Contemporaneous and Post-Program Impacts of a Public Works Program : Evidence from Cote d'Ivoire.

Marianne Bertrand (University of Chicago, Booth School of Business),
Bruno Crépon (CREST), Alicia Marguerie (CREST - University Paris Saclay),
Patrick Premand (World Bank)

[**DRAFT - Please do not cite or circulate without permission from the authors**].
Latest version available here

Abstract

Public works are one of the most popular safety nets and employment policy instrument in the developing world, despite limited evidence on their effectiveness and optimal design features. This paper presents results on contemporaneous and post-program impacts from a public works intervention in Cote d'Ivoire. The program provided 7 months of temporary employment in roads rehabilitation for urban youths. Participants self-selected to apply for the public works jobs, which paid the formal minimum wage and were randomized among applicants. Randomized sub-sets of beneficiaries also received complementary training on basic entrepreneurship or job search skills. During the program, results show limited contemporaneous impacts of public works on the level of employment, but a shift in the composition of employment towards the public works wage jobs. A year after the end of the program, there are no lasting impacts on the level or composition of employment, although positive impacts are observed on earnings through higher productivity in non-agricultural self-employment. Large heterogeneity in impacts are found. Results from machine learning techniques suggest potential trade-offs between maximizing contemporaneous and post-program benefits. Traditional heterogeneity analysis shows that a range of practical targeting mechanisms come close to the machine learning benchmark, leading to stronger contemporaneous and post-program benefits without sharp trade-offs. Overall, departing from self-targeting based on the formal minimum wage leads to strong improvements in program cost-effectiveness.

Acknowledgements

The randomized control trial of the PEJEDEC project public works program was implemented in close collaboration with the government of Cote d'Ivoire (in particular BCPE, Bureau de Coordination des Programmes d'Emploi, and AGEROUTE) and the World Bank. We are particularly grateful to Hamoud Wedoud Abdel Kamil, project team leader at the World Bank; Adama Bamba, Hermann Toualy, Ismahel Abdoul Barry and Fabrice Konan at BCPE; as well as Marius Pokou, Martin Kouakou, Yves N'Cho and Kassi Ernest Bohoussou at AGEROUTE.

High-quality baseline and midline data were collected by ENSEA, under the leadership of Hugues Kouadio, Rosine Addy Mosso, Marie Judith Soro and Nathaniel Gbenro. High-quality endline data was collected by BCPE under the supervision of Ismahel Barry and Alicia Marguerie, with excellent support from Sondo Eloi Somtinda and Jean Awé. Sondoi Eloi Somtinda provided excellent field coordination during baseline, midline and endline data collection.

We thank Stefanie Brodmann, Deon Filmer, Dominique Van de Walle, seminar participants at CREST, ILO, IZA and the World Bank for useful comments and suggestions, as well as Azedine Ouerghi, Stefano Paternostro, Pierre Laporte and Jacques Morisset for strategic policy guidance during the impact evaluation activity. All errors and omissions are our own.

In addition from funding from the PEJEDEC project, the study was also supported by the MESF and DIME i2i Trust Funds at the World Bank.

A detailed impact evaluation report in French is available in: Bertrand, Marianne; Bruno Crépon; Alicia Marguerie et Patrick Premand, 2016. «Impacts à Court et Moyen Terme sur les Jeunes des Travaux à Haute Intensité de Main d'œuvre (THIMO) : Résultats de l'évaluation d'impact de la composante THIMO du Projet Emploi Jeunes et Développement de Compétence (PEJEDEC) en Côte d'Ivoire. » Washington DC : Banque Mondiale et Abidjan : BCP-Emploi.

1 Introduction

Public works programs are an important component of the policy portfolio of decision makers trying to address the social challenges of underemployment and poverty. Under such programs, the government offers temporary employment, typically remunerated at the minimum wage or below, for the creation of public goods, such as road or sewage infrastructure. Unlike welfare programs, such as cash transfers, public work programs transfer cash to their beneficiaries conditional on their meeting work requirements.

There are different types of public works programs. Some are employment guarantee schemes that offer participants a number of days of employment on demand each year and have as primary objective to provide social insurance for the poor. For example, the Mahatma Gandhi National Rural Employment Guarantee program (MGNREGA) in India that guarantees 100 days of work per year per household to all rural households with members willing to do unskilled manual labor at the statutory minimum wage.

Other public programs are implemented to address temporary shocks, such as those induced by an economic downturn, a climatic shock or period of violent conflict, and primarily aim to offer mass public employment as a stabilization instrument. They also include programs that provide temporary employment during the lean agricultural season, to address underemployment and seasonality in agriculture, as well as help households deal with the incidence of shocks and transient poverty. In Sub-Saharan Africa, our context in this paper, labor-intensive public works programs have often been adopted in response to transient negative shocks such as those induced by climatic shocks or episodes of violent conflicts. These programs typically offer temporary (a few months) one-off employment, at the minimum wage or below.

While traditional welfare programs, such as unconditional cash transfers, could also be used to support the most vulnerable, public work programs are often claimed to have a variety of advantages. First, while both “workfare” and welfare programs transfer cash to the poor, workfare programs also contribute to the creation of public assets (e.g. better roads) which may benefit the broader community. This argument is particularly relevant in contexts where physical infrastructure was destroyed or damaged as an outcome of the crisis the programs are aiming to address (e.g. climatic shocks or violent conflict).

Also, “workfare” programs, through skill development or the signaling value of prior work experience, may increase the future employability or productivity of the program beneficiaries.

This benefit can be potentially further improved upon, even if at a higher cost, by adding some complementary productive interventions, such as savings facilitation or training, to the work experience.

Another advantage of workfare programs, as highlighted by Besley and Coate (1992), is that they can solve the difficult problem of efficiently targeting who should be the beneficiaries for the transfers. Indeed, the appropriate targeting of social protections programs is particularly complex in lower income countries because of a lack of robust data, challenges to identify beneficiaries at the bottom of the welfare distribution, as well as weak systems and institutions, leading to potential errors of inclusion or exclusion. Public work programs are appealing as they may solve this problem through self-targeting, in that only the poor would be willing to supply labor in these programs at the stated (low) wage.

Also, and particularly relevant to the post-conflict environment, engaging beneficiaries in work-for-cash rather than simply handing out cash may operate as a social stabilization tool. This might operate through an incapacitation effect: time spent working may displace socially disruptive activities such as crime. Also, work, even if unpleasant, may improve well-being, self-esteem, and overall mental health, a set of non-cognitive assets that may also help beneficiaries become more productively employed.

Even though more mundane but of great practical relevance, another advantage of public work programs compared to traditional welfare programs is that they are often politically more acceptable and sustainable. Political preferences for workfare programs are often linked to (valid or not) concerns about welfare dependency (and how unconditional transfers may disincentive work) as well as a desire to generate immediate visible improvements to employment conditions.

Despite the popularity of public work programs such as those implemented throughout much of Africa, rigorous evidence on their overall effectiveness, as well as evidence of their optimal design features and particular channels through which they may benefit the most vulnerable is limited.

In this paper, we present the results of a randomized control trial designed to assess both the contemporaneous (i.e. during the program) and post-program (i.e. a year after program completion) impacts of public work programs. The particular public work program we evaluate was implemented by the Ivory Coast government in the aftermath of the post-electoral crisis that hit the country in 2010/2011 and was funded by an emergency loan from the World Bank.

The stated objective of the program was to improve access to temporary employment opportunities among lower-skilled young (18-30) men and women in urban or semi-urban areas that were currently unemployed or underemployed, as well as develop the skills of the program participants through this work experience and, for some participants, complementary training.

In particular, participants in the public works program were employed for a period of 7 months in road infrastructure and remunerated at the statutory minimum daily wage, corresponding to about $5 per day (FCFA 2500), approximately $110 per month (FCFA 55 000). Program participants were required to work 6 hours per day, 5 days per week. The road maintenance work, carried out in "brigades" of about 25 youths with a supervisor, was implemented by the National Roads Agency (AGEROUTE) and supervised by BCP-E[1], under the Ministry of Labor and Social Affairs.

All young men and women in the required age range and residing in one of 16 urban localities in Cote d'Ivoire were eligible to apply to the program. Because the number of applicants outstripped supply in each locality, fair access was based on a public lottery, setting the stage for a robust causal evaluation of the impacts of the program. In addition, a randomized sub-set of beneficiaries also received (i) basic entrepreneurship training to facilitate set-up of new household enterprises and entry into self-employment, or (ii) training in job search skills and sensitization on wage employment opportunities to facilitate access to wage jobs (e.g. help in identifying wage job opportunities, CV production, interview skills, etc.).

In addition to a baseline survey of program applicants, we carried rich surveys of youth in the treatments and control groups both during the program (4 or 5 months after the program had started) as well 12 to 15 months after program completion to capture any post-program effects of participation.

Our analysis of contemporaneous effects demonstrates that the program had limited impacts on the level of employment, mostly inducing shifts in the composition of employment. Eighty-six percent of the control group was working 4 to 5 months after the lottery took place, compared to 98 percent of those assigned to the treatments. Moreover, the program also did not substantially raise hours worked, with mean hours worked at about 41 hours in the control group compared to 44 hours for those assigned to the treatments. The value of the program for the modal applicant was therefore not as a way to escape unemployment but more as a way to escape under-employment in low-paying informal activities: monthly earning are about FCFA

[1] Coordination Office for Employment Programs (« Bureau de Coordination des Programmes Emploi »)

20,000 higher in the treatment groups, from a base of FCFA 60,000 in the control group. So, while program participation managed to lift earnings, foregone earnings are quantitatively important, with only about 40 percent of the transfer translating into earnings gain.

These results strongly suggest that self-targeting based on the formal minimum wage did not succeed in this context in getting only the most vulnerable (e.g. those without outside employment opportunities) to benefit from the program. A couple of factors likely explain this failure of self-targeting. First, a job that pays the statutory minimum wage could still be of appeal to many in an environment where informal employment and self-employment are rampant. Second, because the work was only 6 hours per day, many applicants with outside employment opportunities, especially those that allow for more flexible hours, would still see value in applying for the public works program as they could combine it with other activities. Finally, while the unpleasant nature of the work may have discouraged some, it is unclear whether this work is more unpleasant than most informal activities.

Twelve to 15 months after program completion, no impacts are observed either on the level (employment or hours worked) or on composition of employment (salaried work vs. self-employment). However, we do observe sustained positive impacts on earnings (FCFA 5,622 or $11 per month, a 11.6 percent increase compared to the control group), mainly stemming from non-agricultural self-employment activities. These sustained impacts are mostly driven by youths who participated in the public works and complementary basic entrepreneurship training.

Based on these estimates of direct earnings impacts on youths, and given rich data we have collected on program costs, we conclude the public works program under its current form is far from cost-effective, with cost per participants being about 3 times the estimated benefit.

While our results suggest that self-targeting based on the formal minimum wage failed in this context, it is possible that stricter eligibility criteria may have resulted in improved cost-effectiveness. This is the issue we address in the remaining sections of the paper, where we study heterogeneity of program effects, both during the program and post-program.

First, using recent machine learning techniques (Athey and Imbens 2015, Wager and Athey 2015), and relying on the very rich data collected at baseline about each program applicant, we estimate the heterogeneity of program effects both during and after the program. This analysis confirms large differences in predicted impacts across various groups of program participants, both during and after the program. In particular, the average predicted impact in the short-term in the upper quartile of the predicted impact distribution is over FCFA 35,000, compared to -4,800 in the lower quartile of this distribution. Also, the average predicted impact in the longer-term in the upper quartile of the predicted impact distribution is over FCFA 13,000, compared to -1,500 in the lower quartile of this distribution.

While these results suggest that the program effectiveness, both in the short-term and the long-term could be improved through better targeting, they also highlight potential trade-offs between maximizing contemporaneous and post-program benefits. In other words, program participants that benefit most during the program are not systematically those that benefit most after the program has ended. However, a study of the distribution of predicted contemporaneous and post-program impacts in the short-term and long-term does reveal the existence of some common targeting dimensions that may improve effectiveness without sharp trade-offs.

Based on this heterogeneity of causal effects in our data, we can then assess how stricter eligibility rules may have changed the cost-effectiveness of the public works program and how far from maximum predicted effects these rules might have taken the government. Compared to the benchmark scenario with self-targeting based on the formal minimum wage, the cost-effectiveness ratio would improve from 3.2 to between 1.58 and 1.98 based on finer program targeting such as selecting youths with low predicted baseline earnings, self-selection based on a lower offered wage, targeting women only, or targeting based on self-declared baseline earnings. While the analysis cannot decisively indicate which targeting scenario would maximize cost-effectiveness given the confidence intervals around the impact estimates, it does highlight strong improvements in cost-effectiveness when departing from self-targeting based on the formal minimum wage.

2 Framework

The primary objective of workfare programs is to provide income support through temporary jobs. Beyond this short-term goal contemporaneous to the program, another objective is sometimes to facilitate transitions to more productive, higher-earnings occupations after the program. Several potential channels can contribute to such longer-term objectives, for instance the ability to save and invest in ongoing or new activities, or the opportunity to develop skills valued in the labor market.

2.1 Contemporaneous Impacts in the Short Term

Targeting is a key design feature of social safety nets programs seeking to provide income-support in the short term. It is a general issue that each transfer program faces. The specificity of workfare programs is that participants typically self-select into the program. This is a major difference compared to transfer programs, which often select participants based on a screening procedure such as a proxy-means test or a participatory community targeting approach. The mechanism through which workfare programs address the selection problem is to ask for hours of work in exchange for the transfer. This mechanism has several well-known implications. First is the self-selection mechanism: only potential participants for whom the utility level, accounting for earnings and disutility of work, in the program is larger than the utility level should participate. The second important implication is that there are forgone earnings when participating in the program: the time participants spend in the program cannot be used for work on regular activities. Thus the contemporaneous program impact on earnings is the difference between transfers to the participants and forgone earnings. Another consequence is that the impact of participation on income in the short term is heterogeneous and depends on participants' alternative economic opportunities, as well as the extent to which they are able to keep operating these activities while in the program. A reasonable expectation is that the impact is almost zero for the 'marginal' participant. One the other hand, the impact should reach the amount of the transfer for those whose income without the program would have been zero. In this context, the average contemporaneous program impact on self-selected participants depends on the distribution of individual impacts over the population.

The following simple model helps to formalize these ideas. Assume for example that earnings for a program participant are $W_p = w_p h_p$. This is a lump sum. There are fixed hours to work in the program, with no partial participation possible. Let $W_1(h)$ capture participants' earnings for h hours of work on other activities while in the program and let h_1 be the number of hours participants can spend on these activities while in the program. Let $W_0(h)$ capture earnings for h hours of work absent the program and h_0 be the number of hours participants would have spent on these activities absent the program. The impact of the program on earnings in the short term is therefore

$$Impact(Income) = W_p + E_{Part}\ \left(W_1(h_1) - W_0(h_0)\right) = W_p - E_{Part}\ \left(W_0(h_0) - W_1(h_1)\right)$$

where E_{Part} means expectation on participants. It is thus different from the direct transfer received from the program by the amount of forgone earnings: $W_0(h_0) - W_1(h_1)$.
Similarly the impact on the number of hours of work is different from hours spent in the program

$$Impact(hours) = h_p + E_{Part}(h_1 - h_0) = h_p - E_{Part}(h_0 - h_1)$$

A standard ratio to assess the capacity of the program to increase earnings is the ratio of contemporaneous program impact on income to the average income of participants in the control group during the program.

$$\frac{W_p + E_{Part}\ \left(W_1(h_1) - W_0(h_0)\right)}{E_{Part}(W_0(h_0))}$$

One interesting parameter to add is the ratio of the contemporaneous program impact on income to the realized transfer :

$$\frac{W_p + E_{Part}\left(W_1(h_1) - W_0(h_0)\right)}{W_p}$$

In this context, one of the key empirical question for a given program is the extent to which hours of work and earnings change due to participation in the program. What are the changes in the portfolio of activities of program participants? Do we see a substantial change in hours worked? Do we see changes in income during the program?

2.2 Heterogeneity in contemporaneous program impacts

Assuming for a moment that only earnings during the program matter for participation, the basic idea of workfare as a selection device is that participants will self-select into the program. Hours of work are determined by participants seeking to maximize their utility. Assuming earning functions have the same form $W_1(h) = W_0(h) = w * h$ and that there is an increasing convex disutility of effort $c(h)$, there is an optimal number of hours of work $h_0(w)$, leading to a maximum utility level absent the program $V_0(w) = w * h_0(w) - c\big(h_0(w)\big)$. In such a simple setting, there is a threshold in earnings $\underline{w}$ (that is supposedly lower than w_p) such that for $w < \underline{w}, h_0(w) < h_p$. It can be shown that in such a case $h_1(w) = 0$, and that the utility level increases: for potential participants with very few outside employment opportunities, program participation would lead to an increase in the number of hours, in earnings and in the utility level. For example, if we consider people who have no employment opportunity: $w = 0$, hours of work and earnings absent the program would be zero. In such a case, the impact of program participation on hours and earnings would be an increase by respectively h_p and $w_p * h_p$. For individuals with intermediate opportunities, $\overline{w} < w < w_p$, the total number of hours does not change: $h_1(w) = h_0(w) - h_p$ but the utility level increases $V_1(w) = V_0(w) + \big(w_p - w\big)h_p$. Last, for $w > w_p$ the program does not improve the utility level and therefore the individuals would not participate.

This simple setting could be adapted to account for several important practical aspects of public works program. For example, there could be a fixed cost of participation due to the cost of accommodation, or specific disutility for the type of work required by the program, or lost earnings due to the number of hours of transportation,... Notice that, as is well-known, there might be cases when income effects imply that program participation would actually reduce the numbers of hours worked. A general result remains: contemporaneous program impacts on income and utility are heterogeneous in the population. There are 'marginal' participants for whom employment opportunities are large enough for them to be indifferent between participating or not. On the other hand, there are other participants with very few opportunities for whom contemporaneous program impacts are expected to be very large.

It is therefore critical to understanding the heterogeneity of impacts in public works program. We address this issue in this paper by assessing whether there is evidence of heterogeneous program impacts on employment, hours worked and income during the program?

Notice that assuming a disutility of work $c_i(h) = c_i h^2/2$ for individual i, it is possible to show that the intervention is rank preserving for hours worked. This is also the case for earnings if we assume that w_i is the only source of heterogeneity, and thus that disutility of work is homogeneous ($c_i = c$). However, if there is also heterogeneity in c_i, there is no reason for the rank of individuals in the earning distribution to be the same with and without the program. This is important as rank preservation is a property that helps in the identification of the variance in the program impacts.

2.3 Targeting

Targeting is a question related to the issue of heterogeneity. We can consider that a social planner has as an objective to maximize the average utility or income of a target population. Are there different assignment mechanisms that would improve the average impact on utility or earnings? A first idea is to change the characteristics of the workfare contract. Let $I_0(w)$ stands for $W_0(h_0(w))$ and $I_1(w) = h_p * w_p + W_1\big(h_1(w)\big) = h_p * w_p + \big(h_0(w) - h_p\big) * w * 1_{w>\overline{w}}$. Assuming the program is made available to everybody in the population, the average impact on income becomes

$$S(w_p) = \int_{w<w_p} (I_1(w) - I_0(w))f(w)dw$$

$$= \int_{w<w_p} [h_p * w_p + \big(h_0(w) - h_p\big) * w * 1_{w>\overline{w}} - w * h_0(w)]f(w)dw$$

Given the apparent waste due to participants for whom the marginal impact is zero, a natural idea would be to reduce the contract wage. However, this would be ineffective since for the marginal applicant the program effect would remain zero. A marginal reduction in the wage w_p would actually reduce the impact on average income, corresponding to the related reduction in income for each participant (there is no contribution of the newly selected out participants for whom impact of the program is zero):

$$S'(w_p) = \ h_p * F(w_p).$$

However the program is generally not available to everybody in the population. Assume that there is oversubscription. One selection mechanism is to randomly assign applicants to the program so as to meet a budget requirement. The surplus in such a case would write instead

$$S_r(w_p) = \lambda\, S(w_p)$$

with λ such that $\lambda\, F(w_p) w_p * h_p = M$, and M the budget available for the program.

In such a case, reducing the program wage and increasing the share of applicants to meet the budget requirement would have an ambiguous effect on the average impact on earnings. Straightforward computations show the change in the surplus due to a marginal change in w_p would be:

$$S_r'(w_p) = \lambda(\ (h_p * F(w_p) - S(w_p) \frac{w_p f(w_p) + F(w_p)}{w_p * F(w_p)})$$

A reduction in the wage w_p would not lead to a reduction as strong as in the former case, but the effect remains ambiguous.

Another potential assignment mechanism would be to randomly assign individuals to the program with different probabilities depending on some characteristic x, for instance proxies for their productivity or outside opportunities. The assignment probability would be a function $\lambda(w, x)$. Assume that the objective is to maximize the program impact on income:

$$\mathrm{S}(w_p, \lambda) = \int_{(w<w_p)} (I_1(w) - I_0(w)) f(w) \lambda(w, x) dw dx.$$

In this case, the obvious choice for the function $\lambda(w, x)$ is to assign individuals with the highest potential impact on income. It is clear from the computation above that the individual gain

$I_1(w) - I_0(w)$ is decreasing in w. For the same budget constraint, the best allocation would be to assign participants with the least opportunities outside the program. Let $\widetilde{w}$ be the maximum productivity level consistent with full assignment and the budget constraint $F(\widetilde{w})w_p h_p = M$. The assignment mechanism achieving the largest impact would be:

$$\lambda(w, x) = 1_{(w<\widetilde{w})}.$$

It would allow to reach the largest contemporaneous program impacts:

$$\int_{(w<\widetilde{w})} (I_1(w) - I_0(w)) f(w) dw$$

This clearly shows that combining self-selection and screening of applicants is a better way to improve the program compared to simply changing the contract features.

However, when defining an assignment mechanism, it is important to keep in mind that the mechanism is known to potential participants, so that they can make participation decisions. Participation in the program requires that the expected gain from participation is positive. If we consider that application costs an amount $\chi(w, x)$ (for example it takes some time to apply, to register and to participate in the selection process and there are some related forgone earnings), the participation rule is:

$$\lambda(w,x)V_1(w) + \big(1 - \lambda(w,x)\big)V_0(w) - \chi(w,x) > V_0(w) \leftrightarrow V_1(w) - V_0(w) > \chi(w,x)/\lambda(w,x)$$

Changing the assignment rule is likely to change the characteristics of the marginal applicant, unless there is no cost of application[2].

[2] Program advertising frequently only provide the number of slots available. For example, in the case of the public works program in Cote d'Ivoire we discuss in the rest of the paper, the number of slots available in each locality for men and women was known in advance. The program also initially introduced a fixed ratio of slots for men and women. Changing the total number of slots available or the ratio of slots reserved to women is likely to change expectation about λ for men and women. In such a case, the decision to participate might be affected, if application is costly.

2.4 Other aspects of the analysis of contemporaneous program impacts

Another important aspect of the analysis of contemporaneous program impacts relates to use of the transferred income. We would expect an increase in consumption caused by program participation and higher earnings[3]. The allocation of income between consumption and savings also affect post-program impacts. Indeed, the additional income can be used to save or finance investments like training or capital equipment for income-generating activities. Youths may encounter constraints to build up savings, and positive income shocks during the program may be associated with savings accumulation. Savings can have several potential post-program benefits, including precautionary savings to absorb future shocks, or savings to finance investments.

Finally, externalities are another important dimension during the program. Public works programs are likely to have an impact on participants themselves but also on people around them. While externalities are challenging to address empirically, it is nevertheless possible to examine the impact of participation of one member of the household on other members of the household. First there might be an increase in the contribution of participants to expenses at the household level. Second, the participation of one member of the household in the program might have an impact on activities of other members of the household. On the one hand, traditional income effects would imply a reduction in other members' activity. On the other hand, underemployed members of the family could take on part of the forgone activities of the participant.

2.5 Post-program Impacts

A first-order question in the public works literature is about post-program impacts in the medium to long-term. Many public works programs also have the objective to facilitate youth' transition towards more productive occupations after the program. There is little evidence in the literature on such long-term effects, although there are several potential channels. First is

[3] Another question relates to the type of goods for which an increase in consumption should be expected. We could expect that program participation first increases the consumption of goods needed to meet basic needs. However, a common question when it comes to providing cash to young people is whether there are any increase in consumption of temptation goods such as alcoholic beverage, drug or gambling.

the idea of return to capital. Several experiments have proved that returns to capital can be very large for poor households (see Blattman and Ralston, 2015). Common instruments to make capital available to youth have revealed poor performance, for instance micro credit. Transfers made available through public works programs are a way to help alleviate capital constraints for participants. A related empirical issue is whether participation in public works affects savings[4].

However, there are also other possible mechanisms for workfare interventions to have impacts in the long run. Usually, subsidized jobs are seen as a way to improve experience and productivity of participants so that their ability to find a wage job increases. This employment channel is another possibility for the program to have an impact in the long term. Last we can think about behavioral aspects related to program participation. It might be the case that youth do not perceive financial constraints but that they have biased time preferences leading them to undervalue their situation in the future. It might be possible that a program requiring youths to form work habits, like waking up each morning to go to work, may induce lasting behavioral changes.

Lastly, note that it can also be the case that public works programs have negative long-term impacts on participants. This is actually a scenario that has been frequently pointed as the more likely. One channel through which such negative long term impacts may relates to the potential 'stigmatization' of participants, i.e. program participation sending wrong signals to the labor market. Another idea is that the experience provided to participant through the program is of little value or only enhances skills which are not demanded in the regular labor market. Participants may also directly forgo some activities, which may create a form of destruction of capital through program participation. Potential participants might struggle in day to day occupation requiring a lot of search and connection. They could be tempted by the 'easy' way to obtain earnings through a temporary workfare job. However, doing so can induce the loss of capital of connections which might take time to rebuild once the program is terminated.

[4] Notice, also, that it is possible that the program has a temporary impact in the medium. It would be the case if participants have been able to save part of the additional income due to the program but have been unable to use it to start or expand income generating activities. Savings in such a case is mainly used to cope with future income or expense shocks.

3 Intervention and Data

3.1 The PEJEDEC public works intervention

Public works programs were introduced in Côte d'Ivoire in 2008 by a post-conflict assistance project following the 2003-2007 armed conflict.[5] Public works were later included as a component of an Emergency Youth Employment and Skills Development Project[6] (*PEJEDEC*) set-up after the 2010/2011 post-electoral crisis. The PEJEDEC public works program was managed by the Ministry of Social Affairs and Employment, through BCP-E[7], and implemented by the national roads management agency (AGEROUTE). A range of other institutions have been implementing public works interventions with similar features in Cote d'Ivoire[8].

The PEJEDEC public works intervention aims to improve access to temporary employment in road maintenance for low-skilled youths in urban areas. The program targets youths aged 18-30 in 16 localities throughout the country. A 30% quota was initially introduced for women. Participants are offered temporary employment for 6 hours per day and 5 days a week for a total of six months[9]. Participants work in teams of 25 individuals (called "*brigades*"), under the supervision of a team leader and a local supervisor. They perform road maintenance activities such as sweeping streets or cleaning ditches. The jobs are paid FCFA 2,500 (approximately $5) per work day, a wage equal to the legal daily minimum wage in the formal sector. Wages are paid monthly on bank accounts that are set-up for all participants as they start working.

In addition to participating in the public works program, youths are offered various training activities. First, all participants receive a one-week basic life skills training covering issues related to HIV-AIDS, citizenship and hygiene. Second, some participants are offered a complementary *basic entrepreneurship training* to facilitate transition into more productive self-employment upon exit from the program. Third, other participants are offered a *training*

---

[5] *Projet d'Assistance Post Conflit* (PAPC) was implemented by the government of Cote d'Ivoire and supported by the World Bank. It was implemented between 2008 and 2014.

[6] *Projet Emploi Jeune et Développement des Compétences* (PEJEDEC) has been implemented by the government of Cote d'Ivoire (through BCP-E) and supported by the World Bank. It also included intervention for other target groups, including internships, apprenticeships, professional training and entrepreneurship.

[7] Coordination Office for Employment Programs (« Bureau de Coordination des Programmes Emploi »)

[8] Among others, this includes the public works programs implemented by the government of Cote d'Ivoire as part of the C2D project with support from AFD, or as part of a program supported by the African Development Bank.

[9] As explained further below, the wave of the evaluation participated in the program for 7 months, but the standard program is 6 months.

on wage jobs search skills and sensitization to wage jobs opportunities, with the objective to facilitate transition into wage jobs upon exist from the program.

The curricula for the complementary skills training are tailored for low-skill population that may not be able to read and write, in particular by relying on drawings and visuals. Each training lasts approximately 80-100 hours distributed over two two-week periods. They are accompanied with field exercises to be undertaken between the training periods, in parallel to the public works jobs (typically in the afternoons). The trainings are delivered by work brigades, i.e. in groups of 25 youths. Participants do not have to work during the trainings, but still receive their corresponding daily wage.

The *basic entrepreneurship training* aims to build skills to help youth set-up and manage a small non-agricultural micro-enterprise. The training lasts 100 hours and focuses on providing cross-cutting business skills and practical guidance to develop simple business plans for small-scale activities that can be set-up using savings from the public works program. A first phase (40 hours over two weeks) reviews themes related to basic entrepreneurship and business skills. A second phase includes field research for youths to gather information, undertake basic market research and sketch a business plan. A third phase (40 hours over two weeks) includes feedback on youths' basic business plans, and reviews of key related issues from the curriculum. The final phase (20 hours) is an individual post-training follow-up.

The *training on wage jobs search skills and sensitization to wage jobs opportunities* provides information on wage jobs opportunities, skills on jobs search techniques, as well as a more professional environment during the public works programs and skills certification to facilitate signaling upon exit from the program. The training itself lasts 80 hours. The first phase (40 hours over two weeks) reviews how to identify wage jobs opportunities (either locally or through migration), how to search for wage jobs, prepare a CV, apply for a job and participate in a job interview. The second phase includes field exercises to gather information on potential opportunities, identify and visit potential employers or professional networks, etc. The third phase (40 hours over two weeks) provides feedback on field exercises, reviews part of the curriculum and provides additional practical guidance. In addition, supervisors of the brigades offered the wage employment training are trained on how to manage teams and provide feedback to workers, with the objective to mimic the professional experience would have in a more formal wage job. Youths are periodically rated on a range of skills, and these evaluations are later used to issue a work certificate that signals between one and five competencies identified as strength for each participant.

3.2 Experimental design: Enrollment and Randomization

The PEJEDEC public works program was implemented in 16 urban localities throughout Cote d'Ivoire, including Abidjan and cities in the interior[10]. 4 waves were organized between 2012 and 2015, each covering all 16 localities, with a similar number of pre-determined places available for each locality in each wave. In total, 12,666 youths participated in the program. The randomized control trial focuses on the second wave of the program, which took place between July 2013 and February 2014[11]. The identification strategy relies on a two-step randomization process.

The first randomization step involves individual randomization into the program. Before the start of the second wave, and as was the case for the other waves, an intense communication campaign was organized through local newspapers, local radios and public notice boards to invite interested youth to visit a registration office and apply to the program. Enrollment was open for two to three weeks in each locality, between June 2013 and July 2013. Only two eligibility criteria were applied during enrollment: applicants had to be between 18 and 30 years old, and could not have participated to a public works program before.

Once the enrollment period had closed, public lotteries were organized in each locality (separately for men and women, hence stratified by locality and gender) to randomly select beneficiaries among the registered applicants present at the lottery. Remarkably, the public lotteries were put in place at the time of the post-conflict assistance project. Since then, they have been used continuously as a transparent assignment mechanism to allocate limited public works jobs in a way that would be socially acceptable and limit potential tensions. As such, the first step of the randomization protocol was already implemented by the program in its routine operations.

In practice, during the enrollment phase for the second wave of the program, 12,188 individuals applied, of which 10,966 participated in public lotteries where 3,125 beneficiaries were selected

[10] 4 municipalities were covered in Abidjan (Abobo, Yopougon, Koumassi, Marcory) and 12 cities throughout the country (Yamoussoukro, Bouaké, San Pedro, Daloa, Korhogo, Abengourou, Man, Bondoukou, Gagnoa, Séguéla, Daoukro, Dimbokro).

[11] Less than 5% of youths assigned to the public works program did not participate, or participated for less than 3 months. The second wave was extended from six months to seven months for all participants to ensure that the complementary training could be completed during that time for those who were assigned to them (see below).

and assigned to 125 brigades of 25 individuals each (17 men, 8 women)[12]. For the study wave, a waitlist was created so as to protect a control group, although in practice replacements were minimal[13].

The second randomization step involves the randomization of public works brigades into groups receiving different types of complementary training. Specifically, brigades were randomized into three groups: (i) 45 brigades (1,225 individuals) assigned to receive the public works only; (ii) 40 brigades (1,000 individuals) assigned to receive the public works plus the complementary basic entrepreneurship training, and (iii) 40 brigades (1,000 individuals) assigned to receive the public works plus the wage jobs search skills training[14]. This second randomization was stratified by locality, and performed through a lottery held in the project office with implementing partners and a notary public in November 2013. Results remained confidential until two weeks before the start of the trainings.

3.3 Timeline and Data

3.3.1 Timeline and Surveys

The randomized control trial focuses on the second wave of the public works program. Enrollment in the program was open for around three weeks between mid-May and mid-June 2013, following an intense communication campaign led by the implementing agency (AGEROUTE). The public lotteries were held in each locality between the end of June and early July 2013.

A baseline survey was conducted shortly after the public lotteries and before program implementation (between the end of June and mid-July 2013). The study sample comprised all the individuals selected to participate in the program after the first randomization (3,125 individuals), as well as a control group obtained from a (random) sample of 1,035 individuals

[12] Beneficiaries were assigned to brigades within localities based on the number they drew in the public lottery.
[13] Replacement of drop-outs was allowed during the first two-months of the program. Replacements were only possible based on the waiting list, and would have stopped when the waiting list was exhausted. After two months, replacements were not allowed anymore.
[14] All brigades receive a one-week basic life skills training covering issues related to HIV-AIDS, citizenship and hygiene. This training is considered part of the basic public works program.

drawn among the non-beneficiaries not on the waiting list[16]. The data collected included measures of employment and earnings. It also captured a range of other characteristics such as preferences for risk and present, non-cognitive skills, as well as the results of practical tests measuring cognitive, manual and numeracy skills. Attrition at baseline was very small (1.5%).

The public works activities started between early and late July 2013, depending on the locality. In August, participants received the one-week basic life skills training that is considered part of the basic public works program. The second randomization took place in October 2013 in project offices. Brigades' assignment to the various complementary training modalities were not made public until January 2014, in order to limit potential response bias during the midline survey.

In addition to the baseline survey, two surveys were conducted in order to study contemporaneous impacts *during* the program and long-term *post*-program impacts.

To estimate contemporaneous program impacts, a midline survey was conducted on 3,036 individuals (2,001 beneficiaries[17] and the control group) between the end of November 2013 end early January 2014, i.e. 4 to 5 months after the start of the program. Both individuals and households head were interviewed. A two-weeks (random) tracking was implemented in February 2014 to limit attrition, mainly due to the mobility of control individuals[18]. Attrition at midline is limited (2.6%) and balanced across treatment and control groups. The midline questionnaire included very detailed sections on employment (up to three activities), specific information on characteristics of independent activities, a time use module and measures of behavior and well-being.

[16] Individuals randomly assigned to the waiting list are excluded from the sampling frame to prevent contamination of the control group. Sampling for the control was stratified by gender and locality (similar to the randomization procedure).

[17]The 2,001 treated individuals are a sub-sample of the 3,125 beneficiaries stratified by locality, brigade and gender. This sub-sample voluntarily excludes brigades which had been allocated to the wage-employment training. Indeed, their supervisors were following a specific management training at the time of the survey, and we wanted to avoid potential anticipation effects or any behavioral changes that could potentially affect outcomes.

[18] The tracking helped reduce attrition rate from 5.4% (after main data collection) to 2.6%. Before tracking, a small attrition differential was observed between treatment and control groups due to larger mobility out of program localities in the control group. The survey firm had not planned tracking outside the localities. The sample for tracking was randomly selected among the treatment and control groups (stratified by locality and gender) among non-respondents who were alive, not outside Cote d'Ivoire, and excluding individuals that could not be reached since baseline. After tracking, remaining attritors were mainly people impossible to contact or highly mobile individuals, and attrition was balanced between treatment and control group.

The public works program was originally expected to end in January 2014. However, as the complementary trainings were starting in January, participants were given a one-month extension on their contracts, which exceptionally extended the public works duration from 6 to 7 months. This ensured that all 'brigades' selected to participate in one of the trainings could do so while being paid by the program (at the same wage) for the first half of the training, which reduced the opportunity cost of time during the trainings. Complementary trainings were organized between January and mid-March and the second wave of the program ended between early and mid-February 2014 (depending on the locality). Some beneficiaries attended the second half of complementary skills trainings after the end of their contracts, and were given a transport allowance[19].

To evaluate post-program impacts, an endline survey was conducted between March and July 2015, i.e. between 12 to 15 months after the end of the program. The sample included 4,360 individuals. It was comprised of the whole baseline sample of 4,160 individuals in the treatment and control groups, plus 200 individuals randomly selected to be added to the control group[20]. Again, both individuals and household heads were interviewed. A one-and-a-half month (random) tracking phase took place in September 2015. The final attrition rate was 6.2%, and was balanced between treatment and control groups. The endline questionnaire was based on the midline survey and enriched with 'historic' information on job search, independent activities (including past projects) and an employment calendar.

3.3.2 Key outcomes and descriptive statistics

Descriptive statistics

As in many countries in Sub-Saharan Africa, Cote d'Ivoire faces a relatively low unemployment rate, but also a small share of individuals working in wage jobs. A large part of the population is concentrated in informal occupations, mainly in agricultural and non-agricultural self-

[19] Half of the brigades assigned to complementary skills trainings (50% of each type of training) had 25% of their training hours after the end of the public works contract, and received a transportation allowance of FCFA 1500 (the wage was at FCFA 2500). 25% of the brigades had 25% of their training hours (i.e. the second phase of the training) after the end of the public works contract and received the same transportation allowance. The remaining 25% were fully under contract during their trainings. Transportation allowance was paid ex-post in one transfer, based on the number of days of real attendance.

[20] The reinforcement of the control group is explained in section 3.4.1.

employment (Filmer et al, 2014; World Bank, 2015). In addition, most of wage employment takes place in casual and informal jobs without contracts. Overall, many earn less than the legal minimum wage, as the law is only binding for formal private companies and public administration (INS and AGEPE, 2014). Inactivity and unemployment tend to be more frequent for individuals in households in the top of the income distribution, especially those holding a higher education degree. This reflects the fact that the poor and vulnerable often cannot afford not to work. Moreover, strong gender disparities are observed. Women are more likely to be inactive and unemployed compared to men. They are also more likely to be self-employed rather than in wage jobs.

Public works applicants are on average 25 years-old, mostly living in urban areas (93%). They live in households with an average of 6 individuals (with 4 adults). 25% of applicants are head of the household and most of applicants have no more than one child (50% of applicants have no children, 25% have one child). Although three quarters of them attended (at least partially) primary school, around half of the applicants (47%) have no degree[22]. This reflects the fact the program was designed to attract low-skilled youths, although 11% of the applicants have completed secondary school. Less than half of the applicants have attended some form of vocational training, mostly informal apprenticeship. Most of applicants were already working before program implementation (80%), in line with the national (and regional) employment situation marked more by underemployment in low-earning occupations rather than unemployment. Also, although most of applicants declare searching for a wage jobs, most of them aspire to be self-employed in the future. Finally, the data also points to widespread vulnerability, as only half of the applicants have saved money over the last three months and nearly 75% of them report facing constraints for basic needs expenditures.

We compare our evaluation sample to a national sample of 18 to 30 years old individuals living in urban areas[23] to provide insights on public works participants' profile (Table 2). Overall, the program attracts a lower share of inactive and unemployed compared to all youth aged 18-30 in Cote d'Ivoire. Among the employed population, public works applicants are more likely to be in wage employment (for their main occupation) compared to self-employment. The

[22] At the end of primary school students pass a certifying exam (CEPE).

[23] We use the 2013 national employment survey (ENSETE) which data was collected in February 2014 and compare it to our closest dataset, the control group of the midline survey which occurred from November 2013 to January 2014.

distribution by education level is quite similar among the general population and public works applicants. This means that individuals with a relatively high level of education were attracted by the program, even though it had been originally been conceived for low-skilled youth.

Key outcomes

This section describes the main outcomes measured both at midline and endline surveys.

Total monthly earnings are expressed in CFA francs. They are aggregated over up to three (parallel) activities undertook by an individual in the 30 days preceding the survey. They includes payments received in cash and the monetary equivalent for in-kind payments. The variable is winsorized at 99%. Total monthly earnings are decomposed in total (monthly) earnings from wage employment and self-employment (as well as earnings from other occupations, which are not displayed separately).

Has an Activity is a dummy taking a value of 1 if the individual has worked at least one hour over the 7 days preceding the survey, consistent with the employment indicators used at the national level. It takes a value of 0 for inactive and unemployed individuals.

Weekly hours worked capture the total number of hours worked per week is aggregated from up to three (parallel) activities undertook by an individual across all occupations (formal or not, casual or not, wage job or other type of activity). The variable is winsorized at 99%. Weekly hours worked are decomposed in hours worked in wage employment and self-employment (as well as hours worked in other occupations, which are not displayed separately).

Savings stock is the total amount of savings in CFA francs at the time of the survey. It aggregates savings from formal or informal mechanisms. The variable is winsorized at 99%.

A well-being index constitutes an aggregate from 6 measures: measures of happiness and pride in daily activities taken from the time-use module, a self-esteem scale (Rosenberg[24] scale), a 'positive affect' sub-scale (from the CESD[25]), a sub-scale of (positive) attitude towards the

[24] Rosenberg Self Esteem scale is a 10 item instrument to measure self-esteem. We use the validated French version.

[25] Center for Epidemiologic Studies Depression (CESD). The scale was specifically developed to measure depression. It includes an inverted scale that measure positive feelings ("Positive Affects").

future (from the ZTPI[26]), and a sub-scale of (internal) locus of control (the inverted 'fatalist present' sub-scale from the ZTPI).

A behavior index constitutes an aggregate from 6 measures: an inverted measure of anger or frustration in daily activities taken from the time-use module, an inverted measure of impulsiveness, an inverted 'conduct problem' sub-scale (from the SDQ[27]) and a (positive) 'pro-social behavior' sub-scale (from the SDQ).

For both the well-being and behavior index, results are presented in z-score with mean set to zero in the control group, so that estimated coefficients can be interpreted in standard deviations. A positive impact on the well-being index is interpreted as an overall increase in well-being and a positive impact on the behavior index as an overall improvement in attitude.

3.4 Empirical Methodology

3.4.1 Main specifications

We estimate intent-to-treat (ITT) effects for contemporaneous and post-program impacts for the pooled treatment via an ordinary least squares (OLS) regression:

(1)

$$Y_i = \alpha + \beta W_i + \delta X_{i,l} + \epsilon_i$$

where Y is an outcome for individual i, W is an indicator for treatment (being assigned to Public Works at first randomization), and X is a vector of stratification variables (locality l, gender x). Robust standard errors are clustered at brigade level for treated individuals[28].

[26] Zimbardo Time Perspective Inventory (ZTPI). It's an instrument measuring time perspective for individuals in different dimensions. In particular, two dimensions we use are "future" (to have a positive attitude towards future) and "fatalist present" which is very close to the concept of internal locus of control (to feel no control about things happening). We use the validated French version.

[27] Strength and Difficulties Questionnaire. It was initially created to measure behavioral issues for young children and teenagers (3 to 16 years old). We adapted it to our target (18 to 30 years old). We use two sub scales (over the five): "conduct problems" and "pro-social behavior".

[28] We suspect within-brigade error correlation due to the interactions between treated individuals who worked together in the same brigade for months. 'Large' brigade refer to either one brigade (25 individuals) or an aggregation of brigades (from same locality). Brigades are sometimes aggregated to account for the fact that some individuals have been moved across brigades during public works implementation for various reasons: when such movement occurred, we group the different brigades affected.

To estimate post-program ITT effects across treatment arms, we use the following specification:

(2)

$$Y_i = \alpha + \beta_1 W_i + \beta_2 (W_i * T1_i) + \beta_3 (W_i * T2_i) + \delta_1 X_{i,l} + \epsilon_i$$

where $T1$ (respectively $T2$) is an indicator for being assigned to the complementary self-employment training (respectively wage-employment training). Coefficient β_1 estimates the impact of the 'pure' public works while the coefficient β_2 estimates the additional effect of the self-employment training and β_3 the additional effect of the wage-employment training. The sum $\beta_1 + \beta_2$ (respectively $\beta_1 + \beta_3$) estimates the total effect of the program for individuals assigned to public works and complementary self-employment training (respectively wage-employment training). In the results table, we also provide the p-value for the test that this sum is equal to zero.

We analyze heterogeneity in treatment effects by groups G determined by a set of baseline characteristics Z (see discuss in section 5). The specification used in that case is the following:

(3)

$$Y_i = \alpha + \gamma_1 (W_i * G_i) + \gamma_2 (W_i * (1 - G_i)) + \gamma_3 * G_i + \delta_2 \mathrm{X}_{i,l} + \epsilon_\mathrm{i} ,$$

where G is an indicator for belonging to the group determined by Z. We are interested in coefficient γ_1 , which estimates the impact of the program (pooled treatment) for a specific group G. We provide standard errors for $\widehat{\gamma_1}$and the p-value for the test that $\gamma_1 + \gamma_2$ is equal to zero at the bottom of the tables.

For specifications (1) to (3), we use (probability) weights. They are composed to up to four[29] multiplicative weights to account for (i) public lotteries specificities in the first randomization, (ii) locality specificities in the second randomization, (iii) the sub-sampling of non-respondents during tracking surveys, (iv) the sub-sampling of the treated group for midline survey and (v) cases of control individuals who enrolled and eventually participated in later waves of the program[30].

[29] With midline data, we use weights related to (i), (iii) and (iv). With endline data, we use weights related to (i), (ii), (iii) and (v). More details on weight construction provided in Appendix A.

[30] This specific point is detailed in the next paragraphs and in Appendix A.

Potential threats to internal validity

We check for balance across groups in Table 2. Column (3) contains the p-values for the test of difference between treatment and control groups, and column (4) for the test whether all treatment arms are jointly equal to zero. Overall, there is no meaningful differences across groups[31]. We note that collecting the baseline survey after assignment to the program affected a few variables for which control or treatment groups may have had incentives to over or under report. Based on that, we do not add baseline controls on top of stratification controls to our specification.

At midline, compliance to program assignment was high. Only 6 individuals of the sample succeeded in 'cheating' the public lottery (by registering for the program in different locations) and only 2 generated contamination after being selected to the program. Among youth assigned to the public works, take-up was high as 93% of them participated for more than five months out of the seven months of implementation. In total, youth worked an average of 141 days, out of a maximum of the 154 days expected in case of full attendance.

An unforeseen potential issue emerged at endine. Between the end of the phase of the program being evaluated and the time of the endline survey, a few individuals in the control group were able to enroll (and eventually participate) for the third or fourth wave of the program[32]. However, detailed data about enrollment and public lotteries enabled us to observe control individuals' behavior for third or fourth wave (i.e. whether they applied or not to the program, and whether they were selected and participated or not). For post-program impact analysis, control individuals who (later) participated to the program are not included. To adjust for this, control individuals who also applied in future waves but were not selected through the public lotteries are assigned relatively larger wave[33], and 200 individuals were (randomly) added[34] to increase the total size of the control group.

[31] We also checked the balance across groups for both midline and endline respondents.
[32] We identified 140 individuals from our baseline control group (i.e. 13,5%) among next waves beneficiaries (91 for third wave, 49 for fourth wave). Among the 200 individuals randomly added, 30 were also identified to be among third or fourth wave beneficiaries.
[33] Specifically, we use specific weights to account for the behavior of enrollment to the waves and put a zero-weight on control individuals who ever received the program (see Appendix A2).
[34] To keep the same stratification as the initial control group sample, we selected the 20% next ones of each (baseline) lotteries' list.

Finally, the take-up of complementary training is lower than take-up for the ‘pure’ public works, but remains in with take-up levels in similar skills training programs: 72% of individuals assigned to self-employment training and 67,2% of those assigned to wage-employment training attended at least 75% of the training, which correspond to 60 hours over the 80 total hours. For both trainings, only 10% of individuals never attended.

3.4.2 Heterogeneity analysis using machine learning

Machine learning and heterogeneity of treatment effects

The use of machine learning methods in this paper comes from the desire to understand the heterogeneity of treatment effects, in a setup with rich and multi-dimensional baseline data. In particular, we would like to assess how much the treatment effect would vary under alternative assignment rules (based on observable characteristics), and especially if larger treatment effects could be obtained by targeting specific sub-groups compared to average treatment effects estimated across all participants.

In supervised machine learning methods (e.g. regression trees, random forests), the goal is to build a model of the relationship between a set of features (X) and an observed outcome (Y), to be able to make future predictions on the outcome based on observed features in a similar environment. More precisely, a model is first trained on a dataset where (Y, X) are observed and that model is subsequently used to predict Y on a population for which only the characteristics X are observed. Applying that approach to our specific setting, we would like to predict treatment effects conditional on a set of characteristics X that are measured before treatment (at baseline) for each individual in our sample, as our dataset provides us with an observed and realized outcome Y. Machine learning tools can help us detect heterogeneity conditional on a set of characteristics X without making assumptions on which Xs or combinations of Xs might be relevant. Given the prediction model built using such supervised machine learning methods, we would then be able to predict treatment effects for any set of characteristics included in X[35] and simulate treatment effect of the program under alternative selection rules (based on X)[36].

The main challenge in applying machine learning methods to the treatment effect problem, as described above, is that the structure of such algorithms requires to use criteria[37] assessing the quality of predictions made at different steps, which is easy when you can directly compare predictions and realizations for each individual, but less straightforward when you only observe the potential outcomes corresponding to the treatment received (known as the fundamental problem of causal inference). Athey and Imbens (2015) build a framework and propose new algorithms – including causal tree (CT) algorithm, to use machine learning for the specific case of predicting treatment effects conditional on a set of X. Wager and Athey (2015) extend this framework to causal forests and the possibility to make causal inference with the obtained results (see also Wager, Hastie and Efron, 2014).

We rely here on the framework developed by Athey and Imbens (2015). We want to estimate the conditional average treatment effect (CATE): $\tau(x) = E[\, Y_i\,(1) - \; Y_i\,(0) |\, X_i = x]$ with X a p-component vector of features like baseline covariates and Y the outcome of interest. This requires a dataset $(Y_i^{obs}, W_i, X_i), i = 1 \dots N$, considered as an i.i.d. sample drawn from an infinite superpopulation, where W is a binary indicator for treatment and Y^{obs} the realized and observed outcome, that is to say the potential outcome corresponding to the treatment received. We also need to make the assumption of 'unconfoundedness' (e.g. treatment is randomized conditional on observable covariates), which is reasonable given our randomized experiment.

Using causal forests to predict (conditional) treatment effects

In this paper, we implement the causal forests algorithm[38] as defined by Wager and Athey (2015) to have a prediction of $\tau(x)$. In a nutshell, *causal* forests are random forests of *causal*

[35] With some restrictions on the support of the set of X used.
[36] More details on machine learning procedure in Appendix A2.
[37] Actually this is required for two criteria : (1) the 'in-sample goodness-of-fit' measure, which is used in the cross-validation stage to build your model and (2) the 'out-of-sample goodness-of-fit' measure evaluated on a test sample to compare performance of different models.
[38] We thank Wager and Athey for graciously sharing their code with us. Susan Athey, Yanyang Kong and Stefan Wager (2015). causalForest: Causal Trees and Forests. R package version 0.9.0. https://CRAN.R-project.org/package=causalForest.

trees, the latter being standard trees adapted[39] to the case of treatment effect estimation as in Athey and Imbens (2015).

First of all, as in most machine learning methods, our data is split in two separate samples: a training sample used to build the model, and a test sample[40]. This is fundamental as all subsequent analysis will be performed on the test sample, a sample that is not involved at any step in the construction of the prediction model. Given our dataset (small N compared to p), this will be a constraint in terms of sample size.

Then, the model giving a prediction of $\tau(x)$ is built using a training sample set to 90% of the full sample N (standard). In a second step, $\tau(x)$ is estimated on the test sample (set to 10% of N). This is applied to both midline and endline surveys, for our two main outcomes of interest which are total monthly earnings and the well-being index (see section 3.3.2), and with $x \in X$ a rich set of 100 baseline covariates comprising individual and household characteristics, education, employment and savings dimensions, assets held, measures of personality, preferences, ability and cognitive skills.

To discuss heterogeneity of the treatment, we report the mean of $\hat{\tau}(x)$ over each quartile on the test sample denoted $\overline{\hat{\tau}(x)}^{Qi}, i = 1 \dots 4,$ as well as the standard deviation of $\hat{\tau}(x)$ distribution in the test sample. We particularly focus our analysis on the upper quartile range, considering that $\overline{\hat{\tau}(x)}^{Q4}$ provides us with an idea of potential gains on treatment effects when selecting individuals based on a full set of X.

4 Results: ITT Estimates for Public Works Impacts

This section presents ITT estimates for public works impacts on main outcomes for youths, based on the pooled treatment specification in equation (1). The ITT estimates are discussed separately for contemporaneous impacts (3/4 months after the start of the program, while youths are still participating) and post-program impacts (12/15 months after youths have exited from

[39] More specifically, causal trees rely on two new 'in-sample' and 'out-of-sample' goodness of fit measures, developed by Athey and Imbens (2015), thanks to which the algorithm directly leads to an estimation of τ(x).
[40] The test sample is used to assess (or compare) the performance of the model(s) by confronting the predictions made by the model with the 'grounded truth' observed in the test sample. This is justified by the fact that on the sample used to build the model (i.e. training sample), there is overfitting usually leading to 'over confidence' in the prediction quality. The test sample is also the sample used to make causal inference, as it was never involved in the algorithm building process.

the program). The next section discusses mechanisms for post-program impacts and impacts by treatment arms

4.1 ITT Estimates for Main Outcomes

Table 3 (Panel A, columns 1-6) presents contemporaneous ITT estimates on employment and hours worked. 86% of youths in the control group have an activity at midline, with 53% of youths holding a wage job, and 33% self-employed. This is consistent with the employment situation in Cote d'Ivoire and much of Sub-Saharan Africa, where a large numbers of youths are underemployed by working in low-productivity occupations, often in self-employment or informal wage jobs paying less than the minimum wage in the formal sector, and few are formally unemployed. In this context, impacts during the program on the overall employment level are limited for youths in the sample (+12 pp). The stronger employment effects stem from a change in the composition of employment, with strong impacts on the share of youths holding wage jobs (+44 pp) driven by the public works jobs, and a smaller decrease in self-employment (-9 pp). Similar patterns are observed for contemporaneous program impacts on hours worked per week, with a small overall increase in total hours worked (by 3.5 hours from on an average of 41 hours per week in the control group). This is driven by a large increase in hours worked in wage employment (+14 hours) and smaller decrease in hours worked in self-employment (-6.7 hours). Employment in the public works program accounts for approximately 30 hours a week for participants in the treatment group, so that the small increase in overall hours worked in fact hides a large decrease in hours worked in other activities. Youths in the treatment group also become more likely to cumulate various activities[41]. Overall, the observed contemporaneous program impacts on employment raise questions on the effectiveness of the self-targeting mechanism in a context where most youths are working and have to rearrange their activities to make time to access the better public wage jobs offered by the program.

Table 3 (Panel B, columns 1-6) presents post-program ITT estimates on employment and hours worked. Despite strong shifts in youths' employment portfolios during the program, no post-program impacts on employment level, employment composition or hours worked are observed. Youths in the treatment group display a similar employment profile than youths in

[41] It is estimated than the share of youths holding multiple activities increase by 0.2 in the treatment group (results not shown).

the control group, with no statistical difference in the main indicators for employment and hours worked. On the one hand, these results show that the public works program did not lead to "stigmatization" or "scarring" effects for youths. Past participants are not less likely to be employed (including self-employed or holding wage jobs) a year after their exit from the program. This suggests a relatively rapid adjustment back to the pre-program occupations. On the other hand, results also show that the public works does not bring longer-term benefits to youths in terms of employment types or hours worked. An important exception, however, relates to the earnings and productivity in these occupations, to which we now turn

Table 3 (Panel A, columns 7-9) presents contemporaneous ITT estimates of program impacts on earnings. The public works leads to a net increase in earnings of 20,885 FCFA per month (or approximately $42) during the program. The net earnings gains represent 35% increase from the level of earnings in the control group (60,052 FCFA, or $120), or approximately 42% of the average net monthly transfer amount (50,600 FCFA, or $101[42]). As such, the estimated effects point to substantial foregone earnings from activities that youths left or scaled down in order to participate in the program. As for employment patterns, contemporaneous impacts on earnings stem from the strong increase in earnings from the wage jobs offered by the program (+35,385 FCFA, or $71), with a significant decrease in earnings from self-employment (-12,625, or $25).

Table 3 (Panel B, columns 7-9) presents post-program ITT estimates of program impacts on earnings. The public works leads to a small but significant increase in earnings a year after the end of the program (+5,622 FCFA, or $11 per month), a 11.6% increase from the level of earnings in the control group. The increase in earnings is concentrated in self-employment, where earnings increase by 6,223 FCFA ($12.4) per month, or substantial relative increase of 32%. On the other hand, post-program earnings in wage jobs are not statistically different between the treatment and control groups. These results show that, while participants were not more likely to be employed, employed in different occupations, or working longer hours, they were on average running more profitable self-employment activities one year after exiting from the program.

This section goes beyond traditional economic indicators to discuss the impacts of public works on indices of well-being and behavior. The consideration of broader well-being indicators are important as the temporary jobs offered by the program may have indirect benefits or costs

[42] This figure is the average amount transferred over all individuals assigned to the public works (independently of non-compliance and days not worked).

beyond economic dimensions. On the one hand, the public works activities are hard manual labor activities, which some may consider depreciating. On the other hand, there can be a certain status associated with holding a public wage job in the community, in particular a predictable and secured formal wage job. Changes in youths' well-being and behavior are particularly relevant in a post-conflict setting such as Cote d'Ivoire, including as they point to potential program externalities on social cohesion, an issue of strong interest for policymakers.

Table 3 (Panel A, columns 12-13) presents ITT estimates of contemporaneous program impacts on indices of well-being and behavior (see section 3.3.2 for definition). Results show significant contemporaneous program impacts on the well-being index (+0.2 standard deviations), as well as the behavior index (+0.13 standard deviations). Improvements in well-being while youth participate in the program come from a larger share of youths reporting feeling happy and proud, scoring higher on sub-scales for self-esteem, positive affect and positive attitude towards the future, as well as reporting higher present and future life satisfaction[43]. Improvements in behavioral dimensions are more limited, but point to less anger and frustration in daily life, and less impulsiveness, although no related changes in other domains such as pro-social behavior and conduct problems are observed. These results may be associated with the economic gains mentioned earlier. They also raise the possibility that some youths who do not benefit substantially in economic dimensions may nevertheless benefit from the program in psychological or behavioral dimensions, an issue to which we return below.

Table 3 (Panel B, columns 12-13) presents post-program ITT estimates on indices of well-being and behavior. Some lasting improvements are observed on psychological well-being for youths in the medium-term (0.09 standard deviations), although they are more muted than during the program. They are also concentrated in a narrower set of domains such as happiness, self-esteem and present life satisfaction. In contrast, there are no lasting impacts on sub-scales for pride, positive affect, positive attitude towards the future and future life satisfaction. A year after the end of the program, no lasting impacts are observed in the behavior index or any of its subcomponent, suggesting that short-term behavioral gains have faded.

4.2 Mechanisms for Post-Program Impacts and Impacts by Treatment Arms

[43] In contrast, present fatalism is unaffected.

Intermediary Outcomes: Short-term Expenditures and Savings

Table 3 (Panel A, columns 10-11) presents contemporaneous ITT estimates on expenditures and savings, in order to illustrate the mechanisms for the observed impacts on earnings. During the program, the observed increase in earnings (+20,885 FCFA per month, or $42 as mentioned above) translate into an increase in both expenditures and savings. Total monthly expenditures are estimated to increase by 15,085 FCFA ($32), constituting approximately 70% of the nets earnings gains. The overall increase in expenditures can be decomposed in roughly equal shares between youths' own expenditures and their contribution to household expenditures. The additional expenditures are mostly for basic necessities (food, clothes, …), as well as education and training.

Yet, beyond consumption support, youths are able to save a significant share of their net earnings gains. On average, after approximately 4 months in the program, youths in the treatment group have increased their stock of savings by approximately 39,633 FCFA ($79). This impact is of large magnitude as it corresponds to a 182% increase from the average stock of savings in the control group (21,752 FCFA, or $43). The order of magnitude is also consistent with youths saving approximately 30% of their earnings gains. Importantly, youths are not only more likely to save and to save larger amounts, but most of these savings are kept in formal bank accounts. These include accounts in which youths are paid their public works wages. Overall, these substantial contemporaneous increases in savings are likely to contribute to the post-program effects on earnings in self-employment. Indeed, post-program results show that youths are not more likely to be self-employed, but are more likely to operate micro-enterprises with a relatively larger asset stock and scale of operations, pointing to higher productivity. Youths in the treatment group also report higher investments in household enterprises, which likely have been facilitated by savings from the program and are consistent with observed higher earnings in self-employment.

Impacts by Treatment Arms

Table 4 (Panel C) further illustrates the contribution of the various treatment arms to disentangle the causal impacts of complementary skills training from the average post-program impacts documented so far. Overall, little heterogeneity in impacts across treatment arms is observed, suggesting limited value-added of the complementary skills training. Specifically, post-program impacts on employment level, employment composition and hours worked are very consistent across the difference treatment arms[45]. One noteworthy distinction relates to the post-program earnings impact, for which results are mixed. On the one hand, post-program impacts on total earnings are mostly observed for the groups of individuals that were assigned to the basic entrepreneurship or jobs search skills training. In particular, post-program impacts on self-employment earnings mostly come from youths in the treatment group that was assigned to the basic entrepreneurship training. On the other hand, despite being significant for specific treatment arms, the impacts on earnings are not statistically different between arms, so that we cannot reject equality in earnings impacts across groups. The only statistical difference in earnings between arms is when comparing self-employment earnings across the basic entrepreneurship training arms and public works only arms. Ultimately, since there are not statistical differences in impacts on overall earnings across treatment arms, we pool treatment to conduct finer heterogeneity analysis in the rest of the paper.

The limited value-added of the complementary training suggests that skills acquisition through these trainings is not the main mechanism that explains the post-program impacts. In fact, the trainings were effective in improving knowledge in basic entrepreneurship, respectively jobs search skills, as they intended to do. They also led to youths applying these skills in practice, either by intensifying their search for wage jobs (e.g. using a CV for a job search, searching using adds or by applying independently) or their efforts to set-up a new activity (e.g. by undertaking a market study or a preparing a business plan). However, these changes in skills and practices were not sufficient to generate earnings beyond those generated by the basic public works program.

5 Heterogeneity Analysis

The public works program was oversubscribed, with the number of applicants exceeding the number of available program slots by a ratio of 4 to 1. While participation in the program was the result of a random assignment process, which has the advantage of being fair and

[45] The only exception is that hours in self-employment are significantly larger in the jobs search training arms compared to the public works only arm. Still, the coefficient is not statistically different from 0 or from the estimate from the basic entrepreneurship training arms.

transparent, it is possible that the performance of the program might have been improved with a better targeting of the 25 percent of program beneficiaries among applicants.

Whether and by how much alternative targeting might have improved program effectiveness depends on the degree of heterogeneity in program impacts. In theory, given the self-selection mechanism, we would expect heterogeneity in impacts among program applicants, with marginal applicants experiencing very limited benefits from the program compared to infra-marginal applicants with more limited employment opportunities outside the program.

5.1 Quantile Treatment Effects

In practice, detecting heterogeneity of treatment impacts is complicated. Indeed, some parameters of the impact distribution, such as its variance, are not identified. Their identification would require knowledge of the joint distribution of potential outcomes whereas only one potential outcome can be observed at a time for each individual.

To study the potential heterogeneity of treatment impacts in an experimental setting, researchers traditionally look at quantile regressions. Indeed, even if not identified, it is possible to obtain a bound for the variance of impacts (Heckman Smith and Clements, 1997, Djebbari and Smith 2008). Quantile regression results inform us about the lower bound of the variance. When quantile treatment effects are homogeneous, then the lower bound of the variance is zero: a constant treatment effect is consistent with homogeneous quantile treatment effects. Bitler et al. (2006, 2014) provide a well-known application of quantile treatment effects to document heterogeneity of impacts. As stressed in this prior work, quantile treatment effects will help to detect heterogeneity of impacts only if the intervention preserves ranks, or at least does not lead to too much churning in the distribution of the outcomes. In fact, the rank preservation assumption allows one to interpret quantile treatment effects as the 'effect *at* quantile,' hence making it directly informative about heterogeneity of impacts. Under this assumption, observing non homogeneous quantile treatment effects indicates that the lower bound of the variance is strictly greater than zero. However, if the intervention does not preserve ranks, quantile treatment effects are of little help to detect heterogeneity.

Figure 1 graphs quantile treatment effects on earnings during the program (Panel (a)) and after the program (Panel (b)). The horizontal axis in each panel reports the quantile and the vertical axis the estimate of the treatment effect at the corresponding quantile. The shaded area around the estimate provides the 95% confidence interval. As the figure clearly illustrates, the quantile analysis shows substantial heterogeneity of impacts on income during the program. The quantile treatment effect is as large as FCFA 45,000 at the 15% quantile, but only FCFA 15,000 at the 85% quantile. Moreover, the precision of the quantile treatment effects is quite high, strongly suggesting the existence of true heterogeneity rather than just sampling variation.

The quantile analysis of treatment effects on income after the program (Panel (b)) offers a rather different picture. Post-program quantile treatment effects are uniformly small and the small dispersion is within confidence bounds, consistent with sampling variation.

In summary, there appears to be large heterogeneity of quantile treatment effects during the program, but less heterogeneity after the program. This is enough to say that there is heterogeneity in impact on earnings during the program. Moreover, the model derived in the framework (section 2) shows that the intervention possibly preserves ranks, or at least might not induce too much churning in the distribution of contemporaneous effects. Thus, the true variance of impacts might not be too far from the lower bound and therefore the heterogeneity we see in quantile treatment effects is a direct picture of the true underlying heterogeneity during the program. (Put differently, looking at the simple cumulative distribution of earnings in the treatment and control groups reveals almost all of the heterogeneity).

However, this is not necessarily the case for post-program earnings. There might be individual latent components, such as return to capital, that did not contribute to the ranking of individual earnings without the program but contribute to the ranking of individuals' post-program earnings. For example, if there are individuals trapped at the bottom of the earnings distribution without the program, but with high returns to capital (e.g. through setting-up a highly profitable project), these individuals, thanks to their participation in the program, might end up further towards the top of the earnings distribution. This is because the program allows them to save during and implement their latent project after program completion.

5.2. Machine Learning

Given the possible limitations of the quantile regression approach, in particular when it comes to assessing heterogeneity in post-program effects, we turn to a different empirical approach. The technique is based on the identification of (observable) underlying baseline variables contributing to heterogeneity in the treatment effect. In a standard regression framework, this could simply amount to interact the treatment variable with covariates, and to recover predicted impacts conditional on these covariates. Djebarri and Smith (2008) apply this method and show that it allows one to capture a substantial share of the variability of impacts. The machine-learning methods developed in Athey and Imbens (2015) and Athey and Wager (2015) go one step further in identifying the expectation of impacts conditional on a set of covariates. The main innovation of these machine learning methods is that they allow one to uncover the underlying heterogeneity in predicted impacts without making any assumption about where this heterogeneity comes from, or the form that it takes. Although these methods can miss some determinants of heterogeneity (especially if the set of covariates is not rich enough), they offer an alternative and systematic way to explore the heterogeneity of impacts. These methods are particularly applicable to our context given the large number of covariates we have captured in our rich baseline survey. Moreover, these methods permit to detect heterogeneity even if the intervention is not rank preserving.

As already discussed above, implementing these machine learning methods involves a two-step procedure. In the first step, the relevant set of covariates is identified on a "training sample" comprising of a 90% random subsample of the total sample. The second step amounts to estimating impacts conditional on these identified covariates (and interactions of covariates) on a "test sample," the remaining 10% random subsample of the total sample. Of course, this approach can be used to estimate the heterogeneity of predicted impacts both during and after the program; and it can be applied to study heterogeneity in the impacts on earnings, but also on other outcomes, such as the well-being index.

The scatter plot in Figure 2 shows, for each individual in the "test sample," predicted impacts on earnings during the program (x-axis) against predicted impacts on earnings after the program (y-axis). The vertical and horizontal lines on the scatter plot respectively correspond to the average estimated impacts during and after the program. Unsurprisingly given the quantile results reported above, the scatter plot reveals substantive heterogeneity of impact during the program. What is more surprising is the observed variability of impacts after program completion (Figure 2). Indeed, we see a wide range of possible estimated impacts at "endline," ranging from as little as FCFA -10,000 to as much as FCFA 30,000. Estimated impacts on earnings after the program are non-zero for a large share of individuals.

Table 4 reports on features (earnings and well-being) of the measured heterogeneity of estimated impacts, both during and after the program, derived from applying Athey and Imbens

(2015)'s method. Contemporaneous program effects are reported in columns (1) and (2) while post-program effects are in columns (3) and (4). Odd columns display results for earnings, and even for well-being.

Panel (A) of Table 4 presents means and standard errors of the distribution of estimated conditional impacts for each of the 4 outcome variables. Summarizing the evidence from Figure 2, we find that the mean predicted impact on earnings during the program is FCFA 20,108, with a standard deviation of 23,199. The mean predicted impact on earnings after the program is FCFA 5,439, with a standard deviation of 7,643. The mean predicted impact on well-being during the program is .017 with a standard deviation of .07; the mean predicted impact on well-being post-program is .07 with a standard deviation of .08.

Columns (1) of Panel (B) and columns (3) of Panel C offer further confirmation that there is a large amount of heterogeneity in earnings impacts both during and after the program. The average predicted impact on earnings during the program is FCFA -4,823 in the lower quartile of the distribution compared to FCFA 35,072 in the upper quartile. The average predicted impact on earnings after the program is FCFA -1,567 in the lower quartile of the distribution compared to FCFA 13,648 in the upper quartile.

In other words, better targeting could in theory improve program impacts. Replacing randomized assignment into the program of 25% percent of applicants with a targeting rule that "maximize" contemporaneous program impacts on earnings would lead to a 66% increase in impact during the program (FCFA 35,072 compared to FCFA 20,885 achieved by random selection). Replacing randomized assignment into the program of 25% percent of applicants with a targeting rule that "maximize" post-program earnings would more than double post-program impacts (FCFA 13,648 compared to FCFA 5,621 achieved by random selection).

This last result is an important one. It does not come as a surprise that workfare programs have a large and heterogeneous impact on income during the program. This is actually what we expect from the self-selection mechanism, with a fraction of marginal participants almost indifferent between being enrolled or not, and others being infra-marginal. The big uncertainty with workfare programs is whether they have, or can have, impacts after program completion. The potential mechanisms behind the heterogeneity of post-program impacts are less straightforward than for the short-term impacts. They might have to do with experience in the labor market and ability to signal skills and their upgrading. They also might have to do with return to capital and the impact of investments on income generating activities.

While the results in Table 4 so far suggest the possibility of improving program effectiveness both during and after the program through better targeting, a first important question they leave unanswered is whether the targeting that would maximize contemporaneous program impacts would also maximize post-program impacts. In other words, are the covariates that are associated with large predicted impacts during the program associated with large predicted impacts after the program? Similarly, are the covariates that are associated with large predicted impacts after the program also associated with large predicted impacts during the program? The scatter plot in Figure 2 suggests a negative answer in that the correlation between estimated conditional impacts at endline and midline is small and negative: -0.125. The remaining columns in Panel (B) and (C) of Table 4 provide a more formal investigation of these questions.

Specifically, Panel (B) presents mean estimated conditional impacts both during the program and after the program over the quartiles of the distribution of the estimated conditional impacts on earnings during the program; Panel (C) presents mean estimated conditional impacts both during the program and after the program over the quartiles of the distribution of the estimated conditional impacts after the program.

Column (3) of Panel (B) and column (1) of Panel (C) confirm the visual evidence in Figure 2 in that there is no systematic relationship between those that benefit the most during the program and those that benefit the most after the program. Specifically, targeting the program to the 25% of applicants that benefit the most during the program would result in average predicted impacts after the program of FCFA 5,624, close to the average treatment effect. Similarly, targeting the program to the 25% of applicants that benefit the most after program completion would result in average predicted impacts at midline of FCFA 17,928, below the average treatment effect. Columns (2) and (4) of Panels (B) and (C) further reveal that targeting based on highest predicted impacts on earnings during the program or after the program would not result in also targeting those that benefit the most in terms of well-being (either during or after the program). In other words, these results suggest tradeoffs when trying to improve program effectiveness through better targeting. They highlight the challenge in isolating a single targeting rule that might “maximize” all outcomes (during and after the program; income and well-being).

Table 6 proposes another way to investigate the similarities and differences between the two populations selected as the 25% with highest estimated conditional impacts on during-program and post-program earnings. The table considers several baseline characteristics and presents averages over the sample of participants (column (1)), the two upper quartiles of interest (columns (2) and (3)) and for comparison the two bottom quartiles associated (columns (7) and (8)). Although the two upper quartile are defined based on baseline characteristics (they are of the form $\widehat{\tau(x)} > s$ for some s), the non-parametric procedure followed to determine the function $\widehat{\tau(x)}$ makes it difficult to clearly identify differences in characteristics that matter. This table however offers a way to grasp some of the most important differences between the two functions. The table also presents in columns (4) to (6) and (9) to (13) p-values for the test of the assumption of no difference across these groups.

Although not all the characteristics listed are significantly different in the two quartiles (largest predicted impact at midline and largest predicted impact at endline), the table clearly shows that there are significant differences between these two populations. Among them appears first gender: the proportion of women among participants is 32% but we see that this share is larger among the top 25% at "midline" (37.7%) but lower among the top 25% at "endline" (28%). Another interesting difference is the area participants live in: 94.2% of participants live in an urban area, but this is the case for only 80.2% in the top 25% at "endline".[46] Other interesting differences are related to financial characteristics. There is a large difference in savings among the top 25% of impacts on post-program income (FCFA 40,751) and the sample of program participants (FCFA 28,777). Similarly, the share of participants in the top 25% of post-program impacts on earnings saying they have credit constraints is 44.7%, compared to 49.9% in the sample of applicants and 53.4% in the top 25% of impacts on income during the program.

The fact that those that benefit the most after the program appear to be less financially constrained and have a higher stock of saving suggest that this group, while experiencing larger foregone earnings from program participation than those that benefit the most during the program, may nevertheless been able to save a greater share of the higher income induced by program participation, or better able invest these savings into income-generating activities or to run these activities more profitably.

The empirical evidence raises two issues related to this potential mechanism. First, the previous table shows that those for whom the impacts at endline are the largest are not those who have low savings at baseline or those who experience financial constraints. Second, such a mechanism would imply an association between large post-program impacts and large impacts during the program. It would imply that there is first an increase in savings and then investments

[46] The maximum proportion of participants leaving in non-urban area in a 25% draw is (1-0.942)4, so 23.2%. As their actual proportion in the top 25% is 1-0.802=0.198, 0.198/0.232=85.3% of participants living in non-urban areas are in the top 25% of impacts on endline income.

allowing to start new business or improving the profitability of existing businesses. However, a large contemporaneous program impact on income is necessary for an increase in savings to occur. This means that there are also other likely channels. A standard labor market integration as a wage earner might be an alternative channel to explain post-program impacts. Another possibility is an impact of participation in the program on the behavior and aspirations of participants.

5.3. Alternative Targeting Rules

The analysis in this section so far suggests the possibility of targeting the program in alternative ways that would improve contemporaneous program impacts or post-program impacts on income, but also highlight some trade-offs between these two targeting rules. It is also clear that the targeting rules that would emerge from machine learning approaches would be too complicated and expensive to implement, relying on information that is not easily observable or verifiable and involving complex functions based on this information, and hence of limited practical relevance. An important follow-up policy question is therefore whether there are simple targeting rules, or simple changes to the self-targeting procedure, that could be implemented and come close to achieving the predicted impacts in the upper quartiles of the distribution in column (1) of Panel (B) (contemporaneous program effects) and column 3 of Panel C (post-program effects). We address this question in Table 5. In particular, Table 5 presents average contemporaneous program impacts on income (upper panel) and post-program impacts on income (lower panel) for specific sub-populations. For reference, the first two columns of the table reproduce averages over the whole sample of participants and over the top 25% of predicted impact on income.

One of the findings from Table 6 was the greater relative representation of women in the subset of the population that benefit the most during the program. What if only women had been able to apply for the program? Column (3) in Table 5 shows that such a change to the program participation rule would improve impacts both during and after the program. In particular, average impacts on income during (after) the program would be FCFA 3,7150 (FCFA 8,258) if the randomized assignment had been restricted to female applicants. This corresponds to a 77% improvement in average estimated impacts on during-program income and a 46% improvement in post-program impacts compared to randomized assignment of both men and women. While these post-program outcomes are lower than the average impact we estimate in the upper quartile of the distribution of endline impacts, they are in line with the average impacts estimate in the upper quartile of the distribution of impacts during the program. Well-being impacts on women also appear relatively large, especially during the program, but not statistically significantly so.

As we discussed above, one of the reasons why self-targeting might have led many non-poor to apply for the program is that the wage was set at the statutory minimum wage, well above hourly rate for many in the informal sector. What if access to the program had been randomized among those, men and women, willing to participate at a lower wage rate? We can, albeit imperfectly, get at this question. Indeed, our endline survey asks participants whether they would have participated in the workfare program for a FCFA 1,500 salary instead of the offered FCFA 2,500. While it is clearly suboptimal to study heterogeneity of effects based on a variable measured at endline, we note that identical proportions in treatment and control groups agreed to the idea of participating at the lower wage. So, while it is somewhat speculative to run heterogeneity analysis based on such a characteristic, it is interesting to consider because self-selection is intrinsically rooted in workfare programs[47]. Results are displayed in column (2) of the table. We see that here again there is a substantial improvements in both contemporaneous and post-program impacts. Average impacts would have reached FCFA 32,076 at midline and FCFA 8,121 at endline if a random sample of only those willing to work for FCFA 1,500 had been selected. This self-targeting does not significantly raise well-being compared to the current assignment rule.

Importantly, we should note though that this is not what the effect of a workfare program paying FCFA 1,500 would have given. Indeed, these are the earnings impacts, during and after the program, among those willing to participate at FCFA 1,500 but actually being paid FCFA 2,500. Hence, while self-selection would have been better (with lower foregone earnings) had the wage paid to program participants been set below the minimum wage, the value of the program for the participants, at this lower wage, would have been reduced. In other words, while lowering the offered wage improves the self-targeting, it also reduces the transfers, and hence program benefits, to this better targeted group. This is one of the key tradeoff embedded in the self-targeting logic behind workfare programs. If this were a one-shot game, policymakers might have been able to surprise program applicants expecting a FCFA 1,500 wage with a higher

[47] We implemented some robustness checks and examine results based on a prediction of this variable based on baseline covariates and find they are quite similar.

actual wage and hence mimic the results above. Clearly though, in a realistic repeated game setting, such a practice would certainly not be sustainable.

While self-targeting may offer the cheapest targeting approach, what about simple targeting rules based on a few predictors of baseline income? If there is limited churning in the distribution of income over a period of time as short as the length of this program (6-7 months), those with the lowest baseline income are likely to have the lowest income at midline absent program participation. In columns (4) and (5) of Table 5, we experiment with two approaches to directly target the poor. First, we assess impacts during and after the program among those that report baseline earnings in the bottom quartile of the distribution (column (4)). Second, we use the machine learning methods outlined above to predict baseline earnings among program applicants using a limited set of covariates that are both easily observable and not easily manipulated, including gender, age, household characteristics and assets; we then assess program impacts among participants in the lowest quartile of this distribution of predicted baseline income. This second approach is meant to correspond to the proxy means test that are often used for the targeting of benefits to the poor, and more robust to misreporting than self-reported income.

Columns (4) and (5) show that the midline impacts would match predicted impacts in the upper quartile of the predicted impacts distribution at midline under either of these targeting rules. Targeting the 25% with the smallest baseline self-reported income leads to an average expected impact on income during the program of FCFA 32,694; targeting the poor based on the 25% lowest predicted income leads to an average expected impact on income during the program of FCFA 31,968. Targeting based on a proxy test for income leads to estimated post-program impacts that are roughly comparable to restricting program participation to women. Targeting based on predicted baseline income does not result in differential well-being effects. However, targeting based on self-reported income appears to reduce well-being effects during the program but increase them post-program. What is most surprising, and in fact puzzling given the results in Table 4, is the large estimated endline earnings and well-being impacts if the workfare program targets those with the lowest self-reported baseline income.

The previous analysis has shown that the allocation of program slots can be substantially improved with better self-targeting or targeting rules. The machine learning heterogeneity analysis suggests that optimal targeting would differ based on the outcome of interest (contemporaneous vs post-program; income vs. well-being) and that trade-offs will appear. Yet the consideration of alternative targeting rules suggests to sharp trade-offs. In other words, targeting those that will benefit the most on one dimension does not lead to targeting those that will benefit the least on another dimension; instead, it leads to targeting a group that would benefit in a rather average way on those other dimensions. Our results in Table 5, with the exception of the result in column 3, are consistent with this. We can substantially improve impacts during the program through better targeting, at no large costs, but also no strong benefit, in terms of post-program impacts.

6. Cost-effectiveness

The estimated direct economic impacts of the program on youths' earnings can be used to perform cost-effectiveness calculations. The total costs per beneficiary for the basic public works program amounts to FCFA 660,478 ($1321), of which FCFA 354,166 ($708) are direct transfer to beneficiaries, FCFA 255,189 ($510) are other direct costs (material, team leaders and supervisors, basic life skills training), and FCFA 51,123 ($102) are indirect management costs[49]. For the public works only, the contemporaneous impacts on earnings are estimated at approximately FCFA 20,900 ($42) and the post-program impacts on earnings at FCFA 4,100 (8.2$) per month.

Table 6 presents the main results from cost-effectiveness calculations. Assuming that the contemporaneous impacts are constant during the 7 months of the program, and that the estimated post-program impacts are constant for the 13 months after the end of the program[50], the discounted total impacts on earnings 13 months after the end of the program is estimated at FCFA 206,695 ($42). The cost-effectiveness ratio for the intervention is 3.2, meaning that the

[49] Cost-effectiveness calculations are performed for the public works only (without complementary basic entrepreneurship training or jobs search skills training). The comparisons of relative cost-effectiveness between scenarios remain similar if considering cost and impacts from the pooled treatment instead.

[50] We compute a discounted sum of impacts on total earnings from program starts (month 1) up to 13 months after exit from the program (month 20) which corresponds to the moment when post-program impacts are measured. This is conservative in the sense that we make no assumptions for what happen after month 20 (or equivalent to zero-impact assumption). The following assumptions are used for the calculations:
H1 : contemporaneous program impacts β^{During} correspond to the impact on earnings at the end of month 4 and this impact is constant (up to a monthly discount factor) over the program period (month 1 to month 7)
H2 : post-program impacts β^{Post} correspond to the impact on earnings at the end of month 13 and this impact is constant (up to a monthly discount factor) from the end of the program (month 8) to the endline survey (month 20).
H3 : the annual discount rate is equal to $1/\ (1+\delta)$, with $\delta = 10\%$. Using monthly discount factors, $\rho = 1/(1+\delta)^{1/12}$.
Finally, the total discounted impact flow (DIF) is : $DIF\ =\ \sum_{k=1}^{7}(\rho^{k-4} * \beta^{During}) +\ \sum_{k=1}^{13}(\rho^{k-13} * \beta^{Post})$, with β^{During} (respectively β^{Post}) the contemporaneous (respectively post-program) ITT estimates of monthly total earnings impact.

average cost per beneficiary is 3.2 times higher than the average discounted direct impacts on earnings. This poor cost-effectiveness ratio is not surprising given that the net earnings gains are only 42% of the average monthly transfer amounts during the program, and that the cost of transfers only represents 54% of overall program costs. To highlight the poor average cost-effectiveness of the program in another way, the post-program impacts observed after 13 months would need to be sustained for 22.9 years for the program to be cost-effective based on net earnings gains for youths.

This cost-effectiveness analysis should be considered a lower-bound as it does not account for non-economic benefits such as those on psychological well-being or behavior mentioned above, or other externalities from the program, including the indirect benefits of roads rehabilitation. Nevertheless, they provide a benchmark to assess the cost-effectiveness of potential program improvements such as the implementation of alternative targeting mechanisms, in particular if we consider in a first-order approximation that non-economic benefits and externalities are similar across these potential improvements.

Table 6 indicates how adjustments in program targeting would strongly improve cost-effectiveness. In light of the strong observed heterogeneity in impacts, cost-effectiveness improves strongly across the various targeting approaches considered. Compared to the benchmark scenario with self-targeting based on the formal minimum wage, the total discounted total impact on earnings more than double (from FCFA 206,695 to FCFA 418,961, or $413 to $838) under the scenario of selecting youths with low predicted baseline earnings. Large improvements are also observed under other alternative selection criteria, including those that proxy self-selection based on a lower wage, target women only, or target based on self-declared baseline earnings. Overall, the cost-effectiveness ratio would improve from 3.2 to between 1.58 and 1.98 based on finer program targeting. The years needed to sustain post-program impacts for impacts on youths' earnings to justify investments in the program would go down from 22.9 to between 3 and 5.5. While the analysis cannot decisively indicate which targeting scenario would maximize cost-effectiveness given the confidence intervals around the impact estimates, it does highlight strong improvements in cost-effectiveness when departing from self-targeting based on the formal minimum wage.

7. Conclusion

The Cote d'Ivoire public works program we have evaluated shares many of the features of other public works program that have been adopted throughout the developing world in response to transient negative shocks such as those induced by climatic shocks or episodes of violent conflicts, by providing a few months of employment in infrastructure to those willing to work at the stated wage. Our analysis, relying on a randomized control trial as well as the collection of rich data before, during and after the program, has allowed us to assess the effectiveness of these programs in lifting the vulnerable, both economically and psychologically.

Results show that program impacts on employment are limited to contemporaneous shifts in the composition of employment towards the public works wage jobs, with no lasting post-program impacts on the level or composition of employment. However, participation in the program does raise earnings and psychological well-being, both during the program but also, and maybe most importantly, after program completion. Post-program earnings gains are mainly achieved through more vibrant entrepreneurial activities, likely boosted by the additional savings participants were able to secure during the program, but also possibly by other skills that were developed through the workfare and related complementary training.

However, the program as currently implemented is far from cost-effective when benefits are measured solely based on the earnings gains of those participating. This is primarily due to the fairly high indirect cost of implementing public works programs compared to more traditional welfare programs, but also due to the use of self-selection mechanisms based on the formal minimum wage. Many of the individuals who apply to participate in the program are already employed, consistent with the widespread underemployment challenge and limited unemployment in Sub-Saharan Africa. In an environment where informal employment is rampant, many of those who already have some form of occupation self-select into a public works program that offer higher earnings (even by paying the formal minimum wage), as well as potential non-economic benefits on psychological well-being and behavior. In this context, the program has very small average impacts on employment or hours worked, leading to large foregone earnings.

A basic framework to consider self-selection mechanisms shows that a reasonable theoretical expectation is that the impact is almost zero for the 'marginal' participant, and equal to the amount of the transfer for those with no outside employment opportunities. In this context, the distribution of individual impacts over the population is likely to vary substantially. Consistent with this, we demonstrate, using new methods from machine learning, large heterogeneity in program impacts, both during the program but also after the program. In fact, heterogeneity in program impacts on earnings are so large that they suggest that improvements in targeting is a first-order program design question, and perhaps a more critical issue than other program design aspects such as those related to program content like the value-added of complementary skills training.

Results from machine learning techniques suggest potential trade-offs between maximizing contemporaneous and post-program benefits. Yet traditional heterogeneity analysis shows that a range of practical targeting mechanisms come close to the machine learning benchmark, leading to stronger contemporaneous and post-program benefits without sharp trade-offs. This implies that cost-effectiveness could be boosted by departing from self-targeting based on the formal minimum wage. Indeed, we show that a range of potential self-targeting or targeting rules, which could be implemented at minimum additional costs, could substantially raise cost-effectiveness as measured solely based on earnings' gain. Still, even with this improved targeting, post-program impacts would still need to be sustained for at least 3 years for the program to be cost-effective based on participants' earnings gains alone.

Does it mean that public works program are not worth it? While our results so far could be interpreted that way, this is under the important caveat that the cost-effectiveness ratios we currently estimate are based on participants' earning alone, with zero impacts on earnings assumed beyond 13 months after the program. The cost-effectiveness calculations are also conservative as they do not include other social benefits of the public works program, such as the value of the new or upgraded infrastructure or the reduction of negative externalities (for example, incapacitation effects leading to reduction in crime or illegal activities) the program may have generated. These additional benefits are viewed as one of the advantage of workfare programs compared to traditional welfare, although they are rarely formally evaluated. Accounting for such externalities, a public works program with improved targeting may well become cost-effective. It may be particularly socially or politically desirable if social planners put a high weight on externalities and non-economic social benefits related to social cohesion relative to direct economic benefits. In future work, we will attempt to provide a back-of-the envelope calculation as to how much these additional benefits might need to be to justify public works from a cost-benefit perspective.

8. References

Alik-Lagrange, A., & Ravallion, M. (2015). Inconsistent Policy Evaluation: a Case Study for a Large Workfare Program. *NBER Working Paper, 21041*.

Athey, S., & Imbens, G. (2015). Machine Learning for Estimating Heterogeneous Causal Effects *Working Paper*, (No. 3350).

Beegle, K., Galasso, E., & Goldberg, J. (2015). Direct and Indirect Effects of Malawi ' s Public Works Program on Food Security (No. 7505). *Policy Research Working Paper Series*.

Bertrand, Marianne; Bruno Crépon ; Alicia Marguerie et Patrick Premand, (2016). Impacts à Court et Moyen Terme sur les Jeunes des Travaux à Haute Intensité de Main d'œuvre (THIMO) : Résultats de l'évaluation d'impact de la composante THIMO du Projet Emploi Jeunes et Développement des compétences (PEJEDEC) en Côte d'Ivoire. Washington DC : Banque Mondiale et Abidjan : BCP-Emploi.

Besley, T., & Coate, S. (1992). Workfare versus Welfare Incentive Arguments for Work Requirements in Poverty-Alleviation Programs. *American Economic Review*, *82*(1), 249–261.

Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2014). Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment. *NBER Working Paper*, *20142*.

Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2006). What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments. *American Economic Review*, *96*(4), 988–1012.

Blattman, C., & Ralston, L. (2015). Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs. *Working Paper*.

Datt, G., & Ravallion, M. (1994). Transfer benefits from public-works employment: Evidence for rural India. *The Economic Journal*, *104*(427), 1346–1369.

Deininger, K., & Liu, Y. (2013). Welfare and Poverty Impacts of India's National Rural Employment Guarantee Scheme Evidence from Andhra Pradesh. *Policy Research Working Paper*, (WPS6543), 1–31.

Djebbari, H., & Smith, J. (2008). Heterogeneous impacts in PROGRESA. *Journal of Econometrics*, *145*(1–2), 64–80.

Filmer, Deon and Louise Fox. 2014. Youth Employment in Sub-Saharan Africa. *Africa Development Series.* Washington, DC: World Bank.

Gilligan, D. O., Hoddinott, J., & Taffesse, A. S. (2009). The Impact of Ethiopia's Productive Safety Net Programme and its Linkages. *Journal of Development Studies*, *45*(10), 1684–1706.

Heckman, J. J., Smith, J., & Clements, N. (1997). Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *The Review of Economic Studies*, *64*(4), 487–535.

INS, AGEPE (2014). Enquête nationale sur la situation de l'emploi et du travail des enfants (ENSETE 2013).

Jalan, J., & Ravallion, M. (1999). Income Gains to the Poor from Workfare (No. 2149). *Policy Research Working Paper Series*.

Ravi, S., & Engler, M. (2015). Workfare as an Effective Way to Fight Poverty: The Case of India's NREGS. *World Development*, *67*, 57–71.

Rosas, N., & Sabarwal, S. (2016). Public Works as a Productive Safety Net in a Post-Conflict Setting : Evidence from a Randomized Evaluation in Sierra Leone (No. 7580). *Policy Research Working Paper Series*.

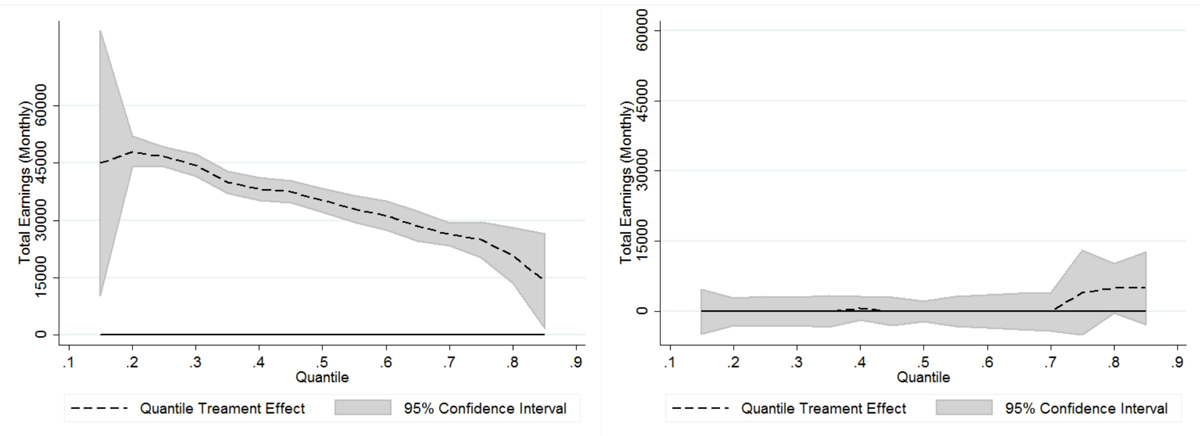
Wager, S., & Athey, S. (2015). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Working Paper*, 1–43.

World Bank. 2015. La force de l'elephant : pour que sa croissance génère plus d'emplois de qualité. Washington, D.C. : World Bank Group.

Table 1: Sample compared to Population

	PW Midline data Control group	National Labor Survey 2013 Youth 18-30, Urban areas
Employment status (primary occupation)		
Inactive population	7,8%	34,9%
Active population	92,2%	65,1%
Unemployed	6,0%	16,0%
Wage employment (including informal)	50,6%	25,1%
Self employment (non agricultural)	26,8%	39,2%
Self employment in agriculture	2,2%	4,7%
Others	14,0%	15,0%
Education (diploma)		
No diploma	47,5%	47,1%
CEPE (completed primary school)	22,8%	21,7%
BEPC (completed middle school)	16,8%	18,5%
BAC or more (completed secondary school)	12,1%	12,7%

Figure 1: Quantile Analysis of Treatment effects on Earnings (Total)



(a) *During* Program

(b) *Post* Program

Table 2: Balance checks and Summary statistics

	(1) Treatment group (pooled)	(2) Control group	(3) Test of (1)-(2) (p-value)	(4) Test between 4 Arms (p-value)	(5) Nb Obs.
Individual characteristics					
Live in urban area	94,2%	92,7%	0,14	0,22	4 099
Age	24,61	24,56	0,58	0,16	4 099
Nationality : Ivorian	95,5%	96,9%	0,04	0,08	4 099
Nb of children	0,80	0,82	0,70	0,90	4 099
Education					
Did not complete primary school (No diploma)	46,9%	48,1%	0,56	0,88	4 098
Has completed primary school (CEPE)	24,8%	22,8%	0,19	0,61	4 098
Has completed middle school (BEPC)	18,0%	16,6%	0,31	0,30	4 098
Has completed secondary school (BAC or +)	10,1%	12,3%	0,06	0,28	4 098
Is a student	4,9%	7,6%	0,00	0,04	4 099
Previous Vocationnal Training	38,8%	41,2%	0,16	0,11	4 096
Including : traditionnal / informal apprenticeship	73,2%	71,2%	0,52	0,67	1 613
Household					
Household size (total number of members)	6,03	6,05	0,92	0,52	4 097
Number of rooms	3,16	3,17	0,87	0,66	4 099
Nb of children (<18 ans)	2,00	1,96	0,51	0,77	4 099
Is head of household	24,8%	23,5%	0,46	0,85	4 099
Share of members working (last 7 days)	55,5%	55,9%	0,62	0,86	4 097
Household Assets (total number) (last 3 months)					
Total	13,86	13,85	1,00	0,18	4 099
Transport	0,73	0,79	0,26	0,11	4 099
Agriculture	4,71	4,64	0,91	0,25	4 099
Household Equipment	1,64	1,65	0,98	0,65	4 099
Communication	6,77	6,78	1,00	0,90	4 099
Savings					
Has Saved (last 3 months)	48,6%	49,8%	0,57	0,79	4 099
Share of formal savings (among those who saved)	25,5%	25,6%	0,57	0,94	1 988
Has a Savings Account	11,2%	10,4%	0,47	0,83	4 099
Savings Stock (FCFA)	28 777,05	26 843,99	0,37	0,83	4 042
Face Constraints to repay loans	19,9%	23,1%	0,03	0,22	4 099
Face Constraints to access credit	49,9%	49,8%	0,96	0,66	4 099
Constraints and expenditures					
Nb of days with no meals (last 7 days)	0,83	0,79	0,46	0,80	4 099
Highly constrained for basic needs expenditures	70,3%	73,8%	0,04	0,15	4 099
Transportation expenditure (last 7 days)	1 924,28	1 772,95	0,15	0,59	4 095
Communication expenditure (last 7 days)	1 731,15	1 623,06	0,36	0,88	4 092
Employment					
Has an activity	79,1%	80,5%	0,28	0,37	4 099
Searched for a job (last 6 months)	76,9%	78,9%	0,17	0,20	4 099
Risk and Time preferences					
Risk aversion level (scale 0 to 10, 0=very averse)	4,69	4,74	0,66	0,43	4 099
Is Risk averse (based on lotteries)	74,0%	71,6%	0,18	0,19	4 099
Patience level (scale 0 to 10, 10=very patient)	3,33	3,42	0,39	0,81	4 095
Preference for present (actualisation rate for 1 month)	0,57	0,57	0,95	0,46	4 099
Cognitive Skills (% success in answers or tasks at each test)					
Raven test (deduction)	23,4%	23,4%	0,92	0,09	4 093
NV7 test (spatial vision)	27,0%	26,4%	0,24	0,11	4 099
Dexterity (sorting nut test)	38,0%	37,4%	0,02	0,03	4 094
Dexterity (nut and bolt test)	33,4%	33,7%	0,21	0,26	4 083

Table 3: Overall results *during* and *post* program

	(1) Employed	(2) Wage Employed (in at least 1 activity)	(3) Self Employed (in at least 1 activity)	(4) Total Hours worked (weekly)	(5) Hours worked in Wage Empl. (weekly)	(6) Hours worked in Self Empl. (weekly)	(7) Total Earnings (monthly)	(8) Earnings in in Wage Empl. (monthly)	(9) Earnings in in Self Empl. (monthly)	(10) Total Expenditures (monthly)	(11) Savings (stock)	(12) Well Being index (z-score)	(13) Behavior index (z-score)
Panel A : *During* program effects (∼ 4,5 months after program starts)													
Public Works Treatment (ITT)	0.12***	0.44***	-0.09***	3.49***	14.04***	-6.71***	20885.31***	35385.33***	-12624.85***	15085.18***	39633.27***	0.20***	0.13***
	(0.01)	(0.02)	(0.02)	(1.19)	(1.21)	(0.89)	(6194.08)	(3699.69)	(4633.09)	(1552.68)	(3086.16)	(0.04)	(0.04)
LocXGender control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.86	0.53	0.33	40.93	22.90	12.19	60051.55	30916.20	25713.13	48043.04	21751.72	0.00	-0.00
Observations	2958	2958	2958	2958	2958	2958	2912	2912	2912	2945	2958	2934	2946
Panel B : *Post* program effects (pooled treatment) (12 to 15 months after program ends)													
Public Works Treatment (ITT)	0.01	0.01	0.01	1.21	-0.27	1.36	5621.62**	-972.88	6223.36***	1916.44	10833.24**	0.09**	0.00
	(0.01)	(0.02)	(0.02)	(1.29)	(1.18)	(1.14)	(2422.04)	(1347.95)	(2125.11)	(1503.10)	(4511.48)	(0.04)	(0.04)
LocXGender control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.87	0.55	0.33	42.27	24.13	13.23	48463.49	25352.70	19718.45	52227.70	54437.32	0.00	-0.00
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3814	3934	3932	3933
Panel C : *Post* program effects (treatment arms) (12 to 15 months after program ends)													
Public Works Treatment (ITT) (PW)	0.01	0.01	0.00	-0.45	-0.33	-0.28	4100.49	-244.47	3736.88*	1810.96	11097.87**	0.12***	0.05
	(0.02)	(0.02)	(0.02)	(1.60)	(1.59)	(1.26)	(2731.38)	(1635.02)	(2213.02)	(1668.30)	(5176.12)	(0.05)	(0.05)
Self-Empl. training (SET)	-0.00	-0.02	0.02	3.07	0.85	2.18	3426.76	-1588.54	6525.02**	-545.44	4776.90	-0.01	-0.06
	(0.02)	(0.03)	(0.03)	(1.90)	(1.91)	(1.54)	(3281.93)	(1676.43)	(3240.86)	(1510.02)	(6098.60)	(0.05)	(0.05)
Wage-Empl. training (WET)	-0.00	0.01	0.01	2.11	-0.66	2.90*	1324.74	-686.35	1249.14	865.97	-5561.65	-0.09**	-0.09*
	(0.02)	(0.02)	(0.03)	(1.78)	(1.65)	(1.63)	(3208.97)	(1543.17)	(3151.02)	(1737.47)	(5831.31)	(0.04)	(0.05)
LocXGender control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.87	0.55	0.33	42.27	24.13	13.23	48463.49	25352.70	19718.45	52227.70	54437.32	0.00	-0.00
p-value PW+SET=0	0.558	0.703	0.470	0.161	0.746	0.244	0.026	0.248	0.002	0.468	0.009	0.025	0.858
p-value PW+WET=0	0.524	0.507	0.644	0.301	0.506	0.109	0.102	0.583	0.118	0.172	0.347	0.433	0.459
p-value SET=WET	0.982	0.332	0.815	0.631	0.390	0.722	0.598	0.584	0.201	0.435	0.109	0.126	0.579
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3814	3934	3932	3933

Robust standard errors clustered at (large) brigade level

Earnings, Expenditures and Savings are in FCFA and winsorized at 99%. Hours also winsorized at 99%.

* $p < .1$, ** $p < .05$, *** $p < .01$

Figure 2: *During* Vs *Post* predicted impacts on Earnings (Causal Forests, Test sample)

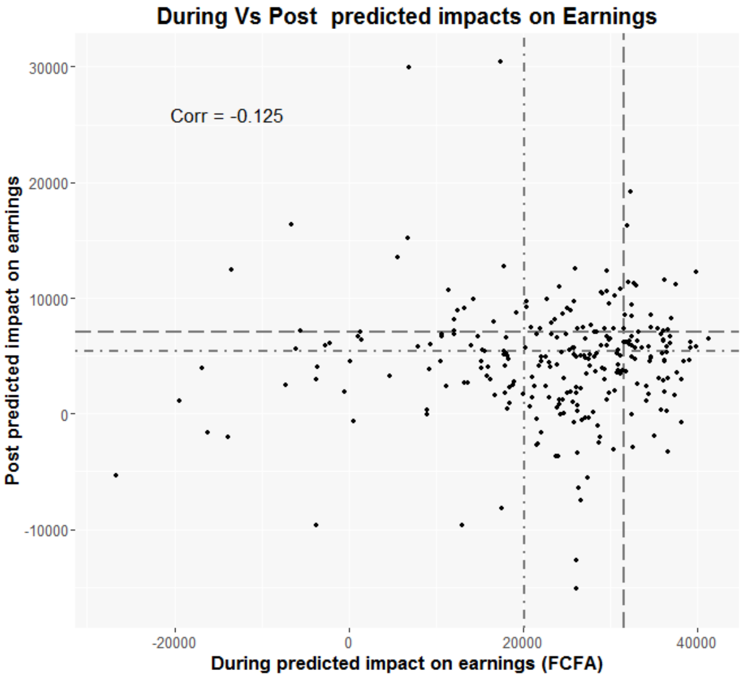


Table 4: Heterogeneity in *During* and *Post* program predicted impacts

	During program		**Post program**	
	(1)	(2)	(3)	(4)
	Total Earnings (monthly)	Well Being index (Z-score)	Total Earnings (monthly)	Well Being index (Z-score)
Panel A : Overall Predicted Impact with Athey - Imbens Causal Forests				
Mean (predicted CATE)	20 108	0,17	5 439	0,07
Standard Deviation (predicted CATE)	23 200	0,07	7 643	0,08
Comparison : ATE (ITT coeff.)	*20 885*	*0,20*	*5 621*	*0,09*
Panel B : By quartile of predicted *During* impact on Earnings				
Mean over Quartile 1 (0 to 25%)	-4 823	0,17	7 475	0,07
Mean over Quartile 2 (25 to 50%)	22 148	0,18	5 111	0,08
Mean over Quartile 3 (50 to 75%)	28 403	0,17	3 514	0,06
Mean over Quartile 4 (75 to 100%)	35 073	0,17	5 624	0,08
Panel C : By quartile of predicted *Post* impact on Earnings				
Mean over Quartile 1 (0 to 25%)	18 961	0,15	-1 568	0,06
Mean over Quartile 2 (25 to 50%)	21 524	0,19	3 727	0,09
Mean over Quartile 3 (50 to 75%)	22 038	0,19	6 050	0,08
Mean over Quartile 4 (75 to 100%)	17 928	0,16	13 648	0,07
Number of observations (Test Sample 10%)	273	273	273	273
Correlation *During* / *Post* program impacts	0.125	0.495		

Table 5: Summary of average impacts under alternative targeting approaches

	Random selection	Machine Learning Pred. cond. on X (Athey Imbens)	Self-selection	Selection based on baseline characteristics X		
	(0) Treated (ITT) (random 25%)	(1) Mean over 4^{th} quartile (test sample) (top 25%)	(2) Low reservation wage for PW	(3) Women	(4) Low baseline Earnings (self-declared) (bottom 25%)	(5) Low baseline Earnings (predicted) (bottom 25%)
During impact on Earnings	20 885***	34 789	32 076***	37 150***	32 695**	36 822***
(s.e)	(6 194)		(6 913)	(6 257)	(7 738)	(9 519)
N	2912	291	2 912	2 912	2 912	2 912
Post impact on Earnings (pooled)	5 622**	11 255	8 121***	8 259**	13 305*	8 845**
(s.e)	(2 422)		(2 511)	(3 556)	(4 381)	(4 506)
N	3934	391	3 934	3 934	3 934	3 934
During impact on Well Being	0,20***	0,26	0,19***	0,28***	0,13***	0,20***
(s.e)	(0,04)		(0,06)	(0,07)	(0,08)	(0,08)
N	2934	293	2 934	2 934	2 934	2 934
Post impact on Well Being (pooled)	0,09**	0,18	0,14***	0,11***	0,19***	0,08***
(s.e)	(0,04)		(0,06)	(0,07)	(0,09)	(0,08)
N	3 932	391	3 932	3 932	3 932	3 932

Column (2) : Outcome variable is a dummy for people who answered they would participate in the program if the wage was set to 1500 rather than 2500 FCFA / day.
45% willing to participate for lower daily wage. Variable measured at endline : balanced across treatment / control + robustness checks done using prediction on the control group.
Column (5) : Prediction using a restricted set of observable and non easily manipulated characteristics (gender, age, household characteristics and assets). Random forest algorithm.
For *Post* estimations (endline survey measure) treatment is pooled.
Standard errors in parentheses. Standard errors clustered at (large) brigade level for treated and individual level for control.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 6: Summary statistics Comparison

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Mean in treatment (reference)	Mean in 4th quartile of predicted impacts *during* program	Mean in 1st quartile of predicted impacts *during* program	P-value (1) - (2)	P-value (2) - (3)	P-value (1) - (3)	Mean in 4th quartile of predicted impacts *post* program	Mean in 1st quartile of predicted impacts *post* program	P-value (1) - (7)	P-value (7) - (8)	P-value (1) - (8)	P-value (2) - (7)	P-value (3) - (8)
Individual characteristics													
Gender	32,0%	37,7%	26,4%	0,00	0,00	0,00	28,0%	34,7%	0,01	0,00	0,05	0,00	0,00
Age	24,61	23,82	25,00	0,00	0,00	0,00	24,38	25,09	0,02	0,00	0,00	0,00	0,53
Nb of children	0,80	0,67	0,98	0,00	0,00	0,00	0,85	0,87	0,20	0,70	0,06	0,00	0,02
Live in urban area	94,2%	97,8%	91,7%	0,00	0,00	0,01	80,2%	99,4%	0,00	0,00	0,00	0,00	0,00
Nationality : Ivorian	95,5%	93,6%	97,9%	0,02	0,00	0,00	97,1%	95,7%	0,00	0,09	0,67	0,00	0,01
Education													
Did not complete primary school (No diploma)	46,9%	46,9%	49,2%	0,97	0,38	0,21	43,1%	50,3%	0,01	0,00	0,03	0,08	0,62
Has completed primary school (CEPE)	24,8%	22,0%	25,7%	0,06	0,10	0,58	21,7%	25,7%	0,01	0,04	0,50	0,88	0,99
Has completed middle school (BEPC)	18,0%	20,0%	16,6%	0,16	0,09	0,27	25,8%	12,9%	0,00	0,00	0,00	0,00	0,02
Has completed secondary school (BAC or +)	6,8%	7,8%	5,8%	0,32	0,13	0,23	7,3%	6,9%	0,56	0,73	0,95	0,63	0,32
Received Vocationnal Training	38,8%	35,0%	46,0%	0,03	0,00	0,00	37,4%	43,5%	0,35	0,01	0,00	0,26	0,25
Is a student	4,9%	7,8%	4,4%	0,00	0,01	0,52	8,4%	4,3%	0,00	0,00	0,33	0,54	0,88
Literate in French	64,7%	64,0%	60,6%	0,68	0,19	0,02	66,6%	62,7%	0,20	0,08	0,16	0,22	0,34
Employment													
Employed	0,79	0,78	0,82	0,33	0,02	0,01	0,84	0,80	0,00	0,07	0,28	0,00	0,23
Wage Employment (At least 1 activity)	0,34	0,35	0,38	0,64	0,21	0,02	0,32	0,38	0,15	0,01	0,02	0,16	0,76
Self-Employment (At least 1 activity)	0,36	0,32	0,41	0,01	0,00	0,00	0,43	0,38	0,00	0,02	0,15	0,00	0,19
Nb of Activities (total)	0,97	0,93	1,04	0,03	0,00	0,00	1,03	1,01	0,00	0,38	0,07	0,00	0,24
Nb of Activities in Wage Employment	0,38	0,39	0,42	0,64	0,27	0,03	0,35	0,42	0,05	0,01	0,02	0,09	0,88
Nb of Activities in Self Employment	0,40	0,35	0,46	0,01	0,00	0,01	0,49	0,43	0,00	0,02	0,19	0,00	0,22
Total Earnings (Monthly) (*)	17 954,84	17 851,23	20 319,38	0,93	0,15	0,04	19 263,56	19 493,94	0,18	0,87	0,10	0,27	0,57
Savings													
Savings Stock (FCFA)	28 777,05	25 107,65	34 534,77	0,08	0,00	0,01	40 750,89	25 252,33	0,00	0,00	0,04	0,00	0,00
Has Saved (last 3 months)	48,6%	47,6%	52,3%	0,57	0,07	0,04	52,9%	48,3%	0,01	0,05	0,85	0,01	0,07
Share if formal savings (among those who saved)	25,5%	24,9%	28,5%	0,73	0,19	0,10	25,8%	28,8%	0,85	0,21	0,04	0,67	0,91
Has a Savings Account	11,2%	11,0%	11,5%	0,86	0,77	0,81	12,3%	11,7%	0,26	0,69	0,62	0,34	0,88
Face Constraints to repay loans	19,9%	20,9%	20,3%	0,53	0,79	0,80	18,8%	23,8%	0,36	0,01	0,00	0,24	0,06
Face Constraints to access credit	49,9%	53,4%	46,4%	0,05	0,01	0,05	44,7%	48,9%	0,00	0,07	0,51	0,00	0,27
Constraints and expenditures													
Transportation expenditure (last 7 days)	1 924,28	1 550,28	2 066,61	0,00	0,00	0,24	2 263,61	1 711,92	0,00	0,00	0,01	0,00	0,01
Communication expenditure (last 7 days)	1 731,15	1 570,88	2 110,85	0,24	0,02	0,01	2 765,52	1 401,12	0,00	0,00	0,00	0,00	0,00
Nb of days with no meals (last 7 days)	0,83	0,94	0,65	0,09	0,00	0,00	0,79	0,85	0,45	0,47	0,73	0,04	0,00
Household													
Household size	6,03	4,91	7,48	0,00	0,00	0,00	6,53	5,25	0,00	0,00	0,00	0,00	0,00
Nb of children	2,00	1,62	2,57	0,00	0,00	0,00	2,15	1,73	0,03	0,00	0,00	0,00	0,00
Is head of household	24,8%	25,7%	23,3%	0,56	0,29	0,34	25,8%	31,0%	0,46	0,01	0,00	0,98	0,00
Nb members who did not complete primary school	1,20	0,99	1,47	0,00	0,00	0,00	1,33	1,08	0,01	0,00	0,00	0,00	0,00
Nb members who completed primary school	1,09	0,91	1,44	0,00	0,00	0,00	1,20	0,92	0,01	0,00	0,00	0,00	0,00
Nb members working	2,23	1,72	2,94	0,00	0,00	0,00	2,54	1,81	0,00	0,00	0,00	0,00	0,00
Household Assets (total number) (last 3 months)													
All Assets	13,86	11,05	19,54	0,00	0,00	0,00	16,35	11,85	0,00	0,00	0,00	0,00	0,00
Including Transportation Assets	0,73	0,25	1,46	0,00	0,00	0,00	1,04	0,58	0,00	0,00	0,00	0,00	0,00
Including Agricultural production Assets	4,71	4,56	6,45	0,80	0,02	0,00	6,61	3,58	0,00	0,00	0,00	0,00	0,00
Including Household equipment	1,64	1,14	2,53	0,00	0,00	0,00	1,56	1,65	0,08	0,23	0,93	0,00	0,00
Including Communication Assets	6,77	5,10	9,10	0,00	0,00	0,00	7,14	6,04	0,01	0,00	0,00	0,00	0,00
Preferences and Cognitive Skills													
Risk averse (based on lotteries)	74,0%	71,2%	0,71	0,09	83,7%	0,04	73,5%	72,9%	0,72	0,75	0,40	0,23	0,28
Raven test (% success in answers)	23,4%	23,0%	22,9%	0,55	0,89	0,39	22,9%	23,7%	0,37	0,39	0,67	0,90	0,34

Table 7: Cost-Effectiveness Analysis

	(1) *Survey measure* *During* Impact on Earnings (β^{During})	(2) *Survey measure* *Post* Impact on Earnings (β^{Post})	(3) Cumulated 'Impact' on Earnings (over 20 mths)	(4) *Administrative data* Total PW cost per beneficiary	(5) Cost Effectiveness Ratio (4) / (3)	(6) Nb of Years after the PW to be break-even
Current program (randomized)	20 885	4 100	206 695	660 478	3,2	23
Alternative Selections :						
(A) Low Reservation Wage for PW	31 926	7 409	332 790	660 478	2,0	6
(B) Women	37 314	7 998	379 191	660 478	1,7	4
(C) Low baseline Earnings (self-declared)	32 492	11 518	397 354	660 478	1,7	3
(D) Low baseline Earnings (predicted)	36 577	11 043	418 941	660 478	1,6	3
Comparison :						
Machine Learning prediction on (Athey Imbens) (top 25%)	34 789	11 358	411 076	660 478	1,61	3

(3) Discounted sum of impacts on total earnings from program starts (month 1) up to 13 months after exit from the program (month 20).
Computed as $\sum_{k=1}^{7}(\rho^{k-4} * \beta^{During}) + \sum_{k=8}^{20}(\rho^{k-13} * \beta^{Post})$, with β^{During} (respectively β^{Post}) the *during* (respectively *post*-program) ITT estimates of monthly earnings impact.
ρ is the monthly discount factor. $\rho = 1/(1+\delta)^{(1/12)}, \delta = 10\%$

(4) : Costs include both administrative implementation costs and the direct transfer made to beneficiaries.

(6) : Assuming (discounted) β^{Post} monthly impact after month 20, this is the number of years until total costs (4) equal cumulated impact on earnings (computed as (3)).

(A) : Group of individuals who answered they would participate in the program if the wage was set to 1500 rather than 2500 FCFA / day.

(C) : Bottom 25% of the distribution of baseline self-declared earnings.

(D) Bottom 25% of the distribution of baseline predicted earnings.

Prediction using a restricted set of observable and non easily manipulated characteristics (gender, age, household characteristics and assets). Random forest algorithm.

A Description of the weights

We describe in this appendix the weights used in the estimations. They depend on the survey used (midline or endline data), the treatment status, the locality. We provide a summary of the weights used with midline and endline data in A.5.

A.1 Randomization weights

A.1.1 First level of randomization : public lotteries

First, we consider weights related to the random selection into the program. Such weights should take into account the specificity of each public lottery held.

Our objective is to estimate an equation of the following type with weight w_i :

$$y_i = a + bT_i + u_i \tag{1}$$

One can easily check that estimator b is obtained as :

$$\widehat{b} = \frac{\sum_{i,T=1} w_i y_i}{\sum_{i,T=1} w_i} - \frac{\sum_{i,T=0} w_i y_i}{\sum_{i,T=0} w_i} \tag{2}$$

There are K different public lotteries [1] with N_k individuals participating to each lottery. Let's note N_{k1} the individuals from lottery k selected in the program (we will call them 'treated') and N_{k0} those who are not selected, with $N_k = N_{k1}+N_{k0}$. Among the N_{k0} , N_{k0s} are randomly selected to be surveyed and constitute the 'control group' of the experiment. The size of the population of lotteries' participants is N_P, with $N_P = \sum_k N_k = N_1 + N_0$. The size of the survey sample for the experiment is $N_E = \sum_k N_{k1} + N_{k0s} = N_1 + N_{0s}$.

We can rewrite equation (2) with weight w_{ki} ($i = 0_s; 1$ according to treatment status) assigned to individuals of the survey sample :

$$\widehat{b} = \frac{\sum_k w_{k1} N_{k1} \overline{y}^{k1}}{\sum_k w_{k1} N_{k1}} - \frac{\sum_k w_{k0s} N_{k0s} \overline{y}^{k0s}}{\sum_k w_{k0s} N_{k0s}} \tag{3}$$

To ensure this estimator can be interpreted as the weighted sum of the impact for each lottery, the following condition should hold :

$$w_{k0s} N_{k0s} = w_{k1} N_{k1} \tag{4}$$

The estimator rewrites as :

$$\widehat{b} = \sum_k \frac{w_{k1} N_{k1}}{\sum_k w_{k1} N_{k1}} \left(\overline{y}^{k1} - \overline{y}^{k0s} \right) \tag{5}$$

[1] In our experiment, public lotteries are held in each locality for each gender separately. Therefore $K = 16 * 2 = 32$

Following condition (4), we take $w_{k1} = 1$ and $w_{k0s} = N_{k1}/N_{k0s} \times N_0/N_1$. We obtain the following estimator:

$$\widehat{b} = \sum_k \frac{N_{k1}}{\sum_k N_{k1}} \left(\overline{y}^{k1} - \overline{y}^{k0s}\right) \tag{6}$$

Note that it means that a higher weight is put on lotteries for which the treated to not-treated ratio is higher[2]. It will be homogeneous across lotteries k (localities x gender) by construction of the survey sample. Note that these weights depend on the quota of treated (N_{k1}) that was assigned to each locality when the program was initially designed. It should be proportionate to the number of disadvantaged youth looking for employment, but such precise figures where not available at the time of design.

A.1.2 Second level of randomization : treatment arms assignment

Then, we want to add weights to take into account the second level of randomization in the experiment : assignment of treated individuals into 3 treatment arms denoted T_a, T_b and T_c. This is relevant when comparing treatment effects across arms, using endline survey data.

Similar to A.1.1, the equation estimated as the following form with a weight w_i:

$$y_i = \alpha + \beta_1 * T_{a,i} + \beta_2 * T_{b,i} + \beta_3 * T_{c,i} + u_i \tag{7}$$

As previously in, β_j estimators ($j = a, b, c$ for 3 options) are :

$$\widehat{\beta_j} = \frac{\sum_{i,T=j} w_i y_i}{\sum_{i,T=j} w_i} - \frac{\sum_{i,T=0} w_i y_i}{\sum_{i,T=0} w_i} \tag{8}$$

We use the same notations as A.1.1 for N_P (whole population), $N_E = N_1 + N_{0s}$ (survey sample population), N_{k1} and N_{k0s} [3].

Brigades of treated individuals (N_1) are assigned across 3 options T_a, T_b and T_c. The number of brigades assigned to the treatment arms varies by locality. We use the following notation : $N_k = N_{a,k} + N_{b,k} + N_{c,k} + N_{0,k}$ with $N_{1,k} = N_{a,k} + N_{b,k} + N_{c,k}$, and $N_P = \sum_k N_k = N_0 + N_a + N_b + N_c$ with $N_1 = N_a + N_b + N_c$.

We put a weight $w_{j,k}$ to treated individuals from lottery k who were assigned to treatment T_j, and weight w_{k0s} for non treated individuals selected in the survey sample. Similar to 3 with subscript $j = a, b, c$, (8) rewrites as :

$$\widehat{\beta_j} = \frac{\sum_k w_{j,k} N_{j,k} \overline{y}^{j,k}}{\sum_k w_{j,k} N_{j,k}} - \frac{\sum_k w_{k0s} N_{k0s} \overline{y}^{k0s}}{\sum_k w_{k0s} N_{k0s}} \tag{9}$$

[2] One could have chosen another option for the weight : $w_{k1} = N_k/N_{k1} \times N_1/N_P$ and $w_{k0s} = N_k/N_{k0s} \times N_0/N_P$. In that case, there will be a higher weight for lotteries where the demand for the program was higher (compared to the quota assigned).

[3] For endline survey, quantity $N_{k0s}$was increased compared to baseline. It affects weights computation through N_{k0s} and N_{0s} but other quantities remain unchanged.

To ensure this estimator can be interpreted as the weighted sum of the impact for each lottery, we need a condition similar to (4) and if this holds, the estimator can be written as :

$$\widehat{\beta}_j = \sum_k \frac{w_{j,k} N_{j,k}}{\sum_k w_{j,k} N_{j,k}} \left(\overline{y}^{j,k} - \overline{y}^{k0s} \right) \tag{10}$$

Similar to A.1.1, we choose the following weights :

- $w_{j,k} = N_{k1}/N_{j,k} \times N_j/N_1$ with $j = a, b, c$ [4]
- $w_{k0s} = N_{k1}/N_{k0s} \times N_0/N_1$

A.2 Sub-sampling weights (midline survey only)

The sample for midline survey ('during' program) includes the control group (N_{0s}) and a sub-sample of the treatment group.

Let's consider that we draw a random sub sample of group l in proportion $P_l = N_l^S/N_l$ (S the drawing variable). Original weights have to be multiplied by S/P_l to take sub-sampling into account.

Therefore, in group $l = k, 1$ for which one draws N_{k1}^S individuals out of N_{k1}, the original weight w_{k1} is updated to $\omega_{k1}^S = w_{k1} \times N_{k1}/N_{k1}^{S_{k1}}$. The weights for the control units, w_{k0s}, are unchanged as there is no sub-sampling of this group for midline survey.

A.3 Control group and potential ex-post enrolment in the program (endline survey only)

Individuals from control group need specific weights when using endline data, because some of them have been able to participate to the following waves of the program[5]. More precisely, control units were allowed to apply to wave 3 (apply meaning to take part to the lotteries) and wave 4. Such behavior could be tracked using administrative data. At the end of the fourth wave of THIMO program, each individual from the control group were in one the following 7 situations (for each locality) :

1. The individual applied to wave 3 (C_3), was selected as 'beneficiary' of wave 3 after public lotteries (T_3) and was therefore not allowed to apply to wave 4 ($\bar{C}_4$). This group is noted $C_3T_3\bar{C}_4$.
2. The individual applied to wave 3 (C_3), was not selected after public lotteries ($\bar{T}_3$), applied to wave 4 (C_4) and was selected as 'beneficiary' of wave 4 after lotteries (T_4). This group is noted $C_3\bar{T}_3C_4T_4$.

[4]Note : $\sum_j w_{j,k} = w_{k1} = 1$, which is the weight used for midline data when there is only one treatment group.

[5]Recall that the study exploits wave 2 (out of 4 waves) of the THIMO program

3. The individual applied to wave 3 (C_3), was not selected after public lotteries ($\bar{T}_3$), applied to wave 4 (C_4) and was not selected after lotteries ($\bar{T}_4$). This group is noted $C_3\bar{T}_3C_4\bar{T}_4$.
4. The individual applied to wave 3 (C_3), was not selected after public lotteries ($\bar{T}_3$) and he did not apply to wave 4 ($\bar{C}_4$). This group is noted $C_3\bar{T}_3\bar{C}_4$.
5. The individual did not apply to wave 3 ($\bar{C}_3$), applied to wave 4 (C_4) and was selected as 'beneficiary' of wave 4 after public lotteries (T_4). This group is noted $\bar{C}_3C_4T_4$.
6. The individual did not apply to wave 3 ($\bar{C}_3$), applied to wave 4 (C_4) and was not selected after public lotteries ($\bar{T}_4$). This group is noted $\bar{C}_3C_4\bar{T}_4$.
7. The individual did not apply to wave 3 ($\bar{C}_3$), and did not apply to wave 4 ($\bar{C}_4$). This group is noted $\bar{C}_3\bar{C}_4$.

This idea is that we don't want to include in the estimations control units that have benefited from further waves of the program (waves 3 and 4), which are precisely groups $C_3T_3\bar{C}_4$, $C_3\bar{T}_3C_4T_4$ and $\bar{C}_3C_4T_4$ following the notations introduced before. To compensate for that, we want to put a higher weight on individuals who had the exactly same behavior (towards wave 3 and 4) but were (randomly) not selected into the program. We introduce a new multiplicative weight for control units ($\widetilde{w}_{k0s,j}$) to control for that.

Intuitively, if there had been only one wave at which individuals could apply (C) and be selected (T), weights would have been :

- $\widetilde{w}_{k0s,\bar{C}} = 1$
- $\widetilde{w}_{k0s,C,\bar{T}} = N_{k0s,C} / N_{k0s,C,\bar{T}}$
- $\widetilde{w}_{k0s,C,T} = 0$

Taking into account the two waves, the weights follow the seven groups previously defined :

- $\widetilde{w}_{k0s,C_3T_3\bar{C}_4} = 0$
- $\widetilde{w}_{k0s,C_3\bar{T}_3C_4T_4} = \frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}} \times 0 = 0$
- $\widetilde{w}_{k0s,C_3\bar{T}_3C_4\bar{T}_4} = \frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}} \times \frac{N_{k0s,C_3\bar{T}_3C4}}{N_{k0s,C_3\bar{T}_3C_4\bar{T}_4}}$
- $\widetilde{w}_{k0s,C_3\bar{T}_3\bar{C}_4} = \frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}} \times 1 = \frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}}$
- $\widetilde{w}_{k0s,\bar{C}_3C_4T_4} = 1 \times 0 = 0$
- $\widetilde{w}_{k0s,\bar{C}_3C_4\bar{T}_4} = 1 \times \frac{N_{k0s,\bar{C}_3C4}}{N_{k0s,\bar{C}_3C_4\bar{T}_4}}$
- $\widetilde{w}_{k0s,\bar{C}_3\bar{C}_4} = 1$

One can easily check that the sum of weights gives the total number of individuals in control group [6]

[6] $(\frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}} \times N_{k0s,C_3\bar{T}_3C_4} + \frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}} \times N_{k0s,C_3\bar{T}_3\bar{C}_4}) + (N_{k0s,\bar{C}_3C_4} + N_{k0s,\bar{C}_3\bar{C}_4})$ That is : $N_{k0s,C_3} + N_{k0s,\bar{C}_3} = N_{k0s}$

A.4 Tracking weights

We want to add a weight taking into account the differential response rate of individuals during each survey (midline and endline. More precisely, one can consider that a given survey consists in two phases a and b :

- The main data collection phase (a), during which the response rate is $R_{a,j}$ for group $j = 1, 0$.
- An additional tracking phase (b), targeting attritors from the first phase. We note $R_{b,j}$ the response rate of the tracking phase for group $j = 1, 0$.

To determine the tracking sample, we first define a sub-sample of 'eligible' attritors[7] $E_{b,j}$ from which a random sub-sample will be drawn in proportion $\pi_j = NE^S_{b,j}/NE_{b,j}$ (j is an index for treatment status x locality).

Individuals who responded (only) during the tracking phase should have a different weight than those who responded during the main survey phase. To take this selection into account, tracking respondants should be weighted by $\omega^T_j = (R^S_{a,j} + \lambda_j s_j R^S_{b_j}(1 - R^S_{a,j})E^S_{b,j}$, with λ_j to be determined, so final weight is $\omega^{S,f}_j = \omega^S_j \times \omega^T_j$.

The sum of the weights on population j is therefore : $\omega_j \times (N^S_{a,j} + \lambda_j NER^S_{s,b_j})$, with NER^S_{s,b_j} the number of individuals who responded during tracking phase (in group S drawn). We make the hypothesis that residual non response $R^S_{b,j}$ is random. The population for which we want to be representative is the respondent population of phase a and the respondent population of phase b. This lead us to take $\lambda_j = NE^S_{b,j}/NER^S_{s,b,j}$

In group j, weights will be set such as [8] :

- $\omega^S_j \times 1$ for phase a respondents
- $\omega^S_j \times NE^S_{b,j}/NER^S_{s,b,j}$ for phase b respondents

Table 8 is a summary for the use of tracking weights. Tracking weights ω^T_j are multiplied to the previous weights.

[7]Among the attritors of phase (a) some individuals are considered 'non eligible' for tracking in order to exclude them from the tracking draw. Non eligible attritors are those considered to be impossible or quasi impossible to reach (which is why we don't want to put extra effort on them) : dead individuals, individuals who migrated to another country, (for endline) individuals who were already impossible to find at baseline 1.5 years ago.

[8]In theory, ω_j should be adjusted so that it does not use correction N_j/N^S_j but rather the correction corresponding to the total of eligibles $N_{a,j} + NE_{b,j}$. However, this number is only known for selected units $S_j = 1$. Therefore we will ignore this aspect, which is fair considering that units where randomly drawn. Finally, it means that we estimate the unknown amount $N_{a,j} + NE_{b,j}$ by $N^S_{a,j} + NE^S_{b,j} \times N_j/N^S_j$

	Eligible for tracking (b)	**Non eligible for tracking (b)**
Respondent during main survey phase (a)	1	1
Non respondent during main survey phase (a)	$NE^S_{b,j}/NER^S_{s,b,j}$ if respondent in phase (b) 0 otherwise	0

Table 8: Weights adjusted as a function of (1) response rate at the end of main survey phase and (2) being eligible for the tracking phase

A.5 Synthesis of the weights used for midline and endline data

Randomization weights w_k		**Sub Sampling weights** ω_k^S		**Tracking weights** ω_j^T	
Application criterion	Weight computation	Application criterion	Weight computation	Application criterion	Weight computation
Treated	$w_{k1} = 1$	Treated	$w_{k,1}^S = N_{k1}/N_{k1}^{S_{k1}}$, k=locality	Respondents main phase ($R_a = 1$)	$\omega^T = 1$
Control	$w_{k0s} = N_{k1}/N_{k0s} \times N_0/N_1$, k=locality x gender	Control	$w_{k,0}^S = 1$	Non Respondents main phase ($R_a = 0$)	$\omega_j^T = NE_{b,j}^S/NER_{s,b,j}^S$ if respondent in tracking phase ($E_b = 1$ et $R_b = 1$), j=locality x treatment status
					$\omega^T = 0$ if non respondent (but sampled) in tracking phase ($E_b = 1$ et $R_b = 0$)
					$\omega^T = 0$ if not sampled for tracking phase ($E_b = 0$)
Final weight: $w_{k,i}^F = w_{k,i} \times \omega_{k,i}^S \times \omega_{k,i}^T$, $i = 0, 1$ (treatment status), $k \in [\![1, 32]\!]$ (locality x gender)					

Table 9: Synthesis of the weights used with midline data

Randomization weights $w_{j,k}$		**Post-enrollment weights $\widetilde{\omega}_{k,j}$**		**Tracking weights ω_j^T**	
Application criterion	Weight computation	Application criterion	Weight computation	Application criterion	Weight computation
Treatment arm T_a, T_b or T_c	$w_{j,k} = N_{k1}/N_{j,k} \times N_j/N_1$, $j = a, b, c$	Selected to participate to wave 3 or 4 (groups $C_3T_3\bar{C}_4$, $C_3\bar{T}_3C_4T_4$ et $\bar{C}_3C_4T_4$)	0	Respondents main phase ($R_a = 1$)	$\omega^T = 1$
Control	$w_{k0s} = N_{k1}/N_{k0s} \times N_0/N_1$, k=locality x gender	Group $C_3\bar{T}_3C_4\bar{T}_4$	$\frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}} \times \frac{N_{k0s,C_3\bar{T}_3C4}}{N_{k0s,C_3\bar{T}_3C_4\bar{T}_4}}$	Non Respondents main phase ($R_a = 0$)	$\omega_j^T = NE^S_{b,j}/NER^S_{s,b,j}$ if respondent in tracking phase ($E_b = 1$ et $R_b = 1$), j=locality x treatment status
		Group $C_3\bar{T}_3\bar{C}_4$	$\frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}}$		
		Group $\bar{C}_3C_4\bar{T}_4$	$\frac{N_{k0s,\bar{C}_3C4}}{N_{k0s,\bar{C}_3C_4\bar{T}_4}}$		$\omega^T = 0$ if non respondent (but sampled) in tracking phase ($E_b = 1$ et $R_b = 0$)
		Group $\bar{C}_3\bar{C}_4$	1		$\omega^T = 0$ if not sampled for tracking phase ($E_b = 0$)
Final weight: $w^F_{k,i} = w_{j,k} \times \widetilde{\omega}_{i,l} \times \omega^T_{k,i}$, $j = 0, a, b, c$ (treatment status), $i = 1, 0s$, l post-enrollment group, $k \in [\![1, 32]\!]$ (locality x gender)					

Table 10: Synthesis of the weights used with endline data