuous. In this paper, we discuss the estimation using linear least squares methods. Wooldridge (2010) also discusses implementations for non-linear models.

As in section 2, we assume that the program implementers had some additional information about the expected size of the treatment effect on the potential beneficiaries, which are unobserved by the researchers, and they carry out the program first for those individuals that have the highest expected impact (selection on gains). As a consequence, the treatment status of individuals coupled with the information on the timing of the intervention delivery will carry additional information for the researchers that we can exploit in the estimation procedure. This information is conveyed by the decrease in the realized treatment effect over time as the intervention is rolled out to beneficiaries in order of the size of the expected impact.

In the case of model (7), we can write the treatment effects of interest as $ATE = E(\beta_i)$ and $ATT = E(\beta_i | D_{iT} = 1)$. For the latter, the conditioning on $D_{iT} = 1$ signifies that we are interested in the ATT for all the beneficiaries that received the intervention during the study period. As discussed above, the value of ATT may change over time as the intervention is rolled out.

We can reformulate (7) by splitting the coefficient of the treatment effect into its mean and the deviation around the mean: $\beta_i = \beta + b_i$ such that $E(b_i) = 0$. After regrouping the terms, we have

$$Y_{it} = \alpha_t + D_{it}\beta + X_{it}\theta + \eta_i + [D_{it}b_i + \varepsilon_{it}]. \tag{8}$$

It is easy to see that $ATE = \beta$ in this case. If the terms in square brackets are uncorrelated with the regressors D and X, in particular if $E(b_i | D, X) = 0$ and $E(\varepsilon_{it} | D, X) = 0$, then we can consistently estimate β using the standard difference-in-difference regression using fixed effects on:

$$Y_{it} = \alpha_t + D_{it}\beta + X_{it}\theta + \eta_i + e_{it} \tag{9}$$

where all terms in the square brackets have been pushed into the error term $e_{it} =$

$D_{it}b_i + \varepsilon_{it}$.

However, if we expect that the intervention (D) was placed based on where it would achieve the highest potential effect (high b_i), then $E(b_i|D) \neq 0$. In order to circumvent the problem caused by endogenous intervention placement, we treat the heterogeneous effect b_i as omitted variable, and include its expected value conditional on regressors D and X in the regression following Mundlak (1978), Chamberlain (1982, 1984) and Wooldridge (2005, 2010). We are interested in controlling for the component of the heterogeneous effect b_i that is correlated with the regressors D and X, and leave the uncorrelated random component as part of the error term.

Notice that when selection on gains occurs, because of the valuation of the implementing organization the value of the intervention variable D contains valuable information about the conditional expectation of the treatment effect, $E(\beta_i|D)$. We assume that the implementers are able to observe relevant characteristics of the target population that are unobservable to the researchers, and use this information to place the interventions in an optimal way.

We calculate the conditional expectation of b_i by allowing the heterogeneous parameters to depend on the regressors in a parametric way. We approximate b_i as a linear combination of the means of regressors D and X over time ($\bar{D}_i$ and $\bar{X}_i$). For ease of exposition, we leave out time-invariant covariates (Z_i in equation 2):

$$E(b_i|D, X) = (\bar{D}_i - \mu_{\bar{D}})\xi + (\bar{X}_i - \mu_{\bar{X}})\psi \tag{10}$$

where $\bar{X}_i = 1/T \sum_{t=1}^{T} X_{it}$ is the individual mean of variables X_{it} over time, and $\mu_{\bar{X}} = E(\bar{X}_i)$ is the expectation of the individual means, which is estimated by the sample mean over all individuals assuming a representative sample: $\hat{\mu}_{\bar{X}} = 1/N \sum_{i=1}^{N} \bar{X}_i$. The variables for D are defined similarly. The demeaning of the right-hand side terms in (10) ensures that $E(b_i) = 0$.[5]

The assumptions made about the distribution of the heterogeneous parameters in (10) are crucial for the identification of β, Therefore, it is useful to spend some time discussing their implications. Assuming a linear approximation in (10) implies monotonicity of the heterogeneous parameter as a function of $\bar{D}_i$ and $\bar{X}_i$, i.e. $E(b_i|\bar{D}'_i, \bar{X}_i) \geq$

---

[5] Wooldridge (2010) points out that when using unbalanced panel sample, the means above have to be calculated over the sample used for estimation.

$E(b_i|\bar{D}_i'', \bar{X}_i)$ or $E(b_i|\bar{D}_i', \bar{X}_i) \leq E(b_i|\bar{D}_i'', \bar{X}_i)$ for all $\bar{D}_i' > \bar{D}_i''$. This means that the intervention reaches the beneficiaries in the order of the expected size of the treatment effects. If this assumption is not reasonable for the data, the identification of β fails based on (10). However, in the context of selective intervention placement, this condition is satisfied by default.

Regarding the structure assumed in (10), note that Chamberlain (1980) assumed a more general structure for the distribution of the heterogeneous parameters using a linear combination of the regressors ($W = \{D, X\}$) for all time periods: $E(b_i|W) = W_{i1}\phi_1 + W_{i2}\phi_2 + \cdots + W_{iT}\phi_T$.[6] Regression (10) uses the group mean of the regressors over time, which is a special case of the general formulation, where each time period receives the same weight ($\phi_1 = \cdots = \phi_T = 1/(T+1)$). Another special case is the specification of Elbers and Gunning (2014), who used the changes in the regressors to predict the heterogeneous parameters ($\phi_1 = -1, \phi_T = 1$). These two sets of assumptions have quite different implications for the interpretation of the model: using means of regressors over time, one assumes that the heterogeneity is driven by the (relative) mean level of individual characteristics, while using the changes in the regressors, one assumes that the heterogeneity is driven by changes in the individual characteristics. For example, if we expect that (for the researchers) unobserved community or household characteristics drive the selection process and not the changes in characteristics, then the natural choice is to use the group means of the regressors, especially the interventions, to predict the value of the heterogeneous parameters as done in (10).[7]

Assuming that (10) is correct, substituting into (8) and taking conditional expectations, we get the correlated random slopes model:

$$
\begin{aligned}
E(Y_{it}|D, X) =& \alpha_t + D_{it}\beta + X_{it}\theta + D_{it} \otimes (\bar{D}_i - \mu_{\bar{D}})\xi + \\
& D_{it} \otimes (\bar{X}_i - \mu_{\bar{X}})\psi + E(\eta_i|D, X) + E(\varepsilon_{it}|D, X)
\end{aligned}
\tag{11}
$$

where $D \otimes X$ denotes all interaction terms between D and X. Assuming that $E(\varepsilon_{it}|D, X) =$

---

[6] Suri (2011) further generalizes this structure and estimates a correlated random coefficient model for technology adoption. Her identification strategy relies on the variation in technology use over time. In the selective intervention placement model of (11), this generalized identification strategy is underidentified.

[7] Another possibility would be to use the baseline values of X to control for treatment effect heterogeneity ($\phi_1 = 1$).

0 or that no selection occurred on time-varying unobservable variables, we can estimate (11) by fixed effects estimation.

Regression (11) allows us to test the presence of correlated random slopes. A Wald-test (or Likelihood Ratio-test) of $H_0 : \xi = 0$ tests for selection on gains in the intervention placement (correlation between the treatment effect and the intervention), while $H_0 : \psi = 0$ tests the heterogeneity of the treatment effect conditional on the mean of observed individual-specific variables (Wooldridge, 2010).

3.2 Treatment effects of interest

Fixed effects estimation of (11) results in consistent estimates for β (ATE) provided that the above discussed conditions and assumptions are satisfied.[8] These requirements are summarized as A1-A4 and C1-C2 in Table 1.

Table 1: Assumptions required for identifying ATE and ATT

Assumptions required	ATE	ATT
A1. $E(\varepsilon_{it}\|D_{i1},\cdots,D_{iT},X_{i1},\cdots,X_{iT},\eta_i,b_i) = 0$ for $t = 1,\cdots,T$ (strict exogeneity)	Yes	Yes
A2. Common trend α_t for all i	Yes	Yes
A3. b_i is time-invariant	Yes	No
A4. $E(b_i\|\bar{D}'_i, \bar{X}_i) \geq (\leq)E(b_i\|\bar{D}''_i, \bar{X}_i)$ for all $\bar{D}'_i > \bar{D}''_i$ (monotonicity)	Yes	No
Further conditions		
C1. $T \geq 2$	Yes, if D is not binary	Yes
C2. $T \geq 3$ and D is gradually rolled out	Yes, if D is binary	No

Note: The table shows whether the above 5 assumptions and 2 further conditions are required for identifying ATE and ATT. The notation is based on equation 8.

These assumptions can be weakened when estimating the ATT as shown in the last column of Table 1. For binary treatment variables, the ATT can be estimated as the coefficient estimate of β ($\hat{\beta}_{DD}$) in the standard difference-in-difference regression (9).

---

[8] Wooldridge (2005) shows that β in (11) estimates the ATE in the eligible population when the model is correctly specified.

However, in the current setting, the ATT also includes the effects of selective intervention placement. Hence, it is not directly comparable to the ATT calculated in Randomized Controlled Trials or other settings where the heterogeneity of the treatment effect (β_i) is independent of the intervention placement (D_{it}). However, Elbers and Gunning (2014) argue that for policymakers the treatment effect of the program inclusive of the effects of selection may be more relevant.[9]

It is important to point out that while the identification of the ATE requires that the model is correctly specified (in particular, the heterogeneity parameters in equation 10),[10] the estimates for ATT are also valid when the model is misspecified (see Table 1). This is because the estimate of ATT depends on the value of $E(D_{iT}\beta_i)$ rather than β. Hence, we do not necessarily need to have a consistent estimate for the ATE in order to estimate the ATT.

When only three rounds of data are available, β in the correlated random slopes regression (11) is just identified. Therefore, it is not possible to test whether the assumptions of this model (in particular, Assumptions A3 and A4 in Table 1) are valid, or that there are other sources of heterogeneity in the DGP (for example, time-varying treatment effects). Hence, in the real world, it is essential to have a good understanding of the implementation of the program, and we need additional information to argue for the validity of one or the other set of identifying assumptions.

Keeping in mind that the selective intervention placement model assumes that the treatment effects change monotonically over time as the intervention is rolled out, it can be useful to plot the data and estimate a difference-in-difference regression with time-varying treatment effects (β_t) first. It is also important to test the validity of the common trend assumption for intervention groups with multiple pre-intervention survey rounds. These can give guidance regarding the validity of the correlated random slopes model given the data.

---

[9] Elbers and Gunning (2014) propose an estimator for $E(D_{iT}\beta_i)$ what they call the Total Program Effect. This can be generalized for more than two time periods using the correlated random slopes model (11) as $TPE(t = T) = \overline{D_{iT}}\beta + \overline{D_{iT} \otimes (\bar{D}_i - \mu_{\bar{D}})}\xi + \overline{D_{iT} \otimes (\bar{X}_i - \mu_{\bar{X}})}\psi$ where $\overline{D_{iT}} = 1/N \sum_{i=1}^{H} D_{iT}$ and $D_{i1} = 0$. In the case of binary intervention variables, the TPE is equivalent to the ATT.

[10] The specification of the correlated random slopes model of (11) is more general than the selection model described in section 2. However, for all underlying data generating processes, (10) must hold in order to identify the ATE.

4 Illustration: Simulations

Next, we look at the performance of the proposed treatment effect estimators for various sets of parameter values for model (1)-(4) presented in section 2. We use Monte Carlo simulations to investigate the effect of changing parameter values on the distribution of the treatment effect estimates for a binary treatment variable.

4.1 Data generating process

We simplify the model presented in section 2 as:

$$Y_{it} = \beta_i D_{it} + \varepsilon_{it} \tag{12}$$

$$\beta_i = \beta + \gamma(Z_i - \mu_Z) + \nu_i \tag{13}$$

$$\tilde{\beta}_i = \beta_i + n_i \tag{14}$$

$$D_i = \begin{cases} (0,1,1) & \text{if } \tilde{\beta}_i > B_2 \\ (0,0,1) & \text{if } B_3 < \tilde{\beta}_i \le B_2 \quad \text{for } t = (1,2,3) \\ (0,0,0) & \text{if } \tilde{\beta}_i \le B_3 \end{cases} \tag{15}$$

For simplicity, we only use one time-invariant control variable (Z_i), which affects the size of the treatment effect. We use the standard normal distribution to generate Z_i and subtract its sample mean in order to guarantee that the $ATE = \beta$.

In the following, we fix $\beta = 1$, $\varepsilon_{it} \sim N(0,1)$, $\nu_i \sim N(0,1)$, $n_i \sim N(0,\sigma_n^2)$ and $B_2 = 2$, and investigate the effect of changes in the other parameters (σ_n^2, γ, B_3) on the performance of the standard difference-in-difference and correlated random slopes estimators of the ATT ($E(\beta_i|D_{iT} = 1)$) and the ATE ($E(\beta_i)$).[11]

4.2 Estimating equations

We use the within transformation to estimate the following two regression equations with $T = 3$ periods and a sample of $N = 1000$ households:[12]

---

[11] All variables are generated independently from the normal distribution.

[12] In the current simulation setting without individual fixed effects, ordinary least squares regression would be more efficient. However, we use the within transformation to mimic the proposed estimation

$$DD : Y_{it} = \beta_{DD} D_{it} + e_{it} \tag{16}$$

$$CRS : Y_{it} = \beta_{CRS} D_{it} + \xi_{CRS} D_{it} (\bar{D}_i - \mu_{\bar{D}}) + \psi_{CRS} D_{it} (Z_i - \mu_Z) + u_{it} \tag{17}$$

From the standard difference-in-difference regression (16) we have $ATT = \hat{\beta}_{DD}$, while the correlated random slopes regression provides the $ATE = \hat{\beta}_{CRS}$.[13]

Equation (15) implies a non-linear relationship between the heterogeneous treatment effect and the intervention variable. The CRS regression provides a first-order approximation of this non-linear relationship due to its reliance on a linear specification in (10) when modeling the correlation between β_i and $\bar{D}_i$.

In the next subsections, we investigate the distribution of these treatment effect estimates using $R = 5000$ replications as we change the parameters that determine the program allocation (D_{it}) and the treatment effect heterogeneity (β_i).

4.3 Noisy inference

First, we investigate how changes in the noise-to-signal ratio in the implementers' observation of the treatment effect ($\tilde{\beta}_i$ in equation 14) affects the treatment assignment and the treatment effect estimates. For simplicity, we set $\gamma = 0$, thereby excluding Z_i from the DGP (and also from the CRS regression (17) with $\psi_{CRS} = 0$). In addition, we also fix the implementers' thresholds for intervention delivery at $B_2 = 2$ and $B_3 = 1$.

The first thing to notice on the right panel of Figure 1 is the high correlation (ρ) between the treatment effect β_i and the intervention placement (solid line). This is influenced by the observations of the implementers in equation (14). The correlation decreases as the noise-to-signal ratio (σ_n^2/σ_ν^2) in the implementers' observations increases. It also changes the share of people receiving the intervention in period 2 (s_2) and 3 (s_3), but the total share of population with the intervention remains stable close to 50%.

The top left panel of the figure shows that the CRS regression estimates the ATE ($\beta = 1$) reasonably as the DGP is in line with the identifying assumption of (10), even

procedure.

[13] It is also possible to calculate the ATT using the CRS regression (17), see footnote 9.

Effect of noise on the treatment effect estimators

Estimates of average treatment effects

β (y-axis)

σ^2_{noise} (x-axis)

Treatment effect — ATE(CRS) - - ATT(DD)

Note: 95 percent confidence interval included around mean

Simulation characteristics

ρ , s_2 , s_3 (y-axis)

σ^2_{noise} (x-axis)

Variable — ρ - - s_2 — s_3

Note: $\rho = cor(\beta_i, \overline{D}_i)$, $s_2 = mean(D_{i2})$ and $s_3 = mean(D_{i3} - D_{i2})$

Share of replications with true treatment effect in 95% CI

Share true TE in 95% CI (y-axis)

σ^2_{noise} (x-axis)

Treatment effect — ATE(CRS) - - ATT(DD)

Parameters: $\beta = 1$, $\gamma = 0$, $\sigma^2_e = 1$, $\sigma^2_v = 1$, $B_2 = 2$, $B_3 = 1$, $N = 1000$, $T = 3$ and $R = 5000$
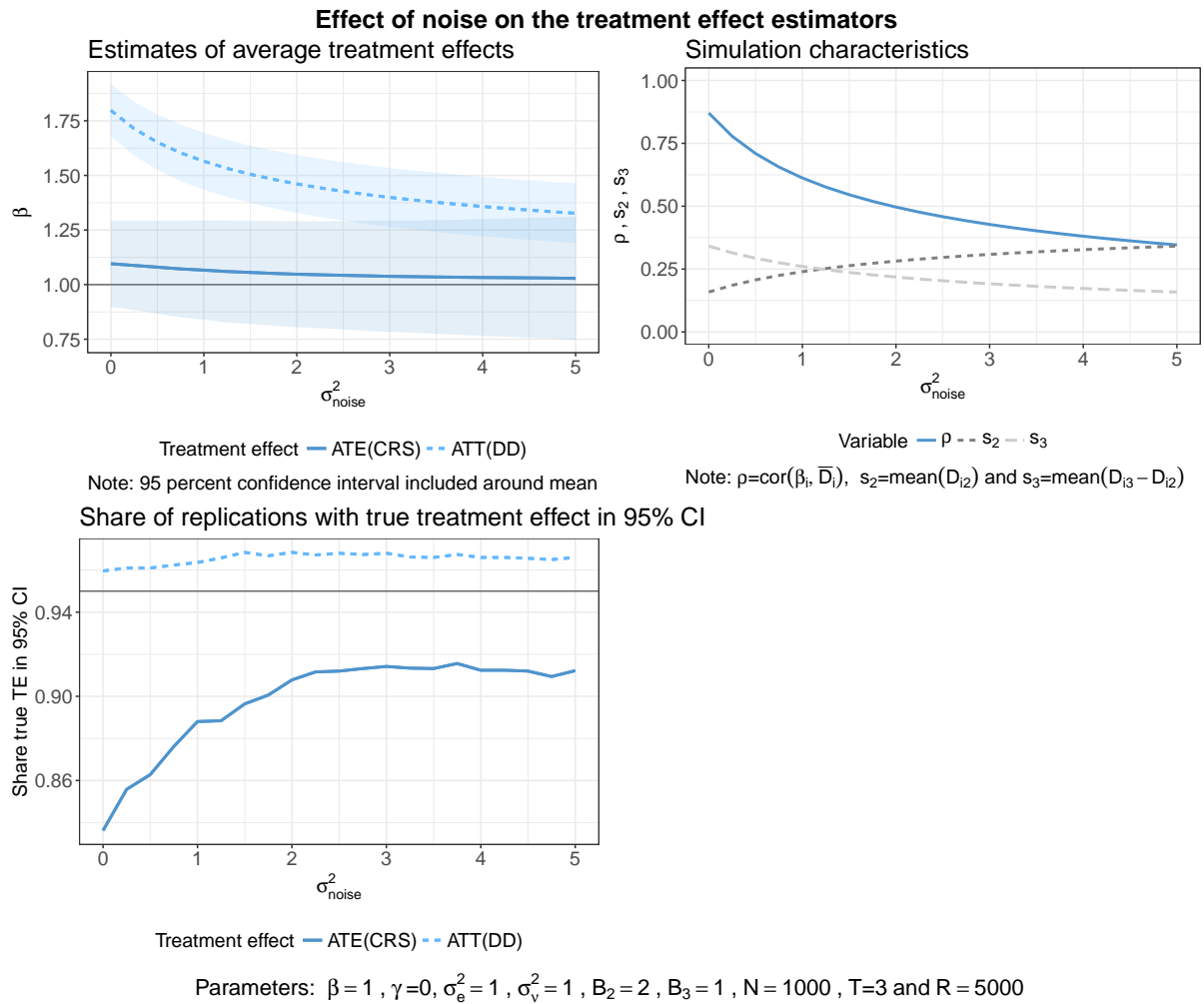
Figure 1: The effect of noisy inference

though it only uses a first-order approximation to correct for the correlation between β_i and $\bar{D}$. The ATT deviates from the ATE due to the large selection effect present in the DGP. However, as the noise in the implementers' observations increases, their ability to select households with the highest treatment effect decreases. As a consequence, the ATT also decreases closer to the ATE. At the extreme, where the implementers only observe noise and the treatment assignment becomes random, the ATT converges to the ATE. Notice also that the standard error of the estimate of ATE is larger than that of ATT, and it increases as the correlation between the intervention variable and the treatment effect decreases.

The bottom left figure shows how often the true ATT and ATE were in the 95% confidence interval of the treatment effect estimates. The black horizontal line shows the expected coverage of 95%. The figure shows that the true ATT lies in the confidence interval of the DD estimate more than 95% of the replications. In the case of the ATE, the confidence interval of the CRS estimate contains the true value of the ATE only in about 90% of the replications, even though its confidence interval is wider than for the DD estimates. This is due to the remaining bias in the (mean) estimates of ATE as seen on the top left figure.

4.4 Changing thresholds

In the following, we fix the noise-to-signal ratio at $\sigma_n^2/\sigma_\nu^2 = 4$, thereby assuming that the implementers observe twice as much noise as signal. In Figure 2, we investigate the effect of changes in the lower threshold for inclusion in the intervention in period 3 (B_3 in equation 15), while keeping the threshold for treatment in period 2 constant at $B_2 = 2$.

The right panel in Figure 2 shows that the share of population receiving the intervention in period 2 is constant $s_2 = 33\%$, while in period 3 the share goes from $s_3 = 64\%$ when $B_3 = -3$ to $s_3 = 0\%$ when $B_3 = B_2 = 2$. Hence, when $B_3 = -3$ there are only very few households that do not receive the intervention, while when $B_3 = 2$ there are no households that receive the intervention in period 3.

When the whole population receives the intervention, there is no more selection effect in period 3, and the ATT approaches the ATE (left panel of Figure 2). The selection effect and the ATT increases as more households are systematically excluded from

Effect of treatment assignment threshold on the treatment effect estimators

Estimates of average treatment effects

Simulation characteristics

Note: 95 percent confidence interval included around mean

Note: $\rho=\mathrm{cor}(\beta_i, \overline{D}_i)$, $s_2=\mathrm{mean}(D_{i2})$ and $s_3=\mathrm{mean}(D_{i3}-D_{i2})$

Share of replications with true treatment effect in 95% CI

Parameters: $\beta=1$, $\gamma=0$, $\sigma_e^2=1$, $\sigma_v^2=1$, $B_2=2$, $\sigma_{noise}^2=4$, $N=1000$, $T=3$ and $R=5000$
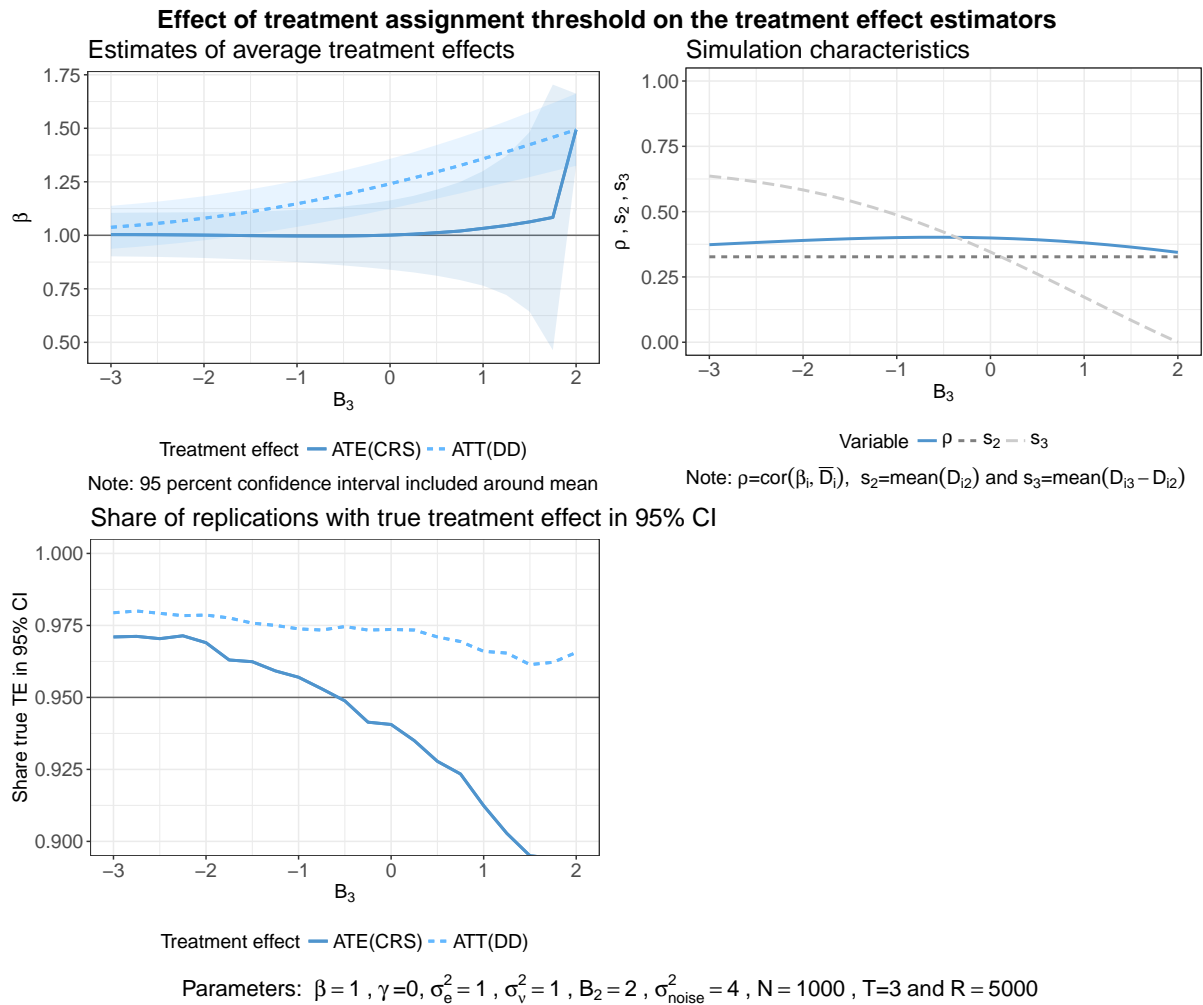
Figure 2: The effect of changing thresholds

the intervention as a result of increasing the threshold for inclusion for period 3 (B_3). However, the DD estimate of the ATT remains unbiased.

The CRS estimate of ATE remains close to its true value until s_3 approaches close to zero. However, its standard error increases as s_3 diminishes. At the extreme of $s_3 = 0$, the estimate the CRS model collapses to the DD estimate because data from period 3 provide no information about the treatment effect (see the discussion in Appendix B).

The bottom left panel of the figure shows that the confidence interval of the CRS estimate includes the true value of the ATE in over 95% of the replications when a substantial share of the population receives the intervention in period 3. As B_3 increases, the accuracy of the CRS estimate decreases. The accuracy of the DD estimate is unaffected by changing the threshold.

4.5   Selection on observables

Now, we turn to discussing the effect of selection on observables. We fix $B_3 = 1$ and $\sigma_n^2/\sigma_\nu^2 = 4$, and investigate the effect of changes in γ in equation (13) using regression equations (16) and (17).

Figure 3 displays the simulation results as γ increases from 0 to 2. The exogenous variable z_i follows the standard normal distribution. The right panel of the figure shows that the correlation between the intervention and the treatment effect (here ρ_β) is quite high due to the relatively low noise-to-signal ratio, which decreases in γ. As γ increases from zero, so does the correlation between the intervention and the control variable Z_i (ρ_Z), which quickly approaches the size of ρ_β. The variance of the treatment effect also increases with γ: $var(\beta_i) = \sigma_\nu^2 + \gamma^2 \sigma_Z^2$. As a consequence, a larger share of the population is exposed to the intervention in period 2, but the overall share of population receiving the intervention remains unchanged around 50% (not shown). The ATT also increases with γ (left panel) as a further consequence. The left panel of the figure shows that the CRS estimate of ATE remains close to its true value. However, the confidence interval of the CRS estimate includes the true ATE under 95% (91-92%) of the replications. This is mostly driven by the value selected for B_3 (not shown), while the size of γ has little effect on the coverage for ATE.

Effect of observable heterogeneity on the treatment effect estimators
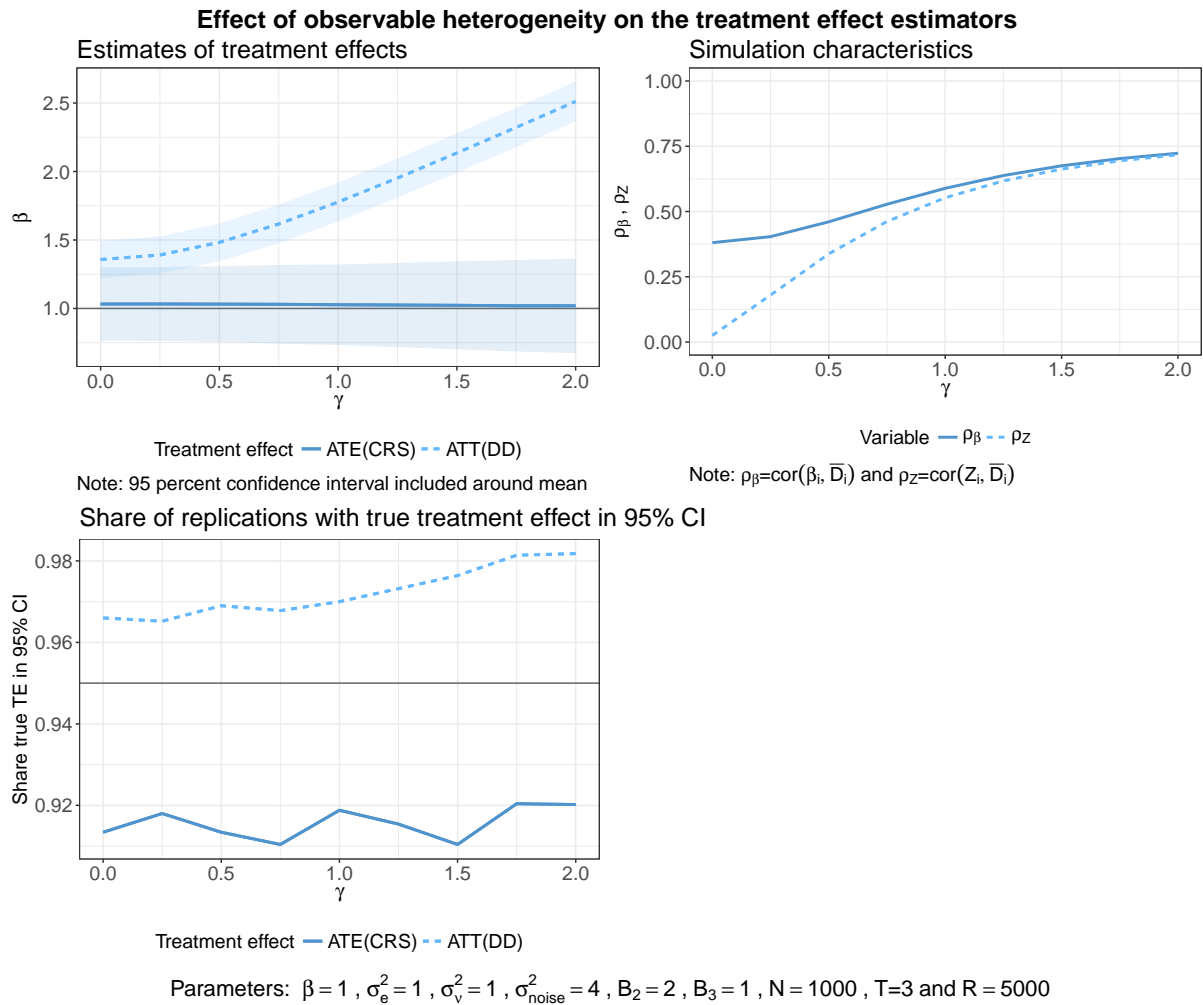
Figure 3: The effect of observable treatment effect heterogneity

4.6 Summary

The above findings using a linear outcome model with normally distributed disturbances suggest that when the model assumptions are satisfied, the correlated random slopes estimates are informative about the ATE in a selective intervention placement setting. The precision of the ATE estimates depends primarily on the share of population exposed to the intervention in period 3. At the same time, the difference-in-difference regression estimates the actual ATT (inclusive of the effect of selective placement) always precisely.

5 Application: One Million Initiative

To demonstrate the use of the two estimators on actual data, we turn to analyzing the effectiveness of a sanitation intervention on latrine ownership and proper handwashing in Mozambique. We use this example because it represents a realistic scenario for the use of the proposed selective intervention placement model.

5.1 The One Million Initiative

The One Million Initiative (OMI) was implemented between 2006-2013 in 3 provinces of Mozambique as a cooperation between UNICEF and the governments of Mozambique and the Netherlands. The program was to reach one million people in poor rural areas (40% of the population in the program area) and to provide them with means to use safe and sustainable drinking water and hygienic sanitation facilities. Specifically, OMI carried out community water supply interventions by creating new boreholes and training water committees on maintenance, and community-based sanitation and hygiene education using the Community-Led Total Sanitation approach (CLTS). The interventions were gradually implemented between 2008 and 2013. Community participation was an essential component of the interventions.

Here, we only look at the effectiveness of the sanitation interventions (CLTS). The aim of CLTS was to eradicate open defecation by triggering communities to build latrines for themselves through awareness raising.[14] No subsidies were provided to pro-

---

[14] See Kar and Chambers (2008) for more information about the implementation of Community-Led Total

Table 2: Number of communities in CLTS intervention arms

	Nr. communities
Phase 1 (2008-2010)	23
Phase 2 (2010-2013)	11
Comparison	22
Total	56

mote latrine construction, which were almost always traditional latrines built from locally available material. The intervention also promoted hand-washing with soap or ash after defecation.

The CLTS interventions were implemented by local NGOs in each of the 18 program districts. The NGOs communicated the program to the communities and decided on the location of the CLTS intervention in agreement with the communities. Annual targets were set for the number of treated communities. The treated communities could apply to certify that they were free of open defecation during the annual Open Defecation Free (ODF) communities evaluation campaign. The NGOs were financially rewarded for the number of ODF communities in their district. Therefore, it was in their best interest to implement the CLTS intervention in communities that had a high likelihood of success.

Regarding the water supply interventions, the communities had to request these through the NGOs. However, the district government decided on the location of these interventions. In practice, there was a substantial overlap between the two types of interventions. For further information about the interventions, see Vigh et al. (2017).

5.2 Data

For the evaluation of the One Million Initiative, survey data was collected in three rounds: August-October 2008 (baseline), August-October 2010 (midline) and July-August 2013 (endline). In each round, data was collected at 1600 households in 80 communities. These communities were not a priori assigned to intervention and control

---

Sanitation. The key elements of the program were: hygiene awareness raising, inciting collective action, transfer of know-how for building safe latrines and rewarding success.

arms. Instead, the implementing organizations decided on the intervention locations and then reported these to UNICEF and the research team. Here, we use the data only on those 34 communities that received the CLTS intervention either between the baseline and midline (Phase 1) or between the midline and endline (Phase 2),[15] and 22 communities that did not receive any CLTS and water supply interventions (Comparison) (Table 2). Among the CLTS intervention communities, 23 also received the water supply intervention.

In general, the CLTS intervention was more effective when implemented together with the water supply intervention. However, here we simplify the analysis by classifying all communities that received the CLTS intervention in the same intervention arm without reference to the water supply intervention.[16] The results are qualitatively similar when we separate the treatment arms by the presence of the water supply intervention.[17]

In the regression analysis, we use data on 1066 households that were interviewed in at least two survey rounds in the 56 communities included in the data analysis (N=2951). For simplicity, we treat this sample as representative and do not apply sampling weights. The tables in Appendix A summarise the community and household characteristics and variable definitions.

5.3 Outcome variables

We present two outcome variables that were expected to be most affected by the CLTS intervention: latrine ownership (LATRINE) and whether adults wash their hands with soap or ash after defecation (HWSD).[18] Both of these variables are defined at the household level. Latrine ownership was verified by interviewers and hand-washing is self-

---

[15] We omitted 7 communities that received the water supply intervention before the midline survey and the CLTS intervention only after the midline survey, because the timing of the interventions could interfere with the effectiveness of the CLTS intervention. The 17 communities that received only the water supply intervention are also omitted from the sample.

[16] The water supply intervention had no effect on latrine ownership and proper hand-washing when implemented without the CLTS intervention.

[17] Results are available upon request.

[18] Almost all latrines are owned and used by a single household. Therefore, latrine use is very highly correlated with ownership. Hand-washing occasions and methods were surveyed for adult women and men in the household separately.

reported.

Figure 4 shows the changes in the two outcome variables for communities with the CLTS intervention and the comparison group. Panel A shows the development in the share of households owning a latrine (left) and washing hands with soap after defecation (right). The outcomes are shown separately for the Phase 1 and Phase 2 intervention groups and the comparison group. Panel B shows the corresponding changes in the outcome variables between the survey periods for the three groups. For an effective intervention, the outcome is expected to increase (more than in the comparison group) from 2008 to 2010 for the Phase 1 intervention group, and from 2010 to 2013 for the Phase 2 intervention group. The outcome on hand-washing follows this pattern. Note that there is also a positive and significant trend in the comparison group for both outcome variables.[19]

The changes in latrine ownership are significantly smaller (diff=-16.5pp, p=0.0001) after the Phase 2 CLTS intervention compared to the Phase 1 intervention group. We argue that this pattern is the one we would expect if the implementing NGOs selected the location of the CLTS interventions based on where they expected the largest treatment effect in terms of the communities becoming open defecation free, as they were incentivized to do.

Our identification strategy requires that the comparison and intervention groups follow the same trend (Assumption A2). In Panel C of Figure 4, we test this assumption for the Phase 2 intervention group on the changes of outcomes from 2008 to 2010. The Phase 2 intervention group had a somewhat higher trend than the comparison group for both outcome variables. However, the difference is not significantly different from zero (p=0.396 and p=0.351) when controlling for observable household characteristics (household size, wealth index and education). Based on this finding, we cannot reject the validity of the common trend assumption. However, we are not able to test, and must assume that the same is true for the Phase 1 intervention group.

---

[19] The positive trend in the comparison group can partially be attributed to other simultaneously running sanitation interventions and by increasing wealth levels.

Panel A. Outcome variables by the timing of the CLTS intervention
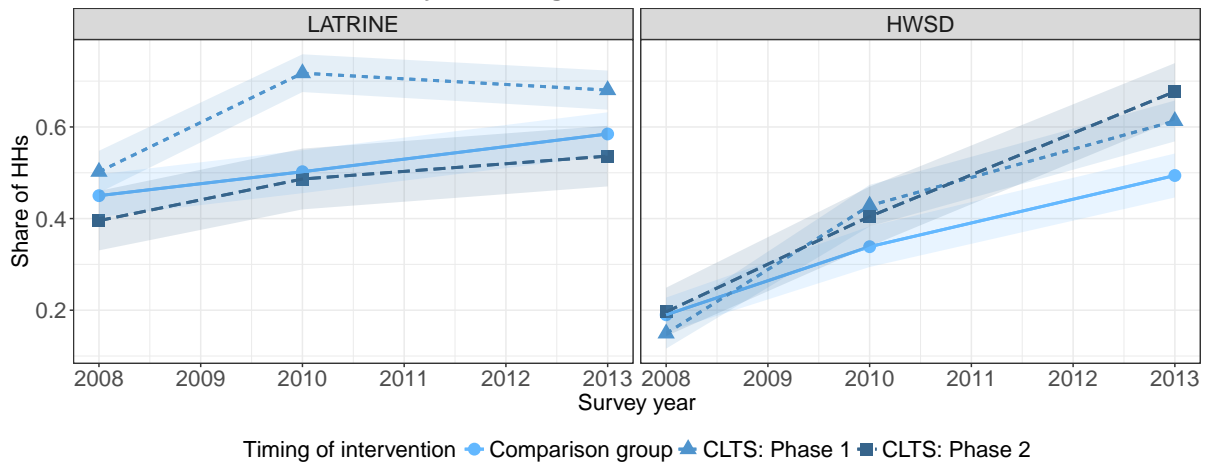
Panel B. Changes in outcome variables by timing of the CLTS intervention
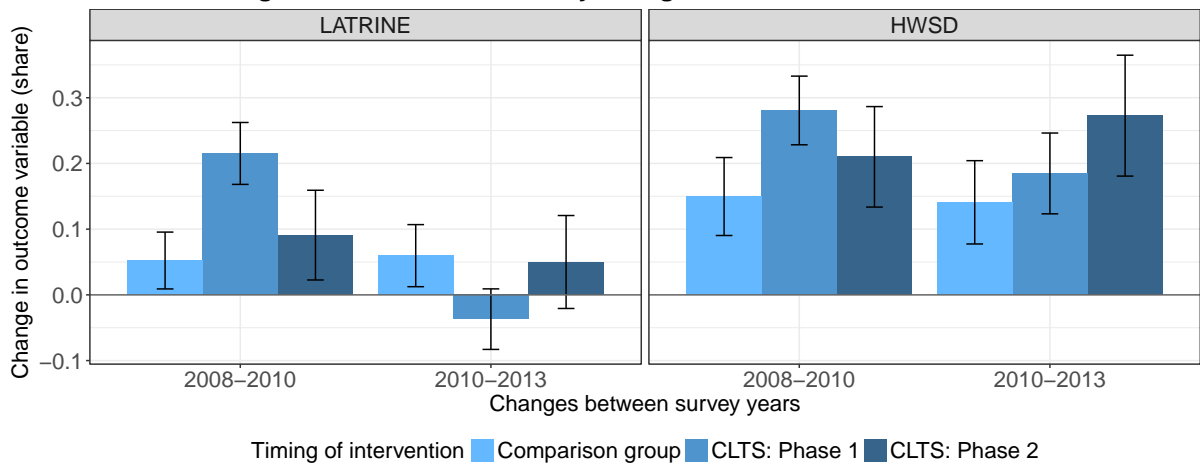
Panel C. Testing common trend assumption for Phase 2 CLTS intervention between 2008–2010
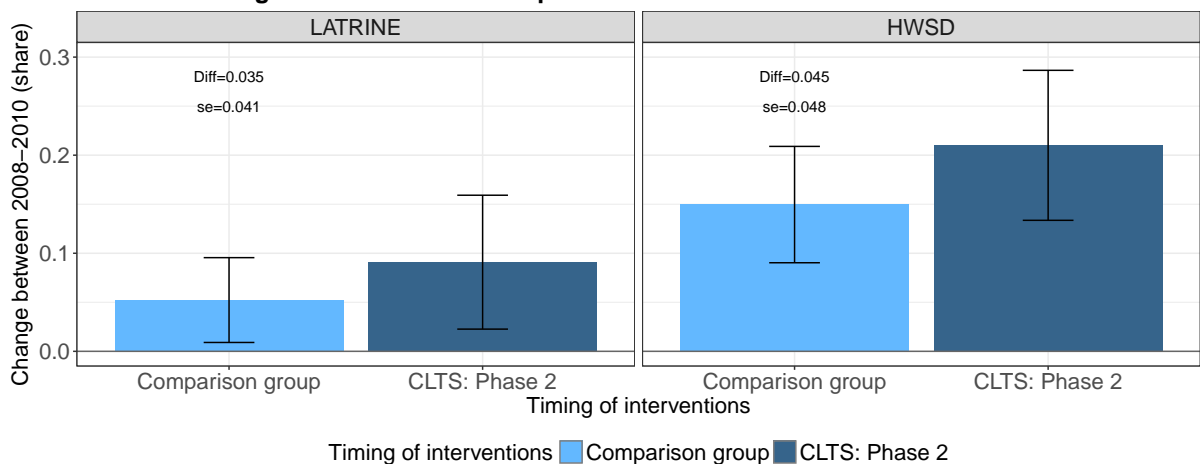
Note: 95 percent confidence interval included around mean

Difference calculated controlling for HH size, wealth index and education. Std.error clustered at community level.

Figure 4: Changes in outcome variables and the timing of the CLTS interventions

5.4 Regression model

The CLTS intervention was implemented at the community level (c). Therefore, we adjust the formulation of (11) to include heterogeneous treatment effect at the community level ($\beta_{c(i)}$) instead of at the household level (i). We estimate the following CRS model:

$$
\begin{aligned}
Y_{it} =& \alpha_t + \beta D_{c(i)t} + \theta X_{it} + \xi D_{c(i)t} \otimes (\overline{D}_{c(i)} - \mu_{\overline{D}}) + \\
& \psi D_{c(i)t} \otimes (\overline{X}_{c(i)} - \mu_{\overline{X}}) + \eta_i + u_{it}
\end{aligned}
\tag{18}
$$

where $D_{c(i)t}$ shows whether there have been a CLTS intervention in the community up until period t in community c of household i. X_{it} consists of time-varying household specific control variables (household size, wealth index and education).

The demeaned variables, $\overline{D}_{c(i)}, \overline{X}_{c(i)}$, are now defined at the community level. Hence, $\overline{X}_{c(i)} = 1/(TN_c) \sum_{t=1}^{T} \sum_{i=1}^{N_c} X_{it}$ is the community mean of X over time with N_c households in community c, and $\mu_{\overline{X}} = 1/N \sum_{c=1}^{C} \sum_{i=1}^{N_c} \overline{X}_{c(i)}$ is the sample mean of the community means in the population.

5.5 Results

The CLTS implementing NGOs were rewarded based on latrine ownership (if all households own and use a latrine) but no incentives were provided based on hand-washing. Therefore, we expect to find that selective intervention placement affected the treatment effects of latrine ownership more than of hand-washing.

Table 3 shows the estimation results for latrine ownership. The reported standard errors are robust to heteroskedasticiy, serial-correlation and clustering at the community level. The first column shows the results of estimating the ATT using the standard difference-in-difference model ($\xi = 0$ and $\psi = 0$). Using this specification, we can conclude that CLTS on average had a positive and significant effect on latrine ownership at the locations selected for the intervention (coef=8.3pp, p=0.036). However, when we control for the correlation between the treatment effect and the intervention placement (column 2), the size of the treatment effect is halved and insignificant (coef=3.4pp, p=0.391). At the same time, we observe a positive and weakly significant

25

Table 3: Estimation results for latrine ownership

	DD	CRS	CRS
	Regression model		
	DD	CRS	CRS
	(1)	(2)	(3)
β:CLTS	0.0828**	0.0341	0.0485
	(0.0394)	(0.0398)	(0.0435)
ξ:CLTS		0.2398*	0.1309
		(0.1345)	(0.1507)
Mean at baseline	0.471	0.471	0.471
Observations	2951	2951	2951
R^2	0.061	0.063	0.070
ψ estimated	No	No	Yes
p-value for F-test: $\psi = 0$			0.112

Notes: *p<0.1; **p<0.05; ***p<0.01
All regressions control for HH size, wealth index, education and year.
HH FE regression results with robust standard errors corrected for
serial correlation, heteroskedasticity and clustering at community level.
Sample includes HHs participating in at least 2 survey rounds.
In (3) ψ contains HH size, wealth index and education.

correlation between the intervention variable and the timing of the intervention as evidence for selective intervention placement (coef=0.240, p=0.075). This is in line with the incentive structure. The last column of the table shows that the treatment effect on latrine ownership may be correlated with some observable time-varying community characteristics (p=0.112 jointly for average household size, wealth index and education). Investigating further, we find that the reduction in the estimate of ξ is driven by the effect of education (share of adults with any formal education in the community) on the treatment effect (coef=0.181, p=0.078).

The estimates of the ATE in column 2 are not significantly different from the standard difference-in-difference estimates of ATT (diff=4.9pp, p=0.214). However, one would draw different conclusions about the effectiveness of the CLTS intervention on latrine ownership based on the ATE and the ATT.

Turning to the effectiveness of the CLTS intervention on hand-washing, Table 4 shows that the standard difference-in-difference method estimates the treatment ef-

Table 4: Estimation results for hand-washing with soap

	Regression model		
	DD	CRS	CRS
	(1)	(2)	(3)
β:CLTS	0.1468***	0.1283*	0.1408*
	(0.0475)	(0.0715)	(0.0750)
ξ:CLTS		0.0919	0.0156
		(0.2046)	(0.2461)
Mean at baseline	0.180	0.180	0.180
Observations	2925	2925	2925
R^2	0.186	0.186	0.187
ψ estimated	No	No	Yes
p-value for F-test: $\psi = 0$			0.433

Notes: *p<0.1; **p<0.05; ***p<0.01

All regressions control for HH size, wealth index, education and year.
HH FE regression results with robust standard errors corrected for
serial correlation, heteroskedasticity and clustering at community level.
Sample includes HHs participating in at least 2 survey rounds.
In (3) ψ contains HH size, wealth index and education.

fect at 14.7 percentage points (p=0.002). Controlling for selective intervention placement, the estimated treatment effect changes only little (coef=12.8pp) as the correlation between the treatment effect and the intervention placement is insignificant (p=0.653). Other observable community characteristics are also uncorrelated with the treatment effect (p=0.433). In this case, the ATE and ATT are not significantly different (diff=1.8pp, p=0.666), and the two treatment effect estimates lead to the same conclusion regarding the effectiveness of the CLTS intervention on hand-washing after defecation.

Summarizing the results, we find evidence of strategic intervention placement on the incentivized outcome (latrine ownership) but not on the non-incentivized outcome (hand-washing).[20] This makes sense from the perspective of CLTS implementing NGOs,

---

[20] Another explanation could be that a lower-cost technology adoption intervention could be expected to have a more homogeneous impact. Because adopting the use of soap requires a much smaller investment compared to building a latrine, it is reasonable to observe less heterogeneity in the treatment

who were incentivized on the overall level of latrine ownership. If the incentives are correctly aligned, then strategic intervention placement is also desirable in terms of the total effect of the program.

Our findings also suggest that there is an optimal stopping point for rolling out the sanitation intervention focused on increasing latrine ownership in the target program area as the treatment effect decreases when the population coverage of the intervention is increased. However, it may be possible to replicate the positive Average Treatment Effect on the Treated in other similar populations relying on the incentive scheme for selective intervention placement.

6 Conclusion

In this paper, we focused on the correlation between the treatment effect and the intervention placement in terms of selection on gains. Using the results of Wooldridge (2005, 2010) and Chamberlain (1980) we discussed how the correlated random slopes model can be used to estimate the Average Treatment Effect. Simulation results from a model of selective intervention placement showed that when the assumptions for the estimation of ATE are satisfied, the CRS regression provides a good linear approximation of the ATE.

The estimation of the ATT relies on less strong assumptions as it estimates the treatment effect inclusive of the effects of selectivity. The resulting treatment effect estimates are the primary interest of policymakers when evaluating the benefits of the program. In addition, for forward looking policymakers, the comparison of ATT ex post to credible estimates of ATE can give guidance with respect to the expected impacts of continuing the program in the same program area and population or replicating it in a different area and population.

Using the One Million Initiative as example, we presented a case where selective intervention placement was present in the intervention design: the implementing NGOs of the CLTS interventions were incentivized to carry out the interventions in communities where they expected the effect on latrine ownership to be the highest. Indeed, we find evidence of selection on gains for this outcome variable. Controlling for selection

---

effect on proper hand-washing.

effects changed our conclusions about the effectiveness of the interventions. However, looking at another relevant outcome variable of the CLTS intervention, which was not incentivized for the NGOs (hand-washing with soap after defecation), we find no effect of selective intervention placement.

From the perspective of the total effect of the program, using an incentive scheme may increase the total impact of the program if it is carefully designed and correctly aligned with the objectives of the policymakers.

References

Chamberlain, Gary (1980), "Analysis of Covariance with Qualitative Data." *The Review of Economic Studies*, 47, 225–238.

Chamberlain, Gary (1982), "Multivariate Regression Models for Panel Data." *Journal of Econometrics*, 1, 5–46.

Chamberlain, Gary (1984), "Panel Data." *Handbook of Econometrics*, Volume 2, 1248–1318.

Chassang, Sylvain, Gerard PadrÃş I Miquel, and Erik Snowberg (2012), "Selective trials: A principal-agent approach to randomized controlled experiments." *American Economic Review*, 102, 1279–1309.

Elbers, Chris and Jan Willem Gunning (2014), "Evaluation of Development Programs: Randomized Controlled Trials or Regressions?" *World Bank Economic Review*, 28, 432–445.

Heckman, James J, Sergio Urzua, and Edward Vytlacil (2006), "Understanding Instrumental Variables in Models with Essential Heterogeneity." *The Review of Economics and Statistics*, 88, 389–432.

Kar, Kamal and Robert Chambers (2008), *Handbook on Community-Led Total Sanitation*.

Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data." *Econometrica*, 46, 69–85.

Suri, Tavneet (2011), "Selection and comparative advantage in technology adoption." *Econometrica*, 79, 159–209.

Vigh, Melinda, Chris Elbers, and Jan Willem Gunning (2017), "Community-Led Interventions to Change Hygienic Behavior: the One Million Initiative in Mozambique." *Mimeo*.

Wooldridge, Jeffrey M. (2005), "Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models." *The Review of Economics and Statistics*, 87, 385–390.

Wooldridge, Jeffrey M. (2010), "Correlated random effects models with unbalanced panels." *Mimeo.*

A   Appendix: Descriptive statistics

Table 5: Community characteristics at baseline in the regression sample

	N obs	Mean Control	Mean CLTS	p-value
Water related diseases (past 6 months)	56	0.314	0.287	0.543
Any cholera cases (past 12 months)	55	0.364	0.545	0.192
Improved water point	56	0.455	0.294	0.229
Primary school with 1-5 grades (EP1)	56	0.864	0.794	0.516
Primary school with 6-7 grades (EP2)	56	0.273	0.353	0.539
Health facility	56	0.182	0.235	0.641
Population: 1-500	56	0.182	0.235	0.641
Population: 501-1000	56	0.227	0.353	0.327
Population: 1001-1500	56	0.182	0.147	0.735
Population: 1501-2000	56	0.136	0.059	0.329
Population: >2000	56	0.227	0.206	0.852

Variables:

Water related diseases (WRD): share of households in the community affected by WRD in the past 6 months.

Any cholera cases: dummy, any reported cholera case in the community in the past 12 months.

Improved water point (IWP): dummy, the community has a functioning IWP.

Primary school: dummy, the community has a primary school.

Health facility: dummy, the community has a health facility.

Table 6: Household characteristics in the regression sample

	Year	N obs	Mean Control	Mean CLTS	p-value
Household size	2008	981	5.549	5.842	0.095
	2010	1052	5.268	5.237	0.836
	2013	918	5.399	5.738	0.056
Children <5 years	2008	981	0.894	0.922	0.650
	2010	1052	0.895	0.833	0.285
	2013	918	0.750	0.744	0.915
Wealth index	2008	981	-0.072	-0.009	0.326
	2010	1052	-0.113	0.060	0.013
	2013	918	0.066	0.089	0.776
Education (age>14)	2008	981	0.368	0.436	0.008
	2010	1052	0.370	0.465	0.000
	2013	918	0.468	0.550	0.003
Latrine ownership	2008	981	0.454	0.482	0.394
	2010	1052	0.509	0.649	0.000
	2013	918	0.601	0.660	0.067
Hand-washing (HWSD)	2008	955	0.198	0.169	0.251
	2010	1052	0.333	0.429	0.002
	2013	918	0.489	0.649	0.000

Variables:

Household size: number of people living in the household (HH).

Children < 5 years: number of children younger than 5 years living in the HH.

Wealth index: calculated using PCA of items owned by the HH, normalized to 0 in the full sample in 2008.

Education (age>14): share of HH members (age>14) with any formal education.

Latrine ownership: the HH owns any type of functioning latrine (verified by interviewers).

Hand-washing (HWSD): adult males and females are reported to wash their hands with soap or ash after defececation (self-reported).

B   Appendix: The importance of the third survey round

In section 3, we stated that our estimation method relies on $T \geq 3$ for identifying the ATE in the presence of heterogeneous treatment effects correlated with unobservable factors affecting treatment assignment. In the following, we demonstrate under which conditions we are able to separate the effect of selectivity (or correlated heterogeneity) from the ATE.

Recall from section 3 that in the estimation regression (11) the term $D_{it} \otimes (\bar{D}_i - \mu_{\bar{D}})\xi$ and $D_{it} \otimes (\bar{X}_i - \mu_{\bar{X}})\psi$ are used to control for the heterogeneity of the treatment effect (b_i). Among the variables D and X, the intervention variables (D) are of interest to us, as these are the variables that are presumably correlated with the heterogeneity in the treatment effect. Hence, in this section we focus the discussion around $D_{it} \otimes \bar{D}_i$.[21]

Let us, first, examine this expression for $T = 2$ given a single binary treatment variable and a baseline at $t = 1$. Then, for the intervention group $D_{i_I} = (0, 1)$ at $t = (1, 2)$ resulting in $\bar{D}_{i_I} = (0.5, 0.5)$, and for the comparison group $D_{i_C} = (0, 0)$ resulting in $\bar{D}_{i_C} = (0, 0)$. Now, observe that $D_{it} \otimes \bar{D}_i = 0.5 D_{it}$.[22] Due to this multicollinearity, we cannot identify the coefficient of $D_{it} \otimes \bar{D}_i$. Hence, without further information about the structure of the heterogeneous effects, we are not able to control for the correlation between intervention placement and the heterogeneous treatment effects.

What if we have an additional observation before or after the interventions were implemented? Assume that we conduct a second follow-up survey after all interventions have been completed before the second survey. Now we have $T = 3$ with $D_{i_I} = (0, 1, 1)$ and $D_{i_C} = (0, 0, 0)$ indicating whether an intervention happened before time t. Importantly, we do not have observations with $D_i = (0, 0, 1)$. Now, $D_i \otimes \bar{D}_i$ is $(0, 0.67, 0.67)$ and $(0, 0, 0)$ for the intervention and comparison groups, respectively. Hence, $D_{it} \otimes \bar{D}_i = 0.67 D_{it}$. As a result, the additional survey round does not allows us to control for the correlation between the intervention placement and the treatment effect.

Finally, let us assume that the intervention is gradually rolled out between the first

---

[21] We omit $D_{ci} \otimes \mu_{\bar{D}}$ because $\mu_{\bar{D}}$ is constant over the sample, and, therefore, this term is perfectly correlated with D_i.

[22] Which follows as, at $t = (1, 2)$, $D_i \otimes \bar{D}_i$ is $(0, 0.5)$ and $(0, 0)$ for the intervention and comparison groups, respectively.

and third round of data collection. Hence, we have observations with $D_i = (0, 0, 1)$. Now, $D_i \otimes \bar{D}_i$ is $(0, 0.67, 0.67)$ for the early intervention locations, $(0, 0, 0.33)$ for the later intervention locations and $(0, 0, 0)$ for the comparison group. Due to the differences in $\bar{D}_i$ for the early and late intervention groups, we are no longer able to express $D_i \otimes \bar{D}_i$ as a linear combination of the intervention variable, time dummies and individual fixed effects. Hence, we are able to control for the selectivity on the treatment effect (selection on gains).

It is important to point out that the above considerations do not apply for multi-valued (continuous or discrete) variables in X. For these variables, the coefficients of $D_{it} \otimes \bar{X}_i$ can also be identified when $T = 2$.