Estimating ICEs in Two Applications

Patrick Lam

September 9, 2013

I now apply the estimation framework and estimate ICEs in two applications from political science and economics. The first application revisits a field experiment from Olken (2007) on the effects of different forms of corruption monitoring on actual corruption. The second application looks at the effects of a national job training program known as JobCorps, using data from a randomized study known as the National Job Corps Study conducted by Mathematica Policy Research, Inc. I follow a similar approach and use the same dataset found in Frumento et al. (2012). The two applications are interesting for estimating ICEs for various reasons. Both are very important substantively and address issues of interest to many scholars. The corruption monitoring study is a unique and interesting field experiment that has made a substantial contribution to the study of corruption. The question of the effect of job training on employment outcomes is perhaps the most widely studied area by economists and statisticians interested in causal inference and program evaluation. In addition, the data available for both applications provide an opportunity to demonstrate the flexibility of the estimation framework and the different ways in which estimating ICEs can increase knowledge and discovery. They incorporate ICE estimation with both binary and continuous dependent variables, binary and continuous treatment variables, single-stage and two-stage estimation, and somewhat randomized and non-randomized treatment assignment settings.

1 Estimating ICEs: A Review

1.1 Framework

Recall that the main idea for estimating the individual causal effects is to estimate or impute the missing potential outcome for each observation. Knowing the missing potential outcome allows us to directly calculate the ICE and any other causal estimand. I use the combination of matching and a Bayesian model to get point estimates and uncertainty intervals for the missing potential outcome.

Under the treatment assignment ignorability and SUTVA assumptions, the distribution of potential

outcomes is identical for observations with the exact same values on the observed covariates. This implies that the distribution of the missing potential outcome for observation i can be approximated with the observed potential outcomes for a set of donor observations with the opposite treatment assignment. Since exact matching is only possible in large samples with discrete covariates, I use predictive mean matching (as described previously) to find donor pools of matches that are similar on the covariate values. To derive the posterior for the ICEs, I incorporate the matching step in a Bayesian model. The Bayesian model captures the uncertainty in the matching process, the donor pool, the parameters of the distributions of missing potential outcomes, and the imputations themselves through the joint posterior.

1.2 Estimation

The general algorithm for estimating ICEs is as follows:

MCMC Algorithm for the Posterior of τ_i

Repeat the following n_{sim} times:

Gibbs Sampler:
1. Draw a matching procedure $\tilde{\mathcal{M}}$ from $p(\mathcal{M})$.
2. Draw $\tilde{\theta}_{\mathcal{M}}$ from $p(\theta_{\mathcal{M}}|Y, X, W, D, \theta^{mis}, \mathcal{M})$.

for (i in 1:N){
3. Determine $\tilde{D}^{(i)}$ from matching procedure. **(matching step)**
4. Draw $\tilde{\theta}_i^{mis}$ to estimate θ_i^{mis}.
}

Draw from PPD and Calculate τ_i:
for (i in 1:N){
5. Draw $\tilde{Y}_i^{mis}$ from $f(\cdot|\tilde{\theta}_i^{mis})$. **(imputation step)**
6. Calculate $\tilde{\tau}_i = W_i(Y_i - \tilde{Y}_i^{mis}) + (1 - W_i)(\tilde{Y}_i^{mis} - Y_i)$.
}

For a binary treatment and continuous outcome variable, I simulate from the posterior through the following steps. Let observation i be a treated (control) observation. Choosing m-to-1 predictive mean matching with m approximately equal to 10% of the smaller treatment arm, I first estimate the parameters of the predictive mean matching $\theta_{\mathcal{M}}$ with a draw $\tilde{\beta}_c$ ($\tilde{\beta}_t$) from the posterior of a Bayesian linear regression of Y_c on X_c (Y_t on X_t). I then calculate a predictive mean score for observation i as $X_i\tilde{\beta}_c$ ($X_i\tilde{\beta}_t$) and also calculate a predictive mean score for all control (treated) observations j as $X_j\tilde{\beta}_c$ ($X_j\tilde{\beta}_t$). I then find the m control (treated) observations with the closest predictive mean score to i and designate them as the donor observations. I then draw $\tilde{\theta}_i^{mis}$ by modeling the donor pool with a

Normal likelihood and Normal prior for a model with mean and variance unknown. Using $\tilde{\theta}_i^{mis}$, I draw an imputation of the missing potential outcome $\tilde{Y}_i^{mis}$ from a Normal distribution and then calculate $\tilde{\tau}_i = W_i(Y_i - \tilde{Y}_i^{mis}) + (1 - W_i)(\tilde{Y}_i^{mis} - Y_i)$. I repeat this process for all observations i for $n_{sim} = 2000$ iterations with a burn-in length of 100.

2 Application 1: Monitoring Corruption

2.1 The Setup and Data

The first application of estimating ICEs comes from a study conducted in Olken (2007) on the effectiveness of corruption monitoring.[1] Corruption is an important topic in both the economics and political science literature, and various ways to combat corruption have been suggested. The Olken study is unique in that it is a randomized field experiment that tested the effectiveness of two types of corruption monitoring in Indonesian villages: top-down monitoring and grassroots bottom-up monitoring. Olken concluded that top-down monitoring is effective in reducing corruption while bottom-up monitoring had little impact. The study is a good example to demonstrate the use of my model because it is a relatively straightforward study that may have heterogenous treatment effects and it also collected data on multiple levels, which I use to demonstrate the flexibility of using the ICE framework.

The setting of the project is 608 villages in the Indonesian provinces of East Java and Central Java between September 2003 and August 2004.[2] Through a national Indonesian government program (Kecamatan Development Project) funded from the World Bank, each village proposes a usually infrastructure related project and is usually given some money for it. The most common type of infrastructure project is a project to surface an existing dirt road with a surface made of sand, rocks, and gravel. The study is limited to villages with such projects.

In order to ensure the proper use of funds, there are various monitoring mechanisms. Each project is associated with a series of approximately three village-level accountability meetings. In the beginning, only 40 percent of the funds are released to the implementation team. At the first village accountability meeting, the implementation team must present an accountability report explaining how the funds were used. Only after the meeting has approved the report would the other 60 percent of the funds be released. These meetings are open to the public but are typically attended by only 30-50 people, most of whom

[1] I obtained the data from the study from Olken's website at `http://economics.mit.edu/faculty/bolken/data`

[2] The following description of the study is mostly taken from Olken (2007).

are members of the village elite.

A second accountability mechanism is the threat of an audit by an independent government development audit agency known as the BPKP. Each project has approximately a 4 percent baseline chance of an audit from the BPKP. The audit process involves auditors checking all financial records and inspecting physical infrastructure. Corruption findings from the audit can lead to officials forcibly returning the money publicly or even criminal action.

In the experimental design for the study, Olken was able to randomize the two types of corruption monitoring. Broadly speaking, the experiment consisted of four treatment conditions: audit, participation either with invitations only or invitations plus comment form, or control. The audit and participation treatments were randomized independently, so a village can possibly receive both an audit treatment and a participation treatment.

- **audit treatment**: The audit treatment is a "top-down" mechanism in which an outside entity (in this case the BPKP) monitors the project for signs of corruption. For the audit treatment, villages were cluster randomized at the subdistrict level to ameliorate spillover effects (all villages within a subdistrict either received an audit treatment or not). The randomization was also stratified or blocked by district and number of years the subdistrict had participated in the program. The audit treatment consisted of increasing the probability of an audit by BPKP from 4 percent to 100 percent. Villages were informed before planning for construction that they would be audited with probability 1 either during or after construction. They were also told that the results of the audit would be presented at a village meeting, so village officials faced a possibility of punishment by the villagers, possible cutoff of funding from future KDP projects, or even criminal action. Of the 608 villages in the study, 283 received the audit treatment and 325 did not.

- **participation treatments**: The participation treatments are intended to be grassroots mechanisms in which local villagers themselves are an integral part of the corruption monitoring. The idea of the participation treatments is to increase village attendance at the village-level accountability meetings, which are open to the public but usually dominated by the village elite. Randomization of the participation treatments was done at the village level, and each village either got the intervention of invitations, invitations and comments, or control. In the invitations intervention, either 300 or 500 invitations were distributed throughout the village prior to each of the three accountability meetings. The invitations were distributed either by sending them home with school children or by asking the heads of hamlets and neighborhood associations to distribute them. The

distribution method and number of invitations were also randomized by village. In the invitations and comments intervention, villages received the invitations exactly as the invitations intervention, but in addition to the invitations, there was a comment form asking for villagers' opinions of the road project. The comment forms are anonymous and summarized by a project enumerator at each accountability meeting. Thus, the comment form produced an additional anonymous avenue through which villagers can monitor corruption without fear of retribution from village leaders. Of the 608 villages in the study, 105 received the invitations intervention, 106 received the invitations and comments intervention, and 114 did not receive a participation intervention.[3]

The corruption and misuse of funds for the projects usually came in the form of either collusion with suppliers to inflate prices or quantities of supplies used or inflated labor costs. Olken and his team measured corruption by doing an independent assessment of the "correct" costs of the project through sampling the materials used in the roads and surveys with suppliers and workers. The difference between this independent assessment and the actual costs of the project is an unbiased measure (with high error) of the corruption. For each village, Olken defined the dependent variable as the log of the reported amount minus the log of the independent assessment amount, which is approximately the *percent expenditure missing*.[4] He reports several different measures of the percent missing variable:

- Percent missing for major items in road project: sand, rocks, gravel, and unskilled labor
- Percent missing for major items in roads and ancillary projects
- Percent missing for materials in road project
- Percent missing for unskilled labor in road project

I consider all four of these continuous measures of corruption in my analyses.

Due to circumstances such as missing data, attrition, or audit treatment randomization at the sub-district level, the treatment assignment in the complete dataset may not be as clean as one would like. Fortunately, Olken also collected a few background covariates at the village level to allow for possible covariate adjustment. The covariates measured include

- Distance to subdistrict

[3]From here on out, I refer to the invitations treatment as "invites" and the invitations and comments treatment simply as "comments".

[4]Due to the noisiness of both the reported amount spent and the independent assessments, the estimates of percent missing are sometimes negative or greater than 1. Such values do not make sense in the context of percent missing so I consider the variable as simply a continuous measure of corruption.

- Education of village head
- Age of village head
- Salary of village head
- Percent of households that are poor
- Village population
- Mosques per 1,000 population
- Mountainous village dummy
- Total village budget
- Number of subprojects

In addition, Olken also collected data on the village-level accountability meetings including attendance levels. I first replicate the results from Olken's initial analyses using ICEs. I then demonstrate the flexibility of the model in estimating other quantities of interest and with different treatment variables and outcome variables.

2.2 The Effect of Monitoring Treatments on Corruption (binary treatments and continuous outcomes)

Olken's main result in the paper is that the audit treatments on average reduce corruption by about 8 or 9 percentage points while the two participation treatments have no consistent statistically significant effect on corruption. The main specification that he uses is a linear regression of the following form:

$$\begin{aligned} \text{PercentMissing}_{ijk} &= \alpha_1 + \alpha_2\, I(\text{Audit})_{jk} + \alpha_3\, I(\text{Invites})_{ijk} \\ &\quad + \alpha_4\, I(\text{Comments})_{ijk} + \epsilon_{ijk} \end{aligned}$$

where i indexes a village, j is a subdistrict, k is a stratum for the audits, and $I(\cdot)$ are indicator variables for whether a village got a specific treatment. The coefficients α_2, α_3, and α_4 are the average treatment effects for the three treatments respectively. Due to the form of the linear regression, Olken's estimated effect for one treatment averages over the distribution of the other treatments in the sample. Specifically, α_1 assumes that the treatment effect of getting the audit treatment versus no audit treatment is the same

Treatment Group	Control Group	Olken Parameter
audit; no participation	no audit; no participation	α_2
audit; invites	no audit; invites	α_2
audit; comments	no audit; comments	α_2
invites; no audit	no invites; no audit	α_3
invites; audit	no invites; audit	α_3
comments; no audit	no comments; no audit	α_4
comments; audit	no comments; audit	α_4

Table 1: Treatment and Control Groups for Average Treatment Effects using ICEs and the Corresponding Regression Parameters from Olken

regardless of whether the village got a participation treatment or not. This assumption may be violated for example, if the effectiveness of an audit is smaller with the presence of a participation treatment as well.

I first demonstrate the flexibility and comparability of estimating ICEs by comparing aggregated average effects from ICEs versus the specification found in Olken. Using the same linear specification above, I run a Bayesian linear regression with improper uniform priors to get the same results as Olken. I then run the ICE algorithm using predictive mean matching on the 10 covariates to get ICEs. To get the various average treatment effects, I simply aggregate the ICEs. There are two important differences between my approach and the original Olken approach. First, I use the covariate adjustment to deal with the less than perfect randomization, which Olken does not include in his specification. Second, I carefully define treatment and control groups to estimate the treatment effects and I allow for treatment effects to differ depending on the presence or absence of other treatment conditions.[5]

Table 1 shows the different treatment and control groups for the seven average treatment effects estimated using ICEs and the corresponding parameters from the linear regression. The rows represent the seven possible interactions between the different treatments. Since Olken did not include interaction terms in his initial model, he constrains the seven possible treatment effect interactions to three treatment effect parameters.

Figure 1 compares the results of both the average treatment effects calculated from the estimated ICEs and the average treatment effects estimated from the regression model for the four different measures of corruption. The red lines indicate the point estimates and 95% credible intervals from the regression method. Note that for each graph, the regression method only produces three distinct estimates corresponding to α_2, α_3, and α_4. The results shown in the graph suggest a few conclusions.

[5]This is equivalent to a linear regression specification with interaction terms between the treatments.

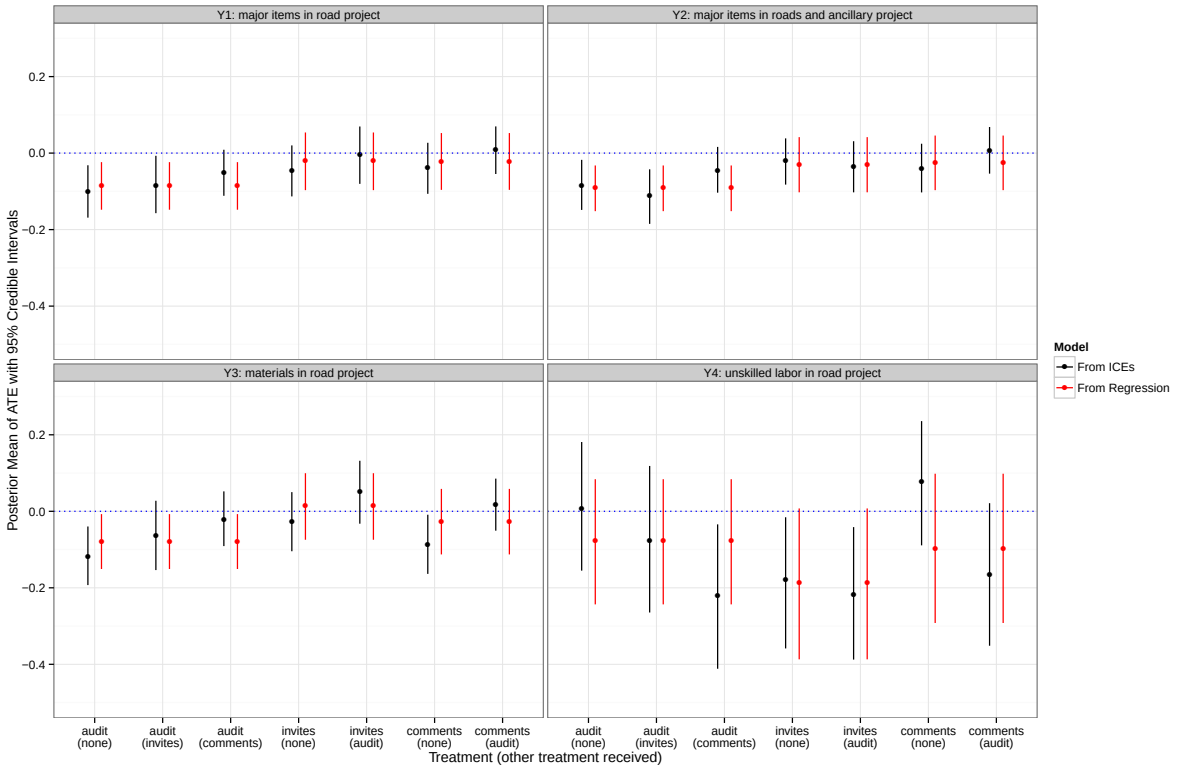

Figure 1: Comparing ICE Average Treatment Effects to Regression

- The treatment effects estimated from the ICEs are relatively close to the ones estimated from the regression method. This likely suggests that the ICE method of aggregating for average effects can recover the same estimates as the regression method, which is known to have good properties given certain assumptions. What this suggests is that the ICE model is giving reasonable answers that are similar to other tried and true methods. The slight differences between the two models are likely due to conditioning and matching on covariates and treatment effect heterogeneity given the presence or absence of other treatment conditions.

- The magnitude of the interactions between the treatments is relatively small, indicating that the presence of one treatment does not dramatically affect the effectiveness of another treatment. From the graphs, it appears that for the same treatment, the red estimates from regression are usually averages of the different black estimates from the ICE method. For example, the audit treatment effect from regression looks to be an average of the three different audit treatment effects from the ICE model. This is not surprising given how the problem was set up. There appears to be weak evidence that the treatments can crowd out one another. For example, looking at the ICE models (black lines) for $Y3$ in the bottom right panel, it is clear that the significant audit treatment in the first column is no longer significant when an invites or comments treatment is added, as made

clear in the second and third columns. Although the differences are themselves small and likely insignificant, this does confirm intuition that multiple monitoring treatments are not necessarily additive.

- The results do also seem to confirm the substantive conclusion that Olken reaches that the audit treatment leads to an approximately 8-9 percentage point decrease in corruption while the participation treatments do not have a consistent effect. However, the results also suggest that the "statistical significance" of the effects from a hypothesis testing standpoint is very tedious, and the presence of multiple treatments can render the results insignificant.

This first result demonstrates the ability of the ICE estimation method to recover various causal quantities accurately when benchmarked against more traditional methods. The estimation process also forces the researcher to think very clearly about what constitutes the treatment and control groups, which leads to a more clear exposition of what the treatment effect represents. Finally, the results presented also show a simple example of how the ICE method can estimate treatment effect heterogeneity in a straightforward manner that mirrors the use of interaction terms in regression. In this case, the treatment effects were estimated separately in the presence and absence of other treatment effects.

2.3 Audit Treatment Effect Heterogeneity

Since the audit treatment seemingly has a significant positive effect, I explore this effect further by looking at treatment heterogeneity and other types of treatment effects using ICEs. I condition on the presence of the other treatments by only comparing observations with the same status on the participation treatments within the matching step. For example, for observations that received the audit treatment and the invites treatment, I only match to observations that do not receive the audit treatment but do receive the invites treatment to estimate the ICEs. I do the same for observations receiving the comments treatment and for those that do not receive a participation treatment. The estimated ICEs are then used to calculate other quantities of interest.

One of the main benefits of the ICE approach is the ability to estimate any treatment effect by simple aggregation. Figure 2 shows the results of average treatment effects for the audit treatment within different subgroups of the data for each of the four measures of corruption.

The first three columns of each panel represent the posterior of the average treatment effect (ATE), average treatment effect for the treated (ATT), and average treatment effect for the controls (ATC) using

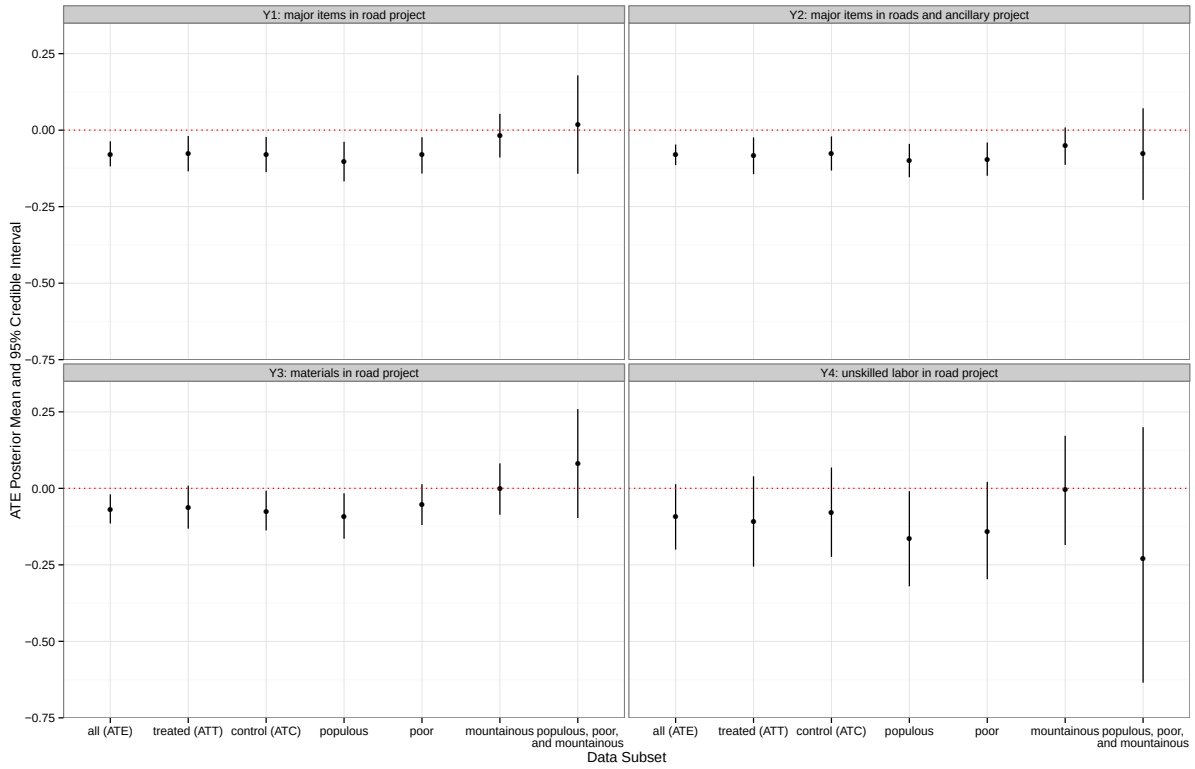

Figure 2: Audit Average Treatment Effects within Subgroups

the ICE estimates. The posteriors are derived simply by averaging the posteriors for all observations, treated observations only, and control observations only respectively. Typically, in observational studies, the ATT and ATE may be different if treatment assignment depended on some covariate that was also correlated with the treatment effects. Since treatment assignment was more or less randomized in this case, it is unsurprising that the ATE, ATT, and ATC are very similar.

The next four columns of each panel show average treatment effects for various subgroups of the data defined by specific covariate values. "Populous" subsets the ATE to villages with population greater than the dataset average. One theory may be that larger villages may be prone to more corruption because it may be harder for citizens to monitor officials due to collective action problems, so an outside audit may be more helpful. "Poor" indicates the ATE for villages with greater percent of households that are poor than the dataset average. One might expect that villages with more poor households may be more susceptible to corruption and thus an outside monitoring mechanism such as an audit may have a greater effect than in wealthy villages. "Mountainous" denotes the ATE for villages that are located in a mountainous region. One can argue that geographically isolated villages have a stronger social bond, which allows for more monitoring within the village, so outside audits may be less helpful. And finally,

"populous, poor, and mountainous" denotes villages that are large, poor, and within a mountainous region. The results show that the ATEs for populous and poor regions is not significantly different from the overall average, but audits seem to have a smaller and insignificant effect in mountainous villages. Subsetting the dataset by all three criteria together renders the sample size too small and the uncertainty intervals become quite wide. The results from Figure 2 suggest that treatment heterogeneity by subgroup may not be a huge problem. It also shows the flexibility of examining treatment effect heterogeneity by simply combining ICEs for various subgroups of observations.

Detecting treatment effect heterogeneity by finding average treatment effects within subgroups is very similar to existing methods and practices. However, estimating ICEs also allows researchers to look at the individuals themselves and look for treatment effect heterogeneity through various graphical methods. As an example, suppose the researcher would like to know whether the audit treatment would have a large effect on specific villages and how that effect differs across villages. One benefit of the Bayesian approach is that it allows the researcher to make probability statements about parameters in a coherent manner. Suppose a large effect for the audit treatment is defined as decreasing the percent missing by 20 percentage points. Then the probability of a large effect for any village is simply the probability of an individual causal effect of less than or equal to -0.2. With simulations from the posterior, this becomes simply the proportion of draws less than or equal to -0.2 for a specific τ_i.

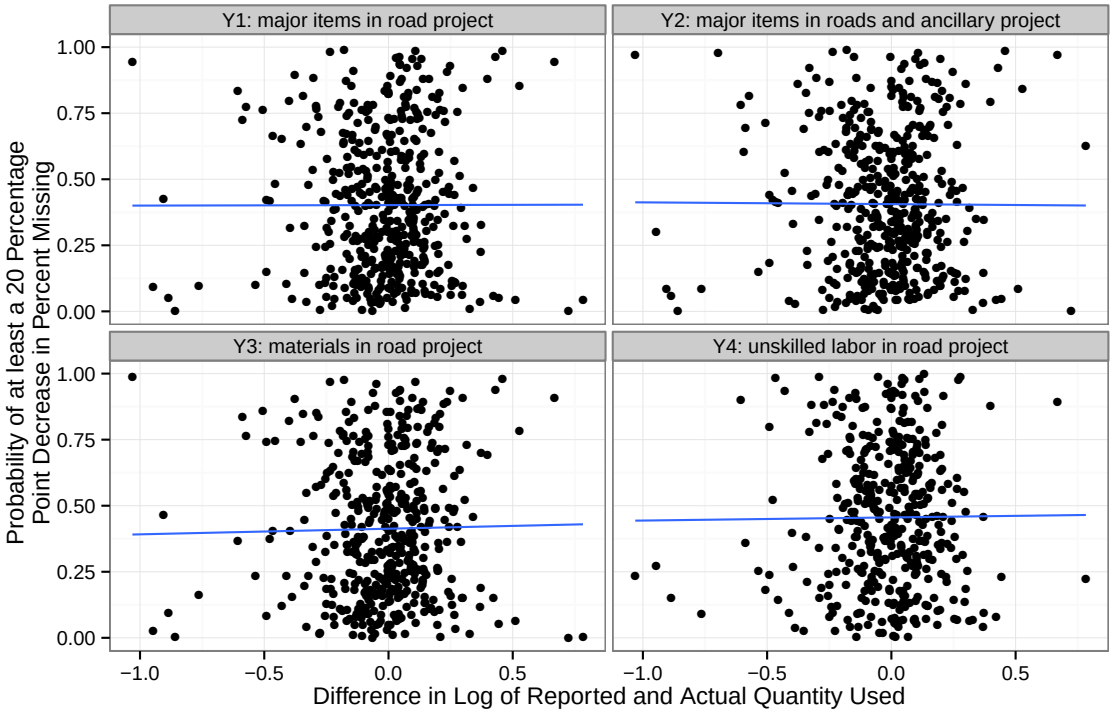

Figure 3: Probability of a Large Audit Treatment Effect by Quantity Overreporting

Figure 3 plots the probability of $\tau_i \leq -0.2$ on the y-axis and the difference in log of reported versus actual quantity of materials or labor used on the x-axis with a best-fit line drawn. Each point on the plot is a single village and each of the four panels on the graph represents one of the four different corruption variables. The y-axis is simply the probability that the audit treatment has a large effect. The x-axis represents how much a village over-reports its materials and labor usage. Recall that corruption can occur through over-reporting of quantity and/or inflating of prices. The results suggest that there is no relationship between how well the audit works and how much quantity over-reporting there is.

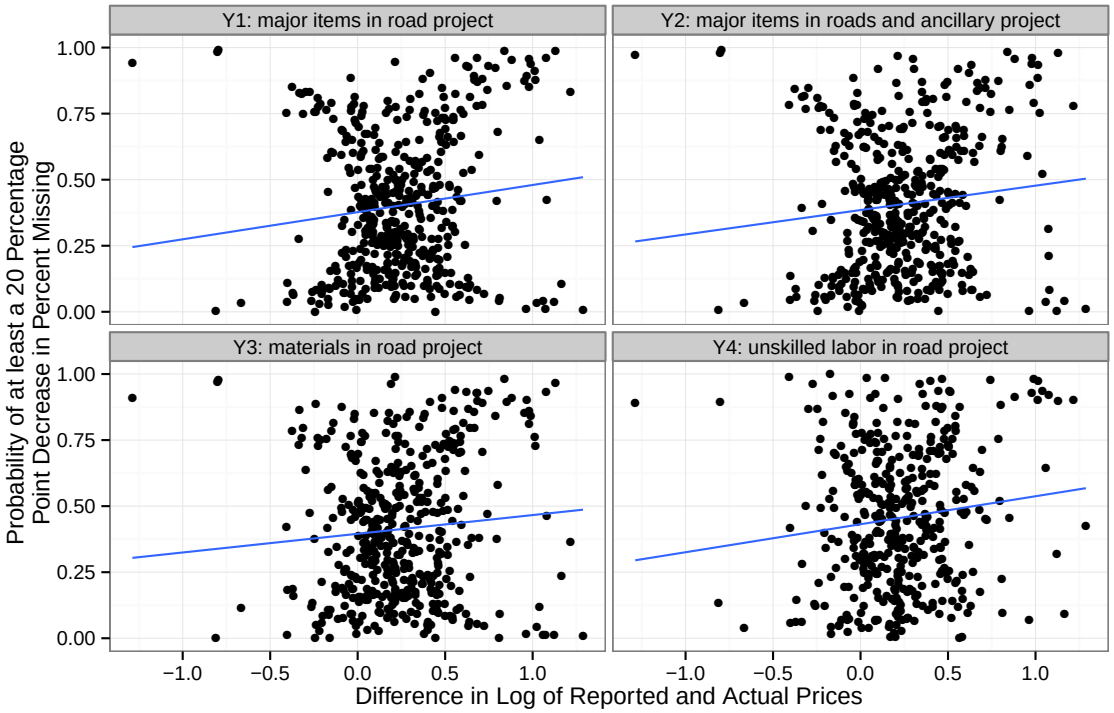

Figure 4: Probability of a Large Audit Treatment Effect by Price Overreporting

However, Figure 4 suggests there may be a relationship between audit treatment effectiveness and price inflation, which is plotted on the x-axis. It seems there is a slightly positive relationship where the probability of a strong audit effect increases with an increase in price inflation. The effect may be even stronger after discarding outliers in the top left of the graphs. The positive relationship suggests that audit treatments may be more effective in villages that over-report their prices. One explanation may be that prices are probably easier to check in an audit by comparing various outside sources, while quantity used may be harder to check in an audit. Therefore, audits work much more effectively in catching price inflation than quantity inflation. Figures 3 and 4 demonstrate one simple graphical way of detecting treatment effect heterogeneity. Given the posteriors of all the individual causal effects, treatment effect heterogeneity is straightforward to examine and researchers can make simple probability statements

about the heterogeneity without resorting to hypothesis testing and the many issues that associated with it.

2.4 Treatment Effect Quantiles

The ICEs from the entire sample form a distribution of causal effects, which researchers may also be interested in. As mentioned before, the ICEs are only in-sample quantities, so any extrapolation from sample quantities to population quantities requires assumptions about how representative the sample is to the population. Nevertheless, the entire distribution of ICEs allows researchers to see what the entire range of effects are and to also look at treatment effect quantiles. However, an important distinction must be made between treatment effect quantiles and quantile treatment effects, the latter of which researchers have tried to develop methods for. A treatment effect quantile refers to the quantiles of the treatment effects whereas a quantile treatment effect refers to the difference of potential outcomes at a specific quantile for each of the two potential outcome distributions. Let $q(\cdot)$ be a quantile function for any quantile. Then

$$\begin{aligned} \text{treatment effect quantile} &= q(Y(1) - Y(0)) \\ \text{quantile treatment effect} &= q(Y(1)) - q(Y(0)) \end{aligned}$$

In the case of average effects, the average treatment effect is equal to the difference in the average of the potential outcome distributions because of the linearity in expectations property. However, in the case of quantiles, the two quantities are different unless strong assumptions about rank order are made. Existing methods such as quantile regression try to estimate the quantile treatment effects, but I argue that treatment effect quantiles are the actual quantities researchers are interested in. Previous methods were unable to estimate treatment effect quantiles due to identification problems.

Figure 5 plots the treatment effect quantiles for the three treatments at the 25th, 50th, and 75th quantiles. The results suggest that the range of individual treatment is quite large and can vary from -0.5 to 0.5. Intuitively, this does not make sense as one would not expect corruption monitoring to increase corruption. There are several possible explanations for this result. The first is that the dependent variables are measured with such noise, with quite a few observations receiving nonsensical values of greater than 1 or less than -1, that the treatment effect quantiles results are driven by such measurement errors. The second explanation may be that quantiles on the extremes of the distribution are estimated with less accuracy as my simulations showed. Therefore, one should consider the treatment effect median

to be more accurate than the other quantiles. Finally, one may consider that there may actually be some cases where monitoring inadvertently leads to more corruption. For example, in the audit treatment, the auditors themselves may be corrupt, and there exists possible collusion or bribery opportunities between the auditor and the project managers, especially since the audits were announced ahead of time. This possible collusion may inadvertently lead to more corruption. Nevertheless, Figure 5 shows that it is possible to get estimates of treatment effect quantiles by looking at the distribution of ICEs.

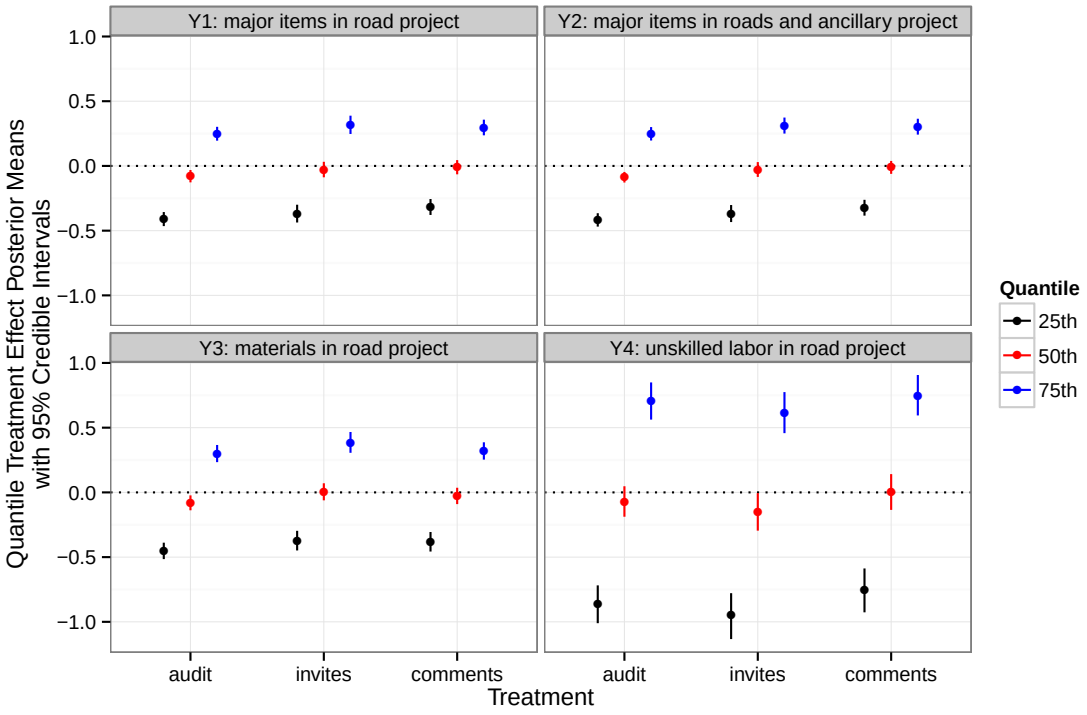

Figure 5: Three Treatment Effects Quantiles

2.5 The Effect of Participation Treatments on Outsider Village Meeting Attendance

Despite the results from above suggesting that only the audit treatment has a significant effect on corruption, I look more closely at the participation treatments and its mechanisms. The participation treatments also provide for an opportunity to demonstrate the flexibility of estimating ICEs because they can be thought of as part of a two-stage data structure. Recall that the participation treatments were theorized to be effective through the grassroots mechanism of increasing non-elite village turnout at village accountability meetings (first stage) and the increased attendance of outsiders should decrease

the likelihood of corruption (second state). It is important to note that this is the only channel through which participation should reduce corruption. Olken was able to record actual attendance data at the three accountability meetings in each village, so I can use this data to estimate the effect of the first stage of treatment on non-elite (outsider) village attendance.

Figure 6 shows the results of the participation treatments on the raw outsider meeting attendance numbers and outsider meeting attendance as a percent of total attendance for each village averaged across three meetings. "Invites" refers to the treatment of sending invitations only, whereas "comments" refers to both an invitation and anonymous comment form, and "participation" lumps the two treatments together into a broad category. Recall also that the treatments were distributed randomly either by sending them home with children at schools or through neighborhood heads. The red and blue lines separate out the two delivery mechanisms. I use the same method to estimate the ICEs as before and the dependent variable is treated as a continuous variable. Since there are a variety of treatments and delivery mechanisms, I focus here only on the average treatment effects for the treated (ATTs) rather than the ATEs. These two quantities of interest should be equal given random assignment of treatment.

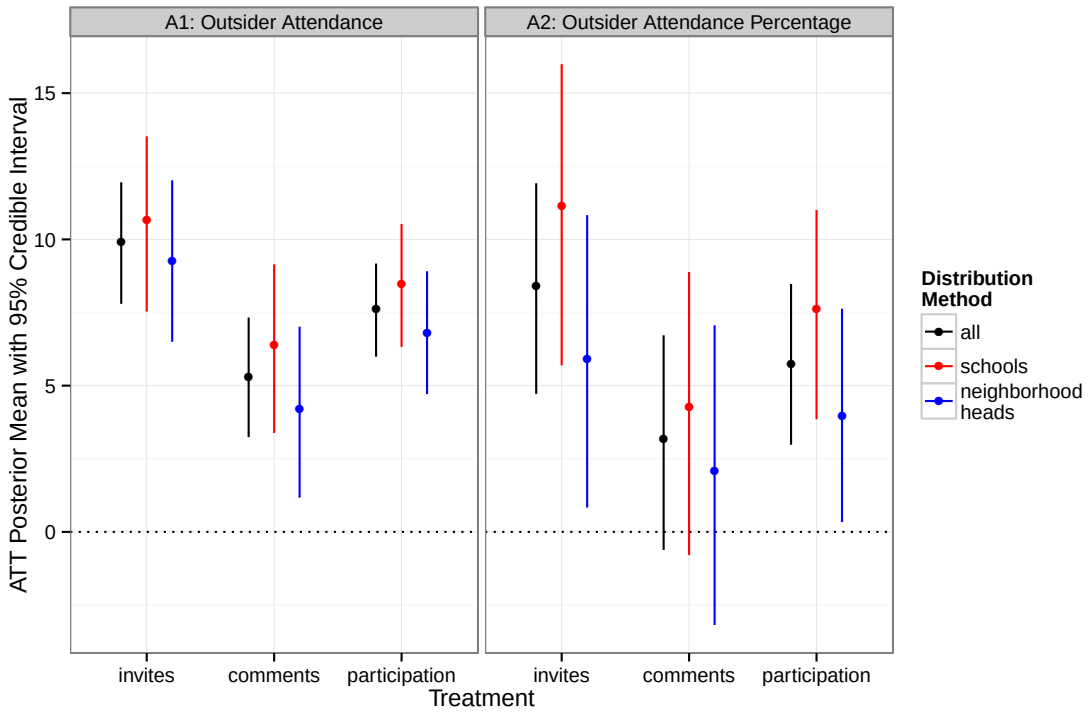

Figure 6: Participation ATTs on Outsider Meeting Attendance

The results from Figure 6 lead to several conclusions. First, it appears that the participation treatments

generally do lead to a significant increase in outsider attendance at the accountability meetings. Receiving a participation treatment in general increases outsider attendance by an average of around 7.5 people or around a 5 percentage point increase of outsiders as a percentage of the audience. Second, it appears that the invitations alone are more effective at increasing outsider attendance than an invitation and an anonymous comment form. This makes sense since the comment forms are a way for villagers to express opinions about the projects without fear or identification and retribution, so they act as a substitute for actually attending the meeting. And finally, it appears that the treatments are slightly more effective when distributed through schools as opposed to through neighborhood heads. This also makes sense since it may be the case that the neighborhood heads are more likely to be corrupt and less likely to have an incentive to increase outsider attendance at the meetings. Overall, the results suggest that the participation treatments actually work as intended in increasing outsider attendance to the accountability meetings.

2.6 The Effect of Outsider Village Meeting Attendance on Corruption (continuous treatments and outcomes)

The previous subsection showed that the participation treatments have a positive and significant average effect on outsider attendance at the village accountability meetings. In this subsection, I look at whether increasing outsider village meeting attendance has the effect of reducing corruption. Here I look at this second stage independently of the first stage. The next subsection will incorporate the two stages together into one model.

This second stage also provides an opportunity to demonstrate how the ICE model I use can be adapted to accommodate non-binary treatments. In this case, both outsider meeting attendance and outsider meeting attendance percentage are considered continuous "treatment" variables, denoted A.[6] The key assumption required for continuous treatments is a linearity assumption, where the effect of continuous treatment A is assumed to be linearly related to the outcome Y. The linear ICE is then simply the effect of increasing A by one unit. The way to conceptualize this is that there are an infinite number of potential outcomes $Y(A)$ since there are an infinite number of possible values for A. The linearity

[6] I refer to the attendance variable as treatment variables here when looking at this second stage independently. They are treatments in the sense that I am interested in their effects on corruption looking only at the second stage. However, in the overall scheme of the study, the treatments are still the participation and audit interventions. I denote these second stage "treatments" with A to avoid confusion.

assumption imposes a structure where the ICE is

$$\tau_i = Y_i(A+1) - Y_i(A);\ \forall A$$

Note that this is equivalent to the previous definition of τ_i for binary treatments if $A = 0$.

To simulate from the posterior for τ_i with continuous treatments, only a few minor adjustments are necessary to the original algorithm.

- Previously, the set of possible donor observations for observation i was all observations with the opposite treatment status. For continuous treatments, the set of possible donor observations for observations i is any observation with a different value on the treatment variable A. Since A is continuous, the set of possible donors is likely to be nearly every other observation in the dataset.
- Denote the counterfactual treatment status[7] for observation i as $A_i + 1$. Then $Y_i^{mis} = Y_i + \tau_i$.
- Once the donor pool has been determined from the matching step, to draw the equivalent of $\tilde{\theta}_i^{mis}$, simply run a linear regression step of Y on A with the donor pool. Let $\tilde{\lambda}_{0i}$ and $\tilde{\lambda}_{1i}$ be the intercept and slope draws from this regression. Then $\tilde{\theta}_i^{mis} = \tilde{\lambda}_{0i} + \tilde{\lambda}_{1i}(A_i + 1)$.
- To draw $\tilde{Y}_i^{mis}$, simply draw from a Normal distribution (for continuous outcome variables) with mean $\tilde{\theta}_i^{mis}$ and the standard deviation equal to $\tilde{\sigma}$ from the regression step above. Then $\tilde{\tau}_i = \tilde{Y}_i^{mis} - Y_i$

The main differences between this and the previous algorithm is simply modeling the donor pool with a linear regression rather than with a Normal model and then specifically defining a counterfactual treatment status. The counterfactual treatment status in the case of binary treatments is already strictly defined as the opposite treatment status whereas in this case, there are an infinite number of potential counterfactual treatment statuses.

Using this algorithm and setup for continuous treatments, Figure 7 shows the results of the effects of outsider attendance and outsider attendance percentage on the four corruption measures.[8] Note that since outsider attendance was not randomly assigned, this second stage analysis resembles an observational study. I calculate the (linear) ATE, which is simply the average of all the ICEs in the data. In

[7] With the linearity assumption, one can really define any counterfactual to estimate the ICE. I use $A_i + 1$ here for simplicity.

[8] In the matching specification for these models, I also include the treatment statuses for the invites, comments, and audit treatments, whether the participation treatments were distributed through schools or neighborhood heads, and total meeting attendance as control variables in addition to the original ten covariates. None of these are post-treatment since the treatment in this case is outsider attendance.

the case of continuous treatments, the "treatment" and "control" groups are not well defined, so ATT and ATC are also not well-defined. The black lines represent the ATE using all observations while the red and blue lines indicate the ATEs for the subgroups of observations that received or did not receive the audit treatment respectively.

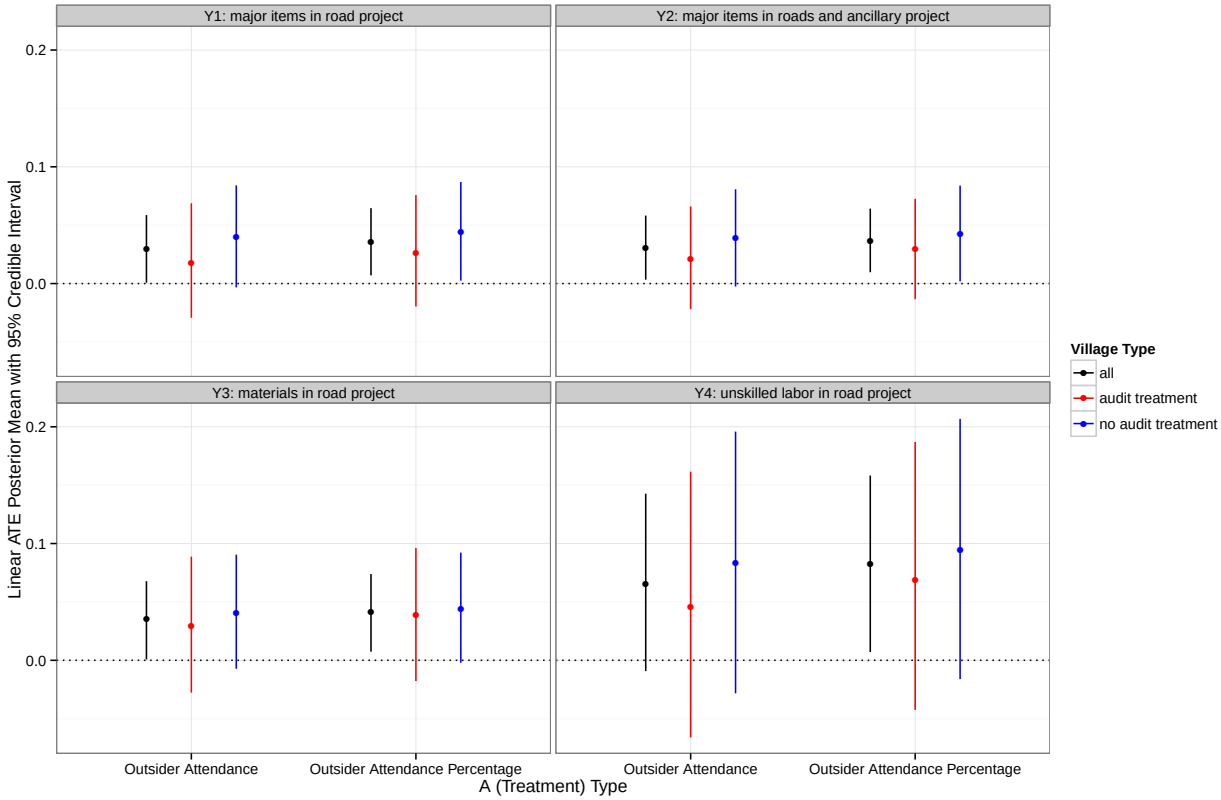

Figure 7: Linear ATE of Outsider Meeting Attendance on Corruption

The results from Figure 7 suggests that increasing outsider attendance by one person or increasing outsider attendance percentage by one percentage point does not really have a significant effect on decreasing corruption. In fact, the point estimates seem to suggest that increasing outsider attendance may actually increase corruption, although the credible intervals often cover zero. With the same caveats about the corruption variables measured with high error, it seems that the grassroots approach to corruption monitoring is ineffective. Although the participation treatments do increase participation, this increase does not appear to lead to a similar increase in accountability.

2.7 Two-Stage Analyses of the Effect of Outsider Meeting Attendance on Corruption

A proper analysis of the effect of increasing outsider meeting attendance should take into account both stages of data. The previous subsection only looked at the effect in the second stage without taking advantage of the randomization of the participation treatments. In this subsection, I demonstrate how to estimate ICEs in a two-stage framework that mirrors existing methods. I consider the participation interventions here to be one intervention without differentiating between invites and comments. There are two ways to conceptualize the two-stage analysis, both based on broad sets of existing methods. The first and more common way to think about the problem is to look at it through the lens of instrumental variables. The second way is to think about it as a problem of identifying causal mechanisms. I use the instrumental approach here, although the framework can be used to identify causal mechanisms as well.

The hypothesized causal pathway is as follows. Villages get assigned to either receive a participation intervention or not. Villages that receive a participation intervention should experience an increase in outsider meeting attendance because of the treatment. The increase in outsider meeting attendance should then result in more corruption monitoring, which should then lead to lower levels of corruption. So far, I have shown that participation interventions do increase outsider meeting attendance on average, but increasing outsider meeting attendance on average does not reduce corruption. However, since both estimates were averages, I have yet to show the effect of outsider meeting attendance on corruption *in those villages where participation increased outsider meeting attendance.* The ICE framework allows me to examine this problem further by specifically linking the two stages together on an individual village level.

In the typical instrumental variables setup, there is a treatment variable of interest where treatment assignment is not ignorable. However, there exists an instrument that has ignorable assignment and is correlated with the treatment variable. The analysis then leverages the ignorable assignment in the instrument to identify the effect for the treatment variable. In this case, the participation treatment would be the instrument and the outsider meeting attendance would be the treatment variable of interest.[9] Under certain assumptions, the instrumental variables analysis can estimate and identify a local average treatment effect (LATE), which is the average treatment effect for compliers. Compliers here are defined as the subgroup of individuals for whom the instrument affects the treatment variable in the hypothesized direction when given the instrument and has no effect when not given the instrument. In our example, a

[9]For this subsection, I only consider the raw outsider meeting attendance number rather than outsider meeting attendance percentage.

village is classified as a complier if outsider meeting attendance increases when receiving the participation intervention and stays the same when not receiving the participation intervention. The LATE is then the effect of outsider meeting attendance on corruption for complier villages.

To identify the LATE in this example (and generally speaking for instrumental variables), the following assumptions must hold:

- **Stable treatment value assumption (SUTVA)**: assumed to hold, although slightly violated by the differing treatments of invites and comments.
- **Ignorable assignment of the instrument**: assumed to hold because of random assignment of participation.
- **Exclusion restriction**: assumes that the participation interventions affect corruption only through the channel of outsider meeting attendance; assumed to hold.
- **Non-zero average causal effect of participation intervention on outsider meeting attendance**: shown to hold in previous sections.
- **Monotonicity**: participation interventions only affect outsider meeting attendance in one direction; assumed to hold although I relax this assumption later.

The key to identifying LATE is to identify which villages are compliers and which are not. If compliance status is known, then LATE would be easy to estimate. However, compliance status is not known, but I can estimate compliance status in the first stage using the ICE framework and then use the ICE framework in the second stage as well to estimate LATE given compliance status.

Consider the following way to use ICEs in an instrumental framework setting. In the first stage, estimate the ICEs for all observations to get the individual effects of the participation intervention on outsider meeting attendance. The posterior of the ICEs represent the uncertainty over compliance status. For each draw from the posterior, consider a village to be a complier village if the ICE is positive and not a complier if the ICE is not positive. For each iteration, classify every village as either a complier or non-complier based on the first stage ICE. The draws from the entire posterior of this first stage characterize the uncertainty over whether or not a village is a complier. The probability of village i being a complier village is simply the proportion of posterior draws greater than 0 in this first stage.

Next, denote the missing potential outcomes from the first stage as A^{mis}. Then, in the second stage, implement the ICE algorithm a second time with the corruption measure as the outcome and outsider

meeting attendance as the treatment. This is the same algorithm as above for continuous outcomes and continuous treatments. However, one key difference is that the counterfactual treatment here is the A^{mis} from the first stage, whereas before, the counterfactual was arbitrarily chosen to be $A - 1$. The idea behind this is that A_i^{mis} is the imputed outsider meeting attendance for observation i if it had received the opposite participation intervention. Then Y_i^{mis} is the potential outcome for corruption given a hypothetical outsider attendance value of A_i^{mis}. A second key difference is that in the potential donor pool at the second stage, donors must be of the same compliance type. So if observation i is drawn as a complier in the iteration, then the donor observations must also be drawn as compliers in that iteration. The ICE algorithm simply imputes two missing potential outcomes for the opposite participation treatment. For each draw of the algorithm, I draw a set of compliers and then draw an estimate of LATE.

The specifications of this two-stage model can vary in several ways. For example, one can include control variables to match either in the first stage or the second stage or both. The assignment for the instrument must be ignorable, so it must be randomly assigned or ignorable after controlling for covariates. In the second stage, including control variables in the matching is optional and may or may not increase the precision of the estimates. One can also choose not to include matching variables, in which case the donor pools in the first and second stages would simply be all observations with a different instrument and treatment statuses respectively.

Another way to alter the specification is to impose the monotonicity assumption. In the specification I initially described, the monotonicity assumption is not strictly necessary and not imposed. It allows the participation intervention to actually decrease outsider meeting attendance. However, if a monotonicity assumption makes sense substantively, imposing it in the algorithm will improve estimates and reduce noise. Let i be an observation that receives the participation intervention. To impose the monotonicity assumption in this example, I must constrain A_i^{mis} produced from the first stage ICE to be less than or equal to the observed A_i. If $A_i^{mis} > A_i$ for any draws of A_i^{mis}, then I simply change the imputation of A_i^{mis} such that $A_i^{mis} = A_i$

Figure 8 presents the results of various specifications of this two-stage model of the raw average outsider meeting attendance number on corruption using the participation intervention as an instrument. I consider four different specifications: two models with the monotonicity assumption, with and without second stage matching, and two models without the monotonicity assumption. I consider two quantities of interest for the four dependent variables: the LATE and the non-complier average treatment effect (NCATE). The LATE considers only compliers whereas the NCATE considers only non-compliers.

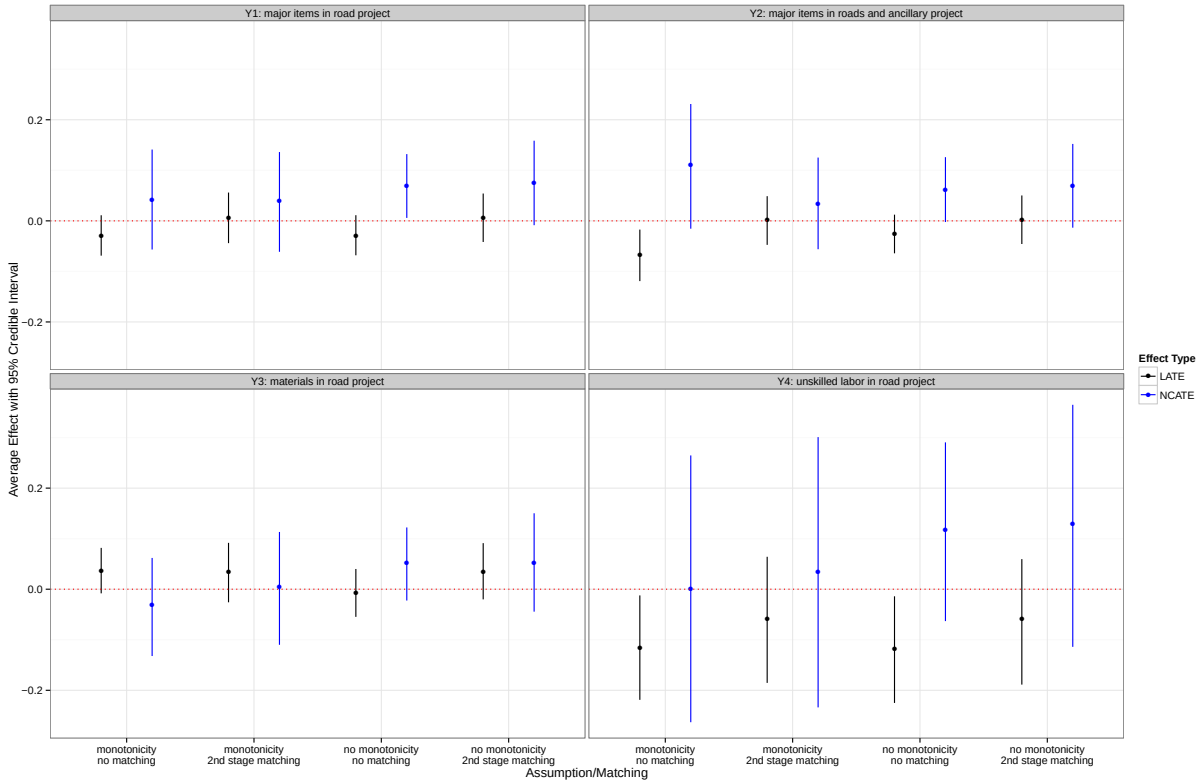

Figure 8: Two-Stage ATEs of Outsider Meeting Attendance on Corruption

The results from Figure 8 lead to several conclusions. First, consider the NCATE estimates. The NCATE is a way to test the validity of the exclusion restriction. Recall that the exclusion restriction states that the instrument only affects the outcome through the treatment. If the exclusion restriction holds, then the NCATE should be zero since the instrument should not be affecting the outcome for non-compliers. The blue lines in Figure 8 confirm that the NCATE is likely zero, suggesting that the exclusion restriction is a valid assumption. The LATE estimates across the specifications and corruption measures suggest that outsider meeting attendance does not have a significant effect on corruption. This confirms the result from before that grassroots monitoring is not very effective in reducing corruption.

3 Application 2: The National Job Corps Study

The second application implements the ICE algorithm on a randomized study of a job training program in the US. The question of whether or not job training programs are effective is one of the most widely evaluated questions in the fields of economics and causal inference. The specific data used here comes from the National Job Corps Study conducted by Mathematica Policy Research, Inc. The job training

program, known as Job Corps, offers job training for disadvantaged youths between the ages of 16 and 24. The study here involved a random sample of all eligible applicants for the program in late 1994 and 1995. I obtained the dataset from Frumento et al. (2012) and closely mirrored the analyses in their paper.

In the original study, 15,386 individuals were sampled and assigned either a treatment (9,409) or control (5,977) intervention. The treatment group was offered the opportunity to enroll in the program while the control group was denied access to the program for three years. Interviews were then conducted with the entire experimental population at baseline and then at 52, 130, and 208 weeks after the random assignment. Due to problems with incomplete baseline interviews, individuals who died during the follow-up, and people who were admitted to the program even though they were assigned to control, the resulting experimental population consisted of 13,987 individuals. Of the individuals that were in the treatment group, not all of them chose to enroll in the program. The treatment group compliance rate (those who were assigned to treatment and then enrolled) was about 68%. The following background covariates were collected on all individuals at baseline:

- Gender
- Age
- Has children
- Years of education
- Mother's years of education
- Father's years of education
- Has job
- Months employed in previous year
- Had job in previous year
- Earnings in previous year
- White or non-white
- With or without a partner
- Ever arrested

- Whether household income > $6000
- Whether personal income > $6000

I deal with missingness in the covariates by using only one imputation from a set of multiple imputations, following the same method as Frumento et al. (2012). They justify using only a single imputation by stating that there was very small variability in the results across multiple imputations. At the follow-up interviews, two outcomes are measured: employment and wages. For the purposes of this application, I only look at the binary employment outcome (employed or not), although future extensions should also look at wages.

Frumento et al. (2012) address three issues with the study in their paper: treatment assignment noncompliance, partially defined wages due to nonemployment, and unintended missing outcomes. Because my focus is on estimating ICEs and showing the flexibility of the model in examining treatment effect heterogeneity, I only address the first problem of noncompliance. I exclude the second problem by looking only at employment rather than wages, and I ignore the third problem by dropping observations with missing outcomes. The latter may induce bias when looking at population estimands, but theoretically poses no problems when limiting the analysis to the sample or individual estimands. I deal with the problem of noncompliance by using principal stratification in a formal two-stage model. The principal strata are defined by estimating ICEs in the first stage. While the first application of monitoring corruption also included a two-stage model, I more formally define the model in the second stage. This application is also different from the first in that all the outcomes and treatments in the two stages are binary variables, which allows for easier notation.

3.1 A Two-Stage Model for the Effect of Job Training on Employment with ICEs

The outcomes I am interested in are the employment statuses of individuals in the experiment at 52 weeks, 130 weeks, and 208 weeks after randomization. Assignment to being offered the choice of enrolling into the job training program is randomized, but actual enrollment in the program is not. Let Z denote the binary treatment assignment and let W denote the binary enrollment in the program indicator. Y denotes any one of the three binary outcome variables. The two-stage model here incorporates the first stage of the effect of Z on W and the second stage effect of W on Y. The setup is a typical instrumental variables study where Z is the instrument. Since W was not randomized, I rely on the randomization

of Z to identify treatment effects. All the typical IV assumptions of SUTVA, monotonicity, exclusion restriction, non-zero average effect of Z on W, and ignorable assignment of Z are assumed here.

Researchers are generally interested in two types of average treatment effects in this setup: the intention-to-treat effect (ITT) of Z on Y and the local average treatment effect (LATE), which is the effect of W on Y for compliers. Compliance here is defined as enrolling in the program if offered and not enrolling if not offered. Due to the nature of the program, I assume that there is only one-sided noncompliance in that individuals can choose not to enroll if offered treatment but they cannot choose to enroll if not offered treatment (monotonicity assumption). Notationally, I define compliance with the potential outcomes framework where W(Z) is the enrollment status given treatment status Z. Then, for compliers,

$$\begin{aligned} W(1) &= 1 \\ W(0) &= 0 \end{aligned}$$

and for non-compliers,

$$\begin{aligned} W(1) &= 0 \\ W(0) &= 0 \end{aligned}$$

Let G be a binary indicator for whether an individual is a complier or not and let $Y(Z)$ denote the potential outcome for Y given treatment assignment Z. Then, the typical treatment effects estimated under this setup are

$$\begin{aligned} \text{ITT} &= E[Y(1)] - E[Y(0)] \\ \text{LATE} &= E[Y(1)|G=1] - E[Y(0)|G=1] \end{aligned}$$

The unbiased estimate of the ITT is simply a difference in means of Y given randomization of Z. Estimating LATE requires first estimating group membership for each individual.

I use principal stratification (Frangakis and Rubin, 2002) and stratify observations given their Z and W indicators. Let $S(Z,W)$ denote a strata of observations with observed values of Z and W. Due to the assumption of only one-sided non-compliance, there are three strata in the data: $S(1,1)$, $S(1,0)$, and $S(0,0)$. The compliance statuses of individuals in $S(1,1)$ and $S(1,0)$ are known. Since individuals not assigned treatment cannot enroll in the program, it must be the case that everybody in $S(1,1)$ are

compliers and everybody in $S(1,0)$ are non-compliers. The only uncertainty in compliance status is with the 5,299 individuals in $S(0,0)$.

I can estimate group membership status using the ICE algorithm in the first stage. Since estimating group membership for $S(0,0)$ is equivalent to estimating $W(1)$, the problem can be considered as one of estimating the ICEs of Z on "outcome" W. This first stage estimation gives the posterior probability of any individual belonging to the compliers group. Using the draws from the first stage, I can then implement a second stage where I estimate the ICEs of W on Y conditional on individuals being drawn as compliers to find complier treatment effects. Simply put, the algorithm is very similar to before. For each iteration of the MCMC, draw a value for W_i^{mis} for $i \in S(0,0)$. Determine compliance status using W_i and W_i^{mis}. For complier treatment effects then, take all individuals labeled as compliers and estimate ICEs with W as treatment and Y as the outcome. In both the first and second stages, matching on covariates is not strictly necessary since Z is randomized. However, using matching can improve estimates by subsetting the potential donor observations to a smaller set of more similar observations rather than using the entire set of observations in the other treatment group. At worst, matching poorly will simply produce a random draw from the potential donor pools. Given the large number of observations in this study, matching done correctly will almost certainly reduce the variance of the estimates.

Recall the original setup for estimating ICEs. Let Y_i^{mis} be the missing potential outcome to be imputed and let θ_i^{mis} be the mean of the distribution for Y_i^{mis}. Let $D_j^{(i)}$ be an indicator for whether the jth observation is a match for observation i. The random component in the model is the outcome when observation j is a match to observation i when i is the individual of interest. Y_i is fixed and therefore not a quantity of interest for modeling. The simplified[10] version of the original posterior was

$$p(\theta|Y,X,W) \propto p(D|\theta)p(Y|D,\theta)p(\theta)$$

where the posterior was augmented with D. In the two-stage model here, the posterior is augmented again with compliance status G.

Let π_j denote the probability of observation j being a complier. For the simple case where compliance status is estimated without matching, the empirical complier proportion can be used:

$$\hat{\pi}_j = \frac{\sum_{i=1}^{N} I(i \in S(1,1))}{\sum_{i=1}^{N}[I(i \in S(1,1)) + I(i \in S(1,0))]}$$

[10] I suppress the notation for the matching to keep things simple.

where $I(\cdot)$ is an indicator variable. Recall the previous complete likelihood[11] for the one-stage ICE model:

$$\begin{aligned}\mathcal{L}_{comp}(\theta^{mis}|Y,D) &= p(Y,D|\theta) \\ &= p(Y|D,\theta^{mis})p(D|\theta) \\ &= \prod_{i=1}^{N}\prod_{j=1}^{N}\left[p(Y_j|\theta_i^{mis})p(D_j^{(i)}|\theta_{\mathcal{M}},\mathcal{M})\right]^{D_j^{(i)}}\end{aligned}$$

With the two-stage model, there is a second data augmentation using compliance status G. If D and G were observed, the complete data likelihood would be

$$\begin{aligned}\mathcal{L}_{comp} &= p(Y|D,G,\theta)p(D|\theta)p(G|\theta) \\ &= \prod_{i=1}^{N}\prod_{j=1}^{N}\left(\left[p(Y_j|\theta_i^{mis})p(D_j^{(i)}|\theta_{\mathcal{M}}^{(G)},\mathcal{M})\right]^{D_j^{(i)}}\pi_j^{G_j}(1-\pi_j)^{(1-G_j)}\right)^{I(G_j=G_i)}\end{aligned}$$

The likelihood here differs from before in that only donor observations within the same compliance status as i contribute information when i is of interest. The likelihood terms for any observation not in the same compliance group as i provide no information for θ_i^{mis} and are dropped. The matching parameters are also estimated separately for each compliance group, as denoted by $\theta_{\mathcal{M}}^{(G)}$. The product over all i's denotes the complete set of ICEs for all observations in the data. Integrating out G in the likelihood involves piecing the likelihood together from the three principal strata. However, like before, the researcher can simply approximate the integrals using Bayesian simulation.

One can complicate the model further by estimating π_j using an ICE step in the first stage, matching[12] and imputing W^{mis}.[13] Let ω^{mis} denote the mean of the distribution W^{mis} drawn by modeling the donor pool in the first stage.[14] The full MCMC algorithm for the two-stage ICE model that I implement contains the following steps.

The parameter θ_i in the two-stage model is the individual intention-to-treat effect. The algorithm also outputs the draws from the distribution of compliance status G. Using the draws of θ_i and G_i, the researcher can calculate any causal effect of interest including the ITT, LATE, and NCATE (non-complier average treatment effect) and explore any treatment heterogeneity. I implement this algorithm on the job training data with predictive mean matching on the 15 covariates with $M = 20$ in both the

[11] Like before, assume the matching parameters are estimated separately and given.

[12] If the covariates are uninformative about compliance status, then the first stage ICE would simply be an approximation of the empirical estimate $\hat{\pi}_j$ from above.

[13] Note that W^{mis} is imputed with certainty for individuals in $S(1,1)$ and $S(1,0)$.

[14] The parameters ω^{mis} are the first stage equivalent of θ^{mis} in the one-stage ICE algorithm.

Two-Stage MCMC Algorithm[a] for the Posterior of τ_i

Repeat the following n_{sim} times:

1. Draw a matching procedure $\tilde{\mathcal{M}}_1$ where the subscript denotes the first stage matching.
2. Draw $\tilde{\theta}_{\mathcal{M}_1}$.

for (i in 1:N){

3. Determine $\tilde{D}_1^{(i)}$ from matching procedure.
4. Draw $\tilde{\omega}_i^{mis}$ to estimate ω_i^{mis}.
5. Draw $\tilde{W}_i^{mis}$ from Bern($\tilde{\omega}_i^{mis}$).
6. Calculate $\tilde{G}_i = Z_i W_i + (1 - Z_i)(W_i^{mis} - W_i)$
7. Draw a matching procedure $\tilde{\mathcal{M}}_2$.
8. Draw $\tilde{\theta}_{\mathcal{M}_2}^{(G)}$ separately for the two compliance groups.
9. Determine $\tilde{D}_2^{(i)}$ from second stage matching conditional on $\tilde{G}$.
10. Draw $\tilde{\theta}_i^{mis}$ to estimate θ_i^{mis}.
11. Draw $\tilde{Y}_i^{mis}$ from Bern($\tilde{\theta}_i^{mis}$).
12. Calculate $\tilde{\tau}_i = Z_i(Y_i - \tilde{Y}_i^{mis}) + (1 - Z_i)(\tilde{Y}_i^{mis} - Y_i)$.

}

[a]Steps 3-5 may be skipped for observations in $S(1,1)$ and $S(1,0)$ since their compliance status is known. The equation in step 6 accounts for this.

first and second stages of the algorithm with an MCMC of length 2000.[15]

3.2 Treatment Effects and Treatment Effect Heterogeneity from a Two-Stage Model

I first estimate three average treatment effects (ITT, LATE, NCATE) across the three survey timepoints of 52, 130, and 208 weeks after randomization. The dependent variable is whether the individual is employed at each timepoint. The average Job Corps participant stays in the training program for 1-2 years, so some of the participants in the program may still be enrolled at the first timepoint of 52 weeks. The ITT measures the average effect of treatment assignment on employment regardless of whether an individual enrolls in the program. The LATE measures the average effect of treatment assignment amongst compliers and the NCATE measures the average effect of treatment assignment amongst non-compliers. To calculate the LATE (NCATE), I take the draws of τ_i for each iteration of the algorithm and average the ones for individuals that were drawn as compliers (non-compliers) within that iteration. This vector consists of draws from the posterior for the LATE (NCATE). I then take the mean and the

[15]The donor pool size $M = 20$ is significantly lower than the 10% of smallest treatment arm number that I used before. Because the dataset is quite large, 10% of the smallest treatment arm would result in $M > 500$. My simulations thus far have not covered such a large dataset so I chose a much smaller number to allow for sufficient variation in the composition of the donor pools.

quantiles of the vector for the point estimate and credible interval.

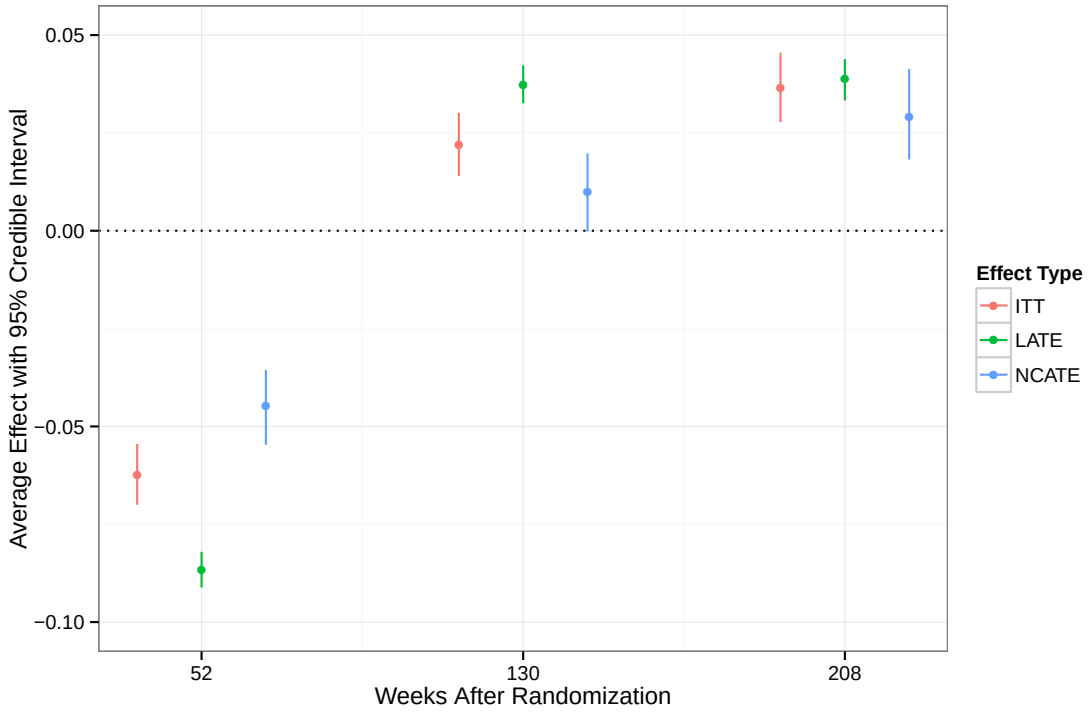

Figure 9: Three Average Treatment Effects at Three Timepoints

Figure 9 presents the results from the posteriors using the two-stage ICE algorithm. At 52 weeks, all the effects are negative, which indicates that job training actually decreases the probability of being employed at 52 weeks. At 130 and 208 weeks, the average effects become positive, suggesting that job training does actually increase employment prospects. There are a few things to note from these results. First, the fact that the effects are negative at 52 weeks is unsurprising. There are at least two possible explanations. The first is that participants in the Job Corps program are likely to still be enrolled in the program and thus have not had an opportunity to search for jobs. Their counterparts that did not enroll probably have higher employment rates since they have spent the 52 weeks looking for jobs. The second explanation is that even if participants have already finished the program, the resulting skills they have acquired lead them to search for higher income jobs, which may take longer to find. The idea is that participants now have a higher "reservation wage", the lowest wage at which they are willing to work. Because I only look at employment outcomes, it can be misleading since those that take job training may demand a higher wage whereas those that did not take job training may be willing to settle for lower-paying jobs. If one imagines the ease of finding a job is inversely related to the wage paid by the job, then lower-paying jobs are easier to obtain and individuals with a higher reservation wage are likely to be unemployed longer. I explore this idea further through exploring treatment effect heterogeneity.

A second finding to note is that the LATE is always stronger in magnitude than the NCATE. This is to be expected as the effect of treatment assignment should be much stronger for those that actually take the treatment than those that do not. However, with the exception of possibly the result in week 130, the NCATE effects, though weaker than LATE, do not seem to be zero. It appears that simply being assigned treatment does actually have an effect on individuals independently of actually enrolling in the job training program. From a methodological perspective, this seems to call into question the validity of the exclusion restriction, which requires that treatment assignment only affects the outcome through actually enrolling in the program. There may be a couple explanations for this. First, it may be the case that individuals who are assigned treatment are given a boost of confidence from simply being offered acceptance into the program. The offer itself may spur the individual to think about the future and to look harder for employment even without enrolling in the program. Second, it may also be the case that individuals who are offered a spot in the program may decide to decline the invitation in favor of another competing job training program or opportunity. Being offered the treatment may simply open their eyes to the opportunities available to them, and they may decide to pursue other opportunities that lead to employment. Nevertheless, a non-zero NCATE may indicate a violation of the exclusion restriction, which likely causes an upward bias in the estimate of the LATE.

The LATE results so far suggest that actual enrollment into the job training program decreases the probability of employment in the beginning and while still in the program, but has a positive effect on employment after completing the program. I now explore treatment effect heterogeneity further using the posterior of the ICEs. The first avenue I explore is whether the LATE is stronger for certain types of individuals characterized by the covariates. I take each of the nine binary covariates in the data and I estimate the LATE for $X = 1$ and $X = 0$.[16] I then take the difference in the LATE for $X = 1$ and $X = 0$. Figure 10 shows the results for the difference in LATEs between the two binary groups for four of the binary covariates.

The way to interpret the lines in Figure 10 is that a positive value on the y-axis indicates that the LATE for $X = 1$ is greater than the LATE for $X = 0$. For example, in the top left corner, at week 52, the LATE for individuals with children is about 5 percentage points greater than the LATE for individuals without children. This suggests that the effect of job training on employment at week 52 is greater on individuals with children. This result may conflate two mechanisms. First, it may be the case that individuals with children are more likely to finish the job training program sooner and thus be employed sooner due to the need for a steady job to raise children. Second, it may also be the case

[16]The process to calculate the LATEs here is similar to before. For each iteration of the MCMC, I identify those individuals drawn to be compliers with $X = 1$ and those draw to be compliers with $X = 0$. I repeat this process for the entire length of the MCMC to get draws from the posteriors for the LATEs for $X = 1$ and $X = 0$.

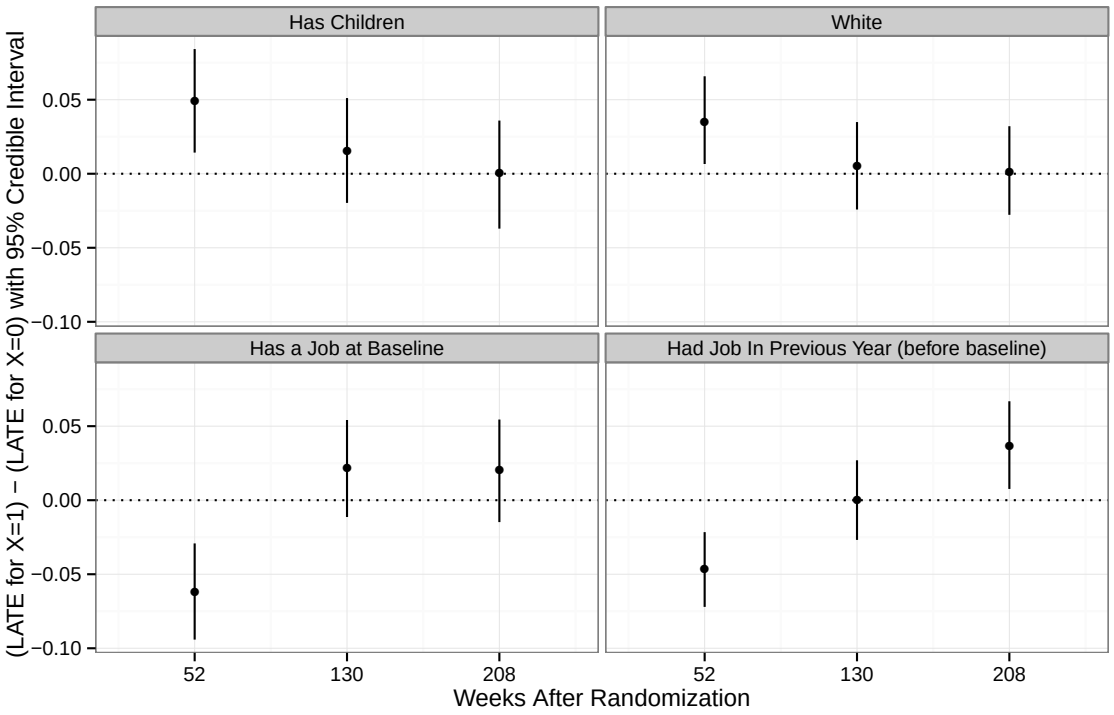

Figure 10: Difference in LATEs for Four Binary Covariates

that individuals with children have a lower reservation wage because they cannot afford to hold out for a higher-paying job and may settle for lower-paying jobs to support their children.

In the top right panel, it appears that the job training works slightly better for whites than for non-whites. There may be numerous explanations for this. Race may be correlated with a large number of other factors which may result in the appearance that whites finish the program faster and/or have an easier and faster time to employment after the program. I control for education and household income in the matching specification, but that may not account entirely for the heterogeneity in effects across races.

The two bottom panels look at the differences in LATE between individuals who were employed before the program and individuals who were not. Employment was measured as having a job when the baseline survey was taken (left) or having a job within the previous year before the baseline survey (right). The two variables are undoubtedly quite highly correlated. In both cases, it appears that having a job before the program significantly decreases the effect of the program on employment at week 52. It may be the case that those already holding jobs beforehand are opting into training for jobs that require more skills, so they must stay in the program longer than others. It may also be the case that their reservation wage after the program is much higher than individuals who had not had a job prior to baseline. Their baseline

jobs may have been of the lower-paying variety, so after the program, they expect an upgrade in their employment whereas those who had not previously held a job may opt to take lower-paying jobs after the program and become employed more quickly. Through evaluating treatment effect heterogeneity by aggregating ICEs, I find some heterogeneity that confirms the two theories of longer duration in the program and higher reservation wages leading to higher initial unemployment.

In Figure 10, I explored treatment effect heterogeneity by looking at LATE for for different groups of individuals based on covariates. The ICE framework also allows for exploring treatment effect heterogeneity in the reverse way by first dividing individuals into groups based on their ICEs and then comparing covariate information for the different groups. The two approaches are slightly different in that the first asks the question "What is the effect of job training for people that look a certain way (based on covariates)?" This second approach asks the question "What do people who benefited/were hurt from job training (in terms of employment) look like?"

In the study, the treatment effects can take on three possible values: 1, 0, and -1.[17] I first classify individuals into one of three effect categories: helped (1), no effect (0), or hurt (-1). I limit the analysis to compliers so that the effects are from the job training program itself. I also limit the analysis here to look only at employment in week 52. I then compare the mean value of the covariates for people in each effect category and see how they differ. To account for the uncertainty in the classifications and in compliance status, I repeat this process for each iteration of the MCMC. Specifically, for each iteration, $\tilde{\tau}_i$ classifies the individual i into an effect category. I then subset to compliers given the drawn compliance status $\tilde{G}$ and record the mean value of the covariates for each of the three effect categories. I repeat this process for all iterations and the result is a series of vectors of covariate means, one vector for each covariate-effect category. The vectors represent the posteriors of the covariate means.

Figure 11 displays the the covariate means by effect category for four of the covariates. Recall that "helped" implies a positive effect of job training on employment at week 52 and "hurt" implies a negative effect. Around 18 percent of those that were helped by job training had children compared to 15 percent for those who were hurt. Those that were hurt by job training also were more likely to have held jobs at baseline and more likely to have had higher earnings in the year before baseline. Finally, those that were helped by job training were also more likely to be white. These results are all consistent with the previous hypotheses that individuals with children are more likely to benefit more quickly from job

[17] Note that given treatment assignment and the outcomes in the data, the possible values each ICE can take is constrained to two values out of 1, 0, and -1. For example, an individual that was assigned treatment and is employed can only have an ICE of either 0 or 1 since treatment assignment could not have hurt employment given that the individual got treatment and is employed.

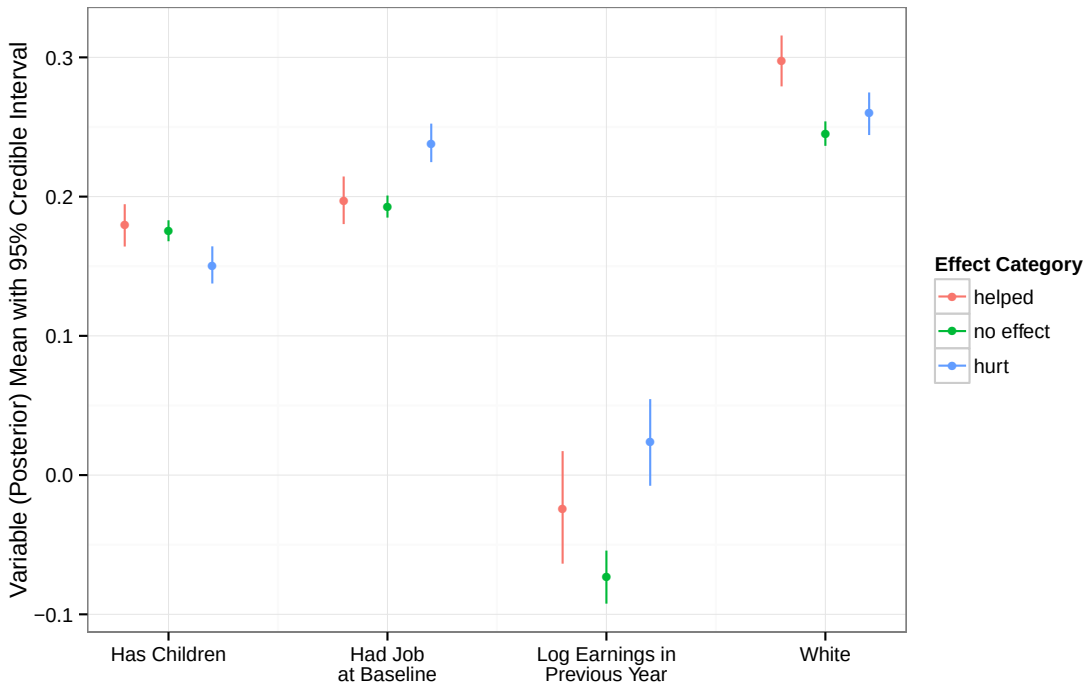

Figure 11: Comparing Covariates by Effect Category for Employment at Week 52

training in terms of finding employment while those with previous jobs are more likely to take longer to become employed, possibly due to staying longer in the program or holding a higher reservation wage.

The results presented here are largely consistent with the idea that the job training program actually works well in the long run in getting people employed. However, there is a short-term cost in terms of immediate employment. The ICE framework allows me to explore this result and find evidence that some individuals are more willing to bear this short-term cost whereas others are more likely to seek immediate employment after finishing the program. Further work can extend the ICE framework to address the issue of wages in conjunction with employment outcomes.

4 Conclusion

I have presented the results from two separate experimental studies related to monitoring corruption and job training. In both cases, the studies originally made a huge contribution to their respective fields. I use the ICE framework that I propose to mostly confirm those results, but I also demonstrate how the framework allows for a more flexible way to approach the problems. I show how to use ICEs to

explore treatment effect heterogeneity and I contribute some interesting results that were not addressed in the original studies. The two applications allowed me to show how to adapt the ICE framework to different types of outcome and treatment variables and to embed it within the framework of instrumental variables and two-stage analyses. The ICE framework can be extended even further to account for all types of data structures and patterns. In the final chapter, I address some further extensions using ICEs and discuss some other remaining issues for future work.

References

Frangakis, Constantine E. and Donald B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58(1):21–29.

Frumento, Paolo, Fabrizia Mealli, Barbara Pacini and Donald B. Rubin. 2012. "Evaluating the Effect of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data." *Journal of the American Statistical Association* 107(498):450–466.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2):200–249.