

Do Workfare Programs Live Up to Their Promises?

Experimental Evidence from Côte d'Ivoire

Marianne Bertrand

Bruno Crépon

Alicia Marguerie

Patrick Premand



WORLD BANK GROUP

Development Economics

Development Impact Evaluation Group

April 2021

Abstract

Workfare programs are one of the most popular social protection and employment policy instruments in the developing world. They evoke the promise of efficient targeting, as well as immediate and lasting impacts on participants' employment, earnings, skills and behaviors. This paper evaluates contemporaneous and post-program impacts of a public works intervention in Côte d'Ivoire. The program was randomized among urban youths who self-selected to participate and provided seven months of employment at the formal minimum wage. Randomized subsets of beneficiaries also received complementary training on basic entrepreneurship or job search skills. During the program, results show limited impacts on the likelihood of employment, but a shift toward wage jobs, higher

earnings and savings, as well as changes in work habits and behaviors. Fifteen months after the program ended, savings stock remain higher, but there are no lasting impacts on employment or behaviors, and only limited impacts on earnings. Machine learning techniques are applied to assess whether program targeting can improve. Significant heterogeneity in impacts on earnings is found during the program but not post-program. Departing from self-targeting improves performance: a range of practical targeting mechanisms achieve impacts close to a machine learning benchmark by maximizing contemporaneous impacts without reducing post-program impacts. Impacts on earnings remain substantially below program costs even under improved targeting.

This paper is a product of the Development Impact Evaluation Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at marianne.bertrand@chicagobooth.edu, bruno.crepon@ensae.fr, amarguerie@worldbank.org and ppremand@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Do Workfare Programs Live Up to Their Promises?

Experimental Evidence from Côte d'Ivoire *

Marianne Bertrand (University of Chicago, Booth School of Business)

Bruno Crépon (CREST)

Alicia Marguerie (World Bank)

Patrick Premand (World Bank)

JEL classification: H53, I38, J24, O12, C93.

Keywords: Workfare, Public Works, Employment, Training, Heterogeneity, Machine Learning, Targeting, Self-Targeting.

* This paper is the result of a collaboration with the government of Côte d'Ivoire. We are very thankful to Adama Bamba, Hermann Toualy, Ismahel Abdoul Barry and Fabrice Konan at BCPE; Marius Pokou, Martin Kouakou, Yves N'Cho and Kassi Ernest Bohoussou at AGEROUTE; as well as Hamoud Wedoud Abdel Kamil at the World Bank. Funding from the MESF and DIME i2i Trust Funds at the World Bank are gratefully acknowledged. We thank Sondo Eloi Somtinda for outstanding field coordination throughout the study; Hugues Kouadio, Rosine Addy Mosso, Marie Judith Soro and Nathaniel Gbenro at ENSEA for leading baseline and midline data collection; as well as Jean Awé and Ismahel Abdoul Barry at BCPE for leading endline data collection. Horacio Vera Cossio, Morgane Hoffmann and Bertille Picard provided excellent research assistance and contributions, in particular on machine learning applications. Morgane Hoffmann drafted the machine learning appendix. We are deeply indebted to Susan Athey, Stefan Wager, Jonathan Davis and Jann Spiess for their suggestions on machine learning implementation. We thank Lori Beaman, Christopher Blattman, Stefanie Brodmann, Deon Filmer, Emanuela Galasso, Jessica Goldberg, Rema Hanna, Clément Imbert, Michael Knaus, Michael Lerner, William Parienté, Anthony Strittmatter, Dominique Van de Walle and seminar participants at CREST, UChicago, CSAE, ILO, IZA, St-Gallen and the World Bank for useful comments and suggestions. Computational reproducibility has been verified by DIME analytics. The findings, interpretations, and conclusions of the paper are those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

1 Introduction

Public works programs are an important instrument in the portfolio of policy makers trying to address the social challenges of unemployment, underemployment and poverty. They offer temporary employment, typically remunerated at the minimum wage or below, for the creation of public goods, such as roads or infrastructure. In Sub-Saharan Africa, our context in this paper, labor-intensive public works programs have been adopted in response to transient negative shocks such as those induced by economic downturns, climatic shocks or episodes of violent conflicts, and often aim to offer public employment as a stabilization instrument.

While traditional welfare programs, such as cash transfers, could also be used to support the poor and vulnerable, workfare programs have some theoretical advantages that could make them superior poverty alleviation tools than welfare programs, both in the short- and medium-run. A first stated advantage of workfare programs, as highlighted by [Besley and Coate \(1992\)](#) or [Ravallion \(1991\)](#), is that they can in principle solve the difficult problem of targeting. The targeting of social protection programs is particularly complex in low-income countries because of a lack of robust data, challenges in identifying beneficiaries at the bottom of the welfare distribution, as well as weak systems and institutions, leading to potential errors of inclusion or exclusion. Public works programs very often rely on self-targeting to select transfer beneficiaries, based on the idea that only the more disadvantaged would be willing to supply labor. How well self-targeting works in reaching the most vulnerable will, however, depend on how broadly (or narrowly) appealing the public works option is (accounting for earnings and disutility of work) compared to the alternative. In environments such as those in many developing countries where a large fraction of the population is under-employed in informal work paying below the legal minimum wage, workfare programs could appeal to a broad cross-section of the population and thus could fail to appropriately target transfers towards the most vulnerable.

A second stated advantage of workfare programs is that they may have longer-lasting positive impacts on individual beneficiaries than standard welfare programs. Regular work, even if unpleasant, may improve skills, behaviors, work habits, or well-being (such as self-esteem or mental health). Particularly relevant to post-conflict environments, engaging beneficiaries in cash-for-work rather than providing cash transfers may also operate as a social stabilization tool through an incapacitation effect: time spent working may displace risky behaviors or socially disruptive activities. All of this may translate into lasting behavioral changes that may increase labor

productivity, such as by building regular work habits that might be especially difficult to learn outside of regular formal employment. Furthermore, through skill development or the signaling value of prior work experience, public works may increase the future employability or productivity of the beneficiaries. Such longer-term benefits can potentially be further enhanced by adding complementary productive interventions, such as savings facilitation or training, to the workfare experience. Whether post-program benefits of participation in public works on employment, behaviors or skills exist, however, largely remains an empirical question. And if such post-program benefits exist, it is unclear whether these benefits are tied to the distinctive “work” part of workfare programs (such as changes in work habits or behaviors), or whether they derive from more generic income support (such as by enabling saving and investing in productive activities), a channel that would also extend to traditional welfare programs. In addition, and of central interest to this paper, if such post-program benefits exist, it is unknown whether there is a trade-off between the shorter-term objectives of public work programs (better targeting cash transfers towards the most vulnerable) and any such longer-term benefits.¹

This paper assesses the extent to which these two main promises of public works programs hold in practice, namely that self-targeting is efficient and that the programs induce immediate and lasting impacts. As such, the paper makes two main contributions. First, we analyze both the contemporaneous and post-program impacts of a randomized public work program on participants’ employment, earnings and behaviors. Second, we leverage machine learning techniques to study the heterogeneity of program impacts, which is key to assess whether departing from self-targeting would improve program effectiveness. Using machine learning as a benchmark, we analyze how program performance would change under alternative self-targeting and targeting approaches.

The public works program we study was implemented by the Côte d’Ivoire government in the aftermath of a post-electoral crisis 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 low-skilled young (18-30) men and women in urban or semi-urban areas who were unemployed or underemployed, as well as to develop their skills through work experience and

¹There are other stated advantages of workfare programs. Workfare programs contribute to the creation or maintenance 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 because of a crisis (e.g. climatic shocks or violent conflict). 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 disincentivize work) as well as a desire to generate immediate visible improvements to employment conditions.

complementary training. Participants in the public works program were employed for a period of 7 months to rehabilitate and clean road infrastructure. Program participants worked 6 hours per day, 5 days per week and were remunerated at the statutory minimum daily wage, corresponding to about \$10 PPP 2014 per day (CFA 2,500), or approximately \$223 PPP 2014 per month (CFA 55,000).²

All young men and women in the required age range and residing in one of 16 urban localities in Côte 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, randomized subsets of beneficiaries were also offered (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 preparation, interview skills, etc.). We carried out rich surveys of youth in the treatment and control groups at baseline, during the program (4 to 5 months after the program had started), and 12 to 15 months after the program ended.

Our results on contemporaneous impacts demonstrate that the program had limited effects on the likelihood of employment, but induced shifts in the composition of employment. 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 earnings are about CFA 27,083 higher in the treatment group, from a base of CFA 42,841 in the control group. While the program lifted earnings, foregone earnings are quantitatively large, with earning gains representing only about 53 percent of the transfer.³ The intervention increased savings and well-being. It also induced changes in work habits and behaviors in the short-term.

Twelve to 15 months after program completion, we do not find impacts on the likelihood of employment, hours worked or the composition of employment (salaried work vs. self-employment). While we find some post-program impacts on earnings, which mostly stem from self-employment activities, they are small and not always robust to alternative specifications. Savings stock and well-being remain higher, but there are no lasting impacts on work habits or behaviors.

²We use an exchange rate of USD 1=XOF 493.757 (official average exchange rate in 2014 (from IMF)) to convert CFA francs to US dollars. We use the PPP conversion factor, GDP (LCU per international \$) of 246.519 for Côte d'Ivoire in 2014 (from the World Bank).

³Datt and Ravallion (1994) and Jalan and Ravallion (2003) estimate the net income gains from public works programs in India and Argentina, finding foregone income ranging between 30% and 50%.

Since our results show that self-targeting based on the formal minimum wage failed in this context, how much would results improve if the offered wage was lowered or if the targeting criteria were adjusted? We use the distribution of predicted impacts derived from machine learning methods to answer both of these questions. First, we find that lowering the wage below the formal minimum wage would not improve program performance. As we explain in a simple theoretical framework, this is because the improvement in self-targeting is offset by lower transfer amounts and because the improvement in self-targeting is itself limited in our empirical context given the small concentration of “marginal” applicants with small predicted program impacts.

Second, we show that there are substantial differences in predicted impacts across participants during the program. The average impact on earnings in the short-term for the top 25% of the predicted distribution (upper quartile) is approximately 2.2 times more than for the bottom 25% (lower quartile). In contrast, we do not detect heterogeneity in post-program impacts on earnings and find no evidence of systematic correlation (positive or negative) between short-term and long-term impacts. We document the characteristics of those who benefit the most from public works, and assess how alternative targeting rules based on these characteristics would improve program effectiveness. Compared to the benchmark scenario with self-targeting based on the formal minimum wage, the cost-effectiveness ratio would improve by 30% to 52% by targeting women only, or by targeting youths with low predicted baseline earnings. In the end, direct impacts on youths’ earnings during and after the program remain substantially below program costs, especially in light of large administrative costs. The cost per participant is about 2 times the estimated direct impacts on earnings up to the endline (12 to 15 months after the program) even under improved targeting.

Despite the popularity of public works programs, experimental evidence on their overall effectiveness remains relatively limited (Subbarao et al., 2012; Gehrke and Hartwig, 2018). Existing evidence mostly comes from quasi-experimental studies on a small number of influential programs, especially from India (Murgai et al., 2015; Imbert and Papp, 2015, 2019; Muralidharan et al., 2016, 2017). Most evidence focuses on short-term economic impacts during the intervention,⁴ and in

⁴Several papers have assessed the role of public works as a short-term safety net or insurance mechanism providing temporary employment and income to vulnerable populations during lean agricultural seasons or after economic shocks. Findings regarding program impacts on welfare and food security are mixed. Galasso and Ravallion (2004) document how a workfare program in Argentina attenuated the negative welfare effects of an economic crisis, and Ravi and Engler (2015) find beneficial impacts of India’s workfare scheme on consumption and food security. On the other hand, Beegle et al. (2017) do not find significant effects on food security in Malawi in one of the few randomized control trial of a public works program so far. Gilligan et al. (2009) also find limited effects of the Ethiopia PNSP program, although households who received larger transfer amounts did

fewer cases impacts on risky behaviors related to conflict or violence (Fetzer, 2020; Amaral et al., 2015). The effectiveness of workfare programs largely depends on whether they have productive impacts (Murgai et al., 2015), and our paper relates to a small literature that assesses whether beneficiaries of public works programs find pathways towards more productive post-program employment in wage jobs or in the informal sector. Ravallion et al. (2005) do not find significant impacts on post-program earnings in Argentina. Alik-Lagrange et al. (2017) find some persistent effects in rural areas of Colombia and suggest participants acquired new skills. Rosas and Sabarwal (2016) document investments from public works beneficiaries in assets and micro-enterprises in Sierra Leone, and Deininger et al. (2016) in agriculture in India. A few studies analyze the effectiveness of complementing public works programs with training or savings facilitation, including Galasso et al. (2004) and Almeida and Galasso (2010). Gilligan et al. (2009) report impacts of the Ethiopia public works program combined with agricultural support on adoption of agricultural technologies and off-farm small businesses. To our knowledge, no study has analyzed whether there are trade-offs between maximizing contemporaneous and post-program benefits from public works.

Our paper further complements the literature on the targeting of social programs, a topic that has garnered substantial policy and research interest (for recent reviews, see Hanna and Olken (2018); Banerjee et al. (2019); Gentilini et al. (2020)). Targeting experiments have predominantly tested how best to rank households in terms of poverty in the context of cash transfer programs, mainly contrasting community-based approaches and statistical methods such as proxy means testing (e.g. Alatas et al. (2012, 2016); Premand and Schnitzer (2020)). Questions have been raised as to whether there are trade-offs between selecting the poorest and maximizing program impacts (Basurto et al., 2020). Earlier studies of workfare programs focused on analyzing the profiles of beneficiaries and benefit incidence patterns.⁵ Following the seminal work of Manski (2004), targeting has also been studied as a statistical decision problem. The approach aims to derive an assignment rule that maximizes welfare after program implementation. Heterogeneous treatment effects are estimated before being plugged into a social welfare objective function to derive the optimal assignment. This is the approach followed by Bhattacharya and Dupas (2012)

see improvements in some measures of food security. A few studies have estimated the impact of public works programs on school enrollment and child labor (Li et al., 2013; Islam and Sivasankaran, 2015; Shah and Steinberg, 2019), also with mixed results.

⁵See for instance Ravallion et al. (1993); Datt and Ravallion (1994); Jalan and Ravallion (2003); Alik-Lagrange and Ravallion (2018).

in a context, close to ours, in which the program allocation involves a budget constraint.⁶ In this paper, we use machine learning techniques as a benchmark to assess alternative targeting and self-targeting options.

On the methodological front, our paper relates to a growing literature applying machine learning to analyze treatment heterogeneity. Since the influential contribution of [Athey and Imbens \(2016\)](#), several recent papers have explored the application of these techniques.⁷ In our application, (i) we assess the extent of heterogeneity in program impacts on earnings, (ii) we assess how a reduction of the offered wage would affect performance based on the distribution of predicted impacts, (iii) we document the profile of individuals with highest predicted impacts, (iv) we assess whether there are trade-offs between maximizing contemporaneous and post-program impacts, (v) we illustrate how machine learning can be used to explore mechanisms for impacts, and (vi) we compare how alternative targeting rules compare to machine learning estimates. One important aspect of our application is that we combine machine learning techniques with the statistical framework for inference developed in [Chernozhukov et al. \(2020\)](#), of which [Breda et al. \(2020\)](#) offer another application.

The rest of the paper proceeds as follows. Section 2 outlines a framework to analyze how self-targeting induces heterogeneity in public works impact. Section 3 describes the intervention design and data. Section 4 presents the empirical strategy and machine learning approach. Section 5 presents results on average impacts, and Section 6 on heterogeneity. Section 7 concludes. Additional material, tables and figures are presented in the Appendix.

2 Framework

2.1 Contemporaneous Impacts: A Simple Framework

In this section, we present a simple framework that clarifies how: 1) one should expect heterogeneity of program impacts under self-targeting; 2) alternative targeting approaches may improve

⁶[Kitagawa and Tetenov \(2018\)](#) develop an alternative approach that avoids the intermediate step of estimating heterogeneous treatment effects and directly identifies the assignment rule. See also [Athey and Wager \(2017\)](#) for an application to observational studies.

⁷For a review, see [Knaus et al. \(2020b\)](#), as well as applications in a variety of context including marketing ([Ascarza, 2018](#)), cash grants for firms ([McKenzie and Sansone, 2019](#)), employment programs ([Knaus et al., 2020a](#)), financial work incentive programs ([Strittmatter, 2018](#)), summer employment programs for disadvantaged youth ([Davis and Heller, 2017, 2020](#)), or role models and educational choices ([Breda et al., 2020](#)).

program effectiveness; and 3) variation in the offered program compensation will impact self-selection and program effectiveness. Appendix Section C provides additional details.

Let θ denote potential participants' hourly earnings in non-program activities, h hours worked, and $c(h)$ the additive disutility of working h hours. Absent the workfare program, individuals have an optimal number of hours worked $h_0(\theta)$ given by $\theta = c'(h_0)$, leading to earnings $W_0(\theta) = \theta h_0(\theta)$ and utility $U_0(\theta) = \theta h_0(\theta) - c(h_0(\theta))$.

Assume that participation in the workfare program provides a transfer T in exchange of h_p hours of work. Further assume that the disutility of time spent in program activities is the same as of time spent in non-program activities. When an individual participates in the program she can also decide to work outside the program. We note $W_1(T, \theta)$ her total earnings in case of program participation and $U_1(T, \theta)$ the corresponding utility.

The impact on individual earnings before the actual decision to participate or not is defined as:

$$W_1(T, \theta) - W_0(\theta) = T - \Delta(T, \theta) = s(T, \theta). \quad (1)$$

where $\Delta(T, \theta)$ represents the earnings that individuals forgo in order to participate in the program, and $s(T, \theta)$ is the net impact on earnings.

Individuals decide to participate in the program if the impact on their utility is positive: $U_1(\theta) - U_0(\theta) > 0$.

There are two cases to consider (see illustration in Appendix Figures B1). The first corresponds to a small transfer T offered for the h_p hours of work: $T < h_p c'(h_p)$. In such a case, only individuals with a low productivity participate and when they participate they only work in the program, leading to earnings $W_1(\theta) = T$ and foregone earnings $\theta h_0(\theta)$. The productivity threshold $\underline{\theta}$ triggering participation is given by $T - c(h_p) = U_0(\underline{\theta})$. Individuals with a productivity larger than $\underline{\theta}$ do not participate.⁸

The second case corresponds to larger transfers: $T > h_p c'(h_p)$. In such a case, there are two types of participants. Individuals with a small productivity ($\theta < \underline{\theta} = c'(h_p)$) only work in the program, again with earnings $W_1(\theta) = T$ and foregone earnings $\theta h_0(\theta)$. Individuals with intermediate productivity ($\underline{\theta} < \theta < \bar{\theta}$) participate in the program but they also work outside the program;

⁸Notice that in such a case the impact on earnings for marginal applicants is $T - \underline{\theta} h(\underline{\theta}) = c(h_p) - c(h(\underline{\theta})) > 0$. There is a discontinuous increase in earnings for marginal participants.

their number of hours worked outside of program activities will be given by $h_0(\theta) - h_p$, as their optimal total number of hours worked is not impacted by program participation. Their earnings in the program will therefore be given by $W_1(\theta) = T + \theta(h_0(\theta) - h_p)$, with foregone earnings θh_p .⁹ Individuals with large productivity ($\theta > \bar{\theta} = T/h_p$) do not participate in the program.¹⁰

Importantly, note that in all cases above, a change in the transfer T does not change forgone earnings, so that $\Delta(T, \theta) = \Delta(\theta)$ and hence $s(T - x, \theta) = s(T, \theta) - x$.

Whichever case applies, given the heterogeneity parameter θ , there will be variation in the program impact on earnings.

Individuals decide whether or not to apply for the workfare program solely based on whether participation will increase their earnings, i.e. if $s(T, \theta) > 0$.¹¹ We call $s_T \equiv s(T, \theta)$ this random variable in the population of those who self-select into the program.

Let B be the total budget for transfers in the program and $N_A(T)$ the number of individuals who self-select into the program when the transfer level is T . Assume, that $B < TN_A(T)$, that is, the program is over-subscribed and a lottery is used to allocate program slots among the pool of self-selected applicants, as in our application. The lottery success rate $\lambda(T)$ is simply given by $B = TN_A(T)\lambda(T)$. The average contemporaneous program impact on earnings over those who self-select (also including those who were randomized out) is:¹²

$$S_{lottery}(T) = \lambda(T)E(s_T) \quad (2)$$

In order to increase program performance, targeting could be introduced to prioritize inframarginal applicants, or the effectiveness of self-targeting could possibly be improved with a lower transfer amount. We discuss both approaches below.

We first look at potential improvements associated with targeting. A growing literature explores

⁹Given the unchanged total number of hours of work for these now marginal applicants, their change in utility is the same as their change in earnings. In such a case, the lower bound of the distribution of the impacts on earnings should be zero.

¹⁰It is possible to show that $\underline{\theta} < \bar{\theta} \Leftrightarrow \underline{\theta} < \underline{\theta}$.

¹¹Marginal applicants are those who have the same utility whether they participate in the program or not. As stressed above, for marginal applicants who work outside the program, the difference in utility is the same as the difference in earnings. In such a case, there should be no discontinuity in earnings for marginal applicants. Our estimation of the distribution of individual treatment effects on earnings shows that this case is likely in our setting.

¹²Expectations are taken over the population of those who select for the transfer T , i.e. they are taken with respect to the distribution $f(\theta|S(T, \theta) > 0)$.

the idea of improving program performance through targeting (see [Manski \(2004\)](#)). Our approach follows [Bhattacharya and Dupas \(2012\)](#), who seek to maximize outcomes under a budget constraint for transfers (B in our case). Consider randomly assigning the program to applicants, with probabilities depending on some observable characteristic z . The assignment probability would be a function $\alpha(z)$. The objective is to find the assignment function $\alpha(\cdot)$ that maximizes program impacts on earnings:

$$S_{targeting}(T, \alpha) = E(\alpha(z)s_T) \text{ s.t. } TN_A(T)E(\alpha(z)) = B \quad (3)$$

Given $E(\alpha(z)s_T) = E(\alpha(z)E(s_T|z))$, the optimal assignment rule is simply $\alpha^*(z) = 1(s_T(z) > \bar{s})$, where $s_T(z) = E(s_T|z)$ and \bar{s} is chosen such that $TN_A(T)P(s_T(z) > \bar{s}) = B$, i.e. $P(s_T(z) > \bar{s}) = \lambda(T)$. In this case, the average contemporaneous program impact on earnings is given by:

$$S_{targeting}(T, \alpha^*) = \lambda(T)E(s_T|s_T > \bar{s}) \quad (4)$$

Such an assignment rule obviously dominates the assignment using a lottery. In the empirical section of the paper, we apply machine learning techniques to estimate the function $s_T(z)$. We then compute the gains associated with this optimized assignment rule, compared both to lottery assignment as well as alternative targeting rules (such as prioritizing women or prioritizing applicants with low self-reported or proxied earnings at baseline).

Second, we consider the effects of changing the transfer amount from T to $T - x$, for a fixed number of hours of work in the program (h_p). As long as the program is oversubscribed, lowering the transfer reallocates program slots from those with lower impact to those with higher impact. While this effect is positive (both in terms of reallocation and in terms of more individuals being served), it comes at the cost of lowering the transfer which negatively affects all participants. Therefore, the change in the average impact on earnings is ambiguous. We derive the expression for the change in the average impact in [appendix C.2](#) and show that it depends on the distribution of s_T .

Indeed, consider all the (potential) distributions achieving the same average impact on earnings. Absent an adjustment of the lottery success rate, a reduction in the transfer causes an equivalent reduction in average impact across all these distributions. However, given a fixed budget for the

program, the lottery success rate can increase; by how much will depend on the distribution of individual program impacts rather than the average program impact. In particular, in distributions that include a large share of applicants with an impact close to zero, a large number of individuals will select-out and thus the lottery success rate will increase more. If the increase in the lottery success rate is large enough, the impact on those who can enter the program more than compensates the initial reduction in average earnings.

More generally, under such an alternative contract, the youth who self-select for the program are those for whom $s(T - x) > 0$ or, given $s(T - x) = s_T - x$, those for whom $s_T > x$. The number of individuals who apply is given by $N_A(T - x) = N_A(T)P(s_T > x)$, and the number of people who can be served $N(T - x)$ is given by $B = (T - x)N(T - x)$.¹³ A lottery is again used to allocate the program among applicants as long as $N(T - x) < N_A(T - x)$. The rate of success of this lottery $\lambda(T - x)$ depends on x and is given in this case by $\lambda(T - x) = N(T - x)/N_A(T - x) = B/(N_A(T)(T - x)P(s_T > x)) = \lambda(T)T/((T - x)P(s_T > x))$ (and by 1 otherwise). Clearly, the lottery rate increases with x : there are fewer individuals who still apply to the program and the amount to distribute per participant is by definition lower. The average contemporaneous program impact on earnings over those who would self-select for the full transfer T (thus including a zero impact for those who select out for the smaller transfer $T - x$ as well as those who are randomized out) can be written as:

$$S_{\text{lottery}}(T - x) = \lambda(T - x)E((s_T - x)1(s_T - x > 0)) = \lambda(T)\frac{T}{T - x}E(s_T - x|s_T > x) \quad (5)$$

Unlike $S_{\text{targeting}}$ in equation 4, there is no direct indication that changing the transfer amount from T to $T - x$ would lead to an improvement in S_{lottery} . Actually, the impact on earnings after a reduction in the transfer depends on the form of the distribution of s_T . Appendix Figure B2 provides two examples that illustrate how $S_{\text{lottery}}(T - x)$ changes with x for different forms of the distribution s_T .¹⁴

The two distributions are chosen to be symmetric around a mean impact of CFA 25,000. They have very different shapes, however.

¹³In the framework, we consider the average earnings over those who initially self-select in the program. We could consider that the social planner seeks to maximize the average of a function of earnings instead of earnings, for example the share of those whose earnings are above a given threshold S . In such a case, it is worth noting that as long as the program is oversubscribed, a reduction in the transfer increases the number of individuals who receive the transfer, which would improve this objective if the transfer is above S .

¹⁴Appendix Section C studies these relationships more formally.

In the first example, the distribution of individual program impacts has a mode close to zero, corresponding to a situation with a large mass of applicants with small program impacts. In such a case, when the transfer is reduced, there are many potential participants who no longer apply (i.e., $P(s - x > 0)$ decreases sharply). This makes it possible to substantially increase the lottery success rate and this increase is large enough to compensate for the reduction in the size of the transfer, so that the average program impact on earnings in the population of initial applicants increases.

In the second example, the mode of the distribution is located at a larger value and the density is almost zero for very low values of s , corresponding to a situation with a small mass of applicants with small program impacts. As a result, when the transfer decreases, there are very few people who do not apply for the program anymore. The lottery rate increases only by a small amount, which is not enough to compensate for the direct negative effect of the reduction in the size of the transfer. As a result, the average program impact on earnings in the population of initial applicants decreases with a reduction in the transfer amount.

The shape of the distribution of s_T therefore plays an important role in overall program performance and in $S_{lottery}(T - x)$, the average contemporaneous program impact on earnings for a reduced transfer. For this reason, the empirical section will study the distribution of predicted impacts on earnings. While this true density function is unknown, we can proxy for it using machine learning estimates of treatment effects on contemporaneous earnings conditional on a rich set of available covariates.

2.2 Post-Program Impacts

A first-order question in the public works literature relates to the existence and size of post-program impacts in the medium to long-term. Indeed, a growing number of public works programs also have the objective to facilitate participants' 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 through which they could unfold. An important consideration relates to the relative allocation of short-term earning gains between consumption and savings, which will affect post-program impacts.

First, public works can help participants overcome capital constraints. Several experiments have

found relatively large returns to capital for poor households (for a review, see [Blattman and Ralston \(2015\)](#)). Common instruments to make capital available to youth, such as micro credit, have not proven very effective. As such, the positive income shock induced by public works programs (but truly any income support program) could alleviate capital constraints, facilitate savings and investments.¹⁵

There are other possible mechanisms for longer-term impacts that are specific to workfare programs. Subsidized employment could be a way to improve experience, skills and productivity of participants, and ultimately increase the likelihood that they find a wage job post-program. Also, there might be behavioral effects related to program participation. For example, a workfare program requiring youths to form work habits, like getting up each morning to go to work, may induce lasting behavioral changes that will improve future employability.

On the other hand, it is also possible that participation in a public works program may have deleterious long-term impacts on participants. First, the work experience provided through the program might be of little value or only enhance skills that are not demanded in the labor market. Negative long-term impacts could also emerge because of the potential "stigmatization" of participants, with program participation sending negative signals to potential future employers. Finally, participants may give up some valuable activities or social connections in order to participate in the program, which may induce a form of capital destruction that may take time to rebuild upon exit from the program.

3 Intervention and Data

3.1 The Public Works Program

The public works intervention we study is part of an Emergency Youth Employment and Skills Development Project (PEJEDEC) set up after the 2010/2011 post-electoral crisis.¹⁶ The public works program aims to provide access to temporary employment in road maintenance (such as

¹⁵Savings can have several potential post-program benefits, including precautionary savings to absorb future shocks, or savings to finance investments, like training or capital for income-generating activities.

¹⁶Public works programs were first introduced in Côte d'Ivoire by a post-conflict assistance project (PAPC) in the aftermath of the 2002-2007 armed conflict. The PEJEDEC public works program built on that experience. It was implemented by the national roads management agency (AGEROUTE), and supervised by BCP, the Coordination Office for Employment Programs ("Bureau de Coordination des Programmes Emploi"), under the Ministry of Labor and Social Affairs.

sweeping roads or cleaning ditches) for low-skilled youths in urban areas. The program targets youths aged 18-30 in 16 localities throughout the country.¹⁷ Participants are offered temporary employment for 6 hours per day and 5 days a week for a total of six months.¹⁸ Participants work in teams of 25 individuals (called “*brigades*”), under the supervision of a team leader and a local supervisor. The jobs are paid CFA 2,500 (approximately \$10 PPP 2014) per workday, 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 upon enrollment. A quota of 30% of program slots was initially reserved for women.

All participants in the public works program receive a one-week basic life skills training covering issues related to HIV-AIDS, citizenship and hygiene. Some participants are also offered a complementary basic entrepreneurship training to facilitate transition into more productive self-employment upon exit from the program. Finally, other participants are offered a training on wage jobs search skills and sensitization to wage jobs opportunities, with the objective to facilitate transition into wage jobs upon exit from the program. Additional information on the complementary training is provided in Appendix D.

3.2 Experimental Design: Enrollment and Randomization

Four waves of the PEJEDEC workfare program were organized between 2012 and 2015, each covering 16 localities, with a similar (pre-determined) number of participant slots 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.

The identification strategy relies on a two-step randomization process. The first step involves individual-level 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 by the implementing agency through local newspapers, local radios and public notice boards to invite interested youths to visit a registration office and apply to the program. Enrollment was open for two to three weeks in each locality. Only two eligibility criteria were applied during enrollment:

¹⁷Four municipalities 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).

¹⁸As explained later, the wave of the program under evaluation lasted 7 months.

applicants had to be between 18 and 30 years old, and they should not have participated in the public works program before.

Once the enrollment period had closed, public lotteries were organized in each locality (separately for men and women) to randomly select beneficiaries among the registered applicants present at the lottery.¹⁹ In practice, 10,966 youths participated in the public lotteries carried out for the second wave of the program, during which 3,125 beneficiaries were selected and assigned to 125 brigades of 25 individuals each (17 men, 8 women).²⁰ Replacement of drop-outs was allowed during the first two months of the program. A waiting list was created to protect the control group, although in practice replacements were minimal.²¹ The public lotteries were held in each locality between the end of June and early July 2013.

The second step involves the randomization of public works brigades into groups receiving different types of complementary training that took place in the seventh and final month of the program. Specifically, brigades were randomized into three groups: (i) 45 brigades (1,225 individuals) were assigned to receiving the public works only; (ii) 40 brigades (1,000 individuals) were assigned to receiving the public works plus the complementary basic entrepreneurship training, and (iii) 40 brigades (1,000 individuals) were assigned to receiving the public works plus the wage jobs search skills training. This second randomization was stratified by locality, and was performed through a lottery held in the project office in the presence of implementing partners and a public notary in November 2013. The results of this lottery remained confidential until two weeks before the start of the complementary training (in January 2014) in order to limit potential response bias during the midline survey.

¹⁹Public lotteries have been used continuously in Côte d'Ivoire as an assignment mechanism to allocate limited slots for jobs since the introduction of public works in the post-conflict period. The transparency of the process makes it socially acceptable and limits potential tensions. As such, the first step of the randomization protocol was already implemented as part of routine program operations.

²⁰Beneficiaries were assigned to brigades within localities based on the number they drew in the public lottery.

²¹Replacements were only possible based on the waiting list, and had to be stopped when the waiting list was exhausted. This ensured that individuals in the control group were not offered the program during its implementation. In practice the waiting lists were never exhausted.

3.3 Timeline and Data

3.3.1 Timeline and Surveys

A baseline survey was conducted shortly after the lotteries. The study sample includes 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 drawn from the non-beneficiaries that were not on the waiting list. The data collected at baseline included information about employment and earnings. It also captured a range of other characteristics such as risk and time preferences, behavioral skills and results of tests measuring skills or manual dexterity. Attrition at baseline was very low (1.5%). The public works activities started between early and late July 2013, depending on the locality. Participants received the one-week life skills training in August 2013.

A midline survey was conducted on 3,036 individuals (2,001 beneficiaries and the control group) between the end of November 2013 and early January 2014, i.e. 4 to 5 months after the start of the program.²² Both treatment and control individuals as well as the heads of their household were interviewed at midline. Attrition at midline was low (2.6%) and balanced across treatment and control groups.²³ The midline questionnaire includes very detailed modules on employment (up to three activities), earnings, time use, well-being, behaviors and list experiments to proxy risky behaviors.

The public works program ended in February 2014. It was originally expected to end in January 2014. However, as the complementary training activities only started in January, participants were given a one-month extension on their contracts, which extended the public works duration from 6 to 7 months. This ensured that all brigades of individuals selected to participate in one of the training programs could do so while being paid by the program (at the same wage) for at least part of their training, which reduced their opportunity cost of time during the training.²⁴

²²The 2,001 treated individuals are a sub-sample of the 3,125 beneficiaries stratified by locality, brigade and gender. We excluded from the midline survey brigades that had been assigned to the wage employment training. This is because their supervisors were following a management training at the time of the survey, and we were wary of any behavioral changes that could potentially affect outcomes.

²³A two-weeks tracking phase was implemented in February 2014 to limit attrition, mainly due to the mobility of control individuals. The tracking helped to reduce attrition rate from 5.4% to 2.6%. 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 Côte d'Ivoire, and excluding individuals that could not be reached since baseline.

²⁴The 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. 75% of the beneficiaries attended

An endline survey was conducted between March and July 2015, i.e. between 12 to 15 months after the program ended. The sample included 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.²⁵ Again, both experiment subjects and household heads were interviewed. A 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 retrospective information on job search, independent activities (including past projects) and an employment calendar.

3.3.2 Descriptive Statistics

Table 1 presents descriptive statistics for the selected applicants (column (1)). Public works applicants are on average 25 years old, and 94% live in urban areas. Applicants live in households with 6 members on average, and 23% head a household. 49% did not complete primary school, and 23% only completed primary school. One third of the applicants have attended some form of vocational training, mostly informal apprenticeships. In line with the national employment situation marked more by underemployment in low-earning occupations rather than unemployment, 78% of applicants were working prior to the program. Finally, applicants have limited financial resources, with only half having saved money over the last three months and 71% reporting being highly constrained for basic needs expenditures.

Table 2 compares our evaluation sample to a national sample of urban youths (between 18 and 30 years old) to provide insights into public works applicants' profiles. The gender breakdown and household asset index of public works applicants are quite similar to that of the national sample, with applicants marginally more male and from marginally poorer households. One main difference is that program applicants have lower educational attainment than the general youth population, and are much less likely to be inactive due to being in school. Because program applicants have left school, they are also more likely to be active and employed than the national population. Among the active population, the program attracts a higher share of applicants that are wage employed rather than self-employed or unemployed. As in many developing countries,

part of the second half of training after the end of their contracts. They were given a daily transport allowance of CFA 1500 (the program wage was CFA 2500) to compensate. The transportation allowance was paid ex-post in one transfer, based on the actual number of days attended. The remaining 25% were fully under contract during their training.

²⁵The replenishment of the control group is further explained in section 4.1.

a large share of youths in Côte d’Ivoire are self-employed and work for themselves. Many are underemployed as they work long hours but have low earnings. However, they are not necessarily searching for wage jobs in a traditional labor market, and unemployment tends to be higher among more educated youths (Christiaensen and Premand (2017)). Overall, Table 2 highlights that the program attracts youths who have left school and are already active in the labor market for wage jobs. These comparisons illustrate that the effectiveness of self-targeting is a priori unclear in contexts with widespread underemployment among low-skill youths: public works are not expected to attract the (more educated) inactive or unemployed youths, but the (less educated) youths are largely already active.

4 Empirical Methodology

4.1 Main Specifications

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

$$Y_i = \alpha + \beta W_i + \delta X_{i,l} + \varepsilon_i \quad (6)$$

where Y is an outcome of interest for individual i , W is an indicator for treatment (being assigned to the public works program at first randomization), and X is a vector of stratification variables (specifically, locality and gender).²⁶ Robust standard errors are clustered at the brigade level for treated individuals.²⁷

To estimate post-program ITT effects by treatment arm, we use the following specification:

$$Y_i = \alpha + \beta_1 W_i + \gamma_1 (W_i \times T1_i) + \gamma_2 (W_i \times T2_i) + \delta_1 X_{i,l} + \varepsilon_i \quad (7)$$

where $T1$ ($T2$) is an indicator for being assigned to the complementary self-employment training

²⁶Specification (6) uses probability weights to account for stratification, sampling of non-respondents during tracking surveys, and enrollment in later waves of the program (see details in Appendix G).

²⁷We suspect within-brigade error correlation due to the interactions between treated individuals who worked together in the same brigade for several months. Some individuals moved across brigades during the program. When such movement occurred, we group the different brigades together into a "broad" brigade for clustering. The results are robust to other definitions of the brigade cluster.

(wage employment training). β estimates the impact of the “pure” public works, while γ_1 (γ_2) estimates the additional effect of the self-employment training (wage employment training).²⁸

We also analyze heterogeneity in treatment effects by group G determined by a set of baseline characteristics Z (see discussion in section 6):

$$Y_i = \alpha + \beta_1(W_i \times G_i) + \beta_2(W_i \times (1 - G_i)) + \gamma * G_i + \delta_2 X_{i,l} + \varepsilon_i \quad (8)$$

We are interested in β_1 , which estimates the impact of the pooled treatment for a specific group G .²⁹

Table 1 presents baseline balance checks between treatment and control groups, with p-values for differences in column (3). Column (4) contains p-values for a test of whether differences between all treatment arms are jointly equal to zero. We focus on baseline respondents interviewed at endline.³⁰ We note that collecting the baseline survey after assignment to the program may have induced some misreporting. Despite this, there are no quantitatively meaningful differences across groups. The few imbalances that are statistically significant are of small magnitude, such as school enrollment, self-reported constraints to repay loans, having an activity or risk aversion.

At midline, compliance to program assignment was high. Only two control individuals ended up in the program by registering in different locations. Among youth assigned to the public works, take-up was high with 93% participating more than five out of seven months. In total, youth worked an average of 141 days out of a maximum of 154 workdays.

The take-up of complementary training was lower than the take-up for the public works intervention: 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 hours. This is in line with take-up observed in other skill training programs.³¹ However, for each training, only 10% of individuals never attended, such that we focus on ITT estimates.

An unforeseen issue emerged at endline. A few individuals from the control group (140) were able to apply (and, for some, participate) in the third or fourth wave of the program. We account

²⁸Specification (7) includes probability weights as in specification (6).

²⁹Specification (8) includes probability weights as in specification (6).

³⁰We use the same weights as for the estimation. We also verified that there is balance across groups for midline respondents, and for baseline respondents.

³¹For instance, in a dual apprenticeship program in Côte d’Ivoire, the take-up was 75% (Crépon and Premand, 2019).

for this issue by excluding these individuals from the post-program impact analysis and assigning larger weights in the post-program analysis to control individuals who also applied in future waves but were not selected through the public lotteries.³² Furthermore, we randomly select 200 additional applicants from the enrollment lists and add them to the endline sample to maintain the total size of the control group used for analysis and related statistical power.

Lastly, recent studies have shown that public works programs can have externalities on labor markets and wages through equilibrium effects. This has mainly been documented in the context of a large-scale program in India (see [Imbert and Papp \(2015, 2019\)](#), [Muralidharan et al. \(2016, 2017\)](#) or [Berg et al. \(2018\)](#)). With 12,666 beneficiaries over 4 years in 16 urban areas, the size of the program we study is small and general equilibrium effects are unlikely in our setting.

4.2 Heterogeneity Analysis with Machine Learning Techniques

The standard heterogeneity analysis from equation (8) relies on the estimation of average treatment effects across sub-groups using a linear interaction in a standard regression framework. The treatment variable is interacted with covariates and predicted impacts can be recovered conditional on these covariates. This approach raises the issue of selecting the dimensions of heterogeneity, which pre-analysis plans ([Casey et al., 2012](#); [Olken, 2015](#)) and multiple hypothesis testing ([List et al., 2019](#)) can help to partly address. The main innovation of machine learning methods is that they require fewer assumptions about the source or form of this heterogeneity. Importantly, they can search for heterogeneity across high-dimensional sets of covariates without assuming a specific functional form.³³ Our motivations to analyze heterogeneity are to find an optimal program assignment rule in the spirit of the approach in [Bhattacharya and Dupas \(2012\)](#) and to study how the average contemporaneous program impact on earnings varies with the transfer. For this purpose, we are interested in identifying a non-parametric estimate of the conditional treatment effect, which requires the use of a flexible estimator.³⁴

³²Specifically, using administrative data from these additional program waves, we were able to identify repeat applicants, and whether they were selected or not based on the public lotteries. We identified 140 individuals from our baseline control group (i.e. 13.5%) among beneficiaries of the next waves of the program. We remove the repeat applicants who were randomly selected for the program from analysis. We then over-weight the non-selected repeat applicants. See details on weights in Appendix G

³³See [Athey and Imbens \(2016, 2017b,a\)](#) for a general discussion of machine learning techniques to analyze heterogeneous treatment effects, or [Athey and Imbens \(2016\)](#) on the use of regression trees.

³⁴A large variety of estimators have been proposed in the literature, including causal forest ([Wager and Athey, 2018](#)), generalized random forest ([Athey et al., 2019](#)) or R-learner ([Nie and Wager, 2017](#)), and many others. [Knaus et al. \(2020a\)](#) offer a general presentation of these algorithms as well as Monte Carlo simulations to study their

We are looking to estimate treatment effects for specific subgroups in the population, defined by their (observable) characteristics z . We would like to estimate the conditional average treatment effect (CATE) for some subgroups, corresponding to $s_0(Z) = E[Y_i(1) - Y_i(0)|Z_i^K = z]$ where Z^K is a vector of K baseline covariates (features) and $Y(k)$ the potential outcome of interest for treatment ($k = 1$) and control ($k = 0$).

A key challenge when using machine learning techniques to study heterogeneity is to derive confidence intervals and perform inference.³⁵ In this paper, we use the inference framework developed by Chernozhukov et al. (2020), who present a general approach using machine learning estimators as a proxy predictor to make inference on *key features* of the CATE function (rather than the whole function).³⁶ This allows us to (i) formally test for the existence of heterogeneity, (ii) compute confidence intervals around the conditional treatment effect for groups of interest (such as those at the top and bottom of the impact distribution), and (iii) compare the characteristics of the population who benefit the most or the least from the program. The approach in Chernozhukov et al. (2020) is “generic” in the sense that it applies to any machine learning algorithm used to estimate heterogeneous treatment effect, including the causal forest and generalized random forest estimators proposed by Wager and Athey (2018) and Athey et al. (2019). In our own implementation, we consider several alternative machine learning algorithms (detailed in Appendix F.2). We present the main results for the best performing algorithm, and provide robustness checks in the Appendix.

When applying machine learning methods, we split our data so that separate sub-samples are used to either build the model (the *auxiliary sample*, on which machine learning predictors are trained and constructed) or make inference (the *main sample*, to which the model is applied, and on which we estimate the different key features of the CATE function). In our application, this procedure is repeated 100 times on random sub-samples.³⁷ Chernozhukov et al. (2020) offer a procedure to aggregate results across simulations and construct valid confidence intervals and p-values.³⁸

performance.

³⁵The issue of detecting true heterogeneity versus noise is also discussed in Davis and Heller (2020) when using causal forest estimators.

³⁶This contrasts with the approach in Wager and Athey (2018) who derive point-wise confidence intervals in the specific case of causal forests.

³⁷Iterations of the data-splitting process are necessary to identify how much variation is induced by specific data splits. It also ensures that each observation will be used on average both for construction and prediction (depending on the data-split), so all the information contained in survey data is used. This is especially important given our rather small sample size.

³⁸Their procedure takes into account the uncertainty coming from both the estimation of the parameters and

We test for the presence of heterogeneity by estimating the β_2 coefficient in the following equation:

$$Y = \alpha_1 + \alpha_2 B(Z) + \beta_1(T - P(Z)) + \beta_2(T - P(Z))(S(Z) - \hat{E}(S(Z))) + \varepsilon \quad (9)$$

Machine learning is used to get $S(Z)$, a relevant *proxy* predictor of $s_0(Z)$, as well as $B(Z)$, a machine learning predictor for $Y(0)$ (both constructed on the auxiliary sample). T is the treatment variable, and $P(Z) = \hat{E}(T|Z)$. We use weights $w(Z) = \{P(Z)(1 - P(Z))\}^{-1}$.

β_1 captures the average treatment effect (ATE) while β_2 is the heterogeneity loading parameter (HET).³⁹ We are particularly interested in β_2 , which offers a test for heterogeneity in treatment effect. Rejecting the null hypothesis that $\beta_2 = 0$ means that (i) there is heterogeneity, and (ii) that our machine learning predictor is a good approximation of $s_0(Z)$. On the contrary, if β_2 is not statistically different from zero, it means either that our machine learning predictor is uncorrelated with $s_0(Z)$ (our predictor is not able to capture heterogeneity correctly), or that there is no heterogeneity. In our application, we test for heterogeneity in impacts on earnings both *during* and *post* program.

Besides detecting heterogeneity, we are also interested in the magnitude of the treatment effects along the distribution. In our application, we consider the top and bottom quartiles of the distribution, corresponding to the 25% of individuals who benefit the most and the least in terms of impacts on earnings. This is because around 25% of total applicants were selected to participate in the program we study.⁴⁰ We recover the parameters of interest $E(s_0(Z)|G_k)$, also referred as Group Average Treatment Effects (GATES), where groups are quartiles of the distribution of predicted treatment effects, through the following weighted linear projection:

$$Y = \alpha_1 + \alpha_2 B(Z) + \sum_{k=1}^4 \gamma_k (T - p(Z)) 1(G_k) + \nu \quad (10)$$

The projection coefficients γ_k are the GATES parameters. The groups are defined as $G_k =$

the data splitting process when aggregating the results (p-values, confidence interval bounds) across simulations. It takes the median of the estimated parameters over all splits, as well as the median of p-values which is then adjusted by a factor of 2. [Breda et al. \(2020\)](#) show that these adjusted p-values can be interpreted as upper bounds. Confidence intervals computed at 95% significance ($\alpha = 0.05$) have to be re-adjusted for split uncertainty. After adjustment, the procedure provides confidence interval bounds at 90%.

³⁹In this framework, the quantity $BLP(Z) = \beta_1 + \beta_2(S(Z) - \hat{E}(S(Z)))$ can be interpreted as the best linear predictor of $s_0(Z)$ based on $S(Z)$. Also $\beta_1 = E[s_0(Z)]$ is the average treatment effect (ATE) and $\beta_2 = \frac{Cov(S(Z), s_0(Z))}{Var(S(Z))}$ is the heterogeneity loading parameter (HET).

⁴⁰[Chernozhukov et al. \(2020\)](#) consider quintiles. We adapted the procedure to quartiles in the context of our application, as the rate of success of the lottery to assign applicants to the program is roughly 25%.

$\{S(Z) \in I_k\}$ with $I_k = [q_{k-1}, q_k)$, where q_k are the quartiles of $S(Z)$, and $q_0/q_4 = -/+ \infty$. We again use weights $w(Z) = \{P(Z)(1 - P(Z))\}^{-1}$. The estimated parameter γ_4 (corresponding to the top quartile of the predicted distribution of impact, group G_4) can be interpreted as the average treatment effect among the 25% of individuals who benefit the *most* from the program. Similarly, γ_1 can be interpreted as the average treatment effect among the 25% of individuals who benefit the *least* from the program (group G_1).

A natural next step is to study the characteristics of the groups of interest (i.e. $E[g(Z)|G_k]$, where $g(Z)$ is the vector of characteristics of an observation). In particular, we can compare baseline characteristics between the top and bottom quartile of the distribution of predicted impacts, namely groups G_4 and G_1 . Although machine learning methods do not allow us to exactly identify which characteristics matter the most for heterogeneous treatment effects, learning about the characteristics of those who benefit the most and the least provides insights about the variables that could be used for targeting.

In the analysis below, we will also assess how belonging to a particular heterogeneity group for a given outcome Y affects treatment effect on another outcome \tilde{Y} . In other words, we seek to identify $E[S_{\tilde{Y}}(Z)|G_{Y,k}]$, where $S_{\tilde{Y}}(Z)$ is the treatment effect on variable \tilde{Y} conditional on Z and $G_{Y,k}$ is the k^{th} heterogeneity group for the treatment effect on the variable Y conditional on Z . This is useful to determine whether there are trade-offs between optimizing selection into the program to maximize during-program impacts and post-program impacts. This is also useful in buttressing our understanding of mechanisms for longer-term impacts: for example, we can assess whether individuals that benefit most from the program in terms of during-program earnings are also those with the greatest post-program savings or post-program well-being. In practice, we can use equation (10) to perform this analysis, replacing Y as a dependent variable with the alternative outcome variable \tilde{Y} .

5 Results

Table 3 presents ITT results. We first display pooled treatment estimates from equation (6), both for contemporaneous impacts measured in the midline survey (4-5 months after the start of the program, while youths are still participating; Panel A) as well as for post-program impacts measured in the endline survey (12-15 months after youths have exited from the program; Panel

B). We then discuss impacts by treatment arm using specification (7) (Panel C).⁴¹ The main outcome variables are employment, type of employment, hours worked and earnings. We also present results for expenditures and savings. In the main specification, we consider continuous variables (such as hours worked and earnings, winsorized at the 97th percentile) to facilitate the interpretation of magnitudes. We also include the logarithm of the main earnings variable. Table 4 presents results for alternative treatment of outliers, including outcomes in logarithm and winsorized at the 99th percentile. Results are generally robust, and we discuss in the text the few cases where there are discrepancies. Table 5 presents results for well-being and behavior indices, as well as for time use variables proxying work habits and engagement in risky behaviors measured from list experiments.⁴²

5.1 Contemporaneous Impacts

Table 3 (Panel A) presents ITT estimates on employment and hours worked during the program. Given the high share of control youths working in some form of activity at midline (85%), the impact of the public works program on the likelihood of employment is rather small (+14 percentage points). A similar pattern is observed for hours worked per week, with a small overall increase in total hours worked (by 4.9 hours) from an average of 39.2 hours per week in the control group. Employment in the public works wage jobs accounts for approximately 30 hours a week for individuals in the treatment group, so that the small increase in overall hours worked hides a large decrease in hours worked in other activities.

In contrast, the intervention had a more substantial impact on the composition of employment, with a large increase in the share of youths holding wage jobs (+48 percentage points, from a base of 49% in the control group) and a decrease in self-employment (-10 percentage points, from a base of 35% in the control group). Correspondingly, we observe a large increase in hours worked in wage employment (+15.6 hours) and a decrease in hours worked in self-employment (-5.7 hours).⁴³ This highlights that youths reorganize their portfolio of activities to participate in the program.

Table 3 (Panel A, columns 7-10) presents estimates of impacts on earnings during the program. Participation in public works leads to a significant net increase in earnings. The magnitude of

⁴¹Appendix Table A1 shows similar results based on specifications that include baseline controls.

⁴²Appendix E provides information on the definition and construction of these variables.

⁴³Youths in the treatment group also became more likely to engage in multiple activities.

the effect amounts to CFA 27,083 per month (or approximately US \$110 PPP 2014). The net earnings gains represent approximately 53% of the average net monthly transfer.⁴⁴⁴⁵ 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. Contemporaneous impacts on earnings stem from a strong increase in earnings from wage employment (+CFA 36,799), which offsets a significant decrease in earnings from self-employment (- CFA 5,715).

These results suggest that self-targeting did not succeed in this context in getting only the most vulnerable (e.g. those with the least outside employment opportunities) to participate in the program. A couple of factors likely explain this failure of self-targeting. First, governments typically cannot legally offer public works programs with wages below the formal minimum wage,⁴⁶ so a job that pays the statutory minimum wage could be of appeal to many in an environment where informal employment and self-employment are rampant. As in many countries in Sub-Saharan Africa, Côte d'Ivoire faces a relatively low unemployment rate, but a large share of individuals working in low-productivity self-employment or informal wage jobs without contracts.⁴⁷ Employment patterns in the control group illustrate this. Overall, many individuals earn less than the legal minimum wage, as the regulations are only binding for formal private companies and public administration. In this setting, the results show that the program induced a reallocation of youths' activities and substantial forgone earnings. Second, because the work was only 6 hours per day, many applicants with outside employment opportunities would still see value in applying for the public works program as they could combine it with other activities, especially those that allow for more flexible hours. 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.

In Table 3 (Panel A, columns 11-12), we assess contemporaneous program impacts on expenditures and savings. The observed increase in earnings CFA 27,083 per month (or approximately US \$110

⁴⁴CFA 50,600 (\$205 PPP 2014) is the average amount transferred over all individuals assigned to the public works (independently of non-compliance and days not worked).

⁴⁵The average treatment effect for the variable measured in *log* (for which we take $\ln(y + 1)$) in column (8) is 2.95 at midline. This is quite large compared to the average treatment effect for the variable in level in column (7) as a percentage of monthly income in the control group ($27083/42841=0.63$). The difference is due to both the large standard error of the dependent variable and the reduction of the standard error between treatment and control group (see formula in footnote 50 below). The standard error of earnings in the treatment group (50483) is substantially smaller than in the control group (65466). This is largely explained by a lower dispersion of earnings, which are more concentrated among participants who receive the program wages and are less likely to have zero earnings.

⁴⁶This is the case for many programs in West Africa, as well as in Ethiopia or India.

⁴⁷For additional discussion, see [Filmer et al. \(2014\)](#); [Christiaensen and Premand \(2017\)](#).

PPP 2014) is associated with an increase in expenditures (CFA 14,529 per month, or \$59 PPP 2014) and savings (roughly CFA 10,000 per month, or \$41 PPP 2014). The increase in total monthly expenditures represents approximately 54% of the earnings gains. It 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 (such as food and clothes), as well as education and training. Youths are also able to save a significant share of their net earnings gains. On average, youths in the treatment group have increased their stock of savings by approximately CFA 39,786 (\$161 PPP 2014) after about 4 months in the program. This large impact represents a 182% increase from the average stock of savings in the control group (CFA 19,250, or \$78 PPP 2014). It is also consistent with youths saving approximately 36% of their earnings gains, or 20% of the public works wages. 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 raise the possibility that youths can invest in job search or self-employment activities after program exit.

Table 5 documents impacts on indices of well-being and behavior, as well as work habits and engagement in risky behaviors.⁴⁸ The consideration of broader well-being indicators is important as temporary public works jobs may have non-monetary benefits or costs. 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. Furthermore, changes in youths' behavior are particularly relevant in a post-conflict setting such as Côte d'Ivoire, as they may point to potential program externalities on social cohesion, an issue of strong interest for local policymakers. We also investigate how participation to public works affects work habits by looking at youths' time spent in rest, leisure or work activities in the morning and in the evening.

The public works program induces substantial improvements in well-being, which increases by 0.2 standard deviation at midline (Table 5, Panel A, column 1). Improvements in well-being stem from a larger share of treated youths reporting feeling happy and proud, scoring higher on sub-scales for self-esteem, positive affect and positive attitude towards the future (see Table A2). The intervention also induces improvements in the behavior index, which increases by 0.12 standard deviation (Table 5, Panel A, column 2). This is driven by reductions in impulsiveness

⁴⁸See Appendix E for definition, and Table A2, Table A3 and Table A4 for effects on components of the indices.

and anger (see Table A3). Youths' work habits also change as they reorganize their days to participate in the program (Table 5, Panel A, columns 3-8). Participants are more likely to be up and working (or travelling to work) at 6am in the morning, and much more likely to be asleep at 10pm at night. Despite changes in work habits and behaviors, we do not observe changes in an aggregate index of youths' engagement in risky behaviors measured through list experiments (Table 5, Panel A, column 9). The estimate is negative, pointing to a reduction in risky behaviors, but not statistically significant.⁴⁹

These results highlight that the public works intervention leads to a re-organization of youths' activities that contribute to substantial forgone earnings. It also induces improvements in non-economic outcomes; while these may be associated with the observed impacts on economic outcomes, it is also possible that some youths who do not benefit substantially in economic terms may nevertheless benefit from the program in other dimensions.

5.2 Post-Program Impacts

Table 3 (Panel B) presents post-program impacts on the same outcomes. Despite strong shifts in youths' employment portfolios during the program, no post-program impacts on the likelihood of being employed, employment composition or hours worked are observed. Overall, while no negative "stigmatization" or "scarring" effects are observed, the post-program results also suggest that the public works does not bring longer-term benefits to youths in terms of employment or hours worked (columns 1-6). Despite an increase in savings during the program, post-program results show that youths are not more likely to be self-employed either.

Table 3 (Panel B, columns 7-10) considers earnings impacts post-program. Overall, the main results show that the public works intervention does not lead to robust changes in earnings at endline. We note that these results are slightly sensitive to the treatment of outliers. While impacts on variables in logarithm are not significant (Table 3, column 8 and Table 4, columns 1-2), small but significant impacts on post-program earnings are found in the level specification (Table 3, column 7 and Table 4, column 3): earnings increase by CFA 4,361, or about 10% compared to the control group. This increase in earnings in the level specification is driven by

⁴⁹When analyzing the components of the index in Table A4, we find a decrease in the share of youths taking drugs or displaying aggressive behaviors, but find an increase in the share of youths reportedly stealing. No changes are found in other indicators such as smuggling, prostitution, or having a firearm at home.

an increase in self-employment earnings. When looking at the characteristics of micro-enterprises that youths operate post-program, we find a relatively larger asset stock (in value) and level of investments (see columns 4-6 in Table A5), discussed further in the next section. Finding impacts on a variable in level but not in logarithm is consistent with potential heterogeneity in impacts,⁵⁰ which we analyze in detail below.

While the intervention does not lead to post-program impacts on employment, hours worked or robust effects on earnings, it does have sustained impacts on savings and psychological well-being. At endline, treated youths have a significantly higher savings stock by CFA 11,505. This represents nearly 25% of savings in the control group, and also approximately 30% of impacts at midline (Table 3, column 12). Post-program improvements in psychological well-being remain significant (0.11 standard deviation) but are also more muted than during the program (Table 5, column 1). They are concentrated in a narrow set of domains such as happiness, self-esteem and less present fatalism; in contrast, there are no lasting impacts on sub-scales for pride, positive affect, or positive attitude towards the future (Table A2, Panel B).

Finally, there is no lasting impact on any dimension of the behavior index, risky behaviors and work habits (Table 5). Overall, these results show that the public works program does not lead to sustained changes in behaviors or work habits, and hence is unlikely to increase productivity or employability via these behavioral channels. In other words, the public work program does not appear to live up to one of its key promises for larger longer-term impact relative to a more standard welfare program.

5.3 Post-Program Impacts by Treatment Arms

Table 3 (Panel C) documents post-program impacts by treatment arms. Overall, we observe little variation in impacts across treatment arms, suggesting limited value-added of the complementary skills training (micro-entrepreneurship or self-employment training and wage job search training). Specifically, post-program impacts on the likelihood of being employed, employment composition and hours worked are very consistent and not statistically different across the different treatment

⁵⁰Indeed, a first order approximation of the log function is that $E(\ln(y)) \approx \ln(E(y)) - 0.5\text{var}(y)/E(y)^2$. Thus $ATE(\ln(y)) \approx ATE(y)/E(y|T=0) - 0.5(\text{var}(y|T=1)/E(y|T=1)^2 - \text{var}(y|T=0)/E(y|T=0)^2)$. As a result, an impact can be detected in level but not in logarithm if the variance is larger in the treatment group than in the control group.

arms (Table 3, Panel C, columns 1-6).⁵¹ No differences in impacts on non-economic outcomes are found between treatment arms in Table 5 either.

Results on total earnings also show no differential impacts for individuals assigned to complementary training (Table 3, Panel C, columns 7-8). The results are robust to alternative treatment of outliers (Table 4, Panel C, columns 1-3). Post-program impacts on self-employment earnings are positive for treated youths assigned to the entrepreneurship training when variables are expressed in level, but the finding is not robust when the variables are expressed in logarithm. In addition, we cannot reject equality of the impacts on total earnings between the public works treatment only and the public works with self-employment training, or between the treatment arms with the self-employment and wage employment training. Since there is no statistical difference in impacts on overall earnings across treatment arms, we pool treatments to conduct the finer heterogeneity analysis in the rest of the paper.

The limited value-added of the complementary training suggests that skills acquisition through training is not a key mechanism to induce post-program impacts. Table A5 shows that the training interventions did improve knowledge as intended: knowledge on basic entrepreneurship increases by 0.11 standard deviation for the self-employment training (column 1, Panel B), and knowledge on job search skills improves by 0.26 standard deviation for the job search training (column 7, Panel B). The training also led youths to apply this knowledge in practice. For instance, the self-employment training increases the share of youths who prepared a business plan for one of their activities by 4 percentage points (column 2, panel A). Also, the wage employment training increased the share of youths who used a CV for a job search by 10 percentage points (column 8, Panel B). However, there are no impacts on job search expenditures or on the likelihood to search for a job. Overall, these changes in skills and practices are small in magnitude and do not appear sufficient to generate earnings beyond those induced by the basic public works program. Interestingly, impacts on the value of business assets (column 4) or on investments in start-up capital for self-employment activities (column 6) are driven by the public works treatment, which again points to savings and investments of public work wages as a key mechanism for post-program impacts. In sum, despite some effects on independent activities, there is no robust impact on profits or self employment earnings.

⁵¹Hours worked are significantly larger in the jobs search training arm 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 arm.

6 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. The allocation of program slots was based on randomized assignment, which had the advantage of being fair and transparent. At the same time, the effectiveness of the program might have been improved with a finer targeting of the 25 percent of beneficiaries among applicants. Recall that the only criteria enforced at enrollment are age (18 to 30) and not being a beneficiary of a previous wave of public works.

Given the self-selection mechanism, we would expect heterogeneity in impacts among program applicants, with marginal applicants experiencing only small gains in earnings compared to inframarginal applicants with fewer employment opportunities outside the program (as outlined in section 2). But by how much alternative targeting might have improved program effectiveness is an empirical question that depends on the magnitude of heterogeneity and the shape of the distribution of program impacts. We now turn to this question. We focus on heterogeneity in the logarithm of total earnings since the distribution of this variable is closer to a normal distribution than the variable in level (which has a higher dispersion even after winsorization). We also show key results for earnings in level for completeness.

6.1 Quantile Treatment Effects

Quantile regressions provide information about the lower bound of the variance.⁵² Heterogeneous quantile treatment effects are always associated with heterogeneity in treatment effects. However, when quantile treatment effects are homogeneous, the only conclusion that can be drawn is that the lower bound of the variance is zero, unless the intervention preserves ranks. Put differently, a constant quantile treatment effect is consistent with homogeneous quantile treatment effects, but the reverse is not true.

Figure 1 presents quantile treatment effects on log earnings during (Panel A) and after (Panel B) the program. 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. The quantile analysis shows substantial heterogeneity in

⁵²See [Heckman et al. \(1997\)](#) or [Djebbari and Smith \(2008\)](#). This lower bound is reached when the intervention preserves rank. In such a case, a quantile treatment effect can be interpreted as an effect at quantile. See [Bitler et al. \(2006, 2017\)](#) for applications.

impacts on earnings during the program (Panel A). The quantile treatment effect is three times larger at the 25% quantile compared with the 75% quantile. The estimated quantile treatment effects are quite precise, suggesting the existence of true heterogeneity rather than just sampling variation. Using a Kolmogorov-Smirnov test, we can reject the null that the distributions under control or treatment are similar. In contrast, post-program quantile treatment effects are uniformly small. The dispersion is within confidence bounds (Panel B), consistent with sampling variation. Although we detect larger quantile treatment effects at the top of the distribution, we cannot reject the null of the Kolmogorov-Smirnov test that the post-program distributions are similar in treatment and control. In summary, there appears to be large heterogeneity during the program, but little heterogeneity after the program.⁵³

The intervention is unlikely to induce churning in the distribution of contemporaneous effects, so that the heterogeneity seen in quantile treatment effects likely points to true underlying heterogeneity during the program. However, this is not necessarily the case for post-program impacts on earnings. There might be individual-level latent factors that may not contribute to the ranking of individual earnings in absence of the program but may contribute to the ranking of individuals' post-program earnings. For example, some individuals might be trapped at the bottom of the earnings distribution absent the program, but because of high latent returns to capital (e.g. through setting-up a highly profitable activity), these individuals might end up higher up in the post-program earnings distribution if the program allowed them to save and implement their latent project. Quantile regressions might thus fail to detect heterogeneity post-program.

6.2 Machine Learning Applications

We now present applications of machine learning techniques based on the approach described in section 4.2. First, we perform a statistical test to detect heterogeneity and then compare the magnitude of the conditional average treatment effects (CATE) across quartiles, with a particular focus on the bottom and top 25% of the distribution. Second, we analyze the characteristics of individuals in the bottom and top quartiles to understand how those who benefit the most differ from those who benefit the least. Third, we use the predicted distribution of conditional impacts to further understand the mechanisms between contemporaneous and post-program impacts. In

⁵³Figure B3 presents the post-program quantile treatment effects by treatment arm. They show similar patterns than those based on the pooled treatment.

particular, we look at post-program impacts on earnings conditional on being in the top quartile of the predicted distribution during program. This is helpful to understand how individuals who benefit the most during the program perform post-program; this is also helpful to assess whether there are trade-offs between maximizing contemporaneous and post-program impacts. Finally, we study the performance of alternative self-targeting (a lower offered wage) and targeting rules using the machine learning estimates as a benchmark.

6.2.1 Existence and Magnitude of Heterogeneity in Impacts on Earnings

Table 6 (panel A) presents results from estimating equation 9, including a test of the statistical significance of β_2 , the coefficient of the heterogeneity loading parameter.⁵⁴ We confirm finding heterogeneous impacts on earnings during the program, as β_2 is statistically different from zero in column (1). However, we do not find evidence of heterogeneity in post-program impacts (the p-value for β_2 is 0.97 in column (2)), which means that either there is no underlying heterogeneity, or the prediction model is not able to detect it. As for ITT estimates, we also run machine learning analysis with the outcome variable in level. Results at midline and endline are consistent in columns 3 and 4. Results are robust to the application of a wide variety of machine learning estimators (Table A6, Panel A), or to the consideration of a larger set of covariates (Table A7, Panel A). Similar results are found when analyzing heterogeneity in post-program impacts on earnings by treatment arm in logarithm (Table A8, Panel A) or in level (Table A9, Panel A).

Panels B and C in Table 6 provide additional insights about the magnitude of heterogeneity by reporting Group Average Treatment Effects (GATES) by quartile, as obtained by estimating equation (10). Panel B displays GATES for contemporaneous program impacts on earnings in log in column (1) and in level in column (3). Panel C presents GATES for post-program impacts on earnings in log in column (2) and in level in column (4). Columns (1) and (3) of Panel B illustrate that there is substantial heterogeneity in impacts on earnings during the program. As shown in section 2, this is expected due to self-selection, with a fraction of marginal participants almost indifferent between being enrolled or not, and others being inframarginal. However, the magnitude of this heterogeneity is noteworthy. The average predicted impact on earnings is CFA 14,660 in the lower quartile of the distribution compared to CFA 31,671 in the upper quartile (column 3). In other words, program impacts are 2.2 times larger in the top quartile than in the

⁵⁴Figure B4 presents the distribution of predicted impacts on earnings during the program and post-program, both in levels and in logarithm.

bottom quartile. In contrast, columns (2) and (4) of Panel C suggest more modest heterogeneity, with no significant difference in post-program impacts between quartiles. Figure 2 illustrates the heterogeneity between groups, during and post-program.⁵⁵

Since we do not observe heterogeneity in post-program impacts, we would further expect that it is possible to improve program effectiveness by maximizing contemporaneous program impacts without losses in post-program impacts. Figure 3 indeed confirms the absence of trade-off between impacts during and after the program.⁵⁶ A high correlation between impacts on earnings during and post program would lead to a concentration of predictions along the diagonal from the top right corner – those who have the largest impacts during and after the program – to the bottom left corner. On the contrary, the scatter plot shows that even within the top quartile of impacts during the program, the post program impacts are widely dispersed on the opposite axis. In Table 6, columns (2) and (4) in Panel B (respectively columns (1) and (3) in Panel C) show predicted impacts at endline (respectively midline) per quartile of predicted treatment effects at midline (respectively endline). They illustrate what would be the impacts on earnings post program (respectively during program) if midline impacts on earnings were maximized (respectively endline impacts were maximized). This further illustrates that there is no systematic relationship between those who benefit the most during the program and those who benefit the most after the program. We can never reject the null that coefficients are equal between groups. In other words, there is no measurable trade-off between short and medium-term impacts when trying to improve program effectiveness through finer targeting.

6.2.2 Patterns of Heterogeneity

Using the Classification Analysis (CLAN) in Chernozhukov et al. (2020), we analyze differences in baseline characteristics between quartiles of the distribution of treatment effects in Table 7. We focus on our two groups of interest, namely the bottom and top quartiles of the distribution of predicted impacts on earnings during the program (columns 1-2). Column (3) reports p-values for a test of equality between the lowest and highest impact groups.

⁵⁵Our focus is on the difference between the top and bottom quartiles. At the bottom of each panel, Table 6 reports a test of equality of the GATES across the four quartiles. We cannot reject the null.

⁵⁶Figure 3 shows predicted impacts on earnings during the program (x-axis) against predicted impacts on earnings after the program (y-axis) (for log earnings, using same machine learning estimator as in Table 6). The solid vertical and horizontal lines on the scatter plot correspond to the average predicted impacts during and after the program. Similarly, the horizontal (respectively vertical) dashed lines represent the 25th and 75th percentile of the distribution of predicted impacts during (respectively post) program.

Table 7 shows clear differences in the profiles of individuals who benefit the least and the most during the program.⁵⁷ The share of women is significantly higher in the upper quartile of predicted impacts (53%) than in the bottom quartile (15%). Several characteristics related to financial status, expenditures, savings and earnings suggest that the lower quartile of predicted impacts was “better-off” at baseline. There is also a very large difference in both propensity to save and baseline savings stock among individuals in the bottom quartile (CFA 65,925) and those in the top quartile (CFA 6,795). Similarly, the share of participants in the bottom quartile reporting they face credit constraints is lower (43%) than the share in the top quartile (58%). The share of individuals working at baseline is substantially higher in the bottom quartile (100%, compared to 44%), as are baseline earnings. Finally, three (six) times as many individuals in the bottom quartile are engaged in self (wage) employment activities at baseline compared to the top quartile. Participants who benefit the least during the program might still be able to save a greater share of their wages, or might be able to better invest these savings into income-generating activities.⁵⁸ It is therefore worthwhile to explore further how treatment effects on other main outcomes differ between the two groups. By doing this, we also highlight how machine learning techniques can help tease out mechanisms explaining program impacts.⁵⁹

Table 8 presents estimates of treatment effects on the main outcomes by quartile of the (predicted) impacts on earnings during the program (panel A), and after the program (panel B). Table 9 analyzes post-program impacts on intermediary outcomes related to productive investments by quartile of (predicted) impacts during the program. At the bottom of each panel, we present a test of equality of treatment effects between the bottom and top quartiles. This allows us to test whether groups with high or low earnings impacts during the program invest differently in their portfolio of income-generating activities or in job searching.

Table 8 confirms that, in the short run, the program has strong impacts on earnings for a more vulnerable group, while also attracting less vulnerable individuals for whom impacts are much more limited. Table 8 (Panel A) shows that the top quartile (corresponding to the top 25% of predicted impacts on earnings during the program) also has the highest impact on the likelihood

⁵⁷As a robustness check, Table A10 shows similar results for earnings in level.

⁵⁸It is also possible that these individuals may particularly benefit from the opportunity to save through the bank accounts set up by the program.

⁵⁹Davis and Heller (2020) also use machine learning to test underlying mechanisms relying on differential treatment response from disadvantaged youth who benefited from summer jobs. They look across types of outcomes while we also analyze outcomes over time.

of employment and wage employment (columns (1) and (2)). This is consistent with results in Table 7, which showed that individuals in the top quartile were less likely to have an activity prior to the program. Total earnings are more than twice higher in the top quartile compared to the bottom quartile (columns (5) or (6)). Column (10) reveals that the savings stock increases significantly for the top quartile, but not more so than for the other groups. Given that the bottom quartile is wealthier at baseline, one could have expected that they, more than other groups, would use program transfers to increase their savings. However, we observe similar impacts on savings across quartiles of predicted impacts on earnings. Lastly, despite variations in impacts on earnings during the program, there is no difference in impacts on well-being across the different quartiles (Table 8, column (11)). This suggests that gains in the well-being dimension are also broadly shared.

Table 8 (Panel B) and Table 9 reveal few differences across several dimensions of post-program impacts between the top and bottom quartiles of predicted impacts during the program. We can never reject the null of equality of treatment effects at endline between quartiles of predicted impacts on earnings at midline. In other words, those who benefit the most during the program perform similarly as others post-program. This illustrates again the lack of a clear trade-off between contemporaneous and post-program impacts along a large number of dimensions.

For completeness, and given the key role that savings during the program would theoretically play in driving any post-program impacts, we also perform the machine learning heterogeneity analysis using saving as the outcome. Indeed, increased savings can be a catalyst for productive investments in activities, the returns of which we might not yet observe at endline. We want to verify that there is no trade-off along this dimension, as seemingly suggested by Table 8. Table A11 shows a similar conclusion, with no significant heterogeneity at endline.⁶⁰ This also means that post-program impacts on savings are broadly spread across participants.

Overall, these results reveal limited heterogeneity in post-program impacts. The most vulnerable individuals who saw the largest gains in earnings and employment during the program do not exhibit relatively larger post-program gains. On the other hand, better-off individuals who saw smaller gains in earnings during the program do not experience higher post-program impacts on investments and earnings either.

⁶⁰In addition, the characteristics of those who benefit the most in terms of earnings and in terms of savings are very close: patterns in Table A12 are similar to those in Table 7.

6.3 Effects of Lowering the Wage

Can the effectiveness of the program be increased by lowering the offered wage? Our estimated Conditional Average Treatment Effects (CATE) together with the framework developed in section 2.1 can shed light on this question.

As we discussed in the conceptual framework, the average impact of a lower transfer given a fixed budget B will depend on the distribution of individual impact in the population (see equation 5 and Appendix C.2). Intuitively, lowering the transfer reallocates program slots from those with lower program impact to those with higher program impact and increases the lottery success rate. While this effect is positive, it comes at the cost of lowering the transfer for all participants. If there is a large share of participants at transfer level T with close to zero program impacts, the first effect will tend to dominate. But more generally, the overall impact of a lowering of the program wage will depend on the shape of the distribution of the individual treatment effect (see Figure B2, which contrasts two examples).

While the true distribution of individual impact is unknown, we can use the distribution of our CATE estimate as a proxy. The upper panel of Figure 4 displays our estimate of the distribution. As the figure shows, the distribution reveals a situation closer to the second example in Appendix Figure B2. The lower bound of the support of the distribution is zero, but there is a small share of participants who experience close to zero program impact, and the distribution function increases slowly up to its mode at around CFA 27,000. This suggests that a reduction in the offered wage may not improve program effectiveness.

This is confirmed in the lower panel of Figure 4, where we quantify the effect of a reduction in the transfer T using equation 5. The figure first shows how the self-selection process is affected as the transfer amount is reduced. The reduction in the number of applicants (dotted black line) is at first slow but becomes more substantial when the reduction in the transfer amount approaches the mode of the distribution of s_T .

The figure also shows (dashed blue line) the effect of the reduction in the transfer amount on the lottery success rate. Naturally, given the fixed budget, the lottery success rate increases as the transfer amount decreases as a) the number of applicants decreases (dotted black line) and b) the size of the transfer per participant is lower. More specifically, if the original total budget for transfers in the program ($0.25T$, or B in the context of our model) is randomly allocated to

those who still apply at the reduced transfer amount $T - x$, the lottery success rate is given by: Lottery rate = $0.25T/((T - x)P(s_T - x > 0))$. The figure reports this lottery rate up to the value where it reaches 1 (\approx CFA 27,500). As the figure shows, given the estimated shape of s_T , the increase in the lottery success rate remains modest for a large range of reduction in the transfer amount.

Last, the solid red line shows how the average impact on earnings varies when the transfer amount is reduced. More precisely, the solid red line displays $E((s - x)1(s - x > 0))/(P(s - x > 0)(T - x))$, which is $S_{lottery}(T - x)$ from equation 5 normalized by $0.25 \times 50,000$ CFA, the amount available for each initial applicant. The figure shows that the average impact on earnings would not increase when the transfer amount diminishes. The index we compute starts at 0.5; it reaches 0.3 for a reduction in the transfer of CFA 15,000 and 0.1 for a reduction in the transfer of CFA 25,000.⁶¹

6.4 Alternative Targeting Rules and Cost-Effectiveness Ratios

The analysis so far suggests that self-targeting based on the minimum wage or a lower wage limits program effectiveness, and that departing from self-targeting could improve contemporaneous program impacts without decreasing post-program impacts. However, targeting rules based on machine learning algorithms would be too complicated and expensive to implement, relying on complex information that is not easily available. Are there alternative targeting rules that could come close to achieving the predicted impacts in the upper quartile of the machine learning estimates? This is the question we take on in the rest of this section.

Table 10 summarizes impacts on earnings during the program (panel A) and post-program (panel B) for specific sub-populations under alternative targeting rules. For reference, the first column displays ITT impacts on (log) earnings for the whole sample of participants selected by randomized assignment. Since maximizing impacts during the program does not reduce post-program impacts, column (2) documents the effect of selecting the observations in the top quartile (Q4) of the distribution of predicted impacts at midline. For comparison, we also report in column (3) results when selecting the bottom quartile (Q1) of the distribution of predicted impacts at midline. For those two columns, Panel A reports group averages (GATES) from the machine learning prediction of midline impacts and Panel B reports the average endline impacts for observations belonging

⁶¹A confidence interval is obtained following the procedure in Chernozhukov et al. (2020).

to each quartile of the predicted impact distribution at midline.⁶² These estimates represent the impacts on earnings that would be achieved by targeting those who benefit the most (respectively the least) in terms of earnings during the program.

We then consider a scenario where only women are targeted. Table 7 showed that women are over-represented in the population that benefit the most during the program. Column (4) (Table 10) shows that targeting only women would improve impacts during the program, with no losses in post-program impacts. In particular, average impacts on income during (after) the program would be CFA 32,097 (CFA 7,168) when targeting only women. This corresponds to a 19% improvement in average estimated impacts on earnings during the program compared to randomized assignment. The point estimate is close to the machine learning benchmark in column (2).

What about targeting on baseline earnings? If there is limited churning in the earnings distribution, those with the lowest baseline earnings are likely to also have the lowest earnings at midline absent program participation. Columns (5) and (6) of Table 10 show results using two approaches to directly target the bottom quartile of the earnings distribution. The first approach targets individuals based on their (self) reported earnings at baseline (column (5)). The second approach uses machine learning techniques to predict (proxy) 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 estimate program impacts for individuals in the lowest quartile of the distribution of baseline income, either reported or predicted. This second approach (column (6)) is meant to mimic proxy means tests, which are often used to target safety nets to the poor and are more robust to misreporting than self-reported income. Columns (5) and (6) show that the contemporaneous impacts under these two approaches would come close to those predicted in the upper quartile of machine learning estimates (column (2)). Targeting individuals with reported baseline income in the bottom quartile leads to an average expected impact on income during the program of CFA 33,954 (column (5)), respectively CFA 32,824 (column (6)) when using predicted (proxied) baseline income. This is a 21%-25% improvement compared to randomized assignment. Post-program impacts are comparable to those obtained when selecting women only, and again not lower than the machine learning benchmark.

These results show that several practical targeting rules could perform better than self-targeting and improve program effectiveness. These alternative targeting rules come close to the machine

⁶²This is similar to Table 6, panel B, column (2).

learning estimates, without trade-offs between maximizing impacts during and post-program.

Panel C of Table 10 shows how program cost-effectiveness ratios vary by targeting rule. The average public works program costs CFA 768,708 (\$1,537.4) per beneficiary.⁶³ In our calculation, we focus on benefits captured by contemporaneous and post-program impacts on earnings under each targeting rule. Contemporaneous impacts are estimated from the midline survey and assumed constant during the 7 months of the program. Post-program impacts are considered constant from the end of the program (month 8) to the endline survey (month 21), and zero thereafter. The cost-effectiveness ratio of the existing program is presented in column (1). The discounted sum of impacts on earnings is CFA 253,920, for a cost-benefit ratio of 3.03. This means that the average cost per beneficiary is 3.03 times higher than the average discounted direct impacts on earnings. This relatively high cost-benefit ratio is driven by the fact that net earnings gains are 53% of the average transfer amounts during the program, and that direct transfers to beneficiaries represent only 46% of overall program costs.

Note that these cost-benefit ratios are conservative. They assume zero impacts beyond what we measure at endline, around 14 months after the end of the program. Program cost-effectiveness clearly depends on the sustainability of post-program impacts. Figure B5 illustrates how long the post-program impacts would need to be sustained for the program to become cost-effective (reaching a ratio of 1 or below) based on impacts on earnings and depending on the assumptions about the dissipation of post-program impacts after the endline (respectively 0, 2 or 5% of dissipation per month). For instance, assuming no dissipation of impacts, the program would become cost-effective when targeting women if post-program impacts are sustained for 7 years, or after 4-6 years when targeting individuals with low baseline earnings (respectively predicted and reported). The time to reach cost-effectiveness increases when assuming that the impact on earnings dissipates over time. Cost-benefit ratios also do not account for non-economic benefits such as those on psychological well-being mentioned above, or other externalities from the program, such as the indirect benefits of roads rehabilitation. They still provide a benchmark to assess 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 or externalities are similar across these potential improvements.

⁶³The total costs can be decomposed as follows: CFA 354,166 (\$717) for direct transfers to beneficiaries, CFA 255,189 (\$517) for other direct costs (material, team leaders and supervisors, basic training), CFA 108,230 (\$219) for skills training, and CFA 51,123 (\$10) are indirect management costs.

Columns (2) to (6) in Table 10 show how adjustments in targeting would affect the cost-effectiveness ratios. Columns (2) and (3) display cost-benefit ratios for individuals in the top and bottom quartiles of predicted impacts based on machine learning techniques. The cost-benefit ratio is nearly three times higher in the bottom quartile (7.94, column (3)) than in the top quartile of predicted impacts (2.44, column (2)). This illustrates the high cost of including marginal applicants. Compared to the cost-benefit ratio of 3.03 for the randomly assigned program, the cost-benefit ratio would improve to 2.33 by targeting directly women (column (5)), or around 2 by targeting individuals with low baseline earnings (between 2 and 2.2 depending on whether they are self-declared or predicted based on other proxies, columns 5-6). While the analysis cannot decisively indicate which targeting scenario would maximize cost-effectiveness given the confidence intervals around the impact estimates, it does highlight potential improvements in cost-effectiveness ranging between 30 and 52 percent when departing from self-targeting based on the formal minimum wage.

7 Conclusion

The Côte d’Ivoire public works program we study in this paper shares many of the features of other public works programs that have been adopted throughout the developing world in response to negative economic, political or climatic shocks. It provided a few months of employment in road rehabilitation to those willing to work at the formal minimum wage. Based on a randomized control trial and rich data collected before, during and after the program, our analysis has allowed us to assess the effectiveness of the program in improving contemporaneous and post-program outcomes among participants.

Results show that program impacts on employment are limited to shifts in the composition of employment towards the public works wage jobs during the program, with no lasting post-program impacts on the likelihood or composition of employment. Public works increase earnings during the program, but post-program impacts on earnings are limited. Savings and psychological well-being improve both during and (to a less extent) post-program. However, we see no long-lasting effects on work habits and behaviors, despite improvements during the program.

The program as currently implemented induces impacts on youths’ earnings that are substantially below program costs. This is primarily due to the limited post-program impacts and a failure of

self-selection: in an environment where informal employment is rampant, a broad cross-section of youth with outside employment opportunities self-select into public works participation. The high cost-benefit ratios also stem from the fairly high indirect cost of implementing public works programs, for instance compared to more traditional welfare programs.

We use recent machine learning techniques to document significant heterogeneity in impacts on earnings during the program, but no significant heterogeneity is found post-program. The results suggest that improvements in self-targeting or targeting are first-order program design questions, and perhaps more critical than other program design aspects related to program content, such as complementary skills training. Given the estimated distribution of individual program impacts, we show that a lower offered wage (and the subsequent change in self-targeting) would have been unlikely to improve program performance. In contrast, we show that a range of practical targeting mechanisms perform as well as the machine learning benchmark, leading to stronger impacts during the program without reductions in post-program impacts. Still, even with this improved targeting, impacts on earnings remain substantially below program costs.

While one might be tempted to conclude from our analysis that public work programs should be de-prioritized by policy makers in favor of welfare programs with more efficient targeting procedures and lower implementation costs, it is important to remember that our cost-effectiveness analysis does not take into account all possible benefits of the program, both for the beneficiaries themselves but also for non-beneficiaries. First, we do observe lasting effects on psychological well-being and savings among beneficiaries that are not included in the cost-benefit ratios. We note, however, that the post-program effects we observe are of relatively small magnitude and it is unclear, especially given the lack of similar sustained impacts on work habits and behaviors, whether the “work” component of the workfare program is responsible for these sustained effects or whether similar effects could be achieved solely with cash transfers. Second, there might be other positive externalities associated with the program, such as a reduction in crime or illegal activities due to an incapacitation effect. While we do not find much evidence of changes in youths’ engagement in risky behaviors, neither during nor after the program, it is still possible that some externalities may arise at the level of the community and hence may be difficult to measure. Lastly, we do not quantify the societal value of the upgraded infrastructure. These two latter potential benefits are a specific feature of public work programs and might be particularly large in post-conflict environments with destroyed physical infrastructure and a high need for

social stabilization. Still, these externalities would need to be very (and likely unrealistically) large for the program to be cost-effective.

References

- ALATAS, V., A. BANERJEE, R. HANNA, B. A. OLKEN, AND J. TOBIAS (2012): “Targeting the poor: Evidence from a field experiment in Indonesia,” *American Economic Review*, 102, 1206–40.
- ALATAS, V., R. PURNAMASARI, M. WAI-POI, A. BANERJEE, B. A. OLKEN, AND R. HANNA (2016): “Self-targeting: Evidence from a field experiment in Indonesia,” *Journal of Political Economy*, 124, 371–427.
- ALIK-LAGRANGE, A., O. ATTANASIO, C. MEGHIR, S. POLANÍA-REYES, AND M. VERA-HERNÁNDEZ (2017): “Work pays: Different benefits of a workfare program in Colombia,” Mimeo.
- ALIK-LAGRANGE, A. AND M. RAVALLION (2018): “Workfare versus transfers in rural India,” *World Development*, 112, 244–258.
- ALMEIDA, R. K. AND E. GALASSO (2010): “Jump-starting self-employment? Evidence for welfare participants in Argentina,” *World Development*, 38, 742–755.
- AMARAL, S., S. BANDYOPADHYAY, AND R. SENSARMA (2015): “Employment programmes for the poor and female empowerment: The effect of NREGS on gender-based violence in India,” *Journal of Interdisciplinary Economics*, 27, 199–218.
- APOSTOLIDIS, T. AND N. FIEULAIN (2004): “Validation française de l’échelle de temporalité: The Zimbardo Time Perspective Inventory (ZTPI),” *European Review of Applied Psychology*, 54, 207–2017.
- ASCARZA, E. (2018): “Retention futility: Targeting high-risk customers might be ineffective,” *Journal of Marketing Research*, 55, 80–98.
- ATHEY, S. AND G. IMBENS (2016): “Recursive partitioning for heterogeneous causal effects,” *Proceedings of the National Academy of Sciences*, 113, 7353–7360.

- (2017a): “The state of applied econometrics: Causality and policy evaluation,” *Journal of Economic Perspectives*, 31, 3–32.
- ATHEY, S. AND G. W. IMBENS (2017b): “The econometrics of randomized experiments,” in *Handbook of Economic Field Experiments*, Elsevier, vol. 1, 73–140.
- ATHEY, S., J. TIBSHIRANI, AND S. WAGER (2019): “Generalized random forests,” *Annals of Statistics*, 47, 1148–1178.
- ATHEY, S. AND S. WAGER (2017): “Efficient policy learning,” Arxiv preprint, arxiv:1702.02896.
- BANERJEE, A., P. NIEHAUS, AND T. SURI (2019): “Universal Basic Income in the Developing World,” *Annual Review of Economics*, 11, 959–983.
- BASURTO, M. P., P. DUPAS, AND J. ROBINSON (2020): “Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi,” *Journal of Public Economics*, 185.
- BEEGLE, K., E. GALASSO, AND J. GOLDBERG (2017): “Direct and indirect effects of Malawi’s public works program on food security,” *Journal of Development Economics*, 128, 1–23.
- BERG, E., S. BHATTACHARYYA, D. RAJASEKHAR, AND R. MANJULA (2018): “Can public works increase equilibrium wages? Evidence from India’s National Rural Employment Guarantee,” *World Development*, 103, 239–254.
- BERTRAND, M., B. CRÉPON, A. MARGUERIE, AND P. 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. AND S. COATE (1992): “Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs,” *The American Economic Review*, 82, 249–261.
- BHATTACHARYA, D. AND P. DUPAS (2012): “Inferring welfare maximizing treatment assignment under budget constraints,” *Journal of Econometrics*, 167, 168–196.
- BITLER, M. P., J. B. GELBACH, AND H. W. HOYNES (2006): “What mean impacts miss: Distributional effects of welfare reform experiments,” *American Economic Review*, 96, 988–1012.

- (2017): “Can Variation in Subgroups’ Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment,” *Review of Economics and Statistics*, 99, 683–697.
- BLAIR, G. AND K. IMAI (2012): “Statistical analysis of list experiments,” *Political Analysis*, 20, 47–77.
- BLATTMAN, C. AND L. RALSTON (2015): “Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs,” SSRN Electronic Journal.
- BREDA, T., J. GRENET, M. MONNET, AND C. VAN EFFENTERRE (2020): “Do Female Role Models reduce the gender gap in science? Evidence from French high schools,” IZA Discussion Paper No. 13163.
- CASEY, K., R. GLENNERSTER, AND E. MIGUEL (2012): “Reshaping institutions: Evidence on aid impacts using a preanalysis plan,” *The Quarterly Journal of Economics*, 127, 1755–1812.
- CHERNOZHUKOV, V., M. DEMIRER, E. DUFLO, AND I. FERNANDEZ-VAL (2020): “Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments,” Arxiv preprint, arxiv:1712.04802.
- CHRISTIAENSEN, L. AND P. PREMAND (2017): “Cote d’Ivoire Jobs Diagnostic : Employment, Productivity, and Inclusion for Poverty Reduction.” World Bank, Washington DC.
- CRÉPON, B. AND P. PREMAND (2019): “Direct and Indirect Effects of Subsidized Dual Apprenticeships,” IZA Discussion Paper No. 12793.
- CÔTÉ, G., P. GOSSELIN, AND I. DAGENAIS (2013): “Évaluation multidimensionnelle de la régulation des émotions : propriétés psychométriques d’une version francophone du Difficulties in Emotion Regulation Scale,” *Journal de Thérapie Comportementale et Cognitive*, 23.
- DATT, G. AND M. RAVALLION (1994): “Transfer benefits from public-works employment: Evidence for rural India,” *The Economic Journal*, 1346–1369.
- DAVIS, J. AND S. B. HELLER (2017): “Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs,” *American Economic Review*, 107, 546–50.
- (2020): “Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs,” *The Review of Economics and Statistics*, 102, 664–677.

- DEININGER, K., H. K. NAGARAJAN, AND S. K. SINGH (2016): “Short-term effects of India’s employment guarantee program on labor markets and agricultural productivity,” Policy Research Working Paper No. 7665, The World Bank, Washington DC.
- DJEBBARI, H. AND J. SMITH (2008): “Heterogeneous impacts in PROGRESA,” *Journal of Econometrics*, 145, 64–80.
- DROITCOUR, J., R. A. CASPAR, M. L. HUBBARD, T. L. PARSLEY, W. VISSCHER, AND T. M. EZZATI (1991): “The Item Count Technique as a method of indirect questioning: A review of its development and a case study application,” in *Measurement errors in surveys*, ed. by P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman, John Wiley & Sons Inc., 185–210.
- FETZER, T. (2020): “Can Workfare Programs Moderate Conflict? Evidence from India,” *Journal of the European Economic Association*, 18, 3337–3375.
- FILMER, D., L. FOX, K. BROOKS, A. GOYA, T. MENGISTAE, P. PREMAND, D. RINGOLD, S. SHARMA, AND S. ZORYA (2014): “*Youth Employment in sub-Saharan Africa*”, World Bank, Africa Development Series.
- GALASSO, E. AND M. RAVALLION (2004): “Social protection in a crisis: Argentina’s Plan Jefes y Jefas,” *The World Bank Economic Review*, 18, 367–399.
- GALASSO, E., M. RAVALLION, AND A. SALVIA (2004): “Assisting the transition from workfare to work: A randomized experiment,” *ILR Review*, 58, 128–142.
- GEHRKE, E. AND R. HARTWIG (2018): “Productive effects of public works programs: What do we know? What should we know?” *World Development*, 107, 111–124.
- GENTILINI, U., M. GROSH, J. RIGOLINI, AND R. YEMTSOV (2020): “Exploring Universal Basic Income : A Guide to Navigating Concepts, Evidence, and Practices,” Washington, DC: World Bank.
- GILLIGAN, D. O., J. HODDINOTT, AND A. S. TAFFESSE (2009): “The impact of Ethiopia’s Productive Safety Net Programme and its linkages,” *The Journal of Development Studies*, 45, 1684–1706.
- GOODMAN, R., H. MELTZER, AND V. BAILEY (1998): “The Strengths and Difficulties Question-

- naire: A pilot study on the validity of the self-report version,” *European Child and Adolescent Psychiatry*, 7, 125–130.
- HANNA, R. AND B. A. OLKEN (2018): “Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries,” *Journal of Economic Perspectives*, 32, 201–26.
- HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): “Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts,” *The Review of Economic Studies*, 64, 487–535.
- IMBERT, C. AND J. PAPP (2015): “Labor market effects of social programs: Evidence from india’s employment guarantee,” *American Economic Journal: Applied Economics*, 7, 233–63.
- (2019): “Short-term Migration, Rural Public Works, and Urban Labor Markets: Evidence from India,” *Journal of the European Economic Association*, 18, 927–963.
- ISLAM, M. AND A. SIVASANKARAN (2015): “How does Child Labor respond to changes in Adult Work Opportunities? Evidence from NREGA,” Mimeo.
- JALAN, J. AND M. RAVALLION (2003): “Estimating the benefit incidence of an antipoverty program by propensity-score matching,” *Journal of Business & Economic Statistics*, 21, 19–30.
- KITAGAWA, T. AND A. TETENOV (2018): “Who should be treated? Empirical welfare maximization methods for treatment choice,” *Econometrica*, 86, 591–616.
- KNAUS, M. C., M. LECHNER, AND A. STRITTMATTER (2020a): “Heterogeneous employment effects of job search programmes: A machine learning approach,” *Journal of Human Resources*, 0718–9615R1.
- (2020b): “Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence,” *The Econometrics Journal*, utaa014.
- KUHN, M. (2008): “Building Predictive Models in R Using the caret Package,” *Journal of Statistical Software*, 28, 1–26.
- KÜNZEL, S., J. SEKHON, P. BICKEL, AND B. YU (2017): “Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning,” *Proceedings of the National Academy of Sciences*, 116.

- LI, T., S. SEKHRI, ET AL. (2013): “The Unintended Consequences of Employment-Based Safety Net Programs,” Manuscript. University of Virginia.
- LIST, J. A., A. M. SHAIKH, AND Y. XU (2019): “Multiple hypothesis testing in experimental economics,” *Experimental Economics*, 22, 773–793.
- MANSKI, C. F. (2004): “Statistical treatment rules for heterogeneous populations,” *Econometrica*, 72, 1221–1246.
- MCKENZIE, D. AND D. SANSONE (2019): “Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria,” *Journal of Development Economics*, 141, 102369.
- MILLER, J. D. (1984): “A new survey technique for studying deviant behavior,” George Washington University, Washington, DC.
- MORIN, A., G. MOULLEC, C. MAÏANO, L. LAYET, J.-L. JUST, AND N. G. (2011): “Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CESD) in French Clinical and Non-Clinical Adults,” *Epidemiology and Public Health/Revue d’Épidémiologie et de Santé Publique*, 59, 327–340.
- MURALIDHARAN, K., P. NIEHAUS, AND S. SUKHTANKAR (2016): “Building state capacity: Evidence from biometric smartcards in India,” *American Economic Review*, 106, 2895–2929.
- (2017): “General equilibrium effects of (improving) public employment programs: Experimental evidence from India,” National Bureau of Economic Research Working Paper No. 23838.
- MURGAI, R., M. RAVALLION, AND D. VAN DE WALLE (2015): “Is Workfare Cost-effective against Poverty in a Poor Labor-Surplus Economy?” *The World Bank Economic Review*, 30, 413–445.
- NIE, X. AND S. WAGER (2017): “Quasi-oracle estimation of heterogeneous treatment effects,” Arxiv preprint, arxiv:1712.04912.
- (2020): “Quasi-oracle estimation of heterogeneous treatment effects,” *Biometrika*, asaa076.
- OLKEN, B. A. (2015): “Promises and perils of pre-analysis plans,” *Journal of Economic Perspectives*, 29, 61–80.

- PREMAND, P. AND P. SCHNITZER (2020): “Efficiency, Legitimacy, and Impacts of Targeting Methods: Evidence from an Experiment in Niger,” *The World Bank Economic Review*, lhaa019.
- RAVALLION, M. (1991): “Reaching the Rural Poor through Public Employment: Arguments, Evidence, and Lessons from South Asia,” *The World Bank Research Observer*, 6, 153–175.
- RAVALLION, M., G. DATT, AND S. CHAUDHURI (1993): “Does Maharashtra’s Employment Guarantee Scheme guarantee employment? Effects of the 1988 wage increase,” *Economic Development and Cultural Change*, 41, 251–275.
- RAVALLION, M., E. GALASSO, T. LAZO, AND E. PHILIPP (2005): “What can ex-participants reveal about a program’s impact?” *Journal of Human Resources*, 40, 208–230.
- RAVI, S. AND M. ENGLER (2015): “Workfare as an effective way to fight poverty: The case of India’s NREGS,” *World Development*, 67, 57–71.
- ROSAS, N. AND S. SABARWAL (2016): “Public Works as a Productive Safety Net in a Post-Conflict Setting : Evidence from a Randomized Evaluation in Sierra Leone,” Policy research working paper no. 7580, World Bank, Washington, DC.
- SHAH, M. AND B. M. STEINBERG (2019): “Workfare and Human Capital Investment: Evidence from India,” *Journal of Human Resources*.
- STRITTMATTER, A. (2018): “What Is the Value Added by Using Causal Machine Learning Methods in a Welfare Experiment Evaluation?” Arxiv preprint, arxiv:1812.06533.
- SUBBARAO, K., C. DEL NINNO, C. ANDREWS, AND C. RODRÍGUEZ-ALAS (2012): *Public works as a safety net: Design, evidence, and implementation*, World Bank, Washington DC.
- VALLIERES, E. F. AND R. J. VALLERAND (1990): “Traduction et Validation Canadienne-Française de L’échelle de L’estime de Soi de Rosenberg,” *International Journal of Psychology*, 25, 305–316.
- WAGER, S. AND S. ATHEY (2018): “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests,” *Journal of the American Statistical Association*, 113, 1228–1242.

Table 1: Baseline Summary Statistics and Balance Checks

	(1)	(2)	(3)	(4)	(5)
	Mean in Treatment group (pooled)	Mean in Control group	Balance Test (p-value) (1)-(2)	Balance Test between 4 arms (p-value)	Observations
Female	0.43	0.42	0.710	0.850	3781
Live in urban area	0.94	0.93	0.337	0.579	3736
Age	24.58	24.67	0.569	0.227	3736
Nationality: Ivorian	0.96	0.97	0.167	0.265	3736
Nb of children	0.90	0.94	0.485	0.997	3736
Education					
Primary education not completed	0.49	0.49	0.944	0.839	3736
Up to primary education completed (CEPE)	0.23	0.22	0.593	0.417	3736
Up to lower secondary education completed (BEPC)	0.18	0.16	0.480	0.725	3736
Upper secondary education completed (BAC or more)	0.09	0.12	0.0724	0.576	3736
Enrolled at school	0.05	0.07	0.0584	0.681	3736
Has participated in vocational training	0.36	0.40	0.128	0.361	3733
Of which: traditional apprenticeship	0.72	0.71	0.716	0.512	1465
Household					
Household size	6.12	6.10	0.915	0.267	3735
Nb of children (< 18 years old)	2.12	2.10	0.911	0.579	3736
Is head of household	0.23	0.23	0.983	0.869	3736
Share of members working (last 7 days)	0.54	0.55	0.290	0.822	3735
Number of rooms in dwelling	3.17	3.10	0.581	0.483	3736
Household assets					
Total Nb of assets	13.53	13.5	0.960	0.404	3736
Nb of transportation assets	0.74	0.83	0.466	0.230	3736
Nb of agricultural assets	4.61	4.52	0.890	0.772	3736
Nb of household durables	1.59	1.60	0.882	0.377	3736
Nb of communication assets	6.60	6.55	0.792	0.639	3736
Savings					
Has saved (last 3 months)	0.49	0.51	0.438	0.619	3736
of which: share of savings (stock) in formal instrument	0.25	0.27	0.516	0.516	1811
Has a savings account	0.11	0.09	0.234	0.936	3736
Savings stock (CFA)	27644.3	26426.1	0.602	0.964	3685
Self-reported constraints to repay loans	0.19	0.23	0.0538	0.967	3736
Self-reported constraints to access credit	0.50	0.50	0.889	0.733	3736
Expenditures					
Nb of days without a meals (last 7 days)	0.80	0.83	0.654	0.706	3736
Self-reported constraints for basic needs expenditures	0.71	0.73	0.187	0.945	3736
Transportation expenditures (last 7 days, CFA)	1774.3	1679.5	0.495	0.724	3732
Communication expenditures (last 7 days, CFA)	1595.4	1540.8	0.721	0.739	3730
Employment					
Has an activity	0.78	0.82	0.0782	0.912	3736
Searched for a job (last month)	0.74	0.76	0.397	0.337	3736
Risk and time preferences					
Risk aversion level (scale 0 to 10, 0=very risk averse)	4.72	4.80	0.620	0.607	3736
Is risk averse (based on hypothetical lotteries)	0.74	0.69	0.0124	0.670	3736
Patience level (scale 0 to 10, 10=very patient)	3.25	3.25	0.999	0.931	3733
Preference for present (actualization rate for 1 month)	0.59	0.59	0.899	0.210	3736
Skills (% of success in answers or tasks at each test)					
Cognitive (deduction, Raven Test)	0.23	0.23	0.930	0.256	3730
Spatial vision (NV7 Test)	0.27	0.27	0.679	0.0912	3736
Dexterity (Nuts Test)	0.39	0.38	0.119	0.263	3731
Dexterity (Bolts Test)	0.33	0.34	0.243	0.246	3721

The table includes all baseline respondents interviewed at endline. Means (columns (1) and (2)) and difference in means (column (3)) estimated using endline estimation weights (see details in Appendix G). Robust standard errors clustered at (broad) brigade level. Assets and savings stock variables winsorized at the 99th percentile.

Table 2: Characteristics of Applicants and National Population of Urban Youths

	Experimental Control Group	All Urban Youths (18-30)
Gender		
Male	55.7	50.8
Female	44.3	49.2
Educational attainment		
Primary education not completed	48.5	38.9
Up to primary education completed (CEPE)	23.2	18.5
Up to lower secondary education completed (BEPC)	16.6	20.0
Upper secondary education completed (BAC or more)	11.7	22.7
Total	100.0	100.0
Asset Index (z-score)	-0.046	0.000
Occupation		
Inactive, at school	2.9	23.6
Inactive, not at school	7.1	11.5
Unemployed	6.4	10.5
Wage-Employed (including informal)	47.0	25.4
Self-Employed (non agricultural)	30.1	26.0
Self-Employed in agriculture	6.5	3.1
Total	100.0	100.0

The first column displays average characteristics of program applicants. Gender, educational attainment and assets are measured at baseline. Employment status is measured at midline. (This is because the midline survey was collected between November 2013 and January 2014, which was closest to the timing of a national employment survey). The second column displays average characteristics of youths between 18 and 30 years old who live in urban areas. This is based on the 2013 National Employment Survey (collected in January 2014). The asset index is based on a principal component analysis of household assets. We include assets that are measured in both surveys: carts, wheelbarrows, bicycles, motorcycles, refrigerators, freezers, air conditioning units, fans, stoves, computers, radio stations, televisions and TV antenna, video players, cell phones, landline phones and cars. Z scores are predicted using the first component, and standardized in the national employment survey.

Table 3: Impacts *during* and *post* program, economic outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Employed	Wage- Employed (in at least 1 activity)	Self- Employed (in at least 1 activity)	Total Hours worked (weekly)	Hours worked in Wage- Empl. (weekly)	Hours worked in Self-Empl. (weekly)	Total earnings in CFA (monthly)	Ln total earnings (monthly)	Earnings in Wage- Empl. in CFA (monthly)	Earnings in Self-Empl. in CFA (monthly)	Total ex- penditures in CFA (monthly)	Savings in CFA (stock)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)												
Public Works Treatment (ITT)	0.14*** (0.015)	0.48*** (0.024)	-0.096*** (0.020)	4.89*** (1.25)	15.5*** (1.29)	-5.69*** (0.94)	27082.9*** (2824.9)	2.95*** (0.19)	36799.0*** (1472.5)	-5715.4*** (1214.6)	14529.3*** (1441.4)	39785.7*** (2389.2)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.85	0.49	0.35	39.18	20.79	11.28	42841.22	7.87	20188.33	12753.65	47233.52	19250.05
Observations	2958	2958	2958	2958	2958	2958	2912	2912	2912	2912	2945	2958
Perm. p-value: no effects	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)												
Public Works Treatment (ITT)	0.015 (0.015)	0.0068 (0.019)	0.010 (0.022)	1.34 (1.28)	-0.61 (1.14)	1.70 (1.11)	4360.6** (1906.5)	-0.037 (0.18)	-452.7 (1002.6)	4005.2** (1790.8)	1361.7 (1406.9)	11505.2*** (3136.5)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.86	0.51	0.36	40.49	22.06	13.26	43481.10	8.42	20706.18	18872.95	50700.71	46348.14
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3934	3814	3934
Perm. p-value: no effects	0.321	0.721	0.647	0.298	0.589	0.129	0.026	0.838	0.652	0.029	0.341	0.000
Panel C: <i>Post</i> program impacts (by treatment arms) (12 to 15 months after program ends)												
Public Works Treatment (ITT)	0.011 (0.019)	0.0081 (0.027)	0.0035 (0.028)	-0.76 (1.70)	-0.71 (1.57)	-0.12 (1.29)	2800.5 (2138.7)	-0.18 (0.22)	312.2 (1260.9)	2168.7 (1852.5)	925.7 (1536.8)	10429.5*** (3410.3)
Self-Empl. training (SET)	0.011 (0.018)	-0.018 (0.028)	0.021 (0.032)	3.42* (1.98)	0.46 (1.80)	2.77 (1.77)	4229.3 (3201.3)	0.22 (0.26)	-1591.8 (1280.0)	5595.5* (2910.0)	278.1 (1383.1)	7169.5 (4729.2)
Wage-Empl. training (WET)	0.0018 (0.019)	0.014 (0.028)	0.000048 (0.032)	3.12 (2.20)	-0.14 (1.65)	2.89 (1.90)	637.5 (2204.3)	0.24 (0.21)	-792.6 (1303.0)	135.8 (2302.8)	1077.6 (1716.7)	-3798.3 (4387.5)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.86	0.51	0.36	40.49	22.06	13.26	43481.10	8.42	20706.18	18872.95	50700.71	46348.14
p-value: PW+SET=0	0.238	0.672	0.375	0.168	0.870	0.153	0.018	0.878	0.289	0.004	0.459	0.000
p-value: PW+WET=0	0.489	0.350	0.906	0.198	0.534	0.113	0.215	0.818	0.704	0.437	0.285	0.142
p-value: SET=WET	0.672	0.219	0.515	0.904	0.698	0.964	0.360	0.963	0.538	0.175	0.627	0.039
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3934	3814	3934
Perm. p-value: PW+SET=0	0.246	0.675	0.384	0.180	0.869	0.164	0.021	0.879	0.294	0.005	0.468	0.000
Perm. p-value: PW+WET=0	0.502	0.364	0.908	0.208	0.544	0.115	0.230	0.816	0.709	0.452	0.302	0.153
Perm. p-value: SET=WET	0.688	0.230	0.513	0.907	0.705	0.967	0.364	0.963	0.548	0.178	0.639	0.042

ITT estimates in panels A and B based on specification in equation 6. Estimates in panel C based on specification in equation 7. Robust standard errors clustered at (broad) brigade level in parentheses. *Hours*, *earnings*, *expenditures*, and *savings* winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y + 1)$. Permutation tests use 10000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 4: Impacts *during* and *post* program, alternative definitions of earnings and savings outcomes

	Total earnings			Self-Employment earnings			Wage-Employment earnings		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Ln, winsorized at 97%	Ln, winsorized at 99%	Level, winsorized at 99%	Ln, winsorized at 97%	Ln, winsorized at 99%	Level, winsorized at 99%	Ln, winsorized at 97%	Ln, winsorized at 99%	Level, winsorized at 99%
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)									
Public Works Treatment (ITT)	2.95*** (0.19)	2.95*** (0.20)	24380.2*** (5998.7)	-1.04*** (0.20)	-1.06*** (0.20)	-11303.4*** (4091.6)	5.93*** (0.25)	5.93*** (0.25)	37181.6*** (4033.5)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	7.87	7.89	54626.16	3.23	3.26	23508.66	4.62	4.64	28163.73
Observations	2912	2912	2912	2912	2912	2912	2912	2912	2912
Perm. p-value: no effects	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)									
Public Works Treatment (ITT)	-0.037 (0.18)	-0.028 (0.18)	7597.5*** (2380.5)	0.22 (0.23)	0.23 (0.23)	7802.7*** (2250.6)	-0.20 (0.19)	-0.20 (0.19)	-542.8 (1145.3)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	8.42	8.43	45690.44	3.56	3.57	20487.91	5.04	5.05	22036.94
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934
Perm. p-value: no effects	0.846	0.880	0.002	0.338	0.326	0.001	0.295	0.304	0.642
Panel C: <i>Post</i> program impacts (by treatment arms) (12 to 15 months after program ends)									
Public Works Treatment (ITT)	-0.18 (0.22)	-0.18 (0.22)	5558.1** (2665.5)	0.13 (0.27)	0.14 (0.28)	5098.3** (2288.8)	-0.22 (0.24)	-0.23 (0.24)	-21.1 (1486.7)
Self-Empl. training (SET)	0.22 (0.26)	0.23 (0.26)	5171.9 (3709.3)	0.28 (0.34)	0.29 (0.35)	7999.6** (3763.3)	-0.065 (0.26)	-0.064 (0.26)	-1327.7 (1531.2)
Wage-Empl. training (WET)	0.24 (0.21)	0.24 (0.21)	1188.7 (3179.3)	-0.0027 (0.32)	-0.0025 (0.32)	439.3 (3699.5)	0.15 (0.27)	0.15 (0.28)	-299.5 (1544.4)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	8.42	8.43	45690.44	3.56	3.57	20487.91	5.04	5.05	22036.94
p-value: PW+SET=0	0.878	0.840	0.003	0.184	0.168	0.000	0.246	0.246	0.318
p-value: PW+WET=0	0.818	0.787	0.049	0.706	0.689	0.165	0.764	0.770	0.826
p-value: SET=WET	0.963	0.972	0.392	0.462	0.451	0.173	0.456	0.452	0.476
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934
Perm. p-value: PW+SET=0	0.877	0.847	0.003	0.194	0.177	0.001	0.254	0.256	0.322
Perm. p-value: PW+WET=0	0.826	0.792	0.053	0.708	0.693	0.166	0.766	0.779	0.836
Perm. p-value: SET=WET	0.962	0.975	0.398	0.468	0.465	0.180	0.469	0.457	0.481

ITT estimates in panels A and B based on specification in equation 6. Estimates in panel C based on specification in equation 7. Robust standard errors clustered at (broad) brigade level in parentheses. For variables (y) in logarithms we take $\ln(y + 1)$. Permutation tests use 10000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 5: Impacts *during* and *post* program, well-being, behaviors and work habits

	(1) Well-being index (z-score)	(2) Behavior index (z-score)	(3) Rest at 6 am of prev. day	(4) Leisure at 6 am of prev. day	(5) Work at 6 am of prev. day	(6) Rest at 10 pm of prev. day	(7) Leisure at 10 pm of prev. day	(8) Work at 10 pm of prev. day	(9) Risky behavior (index)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)									
Public Works Treatment (ITT)	0.20*** (0.046)	0.12*** (0.045)	-0.14*** (0.018)	-0.044*** (0.011)	0.33*** (0.032)	0.071*** (0.019)	-0.054*** (0.015)	-0.028** (0.012)	-0.18 (0.14)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.03	-0.02	0.24	0.09	0.20	0.68	0.18	0.08	0.62
Observations	2934	2946	2955	2955	2955	2953	2953	2953	2956
Perm. p-value: no effects	0.000	0.006	0.000	0.000	0.000	0.000	0.001	0.015	0.340
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)									
Public Works Treatment (ITT)	0.11*** (0.041)	-0.012 (0.040)	0.025 (0.019)	-0.0029 (0.012)	0.0080 (0.016)	-0.000063 (0.018)	0.0011 (0.015)	0.0061 (0.011)	-0.074 (0.10)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.05	0.01	0.26	0.09	0.24	0.68	0.19	0.09	0.63
Observations	3932	3933	3933	3933	3933	3933	3933	3933	3934
Perm. p-value: no effects	0.009	0.774	0.202	0.812	0.619	0.996	0.946	0.572	0.412
Panel C: <i>Post</i> program impacts (by treatment arms) (12 to 15 months after program ends)									
Public Works Treatment (ITT)	0.14*** (0.052)	0.025 (0.050)	0.021 (0.022)	0.0036 (0.014)	0.021 (0.021)	-0.00080 (0.023)	0.0037 (0.019)	0.0041 (0.014)	0.0025 (0.13)
Self-empl.training (SET)	-0.0068 (0.051)	-0.039 (0.045)	-0.0075 (0.032)	-0.0036 (0.013)	-0.016 (0.023)	0.0027 (0.023)	-0.0085 (0.019)	0.0030 (0.012)	-0.096 (0.13)
Wage-empl. training (WET)	-0.075 (0.047)	-0.077 (0.052)	0.022 (0.026)	-0.017 (0.014)	-0.025 (0.022)	-0.00038 (0.022)	0.00041 (0.020)	0.0033 (0.015)	-0.14 (0.15)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.05	0.01	0.26	0.09	0.24	0.68	0.19	0.09	0.63
p-value: PW+SET=0	0.011	0.762	0.648	0.995	0.798	0.929	0.790	0.579	0.451
p-value: PW+WET=0	0.187	0.293	0.076	0.383	0.834	0.957	0.839	0.583	0.324
p-value: SET=WET	0.169	0.410	0.230	0.385	0.652	0.895	0.680	0.981	0.757
Observations	3932	3933	3933	3933	3933	3933	3933	3933	3934
Perm. p-value: PW+SET=0	0.011	0.774	0.656	0.996	0.807	0.930	0.796	0.587	0.474
Perm. p-value: PW+WET=0	0.198	0.295	0.076	0.402	0.842	0.959	0.840	0.597	0.329
Perm. p-value: SET=WET	0.178	0.418	0.239	0.396	0.651	0.899	0.688	0.983	0.766

ITT estimates in panels A and B based on specification in equation 6. Estimates in panel C based on specification in equation 7. The definitions of the *well-being index*, the *behavior index*, and the *risky behavior index* are discussed in Appendix E (the estimation of impacts on the *risky behavior index* is explained in table A4). Tables A2, A3, and A4 present estimates for individual components of these three indices. Robust standard errors clustered at (broad) brigade level in parentheses. Permutation tests use 10000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 6: Heterogeneity in impacts on earnings *during* and *post* program, machine learning results

	(1) Ln total earnings (Monthly)	(2) Ln total earnings (Monthly)	(3) Total earnings (Monthly)	(4) Total earnings (Monthly)
	<i>Midline</i>	<i>Endline</i>	<i>Midline</i>	<i>Endline</i>
Panel A: Predicted average treatment effect and heterogeneity loading parameter				
ATE (β_1)	2.642 (2.336,2.947) [0.000]	-0.0297 (-0.466,0.397) [1.000]	24363.0 (17618.5,31150.3) [0.000]	3314.0 (-1842.7,8523.0) [0.414]
HET (β_2)	1.231 (0.862,1.599) [0.000]	0.111 (-0.515,0.721) [0.970]	0.390 (0.00909,0.777) [0.089]	0.416 (-1.149,5.384) [0.756]
Best ML method	Generalized Random forest	Random forest	Random forest	R-Learner
Panel B: By quartile of predicted impacts on earnings <i>during</i> program (using <i>midline</i>) (GATES)				
Mean in quartile 1 (0 to 25%)	1.465 (0.867,2.062)	-0.306 (-1.322,0.729)	14659.7 (1669.7,28176.4)	460.9 (-11931.0,12389.1)
Mean in quartile 2 (25 to 50%)	2.296 (1.695,2.887)	-0.311 (-1.341,0.729)	23172.3 (9865.0,36605.8)	2793.9 (-9810.5,15283.1)
Mean in quartile 3 (50 to 75%)	2.832 (2.215,3.425)	-0.0372 (-1.076,0.992)	29778.2 (16363.7,42745.5)	5658.1 (-6700.1,17794.3)
Mean in quartile 4 (75 to 100%)	3.966 (3.352,4.560)	-0.252 (-1.289,0.799)	31670.9 (18254.4,44906.7)	4490.7 (-7835.4,16887.3)
P-value all coefficients are equal	0.000	1.000	0.281	1.000
Best ML method	Generalized Random forest	Generalized Random forest	Random forest	Random forest
Panel C: By quartile of predicted impacts on earnings <i>post</i> program (using <i>endline</i>) (GATES)				
Mean in quartile 1 (0 to 25%)	2.860 (2.240,3.474)	-0.0834 (-0.963,0.803)	20120.2 (7275.6,32824.1)	2884.7 (-7540.5,12739.3)
Mean in quartile 2 (25 to 50%)	2.641 (2.036,3.253)	-0.0669 (-0.950,0.791)	26809.6 (14045.8,39322.7)	4020.6 (-6197.8,14214.9)
Mean in quartile 3 (50 to 75%)	2.426 (1.795,3.041)	-0.0609 (-0.915,0.796)	27443.4 (14798.2,40110.1)	3920.1 (-6313.2,14117.8)
Mean in quartile 4 (75 to 100%)	2.652 (2.020,3.277)	0.0360 (-0.823,0.886)	23655.1 (11190.1,36131.8)	2549.3 (-7586.6,13104.9)
P-value all coefficients are equal	0.100	0.888	0.477	1.000
Best ML method	Random forest	Random forest	R-Learner	R-Learner

Heterogeneity analysis based on the approach in Chernozhukov et al. (2020) (see discussion in Section 4.2 and Appendix F.2). Columns (1) and (3) (respectively columns (2) and (4)) focus on outcomes at midline (respectively endline). Estimates in Panel A are based on equation 9. P-value (in brackets) for β_1 tests the hypothesis of no effect. P-value (in brackets) for β_2 tests the hypothesis of no heterogeneity. GATES estimates are based on the specification in equation 10. Panel B (respectively Panel C) shows impacts per quartile of the predicted treatment effects at midline (respectively endline). The best predictions are reported. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$ (see appendix table A6 for comparisons across algorithms). Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted confidence intervals at 90% in parentheses and adjusted p-values for partition uncertainty. *Earnings* variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y+1)$.

Table 7: Baseline characteristics of the bottom and top quartiles of the distribution of predicted impacts on (ln) earnings *during* program

	(1) Mean in 1st quartile	(2) Mean in 4th quartile	(3) Test (1)-(2) (p-value)
Individual characteristics			
Female	0.15	0.53	0
Live in urban area	0.94	0.96	0.608
Age	25.22	24.29	0
Nb of children	0.78	0.86	0.316
Education			
Years of education	11	9.69	0.026
Primary education not completed	0.40	0.48	0.071
Has participated in vocational training	0.52	0.28	0
Household characteristics and assets			
Household size	5.91	6.45	0.060
Is head of household	0.39	0.090	0
Total Nb of assets	0.63	0.43	0
Nb of transportation assets	0.83	0.61	0.030
Nb of agricultural assets	5.39	3.39	0.041
Nb of household durables	1.96	1.63	0.031
Nb of communication assets	7.57	6.68	0.011
Employment			
Has an activity	1	0.44	0
Is Wage-Employed	0.55	0.09	0
Is Self-Employed	0.52	0.14	0
Nb of activities	1.38	0.46	0
Total Earnings (monthly, CFA)	49673.3	653.4	0
Searched for a job (last month)	0.61	0.52	1
Savings, constraints and expenditures			
Has saved (last 3 months)	0.72	0.25	0
Of which: share of formal savings (cond. on savings)	0.33	0.17	0.001
Has a savings account	0.20	0.06	0
Savings Stock (CFA)	65924.9	6795.1	0
Self-reported constraints to repay loans	0.20	0.18	0.513
Self-reported constraints to access credit	0.43	0.58	0
Self-reported constraints for basic needs expenditures	0.66	0.71	0.257
Transportation expenditures (last 7 days, CFA)	3041.8	1021.0	0
Communication expenditures (last 7 days, CFA)	3221.0	655.9	0
Cognitive skills and risk preference			
Cognitive (deduction, Raven Test)	0.24	0.23	0.738
Dexterity (Nuts Test)	0.37	0.39	0.124
Dexterity (Bolts Test)	0.34	0.33	0.458
Positive affect (CES-D scale, Nb positive days)	6.45	6.09	0.084
Positive attitude towards the future (ZTPI scale)	29.33	29.11	0.676
Is Risk averse (based on hypothetical lotteries)	0.71	0.70	1
Patience (scale 0 to 10, 10=very patient)	3.44	3.28	0.872
Preference for present (actualization rate for 1 month)	0.58	0.58	1

Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Column (1) (respectively (2)) displays average characteristics of the bottom (respectively top) quartile of the distribution of predicted impacts on (ln) earnings during the program. Column (3) reports p-values for a test of equality between the top and bottom quartile. Reported results are based on the algorithm with best predictions for midline : Generalized Random forest. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$ (see appendix table A6 for comparisons across algorithms). Means by quartile are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted p-values for partition uncertainty. *Household assets* and *savings stock* variables winsorized at the 99th percentile.

Table 8: Impacts *during* and *post* program on main outcomes, by quartile of predicted impacts on (ln) earnings *during* the program

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Employed	Wage-Empl. (in at least 1 activity)	Self-Empl. (in at least 1 activity)	Total Hours worked (weekly)	Total earnings in CFA (monthly)	Ln total earnings (monthly)	Earnings in Wage-Empl. in CFA (monthly)	Earnings in Self-Empl. in CFA (monthly)	Total expenditures in CFA (monthly)	Savings in CFA (stock)	Well-being index (z-score)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)											
Mean in quartile 1 (0 to 25%)	0.039 (-0.0078, 0.087)	0.35 (0.28, 0.42)	-0.059 (-0.16, 0.046)	0.81 (-4.39, 6.08)	14805 (1639, 27760)	1.46 (0.87, 2.06)	28834 (21922, 35807)	-6859 (-12362, -1345)	15414 (6633, 24276)	35561 (23460, 47674)	0.23 (0.025, 0.44)
Mean in quartile 2 (25 to 50%)	0.093 (0.046, 0.14)	0.40 (0.33, 0.46)	-0.11 (-0.21, -0.0051)	1.12 (-4.09, 6.35)	21937 (8698, 35146)	2.30 (1.70, 2.89)	33479 (26487, 40457)	-6486 (-12019, -936)	10622 (1709, 19388)	32056 (20219, 44174)	0.21 (0.0014, 0.43)
Mean in quartile 3 (50 to 75%)	0.13 (0.082, 0.18)	0.44 (0.37, 0.50)	-0.10 (-0.20, -0.00068)	4.15 (-1.09, 9.28)	30085 (16876, 43013)	2.83 (2.22, 3.42)	36303 (29276, 43299)	-4536 (-9914, 738)	11175 (2444, 19858)	40038 (27897, 51787)	0.11 (-0.10, 0.31)
Mean in quartile 4 (75 to 100%)	0.23 (0.19, 0.28)	0.53 (0.46, 0.60)	-0.12 (-0.22, -0.013)	8.00 (2.63, 13.3)	31604 (18150, 44869)	3.97 (3.35, 4.56)	39464 (32525, 46318)	-4154 (-9726, 1312)	18011 (9471, 26864)	40551 (28577, 52610)	0.21 (0.0047, 0.42)
P-value all interactions equal	0.000	0.002	1.000	0.166	0.228	0.000	0.164	0.911	0.663	1.000	0.984
P-value treatXQ1=treatXQ4	0.000	0.001	0.809	0.104	0.114	0.000	0.068	0.771	1.000	1.000	1.000
Mean in control for Q(1)	0.97	0.85	0.34	48.3	74622.6	10.5	50610.5	14069.4	73954.4	55428.6	0.20
Mean in control for Q(2)	0.97	0.85	0.28	45.0	65875.5	10.3	47993.5	9482.7	59611.2	45068.4	0.16
Mean in control for Q(3)	0.94	0.83	0.25	41.9	60839.7	9.89	46146.6	7528.4	55768.8	42968.5	0.14
Mean in control for Q(4)	0.91	0.79	0.21	37.0	52419.4	9.37	42025.2	5264.7	47311.8	35408.5	0.07
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)											
Mean in quartile 1 (0 to 25%)	-0.020 (-0.098, 0.060)	0.0019 (-0.11, 0.12)	-0.0025 (-0.12, 0.11)	-1.36 (-8.53, 5.80)	1579 (-10755, 13912)	-0.31 (-1.32, 0.73)	3157 (-3737, 10005)	235 (-9751, 10546)	3729 (-5357, 12897)	12399 (-7700, 32587)	0.16 (-0.077, 0.40)
Mean in quartile 2 (25 to 50%)	-0.018 (-0.096, 0.061)	-0.010 (-0.12, 0.10)	0.0047 (-0.11, 0.12)	-1.06 (-8.27, 6.30)	736 (-11458, 13060)	-0.31 (-1.34, 0.73)	593 (-6330, 7529)	836 (-9441, 11045)	1290 (-7596, 10055)	3689 (-16501, 23859)	0.071 (-0.17, 0.31)
Mean in quartile 3 (50 to 75%)	0.018 (-0.061, 0.096)	-0.0028 (-0.12, 0.11)	-0.018 (-0.13, 0.097)	2.32 (-4.89, 9.55)	6456 (-6066, 18844)	-0.037 (-1.08, 0.99)	830 (-6015, 7696)	5076 (-5057, 15254)	-368 (-9472, 8493)	6800 (-13642, 27063)	0.056 (-0.18, 0.30)
Mean in quartile 4 (75 to 100%)	0.012 (-0.067, 0.090)	-0.028 (-0.14, 0.087)	0.049 (-0.064, 0.16)	-0.065 (-7.30, 7.17)	2771 (-9307, 15356)	-0.25 (-1.29, 0.80)	-2661 (-9488, 4389)	5650 (-4708, 16012)	1946 (-7036, 11018)	10040 (-9921, 30277)	0.068 (-0.17, 0.30)
P-value all interactions equal	1.000	1.000	1.000	1.000	1.000	1.000	0.800	1.000	1.000	1.000	1.000
P-value treatXQ1=treatXQ4	1.000	1.000	1.000	1.000	1.000	1.000	0.397	0.855	1.000	1.000	0.979
Mean in control for Q(1)	0.94	0.64	0.40	47.6	64297.7	9.76	30622.6	27727.4	66767.3	77089.8	0.14
Mean in control for Q(2)	0.90	0.56	0.34	44.5	48049.5	8.86	23910.2	19502.6	53638.7	58768.3	0.10
Mean in control for Q(3)	0.88	0.54	0.32	42.6	45313.4	8.37	21591.9	18096.8	49538.7	52302.9	0.10
Mean in control for Q(4)	0.80	0.48	0.32	35.7	34223.3	7.31	16394.0	15165.5	43487.1	42764.8	-0.04

Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Estimates of treatment effects on the main outcomes during the program (Panel A), and after the program (Panel B) by quartile of the (predicted) impacts on earnings during the program. At the bottom of each panel, we present a test of equality of treatment effects between all quartiles respectively between the bottom (Q1) and top (Q4) quartiles. Estimation is based on a specification similar to equation 10 but replacing the dependent variable with the alternative (midline or endline) outcome variable. The best predictions during program are used: Generalized Random Forest. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$ (see appendix table A6 for comparisons across algorithms). Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted p-values for partition uncertainty. *Hours*, *earnings*, *expenditures*, and *savings* winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y + 1)$.

Table 9: Impacts *post* program on intermediate outcomes, by quartile of predicted impacts on (ln) earnings *during* the program

	Investment in Self-Employed Activities				Search for Wage Jobs	
	(1) Total number of income generating activities	(2) Value of productive assets in CFA (stock)	(3) Start-up capital in CFA (main self-empl. act., stock)	(4) Value of investments in CFA (last 3 months)	(5) Searched for a job (last 3 months)	(6) Total spent in job search in CFA (last 12 months)
Mean in quartile 1 (0 to 25%)	0.026 (-0.12, 0.18)	11189 (261, 22223)	14723 (-310, 29499)	538 (-1431, 2560)	0.082 (-0.038, 0.20)	1806 (-3341, 6861)
Mean in quartile 2 (25 to 50%)	0.0093 (-0.14, 0.16)	3843 (-7105, 14666)	2899 (-11880, 18088)	322 (-1654, 2301)	0.0062 (-0.11, 0.13)	-580 (-5819, 4584)
Mean in quartile 3 (50 to 75%)	-0.0043 (-0.16, 0.15)	-1325 (-12576, 9517)	2086 (-13125, 17246)	-625 (-2558, 1338)	0.025 (-0.096, 0.14)	828 (-4278, 5934)
Mean in quartile 4 (75 to 100%)	0.038 (-0.11, 0.19)	-20 (-11026, 10771)	5612 (-9583, 20764)	-138 (-2121, 1916)	0.057 (-0.065, 0.18)	887 (-4198, 5942)
P-value all interactions equal	1.000	0.350	0.893	1.000	0.956	1.000
P-value treatXQ1=treatXQ4	1.000	0.282	0.710	1.000	1.000	1.000
Mean in control for Q(1)	0.55	23944.3	43103.8	3673.3	0.57	11914.3
Mean in control for Q(2)	0.50	18478.1	30054.0	2704.7	0.55	8466.7
Mean in control for Q(3)	0.47	14974.2	26206.3	2242.7	0.55	8269.3
Mean in control for Q(4)	0.40	10781.7	19455.5	974.5	0.56	8413.1

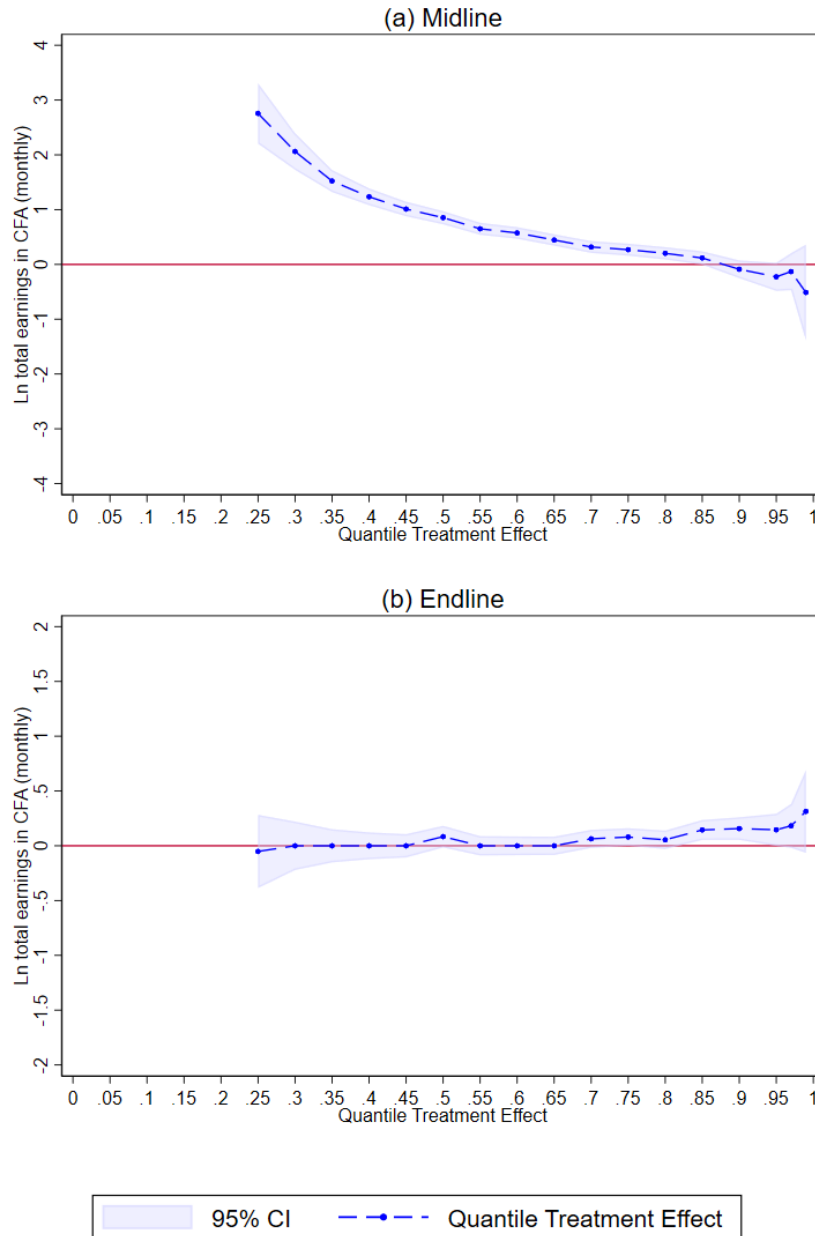
Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Estimates of post-program impacts on intermediary outcomes related to productive investments by quartile of the (predicted) impacts on earnings during the program. At the bottom of each panel, we present a test of equality of treatment effects between all quartiles respectively between the bottom (Q1) and top (Q4) quartiles. Estimation is based on a specification similar to equation 10 but replacing the dependent variable with the alternative (endline) outcome variable. The best predictions during program are used: Generalized Random Forest. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\widehat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$ (see appendix table A6 for comparisons across algorithms). Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted p-values for partition uncertainty. *Value of productive assets, value of investments, start-up capital and total spend in job search* winsorized at the 97th percentile.

Table 10: Impacts on (ln) earnings and cost-benefit ratios under alternative targeting approaches

	Random selection	Machine learning pred. by quartile of pred. impacts <i>during</i> program		Selection on baseline characteristics		
	(1)	(2)	(3)	(4)	(5)	(6)
	Treated (ITT)	Mean in quartile 4 (75% to 100%)	Mean in quartile 1 (0 to 25%)	Women	Low baseline earnings (self-declared) (bottom 25 %)	Low baseline earnings (predicted) (bottom 25%)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)						
Treatment	2.948*** (0.194)	3.966*** (0.307)	1.465*** (0.305)	3.728*** (0.349)	4.035*** (0.384)	3.657*** (0.373)
Observations	2912	2877	2877	2912	2877	2877
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)						
Treatment	-0.037 (0.181)	-0.252 (0.529)	-0.306 (0.525)	-0.050 (0.345)	-0.147 (0.517)	0.020 (0.405)
Observations	3934	3865	3865	3934	3736	3736
Panel C: Cost-benefit Analysis						
Effect <i>during</i> program in CFA [A]	27083	31671	14660	32097	33954	32824
Effect <i>post</i> program in CFA [B]	4361	4491	461	7168	7602	10646
Discounted sum of impacts (<i>during</i> + <i>post</i> program) [C]	253920.50	287958.58	109428.85	330429.84	349837.28	386817.87
Total cost per beneficiary [D]	768708.10	768708.10	768708.10	768708.10	768708.10	768708.10
Cost-benefit ratio (<i>during</i> + <i>post</i> program) [E]	3.03	2.67	7.02	2.33	2.20	1.99

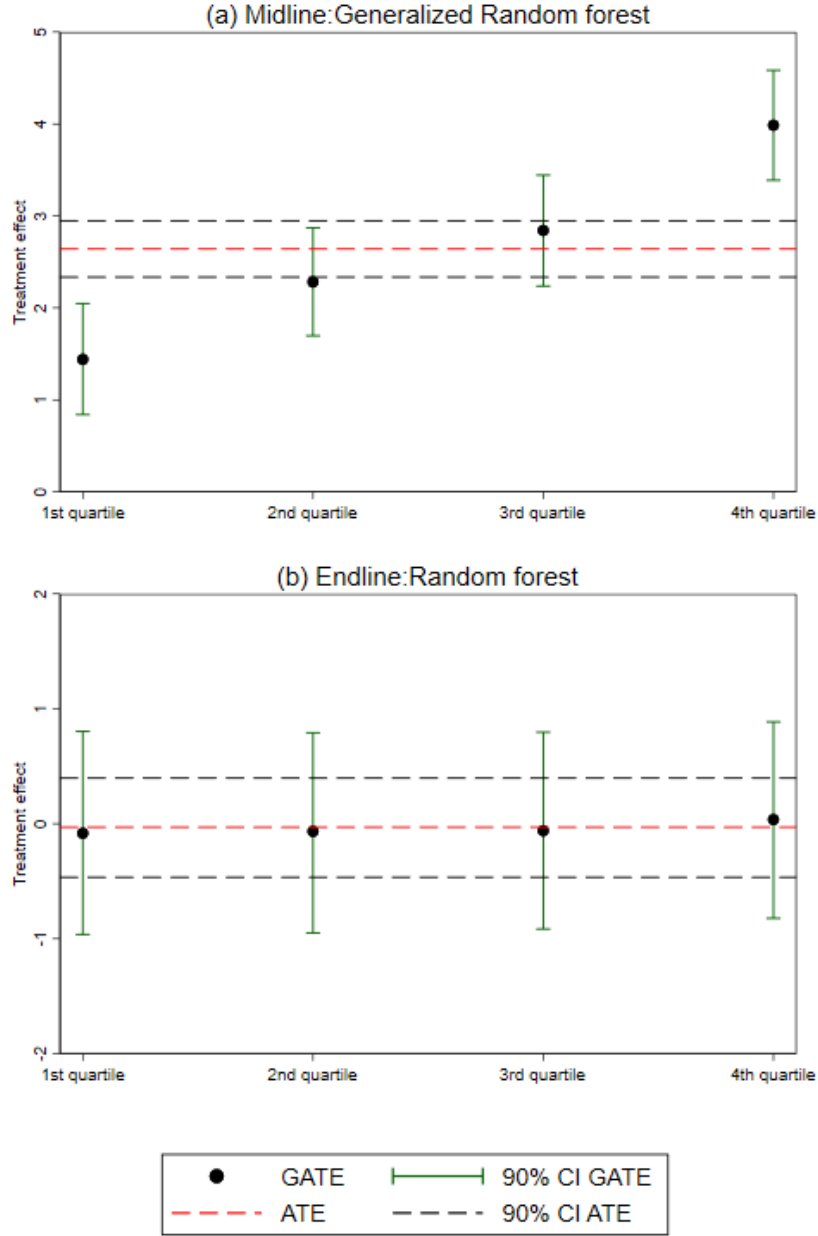
Column (1) is the ITT estimate based on specification in equation 6. Columns (4-6) show the estimated β_1 coefficient from the specification in equation 8. Columns (2) and (3) document the effect of selecting the observations in the top and bottom quartile of the distribution of predicted impacts at midline. Column 2 (respectively column 3) of Panel B reports group averages (GATES) endline impacts for observations in the top quartile (respectively bottom quartile) of predicted impacts at midline. (This is similar to Table 6, panel B, column (2)). Column (6) uses predictions of baseline earnings based on the Elastic Net method. [A] Effect on total monthly earnings *during* program in CFA, in levels, winsorized at the 97th percentile. [B] Effect on total monthly earnings *post* program in CFA, in levels, winsorized at the 97th percentile. [C] Discounted sum of impacts on total earnings from program start (month 1) up to 14 months after program ended (month 21). It is computed as $\sum_{k=1}^7 \rho^{(k-4)} \beta^{During} + \sum_{k=8}^{21} \rho^{(k-7)} \beta^{Post}$, with β^{During} (respectively β^{Post}) the contemporaneous (respectively post-program) ITT estimates of impact on monthly total earnings and ρ the monthly discount factor. $\rho = 1/(1 + \delta)^{1/12}$, $\delta = 10\%$. For column (3), we consider post-program impacts to be 0 when calculating *the discounted sum of impacts* in [C]. [E] = [D] / [C]. Robust standard errors clustered at (broad) brigade level in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

Figure 1: Quantile treatment effects for (ln) earnings *during* and *post* program



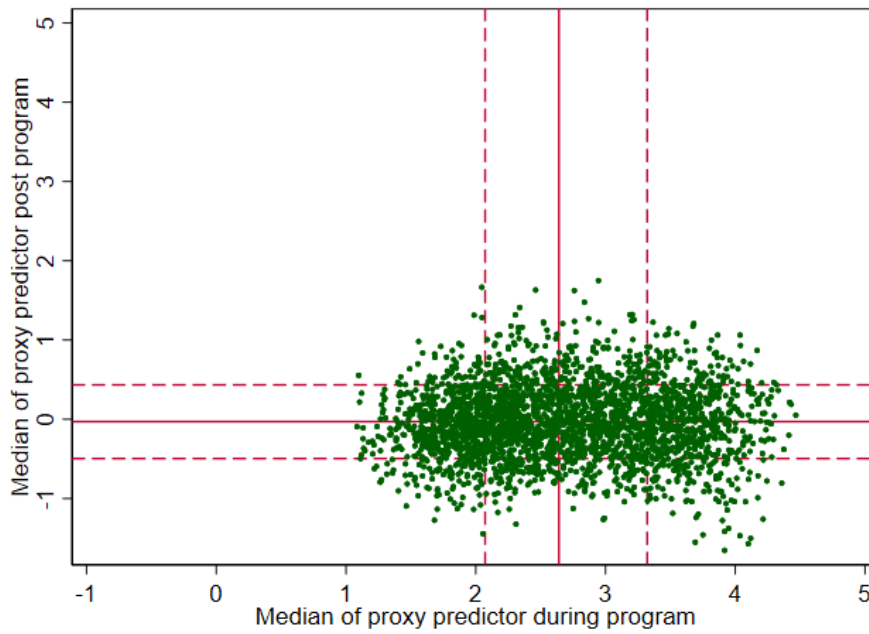
Note: Quantile treatment effects (for non-zero earnings) up to the 99th percentile. Permutation tests use 10000 permutations for each hypothesis. Wilcoxon rank-sum test permutation p-value is 0.000 for panel (a) and 0.466 for panel (b). Kolmogorov-Smirnov permutation p-value is 0.000 for panel (a) and 0.776 for panel (b). Total monthly earnings variable is in CFA. The variable is not winsorized to study the top of the distribution. Results shown up to the 99th percentile.

Figure 2: Group average treatment effects (GATES) for (ln) earnings



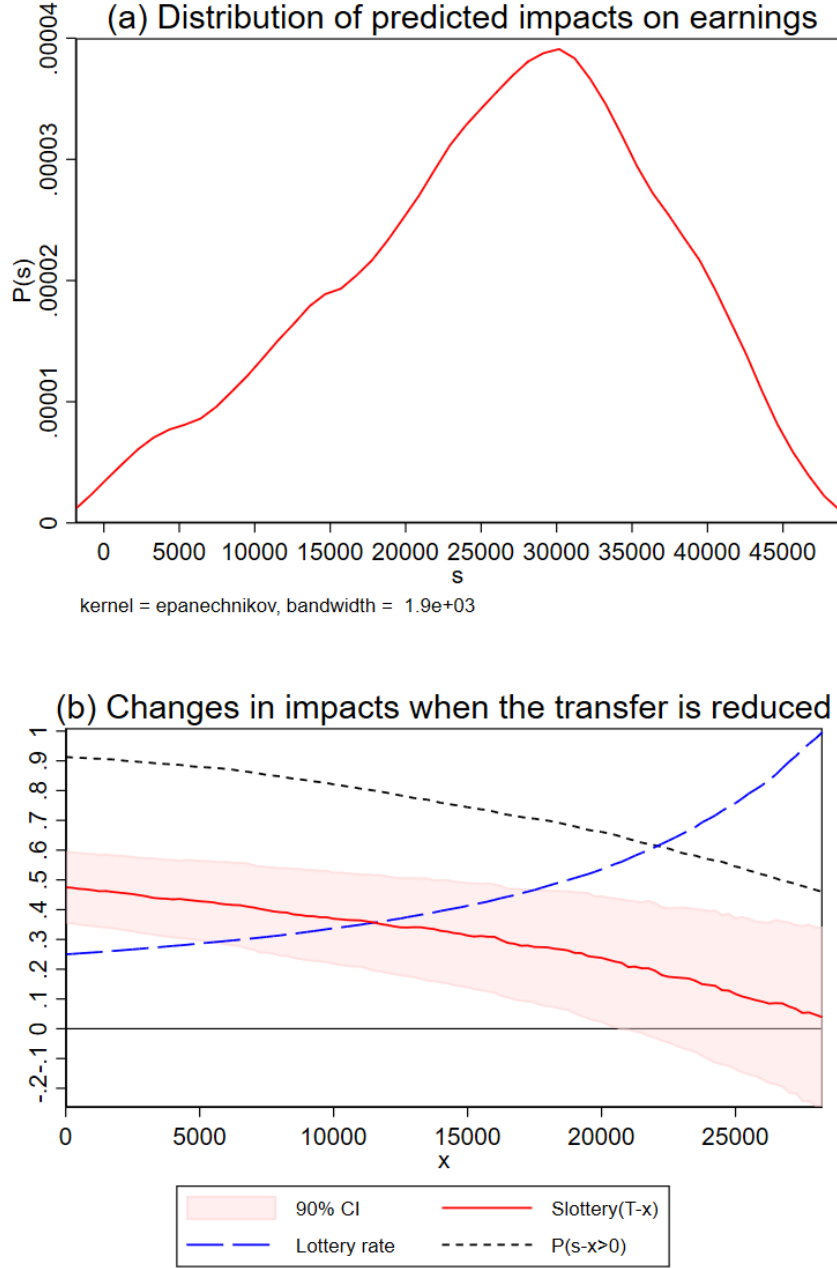
Note: Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). The best predictions are reported. Choice of algorithm is based on Λ and is indicated at the top of each figure (see appendix table A6 for comparisons across algorithms). Total monthly earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y + 1)$.

Figure 3: Predicted impact on (ln) earnings *during* vs *post* program



Note: Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Median across 100 simulations. Solid lines represent the ATE. Dashed lines delimit bottom 25% and top 25% of the distribution. Predictions using Generalized Random Forest for impacts during program. Predictions using Random Forest for post program impacts. Total monthly earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y + 1)$.

Figure 4: Change in impact on total earnings when the transfer is reduced



Panel (a) presents the estimated distribution of our ML estimate of CATE using the best method: Random Forest (see appendix Table A6 for comparisons across algorithms). Panel (b) presents simulations of changes in impacts induced by a reduction in the transfer by x , assuming the distribution of individual impact on earnings in Panel (a) (see conceptual framework in section 2). The dotted black line shows the share of initial applicants who still apply when the transfer is reduced by x . The dashed blue line shows the lottery success rate when the transfer is reduced by x (Lottery rate = $0.25T/((T-x)P(s_T - x > 0))$). The figure reports this lottery rate up to the value where it reaches 1 (\approx CFA 27,500). The solid red line shows the average impact on earnings $E((s-x)1(s-x > 0))/(P(s-x > 0)(T-x))$ when the transfer amount is reduced by x , $S_{lottery}(T-x)$ appearing in equation 5, normalized by $0.25 \times 50,000$, the amount available for each initial applicant.

Appendix A Additional Tables

Table A1: Estimated impacts *during* and *post* program on economic outcomes, with baseline controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Employed	Wage-Employed (in at least 1 activity)	Self-Employed (in at least 1 activity)	Total Hours worked (weekly)	Hours worked in Wage- Empl. (weekly)	Hours worked in Self-Empl. (weekly)	Total earnings in CFA (monthly)	Ln total earnings (monthly)	Earnings in Wage- Empl. in CFA (monthly)	Earnings in Self-Empl. in CFA (monthly)	Total ex- penditures in CFA (monthly)	Savings in CFA (stock)	Well-being index (z-score)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)													
Public Works Treatment (ITT)	0.14*** (0.01)	0.48*** (0.02)	-0.09*** (0.02)	5.06*** (1.26)	15.68*** (1.29)	-6.87*** (1.03)	27485.77*** (2608.25)	2.92*** (0.19)	36799.02*** (1463.69)	-5567.13*** (1196.97)	14431.49*** (1317.32)	40035.35*** (2303.55)	0.19*** (0.05)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Mean in Control	0.85	0.49	0.35	39.69	21.40	12.67	42841.22	7.87	20188.33	12753.65	47233.52	19250.05	-0.03
Observations	2958	2958	2958	2958	2958	2958	2912	2912	2912	2912	2945	2958	2934
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)													
Public Works Treatment (ITT)	0.02 (0.02)	0.00 (0.02)	0.02 (0.02)	0.74 (1.40)	-1.58 (1.26)	2.47** (1.23)	5155.88*** (1902.00)	0.00 (0.19)	-665.49 (1079.42)	4783.59*** (1852.73)	2387.97 (1466.46)	10143.10*** (3316.16)	0.12*** (0.04)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.86	0.51	0.36	41.00	22.50	13.80	43481.10	8.42	20706.18	18872.95	50700.71	46348.14	-0.05
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3934	3814	3934	3932
Panel C: <i>Post</i> program impacts (by treatment arms) (12 to 15 months after program ends)													
Public Works Treatment (ITT)	0.01 (0.02)	0.00 (0.03)	0.02 (0.03)	-1.58 (1.78)	-1.82 (1.65)	0.53 (1.44)	3406.18 (2116.14)	-0.14 (0.22)	69.13 (1319.73)	3071.78 (1901.47)	2131.18 (1596.09)	8598.92** (3644.89)	0.15*** (0.05)
Self-Empl. training (SET)	0.01 (0.02)	-0.02 (0.03)	0.02 (0.03)	3.59* (2.01)	0.74 (1.83)	2.58 (1.80)	4226.69 (3044.51)	0.21 (0.24)	-1668.02 (1287.18)	5203.42* (2826.96)	-207.36 (1422.57)	8143.34* (4479.69)	-0.01 (0.05)
Wage-Empl. training (WET)	0.00 (0.02)	0.01 (0.03)	0.00 (0.03)	3.42 (2.36)	-0.03 (1.65)	3.41 (2.16)	947.80 (1963.81)	0.24 (0.19)	-796.06 (1313.08)	210.73 (2173.59)	881.90 (1599.02)	-3206.44 (4242.97)	-0.08* (0.05)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.86	0.51	0.36	41.00	22.50	13.80	43481.10	8.42	20706.18	18872.95	50700.71	46348.14	-0.05
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3934	3814	3934	3932

ITT estimates in panels A and B based on specification in equation 6. Estimates in panel C based on specification in equation 7. The set of baseline controls differs for each outcome. Variables are selected from a pool of 1312 covariates using post-double selection lasso. Control variables include information about individual characteristics, education, household composition, experience of violence, household expenditure, asset ownership, and access to infrastructure. Robust standard errors clustered at (broad) brigade level in parentheses. Hours, earnings, expenditures, and savings winsorized at the 97th percentile. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A2: Estimated impacts *during* and *post* program on well-being index components

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Well-being index (z-score)	Self-esteem (Rosenberg scale) (z-score)	Positive Affect (CES-D sub scale) (z-score)	Positive attitude towards the future (ZTPI sub scale) (z-score)	Present fatalism (ZTPI sub scale) (z-score)	Happiness in daily activities (z-score)	Pride in daily activities (z-score)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)							
Public Works Treatment (ITT)	0.20*** (0.05)	0.14*** (0.04)	0.18*** (0.04)	0.086** (0.04)	0.021 (0.04)	0.15*** (0.05)	0.15*** (0.05)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.03	-0.02	-0.05	-0.01	0.03	0.01	-0.00
Observations	2934	2951	2958	2951	2955	2950	2949
Perm. p-value: no effects	0.000	0.001	0.000	0.035	0.606	0.002	0.001
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)							
Public Works Treatment (ITT)	0.11*** (0.04)	0.10** (0.04)	0.041 (0.04)	0.061 (0.05)	-0.093** (0.04)	0.076* (0.04)	0.053 (0.04)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.05	-0.05	-0.03	-0.01	0.05	-0.02	-0.02
Observations	3932	3933	3932	3933	3933	3933	3933
Perm. p-value: no effects	0.010	0.022	0.314	0.197	0.041	0.074	0.201
Panel C: <i>Post</i> program impacts (by treatment arms) (12 to 15 months after program ends)							
Public Works Treatment (ITT)	0.14*** (0.05)	0.12** (0.05)	0.063 (0.05)	0.094* (0.05)	-0.052 (0.05)	0.11** (0.05)	0.089* (0.05)
Self-Empl. training (SET)	-0.0068 (0.05)	-0.044 (0.06)	-0.0078 (0.05)	-0.11* (0.06)	-0.11** (0.05)	-0.018 (0.05)	-0.025 (0.05)
Wage-Empl. training (WET)	-0.075 (0.05)	-0.019 (0.05)	-0.062 (0.05)	0.0034 (0.05)	-0.020 (0.05)	-0.082* (0.05)	-0.087* (0.05)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.05	-0.05	-0.03	-0.01	0.05	-0.02	-0.02
p-value: PW+SET=0	0.011	0.165	0.275	0.816	0.001	0.059	0.177
p-value: PW+WET=0	0.187	0.065	0.977	0.046	0.201	0.598	0.972
p-value: SET=WET	0.169	0.703	0.232	0.038	0.073	0.138	0.216
Observations	3932	3933	3932	3933	3933	3933	3933
Perm. p-value: PW+SET=0	0.014	0.175	0.277	0.820	0.000	0.064	0.188
Perm. p-value: PW+WET=0	0.201	0.066	0.977	0.052	0.211	0.609	0.970
Perm. p-value: SET=WET	0.178	0.706	0.239	0.040	0.081	0.141	0.224

ITT estimates in panels A and B based on specification in equation 6. Estimates in panel C based on specification in equation 7. The definition of the *well-being index* and variables entering the index is detailed in Appendix E. *Present fatalism* enters as an inverted measure in the index (a negative impact in column (5) is associated with a positive impact on the index in column (1)). Robust standard errors clustered at (broad) brigade level in parentheses. Permutation tests use 10000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A3: Estimated impacts *during* and *post* program on behavior index components

	(1) Behavior index (z-score)	(2) Conduct problems (SDQ sub scale) (z-score)	(3) Pro-social behavior (SDQ sub scale) (z-score)	(4) Impulsiveness (DERS sub scale) (z-score)	(5) Anger in daily activities (z-score)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)					
Public Works Treatment (ITT)	0.12*** (0.04)	-0.031 (0.04)	0.023 (0.04)	-0.095** (0.04)	-0.13*** (0.04)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.02	0.01	-0.00	0.02	-0.01
Observations	2946	2957	2956	2954	2950
Perm. p-value: no effects	0.005	0.459	0.569	0.034	0.003
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)					
Public Works Treatment (ITT)	-0.012 (0.04)	0.013 (0.04)	-0.0032 (0.04)	0.0050 (0.04)	0.014 (0.04)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.01	-0.01	-0.01	-0.01	-0.03
Observations	3933	3933	3933	3933	3933
Perm. p-value: no effects	0.772	0.767	0.941	0.894	0.731
Panel C: <i>Post</i> program impacts (by treatment arms) (12 to 15 months after program ends)					
Public Works Treatment (ITT)	0.025 (0.05)	0.034 (0.05)	0.0066 (0.06)	-0.051 (0.04)	-0.013 (0.05)
Self-Empl. training (SET)	-0.039 (0.04)	-0.062* (0.04)	-0.054 (0.06)	0.054 (0.04)	0.012 (0.05)
Wage-Empl. training (WET)	-0.077 (0.05)	-0.0041 (0.05)	0.024 (0.07)	0.12*** (0.04)	0.072 (0.05)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.01	-0.01	-0.01	-0.01	-0.03
p-value: PW+SET=0	0.762	0.530	0.317	0.939	0.978
p-value: PW+WET=0	0.293	0.574	0.546	0.147	0.222
p-value: SET=WET	0.410	0.197	0.066	0.165	0.229
Observations	3933	3933	3933	3933	3933
Perm. p-value: PW+SET=0	0.767	0.534	0.322	0.940	0.980
Perm. p-value: PW+WET=0	0.300	0.580	0.549	0.155	0.236
Perm. p-value: SET=WET	0.418	0.212	0.079	0.170	0.240

ITT estimates in panels A and B based on specification in equation 6. Estimates in panel C based on specification in equation 7. The definition of the *behavior index* and variables entering the index is detailed in Appendix E. *Conduct problems*, *impulsiveness* and *anger in daily activities* enter as inverted measures in the index (a negative impact in columns (2), (4) or (5) is associated with a positive impact on the index in column (1)). Robust standard errors clustered at (broad) brigade level in parentheses. Permutation tests use 10000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A4: Estimated impacts *during* and *post* program on risky behaviors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Stealing	Assaulting someone	Believing smuggling is necessary (to earn a living)	Prostituting	Threatening someone	Taking illicit drugs	Smuggling stolen objects	Ties with a smuggling network	Keeping fire arms at home
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)									
Public Works Treatment (ITT)	0.09*	-0.11**	0.00	0.01	-0.02	-0.11***	-0.05	0.01	0.01
	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Mean in Control	0.05	0.18***	0.10***	0.07*	0.11***	0.03	0.05*	0.02	0.01
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value: no design effect in list A	1.000	1.000	0.446	0.008	0.156	0.015	0.314	0.485	0.152
p-value: no design effect in list B	1.000	1.000	0.655	0.373	1.000	0.425	0.645	0.107	0.856
Impact in list A	0.09	-0.19	-0.07	0.03	-0.11	-0.08	-0.14	0.04	-0.01
Impact in list B	0.08	-0.03	0.08	-0.02	0.07	-0.15	0.04	-0.01	0.03
p-value: Impact A= Impact B	0.930	0.097	0.126	0.694	0.070	0.501	0.046	0.587	0.677
Observations	2956	2956	2956	2956	2953	2955	2954	2954	2955
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)									
Public Works Treatment (ITT)	0.03	-0.01	0.01	-0.01	-0.01	0.01	-0.09***	0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Mean in Control	0.04	0.09***	0.08***	0.10***	0.08***	0.03	0.09***	0.06**	0.05**
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value: no design effect in list A	0.083	—	0.317	—	1.000	0.157	0.063	0.510	0.981
p-value: no design effect in list B	—	1.000	1.000	0.083	0.214	0.171	0.612	0.425	0.611
Impact in list A	0.01	0.04	0.02	-0.02	-0.01	-0.04	-0.09	0.07	0.05
Impact in list B	0.05	-0.07	-0.00	-0.00	-0.01	0.06	-0.08	-0.07	-0.04
p-value: Impact A=Impact B	0.592	0.108	0.752	0.743	0.945	0.144	0.915	0.058	0.190
Observations	3933	3933	3933	3933	3933	3932	3933	3933	3933
Panel C: <i>Post</i> program impacts (by treatment arms) (12 to 15 months after program ends)									
Public Works Treatment (ITT)	0.01	0.03	0.03	-0.05	0.00	0.02	-0.06	0.01	0.01
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Self-Empl. training (SET)	0.03	-0.08	-0.04	0.10***	0.02	-0.05	-0.07*	0.01	-0.02
	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)
Wage-Empl. training (WET)	0.01	-0.06*	-0.02	0.01	-0.06	0.01	-0.01	-0.03	0.01
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Mean in Control	0.04	0.09***	0.08***	0.10***	0.08***	0.03	0.09***	0.06**	0.05**
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value: PW+SET=0	0.275	0.344	0.662	0.197	0.534	0.498	0.001	0.708	0.681
p-value: PW+WET=0	0.539	0.419	0.820	0.316	0.194	0.450	0.072	0.571	0.668
p-value: SET=WET	0.667	0.689	0.613	0.008	0.077	0.188	0.185	0.416	0.388
Observations	3933	3933	3933	3933	3933	3932	3933	3933	3933

Variables measured using a double list experiment (Droitcour et al., 1991), whereby each respondent was assigned to a list A with sensitive items and a list B without sensitive items, or vice versa (see Appendix E). Difference-in-means estimation (Miller, 1984) was used to estimate the mean in control and treatment effects. In Panels A and B, the specification $Y_i = \alpha + \gamma_1 L_i + \gamma_2 W_i + \gamma_3 (L_i \times W_i) + \delta_i X_{i,l} + \varepsilon_i$ was used; where γ_1 is the mean in the control group, γ_3 is the treatment effect, and $X_{i,l}$ is a vector of stratification variables. Similarly, in Panel C we used $Y_i = \alpha + \gamma_1 L_i + \gamma_2 W_i + \gamma_3 (W_i \times T1_i) + \gamma_4 (W_i \times T2_i) + \gamma_5 (L_i \times W_i) + \gamma_6 (L_i \times W_i \times T1_i) + \gamma_7 (L_i \times W_i \times T2_i) + \delta_i X_{i,l} + \varepsilon_i$; where γ_1 is the mean in the control group, γ_5 the effect of "pure" public works, and γ_6 (γ_7) the additional effect of self-employment training (wage employment training). The test for the presence of design effects is based on the likelihood ratio test (Blair and Imai, 2012). The null hypothesis is no design effect. We report Bonferroni adjusted p-values. Weights are used for estimation but not for the design effect test (because it is not supported). The dash symbol indicates that the test statistics could not be processed due to a lack of variance in estimated probabilities: $P(C = 4, S = 1) = P(C = 4, S = 0) = 0$. However, none of the point estimates of joint probabilities were negative in such cases. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A5: Estimated impacts *post* program on skills, investments in self-employed activities and search for wage jobs

	Investment in self-employed activities						Search for wage jobs			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Self-Empl. Quiz (z-score)	Prepared a business plan	Total number of income generating activities	Value of productive assets in CFA (stock)	Value of investments in CFA (last 3 months)	Start-up capital in CFA (main self-empl. act., stock)	Wage-Empl. Quiz (z-score)	Used a CV for job search	Total spent in job search in CFA (last 12 months)	Searched for a job (last 3 months)
Panel A: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)										
Public Works Treatment (ITT)	0.12*** (0.04)	0.024** (0.01)	0.048* (0.03)	4159.1*** (1529.97)	347.4 (292.63)	8712.6*** (2071.94)	0.12*** (0.04)	0.047*** (0.01)	-468.4 (823.02)	0.026 (0.02)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.07	0.05	0.48	12570.82	1910.81	22198.36	-0.09	0.15	9026.08	0.52
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3934
Perm. p-value: no effects	0.005	0.021	0.076	0.011	0.248	0.000	0.004	0.002	0.582	0.225
Panel B: <i>Post</i> program impacts (by treatment arms) (12 to 15 months after program ends)										
Public Works Treatment (ITT)	0.073 (0.06)	0.013 (0.01)	0.040 (0.03)	4878.0** (1956.43)	580.0 (404.36)	8282.5*** (2620.46)	0.019 (0.05)	0.016 (0.02)	-411.6 (915.44)	0.023 (0.03)
Self-Empl. training (SET)	0.11* (0.06)	0.041*** (0.01)	0.0058 (0.04)	-661.3 (2282.23)	-428.1 (383.22)	2534.1 (3330.95)	0.048 (0.05)	-0.0064 (0.02)	-324.3 (863.67)	-0.0023 (0.03)
Wage-Empl. training (WET)	0.033 (0.05)	-0.0058 (0.01)	0.021 (0.04)	-1576.7 (2040.88)	-296.7 (450.91)	-1187.7 (3306.09)	0.26*** (0.05)	0.10*** (0.02)	146.5 (942.35)	0.011 (0.03)
Strata f.e. control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	-0.07	0.05	0.48	12570.82	1910.81	22198.36	-0.09	0.15	9026.08	0.52
p-value PW+SET=0	0.001	0.000	0.217	0.046	0.625	0.000	0.162	0.546	0.449	0.458
p-value PW+WET=0	0.026	0.568	0.076	0.064	0.494	0.018	0.000	0.000	0.797	0.300
p-value SET=WET	0.107	0.001	0.695	0.648	0.764	0.236	0.000	0.000	0.591	0.746
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3934
Perm. p-value PW+SET=0	0.001	0.000	0.223	0.051	0.631	0.000	0.172	0.561	0.464	0.467
Perm. p-value PW+WET=0	0.029	0.575	0.081	0.067	0.502	0.018	0.000	0.000	0.804	0.304
Perm. p-value SET=WET	0.113	0.000	0.699	0.660	0.770	0.246	0.000	0.000	0.599	0.747

ITT estimates in panels A and B based on specification in equation 6. Estimates in panel C based on specification in equation 7. *Value of productive assets, value of investments, start-up capital and total spend in job search* winsorized at the 97th percentile. Robust standard errors clustered at (broad) brigade level in parentheses. Permutation tests use 10000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A6: Comparison of Machine Learning algorithms to predict impacts on earnings *during* and *post* program

	Estimates in logs					Estimates in levels				
	(1) Elastic net	(2) Generalized Random forest	(3) Gradient boosting	(4) R-Learner	(5) Random forest	(6) Elastic net	(7) Generalized Random forest	(8) Gradient boosting	(9) R-Learner	(10) Random forest
Panel A: Predicted average treatment effect and heterogeneity loading parameter <i>during</i> program										
ATE (β_1)	2.631 (2.9,2.3) [0.000]	2.642 (2.9,2.3) [0.000]	2.643 (2.9,2.3) [0.000]	2.635 (2.9,2.3) [0.000]	2.634 (2.9,2.3) [0.000]	24577.6 (31323.2,17831.5) [0.000]	24351.6 (31214.6,17669.0) [0.000]	24238.5 (30976.0,17471.1) [0.000]	24549.6 (31423.5,17789.9) [0.000]	24363.0 (31150.3,17618.5) [0.000]
HET (β_2)	0.980 (1.3,0.6) [0.000]	1.231 (1.6,0.9) [0.000]	0.420 (0.6,0.2) [0.000]	0.920 (1.2,0.6) [0.000]	0.849 (1.1,0.6) [0.000]	0.405 (0.9,-0.07) [0.157]	1.070 (2.1,-0.05) [0.121]	0.288 (0.7,-0.03) [0.160]	0.429 (1.3,-0.2) [0.353]	0.390 (0.8,0.009) [0.089]
Λ	0.9	1.0	0.7	0.9	0.9	6007.2	6401.7	5988.1	4621.1	6785.3
Panel B: Predicted average treatment effect and heterogeneity loading parameter <i>post</i> program										
ATE (β_1)	-0.0551 (0.4,-0.5) [1.000]	-0.0508 (0.4,-0.5) [1.000]	-0.0484 (0.4,-0.5) [1.000]	-0.0570 (0.4,-0.5) [1.000]	-0.0297 (0.4,-0.5) [1.000]	3217.5 (8428.4,-1918.3) [0.439]	3527.3 (8737.3,-1706.6) [0.374]	3546.3 (8685.8,-1630.9) [0.358]	3314.0 (8523.0,-1842.7) [0.414]	3474.5 (8642.7,-1641.7) [0.370]
HET (β_2)	-0.0581 (0.6,-0.7) [1.000]	0.156 (1.6,-1.3) [1.000]	0.0376 (0.3,-0.2) [1.000]	-0.00219 (3.1,-2.8) [1.000]	0.111 (0.7,-0.5) [0.970]	0.0837 (0.6,-0.5) [1.000]	0.391 (1.8,-1.0) [1.000]	0.0220 (0.4,-0.3) [1.000]	0.416 (5.4,-1.1) [0.756]	0.0800 (0.5,-0.4) [1.000]
Λ	0.10	0.1	0.1	0.1	0.2	1625.5	1742.1	1679.7	2182.6	1746.7

Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Estimates are based on equation 9. P-value (in brackets) for β_1 tests the hypothesis of no effect. P-value (in brackets) for β_2 tests the hypothesis of no heterogeneity. Panel A (respectively Panel B) shows estimates of β_1 and β_2 at midline (respectively endline). The Λ (lambda) statistic is displayed at the bottom of each panel: the larger Λ gets, the stronger the correlation between $\widehat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$. Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted confidence intervals at 90% in parentheses and adjusted p-values for partition uncertainty. Earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y+1)$.

Table A7: Heterogeneity in impacts on earnings *during* and *post* program, machine learning results for an extended set of covariates

	(1) Ln total earnings (Monthly)	(2) Ln total earnings (Monthly)	(3) Total earnings (Monthly)	(4) Total earnings (Monthly)
	<i>Midline</i>	<i>Endline</i>	<i>Midline</i>	<i>Endline</i>
Panel A: Predicted average treatment effect and heterogeneity loading parameter				
ATE (β_1)	2.311 (1.918,2.714) [0.000]	-0.136 (-0.723,0.449) [0.928]	27252.0 (17338.2,37013.8) [0.000]	4011.4 (-4014.3,11850.4) [0.631]
HET (β_2)	2.336 (1.156,3.577) [0.000]	0.272 (-0.306,0.814) [0.637]	2.297 (0.387,4.180) [0.032]	0.136 (-1.855,2.883) [0.888]
Best ML method	Generalized Random forest	Random forest	Generalized Random forest	R-Learner
Panel B: By quartile of predicted impacts on earnings <i>during</i> program (using <i>midline</i>) (GATES)				
Mean in quartile 1 (0 to 25%)	1.430 (0.596,2.233)	-0.695 (-2.047,0.682)	9616.7 (-9485.5,28647.1)	1309.1 (-17605.5,20036.9)
Mean in quartile 2 (25 to 50%)	2.063 (1.260,2.854)	-0.589 (-1.953,0.810)	31539.2 (12393.6,51035.5)	3483.3 (-15126.9,22326.7)
Mean in quartile 3 (50 to 75%)	2.602 (1.763,3.397)	-0.126 (-1.515,1.306)	32238.6 (12793.6,51515.6)	6594.5 (-11856.0,24930.0)
Mean in quartile 4 (75 to 100%)	3.312 (2.521,4.094)	-0.0495 (-1.417,1.326)	33266.5 (13826.0,52380.5)	8086.5 (-11723.2,27267.8)
P-value all coefficients are equal	0.007	1.000	0.292	1.000
Best ML method	Generalized Random forest	Generalized Random forest	Generalized Random forest	Generalized Random forest
Panel C: By quartile of predicted impacts on earnings <i>post</i> program (using <i>endline</i>) (GATES)				
Mean in quartile 1 (0 to 25%)	1.935 (1.102,2.761)	-0.338 (-1.512,0.819)	21497.8 (3457.5,39527.0)	2309.3 (-12952.7,18286.3)
Mean in quartile 2 (25 to 50%)	1.989 (1.164,2.868)	-0.359 (-1.520,0.828)	27897.3 (10231.8,45734.2)	5194.6 (-10461.8,20933.5)
Mean in quartile 3 (50 to 75%)	2.246 (1.456,3.038)	-0.432 (-1.562,0.736)	28716.5 (10596.6,46321.6)	3064.9 (-12565.0,18074.0)
Mean in quartile 4 (75 to 100%)	2.555 (1.768,3.370)	0.368 (-0.800,1.488)	29256.8 (11482.7,46980.5)	5016.5 (-10414.7,21248.5)
P-value all coefficients are equal	0.093	0.682	0.685	0.998
Best ML method	Random forest	Random forest	R-Learner	R-Learner

Heterogeneity analysis based on the approach in Chernozhukov et al. (2020) (see discussion in Section 4.2 and Appendix F.2). All baseline variables in the balance check table (Table 1) are used as covariates. Columns (1) and (3) (respectively columns (2) and (4)) focus on outcomes at midline (respectively endline). Estimates in Panel A are based on equation 9. P-value (in brackets) for β_1 tests the hypothesis of no effect. P-value (in brackets) for β_2 tests the hypothesis of no heterogeneity. GATES estimates are based on the specification in equation 10. Panel B (respectively Panel C) shows impacts per quartile of the predicted treatment effects at midline (respectively endline). The best predictions are reported. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$ (see appendix table A6 for comparisons across algorithms). Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted confidence intervals at 90% in parentheses and adjusted p-values for partition uncertainty. Earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y+1)$.

Table A8: Estimated impacts on (ln) earnings *post* program, by treatment arms

	(1) Public Works only (PW)	(2) PW and Self-Employment Training (SET)	(3) PW and Wage-Employment Training (WET)
	<i>Endline</i>	<i>Endline</i>	<i>Endline</i>
Panel A: Predicted average treatment effect and heterogeneity loading parameter			
ATE (β_1)	-0.0753 (-0.601,0.447) [1.000]	-0.0116 (-0.545,0.529) [1.000]	0.120 (-0.406,0.648) [0.930]
HET (β_2)	0.0866 (-0.211,0.382) [0.818]	-0.572 (-1.431,0.299) [0.425]	0.450 (-0.595,1.935) [0.722]
Best ML method	Gradient boosting	Elastic net	R-Learner
Panel B: By quartile of predicted impacts on earnings <i>post</i> program (<i>using endline</i>) (GATES)			
Mean in quartile 1 (0 to 25%)	-0.220 (-1.270,0.826)	0.408 (-0.663,1.487)	-0.0540 (-1.112,1.004)
Mean in quartile 2 (25 to 50%)	-0.141 (-1.188,0.923)	0.132 (-0.958,1.206)	0.138 (-0.913,1.187)
Mean in quartile 3 (50 to 75%)	-0.172 (-1.217,0.873)	-0.116 (-1.193,0.962)	0.238 (-0.805,1.279)
Mean in quartile 4 (75 to 100%)	0.148 (-0.909,1.214)	-0.522 (-1.587,0.543)	0.168 (-0.912,1.223)
P-value all coefficients are equal	0.902	0.478	0.950
Best ML method	Gradient boosting	Elastic net	R-Learner

Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Columns (1-3) show estimated impacts for each treatment arm compared to the control group. Estimates in Panel A are based on equation 9. P-value (in brackets) for β_1 tests the hypothesis of no effect. P-value (in brackets) for β_2 tests the hypothesis of no heterogeneity. Estimates in panel B are based on the specification in equation 10. They show impacts per quartile of the predicted treatment effects at endline. The best predictions are reported. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$. Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted confidence intervals at 90% in parentheses and adjusted p-values for partition uncertainty. Earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y + 1)$.

Table A9: Estimated impacts on earnings (in levels) *post* program, by treatment arms

	(1)	(2)	(3)
	Public Works only (PW)	PW and Self-Employment Training (SET)	PW and Wage-Employment Training (WET)
	<i>Endline</i>	<i>Endline</i>	<i>Endline</i>
Panel A: Predicted average treatment effect and heterogeneity loading parameter			
ATE (β_1)	2155.1 (-3648.0,8097.0) [0.799]	4166.5 (-2085.9,10463.6) [0.397]	3970.3 (-2233.2,10007.5) [0.412]
HET (β_2)	0.711 (-0.793,2.217) [0.645]	0.167 (-2.294,3.175) [0.822]	0.859 (-0.383,2.666) [0.359]
Best ML method	Generalized Random forest	R-Learner	R-Learner
Panel B: By quartile of predicted impacts on earnings <i>post</i> program (<i>using endline</i>) (GATES)			
Mean in quartile 1 (0 to 25%)	-804.5 (-12141.7,10585.9)	4252.3 (-8217.4,16697.1)	954.7 (-11358.2,13202.6)
Mean in quartile 2 (25 to 50%)	314.7 (-11327.4,12027.1)	5075.0 (-7721.7,17162.5)	3213.2 (-8929.3,15509.9)
Mean in quartile 3 (50 to 75%)	2580.6 (-9034.2,14329.5)	4686.6 (-8013.1,17365.6)	3843.2 (-8184.3,15983.8)
Mean in quartile 4 (75 to 100%)	6306.2 (-5642.9,17819.6)	3651.1 (-8919.7,15948.0)	5448.1 (-6760.2,17345.3)
P-value all coefficients are equal	0.715	0.772	0.828
Best ML method	Generalized Random forest	R-Learner	R-Learner

Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Columns (1-3) show estimated impacts for each treatment arm compared to the control group. Estimates in Panel A are based on equation 9. P-value (in brackets) for β_1 tests the hypothesis of no effect. P-value (in brackets) for β_2 tests the hypothesis of no heterogeneity. Estimates in panel B are based on the specification in equation 10. They show impacts per quartile of the predicted treatment effects at endline. The best predictions are reported. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$. Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted confidence intervals at 90% in parentheses and adjusted p-values for partition uncertainty. Earnings variable is in CFA, winsorized at the 97th percentile.

Table A10: Baseline characteristics of the bottom and top quartiles of predicted impacts on earnings (in levels) *during* program

	(1) Mean in 1st quartile	(2) Mean in 4th quartile	(3) Test (1)-(2) (p-value)
Individual characteristics			
Female	0.18	0.43	0
Live in urban area	0.91	0.97	0
Age	24.91	24.41	0.012
Nb of children	0.84	0.80	0.065
Education			
Years of education	10.16	10.16	0.065
Primary education not completed	0.45	0.46	0.095
Has participated in vocational training	0.53	0.27	0
Household characteristics and assets			
Household size	6.75	5.52	0
Is head of household	0.28	0.20	0.005
Total Nb of assets	0.59	0.51	0
Nb of transportation assets	1.22	0.42	0
Nb of agricultural assets	7.74	2.42	0
Nb of household durables	2.31	1.26	0
Nb of communication assets	8.42	5.70	0
Employment			
Has an activity	0.91	0.66	0
Is Wage-Employed	0.42	0.28	0
Is Self-Employed	0.52	0.22	0
Nb of activities	1.23	0.72	0
Total Earnings (monthly, CFA)	34385.3	8298.2	0
Searched for a job (last month)	0.62	0.52	0.519
Savings, constraints and expenditures			
Has saved (last 3 months)	0.52	0.45	0.033
Of which: share of formal savings (cond. on savings)	0.28	0.24	0.467
Has a savings account	0.15	0.10	0.006
Savings Stock (CFA)	51724.9	14685.6	0
Self-reported constraints to repay loans	0.29	0.11	0
Self-reported constraints to access credit	0.41	0.59	0
Self-reported constraints for basic needs expenditures	0.69	0.72	0.510
Transportation expenditures (last 7 days, CFA)	3066.3	1065.9	0
Communication expenditures (last 7 days, CFA)	2810.9	895.8	0
Cognitive skills and risk preference			
Cognitive (deduction, Raven Test)	0.23	0.23	0.798
Dexterity (Nuts Test)	0.37	0.38	0.104
Dexterity (Bolts Test)	0.34	0.33	0.900
Positive affect (CES-D scale, Nb positive days)	6.31	6.22	0.856
Positive attitude towards the future (ZTPI scale)	29.32	29.20	0.772
Is Risk averse (based on hypothetical lotteries)	0.71	0.72	1
Patience (scale 0 to 10, 10=very patient)	3.30	3.34	0.945
Preference for present (actualization rate for 1 month)	0.57	0.58	0.917

Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Column (1) (respectively (2)) displays average characteristics of the bottom (respectively top) quartile of the distribution of predicted impacts on earnings during the program. Column (3) reports p-values for a test of equality between the top and bottom quartile. Reported results are based on the algorithm with best predictions for midline : Generalized Random forest. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$ (see appendix table A6 for comparisons across algorithms). Means by quartile are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted p-values for partition uncertainty. *Household assets* and *savings stock* variables winsorized at the 99th percentile.

Table A11: Estimated impacts *during* and *post* program on savings

	(1)	(2)	(3)	(4)
	Ln Savings (Monthly)	Ln Savings (Monthly)	Savings in CFA (Monthly)	Savings in CFA (Monthly)
	<i>Midline</i>	<i>Endline</i>	<i>Midline</i>	<i>Endline</i>
Panel A: Predicted average treatment effect and heterogeneity loading parameter				
ATE (β_1)	3.543 (2.981,4.100) [0.000]	0.472 (-0.0260,0.977) [0.127]	36768.4 (30901.6,42641.1) [0.000]	9492.9 (1192.6,17761.4) [0.050]
HET (β_2)	1.303 (0.454,2.169) [0.005]	0.247 (-1.230,2.565) [0.809]	0.440 (-0.0100,0.889) [0.110]	0.193 (-0.101,0.505) [0.388]
Best ML method	Generalized Random forest	R-Learner	Random forest	Gradient boosting
Panel B: By quartile of predicted impacts on savings <i>during</i> program (GATES)				
Mean in quartile 1 (0 to 25%)	2.481 (1.355,3.611)	0.748 (-0.440,1.920)	29852.0 (18249.3,41593.9)	10275.5 (-9226.9,29636.8)
Mean in quartile 2 (25 to 50%)	3.174 (2.052,4.286)	0.302 (-0.870,1.471)	33873.9 (21920.6,45466.9)	7074.7 (-12459.8,27015.8)
Mean in quartile 3 (50 to 75%)	3.785 (2.671,4.889)	0.563 (-0.619,1.746)	38335.9 (26568.8,49862.4)	10085.7 (-9870.0,29882.8)
Mean in quartile 4 (75 to 100%)	4.778 (3.659,5.896)	0.316 (-0.885,1.497)	45983.1 (34465.3,57655.4)	9897.8 (-9570.1,28961.8)
P-value all coefficients are equal	0.031	1.000	0.269	1.000
Best ML method	Generalized Random forest	Generalized Random forest	Random forest	Random forest
Panel C: By quartile of predicted impacts on savings <i>post</i> program (GATES)				
Mean in quartile 1 (0 to 25%)	3.615 (2.557,4.647)	0.316 (-0.703,1.346)	36489.5 (26070.4,47031.0)	3831.7 (-12411.8,19889.7)
Mean in quartile 2 (25 to 50%)	3.551 (2.488,4.582)	0.336 (-0.705,1.367)	37213.5 (26671.1,47659.7)	7044.9 (-9401.7,23694.9)
Mean in quartile 3 (50 to 75%)	3.574 (2.503,4.652)	0.483 (-0.527,1.506)	37366.7 (26941.2,47912.0)	9436.6 (-7461.4,25802.8)
Mean in quartile 4 (75 to 100%)	3.585 (2.532,4.646)	0.731 (-0.286,1.740)	39460.8 (28965.9,50086.8)	16964.8 (816.3,33311.6)
P-value all coefficients are equal	0.507	0.909	1.000	0.377
Best ML method	R-Learner	R-Learner	Gradient boosting	Gradient boosting

Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Columns (1) and (3) (respectively columns (2) and (4)) focus on outcomes at midline (respectively endline). Estimates in Panel A are based on equation 9. P-value (in brackets) for β_1 tests the hypothesis of no effect. P-value (in brackets) for β_2 tests the hypothesis of no heterogeneity. GATES estimates are based on the specification in equation 10. Panel B (respectively Panel C) shows impacts per quartile of the predicted treatment effects at midline (respectively endline). The best predictions are reported. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$ (see appendix table A6 for comparisons across algorithms). Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted confidence intervals at 90% in parentheses and adjusted p-values for partition uncertainty. *Savings* variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y+1)$.

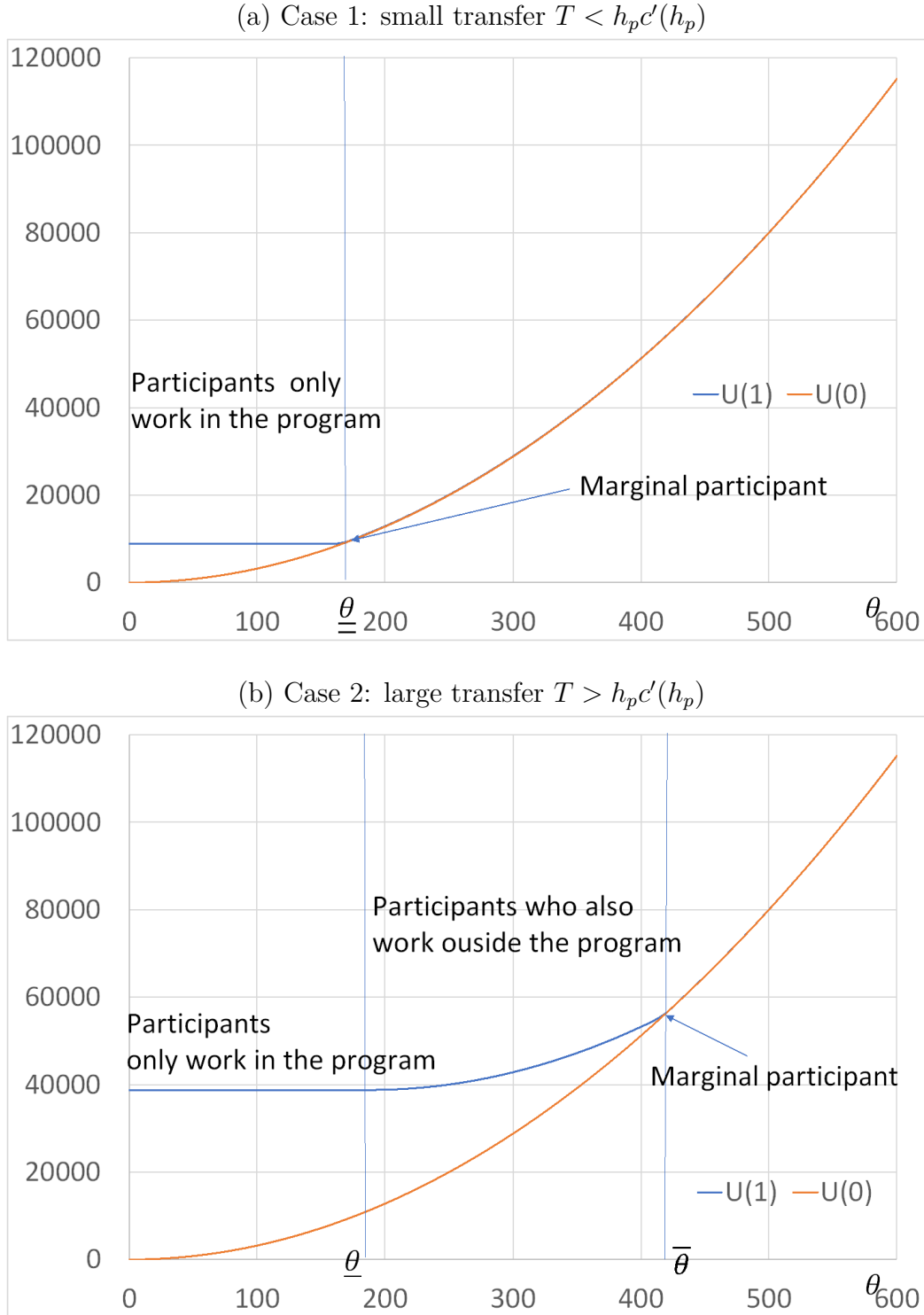
Table A12: Baseline characteristics of the bottom and top quartiles of predicted impacts on (ln) savings *during* program

	(1) Mean in 1st quartile	(2) Mean in 4th quartile	(3) Test (1)-(2) (p-value)
Individual characteristics			
Female	0.14	0.46	0
Live in urban area	0.95	0.95	0.598
Age	24.65	24.29	0.080
Nb of children	0.66	0.81	0.070
Education			
Years of education	11.76	9.71	0
Primary education not completed	0.36	0.49	0
Has participated in vocational training	0.47	0.34	0.001
Household characteristics and assets			
Household size	5.35	6.46	0
Is head of household	0.37	0.11	0
Total Nb of assets	0.65	0.44	0
Nb of transportation assets	0.68	0.70	0.331
Nb of agricultural assets	3.88	4.44	0.402
Nb of household durables	1.75	1.70	0.125
Nb of communication assets	6.95	6.72	0.014
Employment			
Has an activity	1	0.45	0
Is Wage-Employed	0.55	0.12	0
Is Self-Employed	0.48	0.14	0
Nb of activities	1.30	0.49	0
Total Earnings (monthly, CFA)	38482.1	658.7	0
Searched for a job (last month)	0.63	0.53	0.756
Savings, constraints and expenditures			
Has saved (last 3 months)	0.74	0.24	0
Of which: share of formal savings (cond. on savings)	0.34	0.18	0.002
Has a savings account	0.19	0.06	0
Savings Stock (CFA)	63881.6	5518.0	0
Self-reported constraints to repay loans	0.22	0.20	0.709
Self-reported constraints to access credit	0.39	0.57	0
Self-reported constraints for basic needs expenditures	0.66	0.72	0.252
Transportation expenditures (last 7 days, CFA)	2754.9	967.5	0
Communication expenditures (last 7 days, CFA)	2760.5	820.6	0
Cognitive skills and risk preference			
Cognitive (deduction, Raven Test)	0.24	0.23	0.781
Dexterity (Nuts Test)	0.37	0.38	0.300
Dexterity (Bolts Test)	0.34	0.33	1
Positive affect (CES-D scale, Nb positive days)	6.42	6.16	0.212
Positive attitude towards the future (ZTPI scale)	29.34	29.10	0.586
Is Risk averse (based on hypothetical lotteries)	0.73	0.71	1
Patience (scale 0 to 10, 10=very patient)	3.39	3.31	1
Preference for present (actualization rate for 1 month)	0.57	0.59	0.507

Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Column (1) (respectively (2)) displays average characteristics of the bottom (respectively top) quartile of the distribution of predicted impacts on (ln) savings during the program. Column (3) reports p-values for a test of equality between the top and bottom quartile. Reported results are based on the algorithm with best predictions for midline : Generalized Random forest. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$ noted $S(Z)$ and $s_0(Z)$. Means by quartile are estimated for each sample split, reported values are the medians across 100 sample splits. Adjusted p-values for partition uncertainty. *Household assets* and *savings stock* variables winsorized at the 99th percentile.

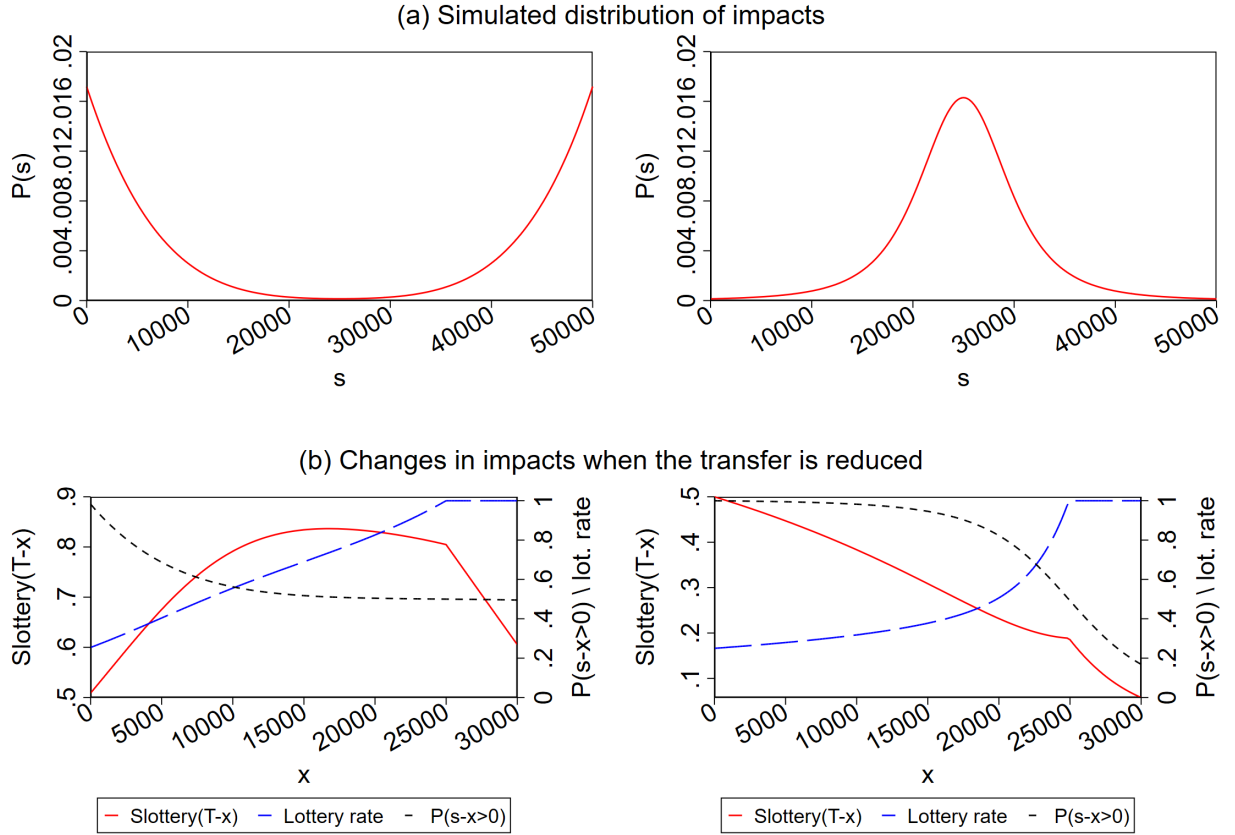
Appendix B Additional Figures

Figure B1: Participation in the program under different transfer amounts



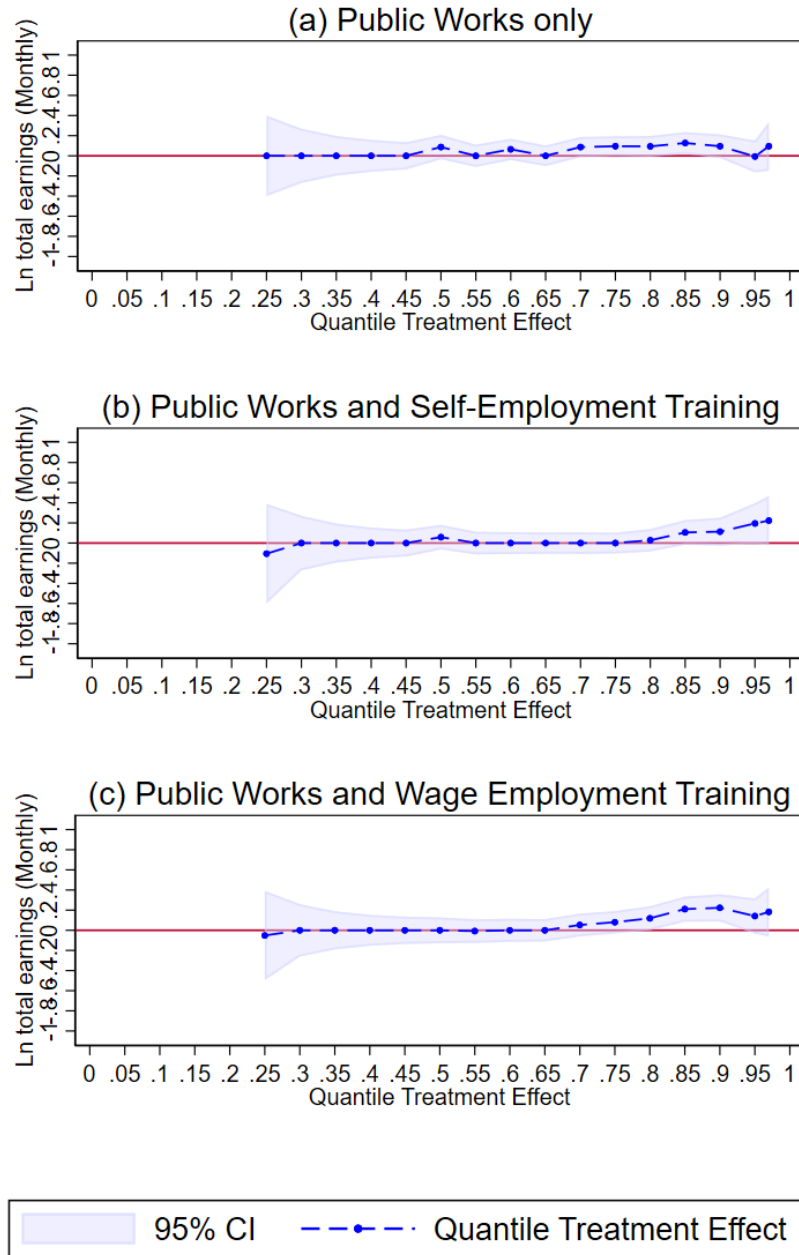
This figure presents two different hypothetical distributions of individual treatment effects. The two panels present the utility level reached by individuals when they are offered, respectively not offered the program. Panel (a) considers the case of a *small* transfer, in such a case, those who decide to participate only work in the program. Panel (b) considers the case of a *large* transfer and shows that the most productive participant also work outside the program. For details, see conceptual framework in section 2.

Figure B2: Distribution of individual treatment effects and impact of a reduction of the transfer amount



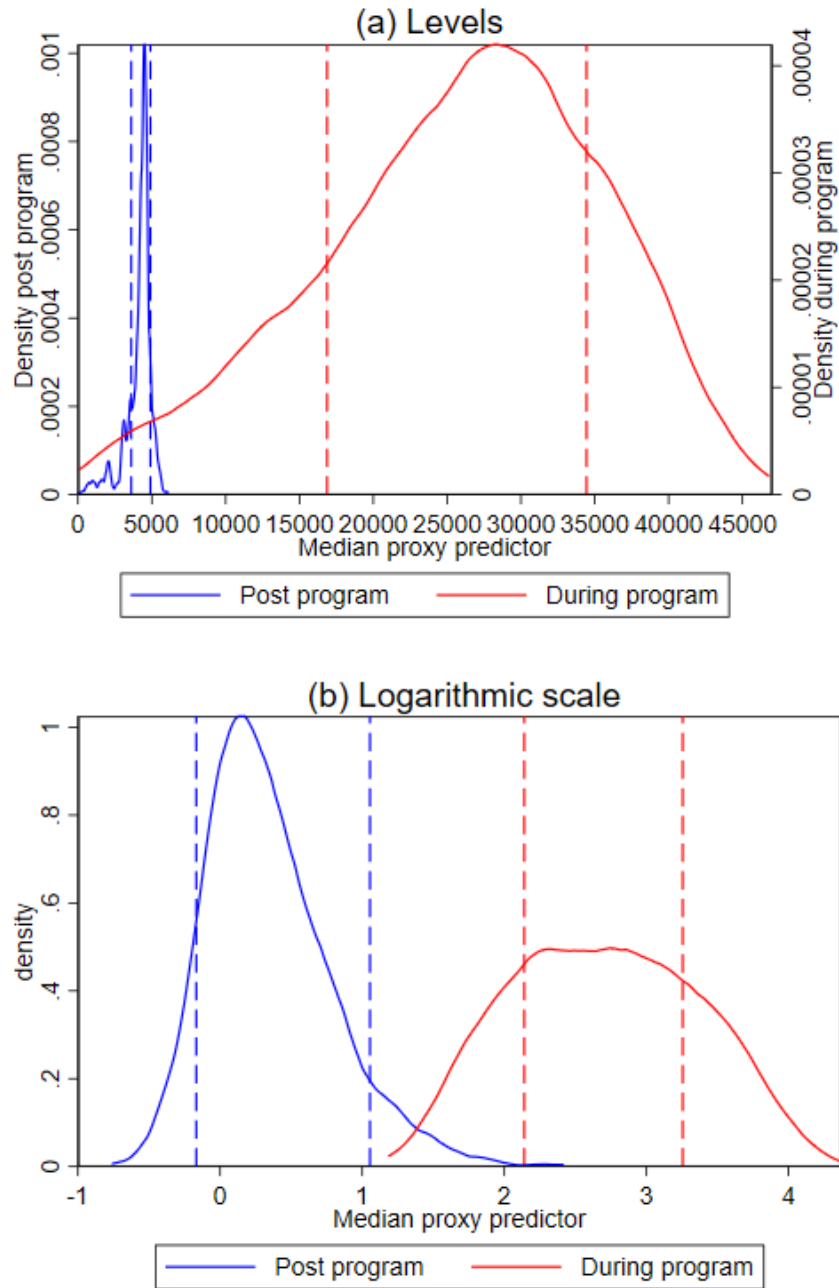
Panel (a) presents two different hypothetical distributions of individual treatment effects. Panel (b) presents simulations using the conceptual framework in section 2. It illustrates the changes induced by a reduction in the transfer by x , assuming the distribution of individual impact on earnings in Panel (a). The dotted black line shows the share of initial applicants who still apply when the transfer is reduced by x . The dashed blue line shows the lottery success rate when the transfer is reduced by x (Lottery rate = $0.25T/((T-x)P(s_T - x > 0))$). The solid red line shows the average impact on earnings $E((s_T - x)1(s_T - x > 0))/(P(s_T - x > 0)(T - x))$ when the transfer amount is reduced by x , $S_{lottery}(T - x)$ from equation 5, normalized by $0.25 \times 50,000$, the amount available for each initial applicant.

Figure B3: Quantile treatment effects for (ln) earnings *post* program, by treatment arm



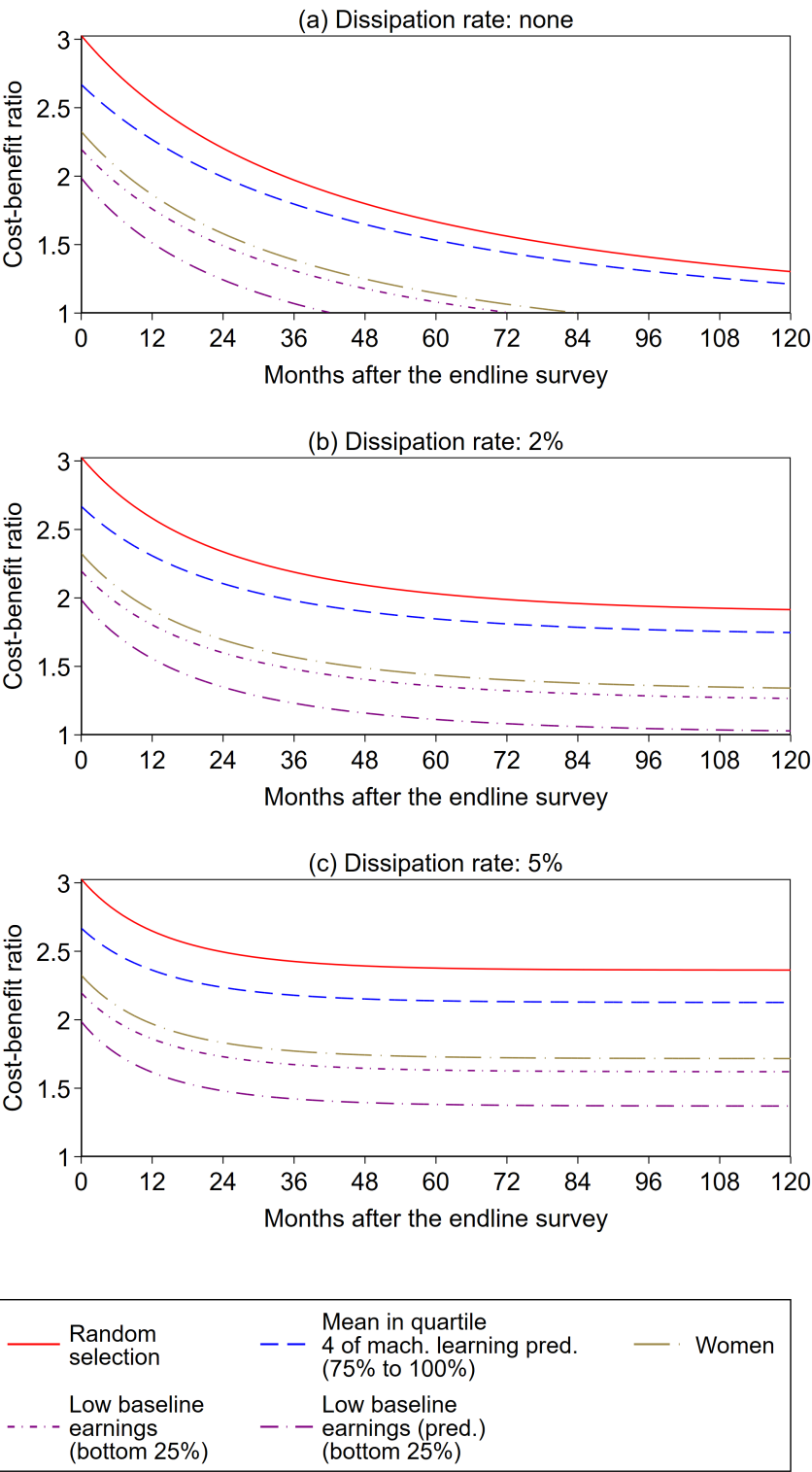
Note: Quantile treatment effects (for positive earnings) up to the 97th percentile. Permutation tests use 10000 permutations for each hypothesis. Wilcoxon rank-sum test permutation p-value is 0.471 for panel (a) 0.464 for panel (b) and 0.459 for panel (c). Kolmogorov-Smirnov permutation p-value is 0.751 for panel (a) 0.776 for panel (b) and 0.750 for panel (c). Total monthly earnings variable is in CFA. For variables (y) in logarithms we take $\ln(y + 1)$.

Figure B4: Distribution of predicted impacts on earnings



Note: Heterogeneity analysis based on the approach in [Chernozhukov et al. \(2020\)](#) (see discussion in Section 4.2 and Appendix F.2). Estimated distribution of individual treatment effects (across 100 simulations). Dashed lines delimit bottom 25% and top 25% of the distribution. Predictions use Random Forest for impacts during program in levels, and Generalized Random Forest for impacts during program in logarithm. Predictions use R-Learner for post-program impact in levels, and Gradient Boosting for post-program impact in logarithm. Total monthly earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y + 1)$.

Figure B5: Cost-effectiveness ratios over time under alternative targeting rules, depending on the sustainability of post-program impacts



Note: The discounted sum of post-program impacts is assumed to continue beyond what we measure at the endline survey, 15 months after the end of the program. The top panel assumes no dissipation of impacts. The middle and bottom panels assume a dissipation rate on top of the discount rate: a 2% monthly dissipation rate (respectively 5%) is equivalent to a 22% decrease in impact in one year (respectively 49%).

Appendix C Additional details on the framework

C.1 Individuals impacts

Let θ denote potential participants' hourly earnings in non-program activities, h hours worked, and $c(h)$ the additive disutility of working h hours. Absent the workfare program, individuals have an optimal number of hours worked $h_0(\theta)$ given by $\theta = c'(h_0)$, leading to earnings $W_0(\theta) = \theta h_0(\theta)$ and utility $U_0(\theta) = \theta h_0(\theta) - c(h_0(\theta))$. Notice we have $U'_0(\theta) = h_0(\theta)$.

Assume that participation in the workfare program provides a transfer T in exchange of h_p hours of work. Further assume that the disutility of time spent in program activities is the same as for time spent in non-program activities. Individuals seek to maximize $T + \theta(h - h_p) - c(h)$. This leads to the first-order condition $\theta = c'(h)$. It corresponds to hours of work $h_1(\theta) > h_p$ if $c'(h_1(\theta)) > c'(h_p)$, i.e. $\theta > \underline{\theta}(h_p) \equiv c'(h_p)$. Notice that in this case $h_1(\theta) = h_0(\theta)$.

When $\theta > \underline{\theta}(h_p)$, the solution of the previous maximization leads to: $h_1(\theta) = h_0(\theta)$. The objective reaches the value $U_1(T, h_p, \theta) = T + \theta(h_0(\theta) - h_p) - c(h_0(\theta)) = U_0(\theta) + T - \theta h_p$. The corresponding earnings are $W_1(T, h_p, \theta) = T + \theta(h_0(\theta) - h_p) = W_0(\theta) + T - \theta h_p$. Individuals will however participate in the program only if $U_1(\theta) > U_0(\theta)$, which is equivalent to $T - \theta h_p > 0$. Thus individuals participate in the program if $\theta < \bar{\theta}(T, h_p) \equiv T/h_p$.

When $\theta < \underline{\theta}(h_p)$, the maximization leads to the constrained solution $h_1(\theta) = h_p$. The objective reaches the value $U_1(T, h_p, \theta) = T - c(h_p) = U_0(\theta) + T - \theta h_0(\theta) - c(h_p) + c(h_0(\theta))$. The corresponding earnings are $W_1(T, h_p, \theta) = W_0(\theta) + T - \theta h_0(\theta)$. Individuals will participate in the program if $T - c(h_p) > \theta h_0(\theta) - c(h_0(\theta)) = U(\theta)$. Thus individuals participate in the program if $\theta < \underline{\underline{\theta}}(T, h_p) \equiv U_0^{-1}(T - c(h_p))$.

We can define the difference in earnings $s(T, \theta) = W_1(\theta) - W_0(\theta)$ absent the final decision to participate:

$$s(T, \theta) = T - \theta (h_p 1(\theta > \underline{\theta}(h_p)) + h_0(\theta) 1(\theta < \underline{\theta}(h_p)))$$

We can show that $\underline{\theta}(h_p) < \bar{\theta}(T, h_p) \Leftrightarrow \underline{\theta}(h_p) < \underline{\underline{\theta}}(T, h_p) \Leftrightarrow T > h_p c'(h_p)$.

Notice first that $\underline{\theta}(h_p) \equiv c'(h_p)$ and $\bar{\theta}(T, h_p) \equiv T/h_p$. Hence, $\underline{\theta}(h_p) < \underline{\underline{\theta}}(T, h_p)$ is directly equivalent

to $c'(h_p) < T/h_p$. Notice also that $c'(h_0(\underline{\theta}(h_p))) = \underline{\theta}(h_p) = c'(h_p)$, thus $h_0(\underline{\theta}(h_p)) = h_p$. Then:

$$\begin{aligned}
& \underline{\theta}(h_p) < \bar{\theta}(T, h_p) \\
& \Leftrightarrow \underline{\theta}(h_p) < T/h_p \\
& \Leftrightarrow \underline{\theta}(h_p)h_p - c(h_p) < T - c(h_p) \\
& \Leftrightarrow U(\underline{\theta}(h_p)) < U(\underline{\theta}(T, h_p)) \\
& \Leftrightarrow \underline{\theta}(h_p) < \underline{\theta}(T, h_p)
\end{aligned}$$

There are thus two situations (see Appendix Figures B1):

$T > h_p c'(h_p)$: individuals with $\theta < \underline{\theta}(h_p)$ participate in the program and they only work in the program. Their hours of work are given by $h_1(\theta) = h_p > h_0(\theta)$. Individuals with $\underline{\theta}(h_p) < \theta < \bar{\theta}(T, h_p)$ participate in the program and also work outside the program. Their hours of work are given by $h_1(\theta) = h_p = h_0(\theta)$. Marginal participants are individuals with $\theta = \bar{\theta}(T, h_p)$. Because total hours of work are the same, for marginal applicants $W_0(\theta) = W_1(T, h_p, \theta)$.

$T < h_p c'(h_p)$. Only individuals with $\theta < \underline{\theta}(T, h_p)$ participate in the program and they only work in the program: $h_1(\theta) = h_p > h_0(\theta)$. Marginal participants are individuals with $\theta = \underline{\theta}(T, h_p)$. Because $h_1(\theta) = h_p > h_0(\theta)$, for marginal applicants $W_0(\theta) < W_1(T, h_p, \theta)$.

C.2 Overall impact and changes in the amount of the transfer T

As discussed in Section 2.1 the program is oversubscribed and a lottery is used to allocate available slots among applicants. We discuss here some aspects related to the variation of the overall impact when the amount of transfer is changed from T to $T - x$. In Section 2.1 we derive the expression of the average contemporaneous impact:

$$S_{\text{lottery}}(T - x) = \lambda(T - x)E((s_T - x)1(s_T - x > 0)) = \lambda(T)\frac{T}{T - x}E(s_T - x | s_T > x) \quad (11)$$

where $\lambda(T - x)$ is the rate of success of the lottery and is given by $B = \lambda(T - x)(T - x)N_A(T)P(s_T > x)$.

Assume we start from an initial T larger than $h_p c'(h_p)$, and progressively increase x . Applicants are those for whom $U_1(T - x, h_p, \theta) > U_0(\theta)$, and marginal applicants are those with productivity $\theta_m(x)$

such that $U_1(T - x, h_p, \theta_m(x)) - U_0(\theta_m(x)) = 0$. As emphasized before, as long as $T - x > h_p c'(h_p)$, we also have $W_1(T - x, h_p, \theta_m(x)) - W_0(\theta_m(x)) = 0$: marginal applicants have the same earnings as if they did not participate in the program. This holds as long as $T - x < h_p c'(h_p)$. All individuals participating in the program work longer hours than absent the program and thus $W_1(T - x, h_p, \theta_m(x)) - W_0(\theta_m(x)) = c(h_p) - c(h_0(\theta_m(x))) > 0$, but marginal applicants are still defined by $U_1(T - x, h_p, \theta_m(x)) - U_0(\theta_m(x)) = 0$.

We can also obtain the derivative of this objective with respect to x . How this function varies with x depends on the shape of the density function of s_T , the density of treatment effects under transfer T . A change dx of x leads to a change $dS_{lottery}(T - x) = \lambda(x)dE((s_T - x)1(s_t - x > 0)) + E((s_T - x)1(s_t - x > 0))d\lambda(x)$. The first component $\lambda(x)dE((s_T - x)1(s_t - x > 0))$ simply writes as $-\lambda(x)P(s_T > x)dx$: the increase in x has a direct negative impact because the transfer is smaller, and this applies to all of those who receive the transfer. The second component is driven by a change in the lottery rate. Using the equation defining the lottery rate, we see that $d\lambda(x) = \lambda(x)(1/(T - x) + f(x)/P(s_T > x))dx$. The derivative is unambiguously positive. The savings made because some former participants would no longer apply (the term with $f(x)$) and because the transfer per participant is reduced allows to increase the share of applicants that will be served. The derivative is positive, as long as $\lambda(x) < 1$. The negative initial impact due to the reduction in the transfer is mitigated by an increase in the lottery rate. The overall change is thus

$$\frac{dS_{lottery}(T - x)}{dx} = \lambda(x) \left(\left(\frac{1}{T - x} + \frac{f(x)}{P(s_T > x)} \right) E((s_T - x)1(s_t - x > 0)) - P(s_T > x) \right) \quad (12)$$

Appendix D Description of complementary training

Randomized subsets of beneficiaries received complementary training on basic entrepreneurship or job search skills. Each training lasted approximately 80-100 hours over two two-week periods. Field exercises were undertaken between the training periods, in parallel to the public works jobs (typically in the afternoons). The training was delivered by work brigades, i.e. in groups of 25 youths. Participants did not have to work during the training, but still received their corresponding daily wage.⁶⁴ The curricula for the complementary skills training were tailored for low-skill populations that may not be able to read and write, in particular by relying on drawings

⁶⁴Some youths were offered the second half of the training after their exit from the public works program. While these youths were not paid during that time, they received a small stipend to cover transportation costs.

and visuals.

The basic entrepreneurship training aimed to build skills to help youth set-up and manage a small non-agricultural micro-enterprise. The training lasted 100 hours and focused 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) covered topics related to basic entrepreneurship and business skills. A second phase included field research for youths to collect information, undertake basic market research and outline a business plan. A third phase (40 hours over two weeks) included feedback on youths' basic business plans, and reviewed related topics from the curriculum. The final phase (20 hours) included post-training follow-up.

The training on wage jobs search skills provided 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 lasted 80 hours. The first phase (40 hours over two weeks) discussed 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 included field exercises to collect information on potential opportunities, identify and visit potential employers or professional networks, etc. The third phase (40 hours over two weeks) provided feedback on field exercises, reviewed part of the curriculum and provided additional practical guidance. In addition, supervisors of the brigades who were offered the wage employment training were also trained on how to manage teams and provide feedback to workers, with the objective to mimic the professional experience one would have in a more formal wage job. Youths were periodically rated on a range of skills, and these evaluations were later used to issue a work certificate that signaled between one and five competencies identified as strengths for each participant.⁶⁵

Appendix E Definition of key outcome variables

Total monthly earnings are expressed in CFA francs. They are aggregated over up to three activities undertaken by an individual in the 30 days preceding the survey. They include payments received in cash and the monetary equivalent for in-kind payments. The variable is winsorized at

⁶⁵The evaluation policy report contains additional details on the trainings (Bertrand et al., 2016).

97% (unless stated otherwise). Total monthly earnings are decomposed in total (monthly) earnings from wage employment and self-employment (as well as earnings from other occupations, which are generally small hence not shown separately). When shown in log, the log transformation is applied to earnings plus one.

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 official employment indicators used in Côte d'Ivoire. We assign a value of 0 for inactive and unemployed individuals. To provide information on the composition of employment, we also analyze having at least one wage job (*Wage employed*), or at least one self-employment activity (*Self-employed*).

Weekly hours worked capture the total number of hours worked over the 7 days preceding the survey. It aggregates information from up to three activities undertaken by an individual across all occupations (wage employment, self-employment or other types of activity). The variable is winsorized at 97%. Weekly hours worked are decomposed in (*hours worked in wage employment*) and (*hours worked in self-employment*) (as well as hours worked in other occupations, which are generally small and 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 sources. The variable is winsorized at 97%. When shown in log, the log transformation is applied to savings plus one.

Total expenditures is expressed in CFA francs and aggregates several types of expenditures, both for the individual and for other household members. It includes basic expenditures (health, clothing, sanitation, and accommodation), communication expenditures (mobile, internet, and medias), investments (education, training, maintenance of assets), transportation expenditures, temptation goods (alcohol, tobacco, gambling, and luxury goods) and social expenditures (celebrations and charity). The variable is winsorized at 97%. When shown in log, the log transformation is applied to expenditures plus one.

The *well-being index* aggregates 6 measures: two measures of *happiness* and *pride* in daily activities from a time-use module,⁶⁶ the Rosenberg *self-esteem* scale,⁶⁷ the *positive affect sub-scale*

⁶⁶The time use module measured which activities the respondent performed at different times of the last “business day” (at 6am, 10am, 3pm, 7pm and 10pm). Respondents were also asked whether they felt happy, proud or angry while performing those activities. The measure of happiness (respectively pride) is the number of times (out of the 5 times in the last day) respondents reported feeling happy (respectively proud). A z-score of the measure is included in the well-being index.

⁶⁷The Rosenberg self-esteem scale includes 10 items that measure self-esteem or self-worth. We use a validated

from the CESD scale,⁶⁸ the sub-scale of *(positive) attitude towards the future* and the inverted sub-scale of *present fatalism* from the ZTPI scale.⁶⁹ The well-being index is a z-score, with a mean of zero and a standard deviation of one 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 improvement in well-being.⁷⁰

The *behavior index* aggregates 4 measures: an inverted measure of *anger* in daily activities taken from the time-use module,⁷¹ an inverted measure of *impulsiveness* from the DERS scale,⁷² the *conduct problems sub-scale* (inverted) and the *pro-social behavior sub-scale* from the Strengths and Difficulties Questionnaire (SDQ).⁷³ As for the well-being index, the behavior index is a z-score with a mean of 0 and standard deviation of 1 in the control group, so that estimated coefficients can be interpreted in standard deviations. An increase in the index corresponds to an overall improvement in behavior and attitude.⁷⁴

The *risky behavior index* is the mean of 9 risky behaviors measured through list experiments. They include stealing, assaulting someone, believing smuggling is necessary (to earn a living), prostitution, threatening someone, taking illicit drugs, smuggling, having ties with a smuggling network, and having firearms at home. Because respondents may not respond truthfully to direct questions about these sensitive behaviors, we used list experiments instead. Rather than asking directly a sensitive question about a risky behavior (e.g. stealing), 5 affirmations are read to respondents, and respondents are asked how many of these affirmations (between 0 and 5) are

version of the instrument in French (Vallieres and Vallerand, 1990).

⁶⁸The Center for Epidemiologic Studies Depression (CESD) scale includes an inverted subscale that measures positive feelings (“Positive Affects”). We use a validated version in French (Morin et al., 2011).

⁶⁹The Zimbardo Time Perspective Inventory (ZTPI) captures different dimensions of time perspectives. We use the two subscales of “future” (to have a positive attitude towards future) and “present fatalism” which is very close to the concept of external locus of control, in the sense that one feels *no* control over life events. The inverted “present fatalism” measure is therefore similar to a measure of internal locus of control. We use a validated version of the instrument in French (Apostolidis and Fieulaine, 2004).

⁷⁰The index adds up the 6 measures described above, out of which one is inverted (*present fatalism*). Therefore a negative impact on the *present fatalism* measure induces an improvement in the overall well-being index, corresponding to greater well-being.

⁷¹This was built as in footnote 66.

⁷²The Difficulties in Emotion Regulation Scale (DERS) is used to measure socio-emotional regulation, in particular the difficulties of regulation of emotions in adults. Three of the six questions of the “difficulties to control impulsive behavior” scale were retained, based on a validated French version of the instrument (Côté et al., 2013).

⁷³The Strengths and Difficulties Questionnaire (SDQ) measures behavioral difficulties in young people, initially among children and adolescents from 3 to 16 years old (Goodman et al., 1998). The instrument was slightly adapted for an older age group 18 to 30 years old. We use two of five sub-scales from a validated questionnaire in French available at www.sdqinfo.com.

⁷⁴The index adds up the 4 measures described above, which are all inverted in the index except *pro-social behavior* measure. A negative impact on inverted measures, for example *conduct problems*, corresponds to a positive behavior and leads to an improvement in the overall behavior index.

true for them. To estimate the proportion of individuals for which the sensitive question is true in a sample, the sample is (randomly) assigned to two lists. The first list includes 5 affirmations including the risky behavior, and the second list only includes the other 4 affirmations (without the risky behavior). We implemented a “double” list experiment to avoid losing statistical power: each half of the sample answered both a list with sensitive questions, and a (different) list with control questions corresponding to the other sample. List experiments were piloted extensively to ensure a good understanding by respondents. In the analysis, we use the likelihood ratio test introduced by [Blair and Imai \(2012\)](#) to test for the existence of design effects.

Appendix F Machine Learning Application to Study Heterogeneity in Treatment Effects

To complement Sections [4.2](#) and [6.2](#), this appendix provides additional details on the application of machine learning methods to analyze treatment effect heterogeneity. The application is based on [Chernozhukov et al. \(2020\)](#). Section [F.1](#) describes the sample used to train the models and make predictions. Section [F.2](#) presents the machine learning algorithms and their parameters. Finally, Section [F.3](#) describes how we adapted the procedure in [Chernozhukov et al. \(2020\)](#) to our experimental setting.

We use similar notations as in [Chernozhukov et al. \(2020\)](#). For clarity, the Baseline Conditional Average (BCA) writes:

$$b_0(Z) := E[Y(0)|Z]$$

And the Conditional Average Treatment Effect (CATE) is defined as :

$$s_0(Z) := E[Y(1)|Z] - E[Y(0)|Z]$$

F.1 Sample for Machine Learning Implementation

Supervised machine learning algorithms require samples for which both covariates (features) and outcomes are observed. In our case, this requires baseline covariates (a set of K covariates, Z^K) and midline or endline outcome of interest (Y). As discussed in the text, our study data

has two specificities. First, our midline sample is a subsample of the baseline, while the full baseline sample is included at endline. Second, some control individuals entered subsequent waves of the public works program between midline and endline surveys, and are excluded from the endline sample used for analysis. As a result, the algorithms can use three potential samples: a ‘midline’ (Z_i^K, Y_i^{During}, W_i) (respectively ‘endline’ (Z_i^K, Y_i^{Post}, W_i)) sample can be used to build and apply the model to predict ‘during’ (respectively ‘post’) conditional treatment effects, where W corresponds to the treatment variable. A third (marginally smaller) sample can be used to study how effects vary ‘during’ and ‘post’ program by taking the intersection of non-attriters and non-missing outcomes for both surveys.⁷⁵ When applying the algorithm on the endline data, we drop control individuals who applied to a later wave of the public works program (as in the main analysis).⁷⁶ The final sample size depends on the number of missing variables for the outcome considered. The total sample we use ranges between 2,884 and 2,958 units for midline and between 3,745 and 3,910 units for endline.

We use a set (Z^K , with $K = 21$) of features (covariates) measured at baseline (Table A13). They include both individual and household characteristics, as well as main indicators on employment, financial situation and self-reported constraints on basic needs expenditures. We also show the robustness of the main results to the inclusion of all baseline variables in the balance check table (Table 1).

F.2 Machine Learning Algorithms

We consider five alternative machine learning algorithms to estimate the proxy predictors and apply the procedure in Chernozhukov et al. (2020): elastic net, boosted trees, random forest, Rlearner (based on elastic net and proposed by Nie and Wager (2020)) and generalized random forest (proposed by Wager and Athey (2018)). All algorithms are implemented in **R**, and we adapted the codes provided by the authors.⁷⁷ These machine learning methods can be divided in two groups based on the way they approach the CATE function (Künzel et al. (2017)).

⁷⁵There is also some attrition between survey rounds, and some missing values in baseline covariates. We exclude from each sample the attriters from follow-up surveys (since outcomes are not observed). Missing values among baseline covariates are replaced by the mean in the same strata. Individuals with missing values for the outcome of interest (among nonattriters) are dropped from the sample.

⁷⁶Recall that 200 individuals were sampled to be added to the control group at endline to compensate for these observations: because these individuals were not part of the baseline survey, the machine learning model cannot be applied to them since predictions rely on observed Z^K .

⁷⁷<https://github.com/demirermert/MLInference>

1. Two learners

The first group of machine learning methods includes Elastic Net, Random Forest and Boosted Trees. They predict separately $E[Y_i(1)]$ in the treatment group and $E[Y_i(0)]$ in the control group. In practice, a first model is fitted on the treatment group and a second on the control group, using an auxiliary sample.⁷⁸ The two fitted models are then used to predict potential outcomes $\hat{Y}_i(1)$ and $\hat{Y}_i(0)$ for each individual in the remaining sample (main sample). In order to obtain $S(Z)$, the difference between the two predictions for each individual are computed. All models are implemented using the **caret** package (Kuhn (2008)) (respectively named glmnet, ranger and gbm).

Tuning parameters

For each split, the tuning parameters are chosen separately for the model on the control and the treatment group. There are no set rule to choose these parameters. In our case, we let **caret** define a default search grid and we set a relatively high tuning length for all models based on our computational capacities. Tuning parameters were all selected based on the mean squared error estimates and 5-folds cross validation. For all methods, we pre-process outcomes and covariates and center-scale them before feeding the model.

For each method we have the following tuning parameters :

- Elastic net : *alpha* (Mixing Percentage), *lambda* (Regularization Parameter)
- Boosted trees : *n.trees* (# Boosting Iterations), *interaction.depth* (Max Tree Depth), shrinkage (Shrinkage), *n.minobsinnode* (Min. Terminal Node Size),
- Random Forest : *mtry* (# Randomly Selected Predictors), *splitrule* (Splitting Rule), *n.minobsinnode*

2. Single learners

The two alternative models we consider are Rlearner and Generalized Random Forest (with their variations). They are “single learners” and use a different approach to approximate $s_0(Z)$. Instead of fitting a model on the treatment and on the control group separately to estimate $E[Y_i(1)]$ and $E[Y_i(0)]$, they directly fit a model to estimate $E[Y_i(1)] - E[Y_i(0)]$.

⁷⁸The sample is split between an auxiliary sample where machine learning predictors are trained and constructed and a main sample where they are used for prediction and on which we estimate the different key features of the CATE function.

Athey and Imbens (2016) discuss the benefits of this approach compared to the two-learners approach. One remaining quantity, the Baseline Conditional Average $b_0(Z)$, is needed. For Rlearner with boosting, we use boosted trees fitted on the control group to estimate $b_0(Z)$ and symmetrically elastic net for Rlearner based on lasso. For Generalized Random Forest we predict $b_0(Z)$ using the random forest already fitted on the control group. We rely on the **grf** package to implement Generalized Random Forest and on the **rlearner**⁷⁹ package for Rlearner.

Tuning parameters

For each split of the data, we choose the tuning parameters separately for $S(Z)$ and $B(Z)$. Again, there is no theoretical basis to determine the choice of search grid parameters. We keep the package default grid and put a convenient length of parameters combination according to our computational capabilities. Parameters were selected based on the mean squared error estimates and 5-folds cross validation.

F.3 Adaption to the Experimental Setting

In our application, we repeat $S = 100$ times the procedure developed by Chernozhukov et al. (2020).⁸⁰ The first step of the method requires partitioning our dataset into an auxiliary and a main sample. We adapted the algorithm so that the sample splits are stratified by our randomization blocks (*locality * gender*), which represents 32 strata. This is important to preserve the identification strategy when estimating directly the CATE for the single learners, since they fit a model on different splits of the data.

Lastly, we introduce two adjustments in the linear projections of Best Linear Predictor (BLP) and Group Average Treatment Effects (GATE) along with predicted baseline effect $B(Z)$ and predicted treatment effect $S(Z)$. We add locality-gender fixed effects, corresponding to the randomization stratification variables. We also adjust the weights used. In the main specification of the paper, we use weights to take into account randomized assignment by lotteries, survey

⁷⁹<https://github.com/xnie/rlearner>

⁸⁰Appendix figure B6 shows the scatter plot of earnings y_i and \hat{y}_i for midline and endline as well as the regression line of y on \hat{y} . The figure shows that the slope coefficients are close to 1 and that the R^2 remain low. Note that a low R^2 on $y(1)$ and $y(0)$ does not mean that our algorithm cannot identify heterogeneity in the treatment effects, which is related to difference between $y(1)$ and $y(0)$. This is illustrated in the paper where we find heterogeneity in impacts on midline earnings.

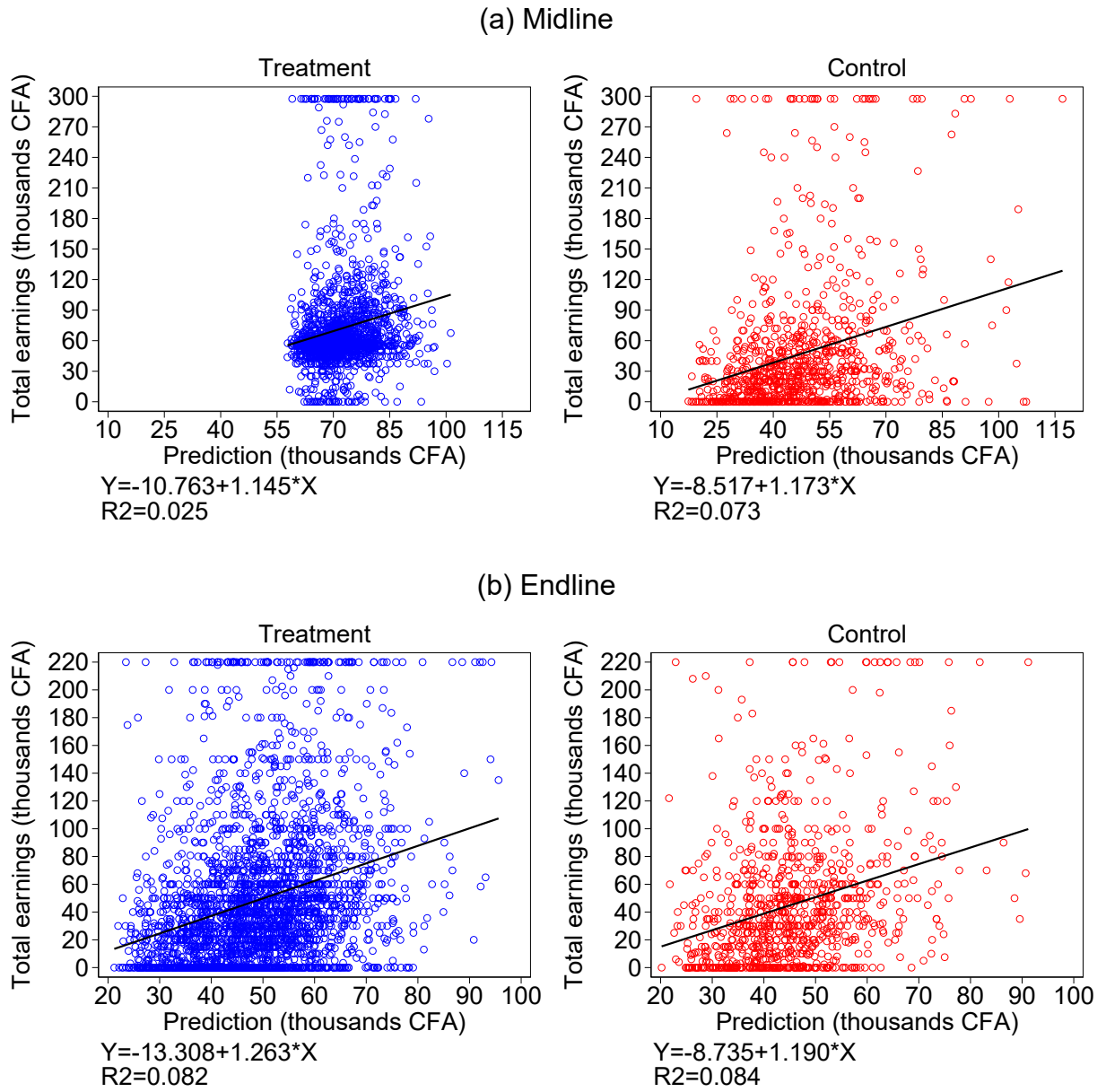
attrition, and sub-sampling at midline. Since our survey weights will be multiplied by inverse propensity score weights as recommended by [Chernozhukov et al. \(2020\)](#), we make sure not to incorporate two times the inverse propensity score.

Table A13: Baseline variables used in Machine Learning algorithms

Variable description	Type
<i>Individual characteristics</i>	
Female	Binary
Age	Continuous
Nb of children	Continuous
Live in urban area	Binary
<i>Education</i>	
Education (total number of years)	Continuous
Has participated in vocational training	Binary
<i>Household characteristics</i>	
Household size	Continuous
Is head of household	Binary
<i>Household assets</i>	
Total nb of assets ¹	Continuous
<i>Employment</i>	
Total nb of activities	Continuous
Total nb of wage-employment activities	Continuous
Total nb of self-employment activities	Continuous
Is engaged in (at least one) casual activity	Binary
Total Earnings (monthly)	Continuous
<i>Savings, Expenditures and Constraints</i>	
Has Saved (last 3 months)	Binary
Savings Stock (FCFA)	Continuous
Has a Savings Account	Binary
Self-reported constraints to repay loans	Binary
Self-reported constraints to access credit	Binary
Transportation expenditures (last 7 days, CFA)	Continuous
Communication expenditures (last 7 days, CFA)	Continuous

[1] Assets include livestock, chicken, other animals, plows, field sprayer, carts, wheelbarrows, bicycles, motorcycles, pirogues, refrigerators, freezers, air conditioning units, fans, stoves, computers, radios, television, TV antenna, video players, landline, mobile phones, cars.

Figure B6: Relation between predictions and actual earnings (in level) (Random forest)



Note: Predictions estimated using equation $\hat{Y} = \hat{B}(Z) + T * \hat{\tau}(Z)$. Each value plotted is the median across 100 sample splits.

Appendix G Weights

This appendix describes the weights used in the analysis. Table A14 summarizes the weights used with midline data, and Table A15 with endline data. In general results are robust if weights are not included.

Randomization weights

We consider two sets of randomization weights. First, for both midline and endline, we consider weights that account for variations in selection probability by lottery location and gender. There are K different public lotteries ($K = 32$) with N_k individuals participating to each lottery. Denote N_{k1} the individuals from lottery k selected in the program ('treated') and N_{k0} those who are not selected, with $N_k = N_{k1} + N_{k0}$. Among the N_{k0} , N_{k0s} are randomly drawn to be surveyed and constitute the 'control group'. The size of the population of lottery participants is N_P , with $N_P = \sum_k N_k = N_1 + N_0$. The size of the survey sample is $N_E = \sum_k N_{k1} + N_{k0s} = N_1 + N_{0s}$. We use weight w_{ki} ($i = 0_s; 1$ according to treatment status) for individuals in the survey sample, with $w_{k1} = N_k / N_{k1} \times N_1 / N_P$ and $w_{k0s} = N_k / N_{k0s} \times N_0 / N_P$. This means that we put a higher weight on lotteries where the demand for the program (total population participating in the lottery) was higher, compared with other lotteries.

Second, when estimating treatment effects by arm using endline survey data, we also consider that the number of brigades assigned to each treatment arm varies by locality. Brigades of treated individuals (N_1) are assigned to 3 treatment options T_a , T_b and T_c . 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 a weight w_{k0s} for control individuals in the survey sample, with:

- $w_{j,k} = N_{k1} / N_{j,k} \times N_j / N_1$ with $j = a, b, c$ ⁸¹
- $w_{k0s} = N_{k1} / N_{k0s} \times N_0 / N_1$

Sub-sampling weights (midline survey only)

The sample for the midline survey includes the control group (N_{0s}) and a sub-sample of the treatment group. Consider that we draw a random sub sample of group l in proportion $P_l = N_l^S / N_l$. To take sub-sampling into account, original weights are multiplied by S / P_l . Therefore,

⁸¹Note : $\sum_j w_{j,k} = w_{k1} = 1$, which is the weight used for midline data when there is only one treatment group.

in group $l = k, 1$ we draw N_{k1}^S individuals out of N_{k1} , and the original weight w_{k1} becomes $\omega_{k1}^S = w_{k1} \times N_{k1}/N_{k1}^{S_{k1}}$. All control units are included in the midline sample so that their weights w_{k0s} are unchanged.

Control group and subsequent enrollment in the program (endline survey only)

When using endline data, we adjust weights for control individuals because some of them were able to apply (and sometimes get selected) in waves 3 and 4 of the program, as discussed in section 4.⁸² Weights for control individuals depend on their status in wave 3 and wave 4, which is one of the following 7 situations:

1. Group $C_3T_3\bar{C}_4$: 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).
2. Group $C_3\bar{T}_3C_4T_4$: 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).
3. Group $C_3\bar{T}_3C_4\bar{T}_4$: 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).
4. Group $C_3\bar{T}_3\bar{C}_4$: Applied to wave 3 (C_3), was not selected after public lotteries (\bar{T}_3) and did not apply to wave 4 (\bar{C}_4).
5. Group $\bar{C}_3C_4T_4$: 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).
6. Group $\bar{C}_3C_4\bar{T}_4$: 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).
7. Group $\bar{C}_3\bar{C}_4$: Did not apply to wave 3 (\bar{C}_3), and did not apply to wave 4 (\bar{C}_4).

We introduce a new multiplicative weight for control units ($\tilde{w}_{k0s,j}$). We do not include control units that have benefited from subsequent waves of the program (waves 3 and 4) in the estimation. This means we assign a weight of 0 to groups $C_3T_3\bar{C}_4$, $C_3\bar{T}_3C_4T_4$ and $\bar{C}_3C_4T_4$.⁸³ To compensate, we put a higher weight on individuals who also applied in subsequent phases (waves 3 and 4) but were not selected during the lotteries. The weights for the remaining four groups are:

$$\bullet \tilde{w}_{k0s,C_3\bar{T}_3C_4\bar{T}_4} = \frac{N_{k0s,C_3}}{N_{k0s,C_3\bar{T}_3}} \times \frac{N_{k0s,C_3\bar{T}_3C_4}}{N_{k0s,C_3\bar{T}_3C_4\bar{T}_4}}$$

⁸²Recall that the study focuses on wave 2 (out of 4 waves) of the public works program

⁸³Hence $\tilde{w}_{k0s,C_3T_3\bar{C}_4} = 0$; $\tilde{w}_{k0s,\bar{C}_3C_4T_4} = 1 \times 0 = 0$; $\tilde{w}_{k0s,C_3\bar{T}_3C_4T_4} = \frac{N_{k0s,C_3}}{N_{k0s,C_3\bar{T}_3}} \times 0 = 0$.

- $\tilde{w}_{k0s,C_3\bar{T}_3\bar{C}_4} = \frac{N_{k0s,C_3}}{N_{k0s,C_3\bar{T}_3}} \times 1 = \frac{N_{k0s,C_3}}{N_{k0s,C_3\bar{T}_3}}$
- $\tilde{w}_{k0s,\bar{C}_3C_4\bar{T}_4} = 1 \times \frac{N_{k0s,\bar{C}_3C_4}}{N_{k0s,\bar{C}_3C_4\bar{T}_4}}$
- $\tilde{w}_{k0s,\bar{C}_3\bar{C}_4} = 1$

Tracking weights

Lastly, we add a weight taking into account the differential response rate of individuals during each survey (midline and endline). More precisely, each survey consisted in two phases a and b :

- A 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 main 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.⁸⁴ $E_{b,j}$ from which a random sub-sample is drawn in proportion $\pi_j = NE_{b,j}^S / NE_{b,j}$ (j is an index for treatment status \times locality). Individuals interviewed during the tracking phase take a different weight than those interviewed during the main survey phase. Tracking respondents are weighted by $\omega_j^T = (R_{a,j}^S + \lambda_j s_j R_{b,j}^S (1 - R_{a,j}^S) E_{b,j}^S)$, with λ_j , so that the final weight is $\omega_j^{S,f} = \omega_j^S \times \omega_j^T$.

The sum of the weights on population j is therefore : $\omega_j \times (N_{a,j}^S + \lambda_j NE_{b,j}^S)$, with $NE_{b,j}^S$ the number of individuals from the tracking sample who responded during tracking phase. We make the hypothesis that residual non-response $R_{b,j}^S$ is random. We seek to be representative of the respondent population of phases a and b . This lead us to take $\lambda_j = NE_{b,j}^S / NE_{b,j}^S$

In group j , weights will be set such as:⁸⁵

- $\omega_j^S \times 1$ for phase a respondents
- $\omega_j^S \times NE_{b,j}^S / NE_{b,j}^S$ for phase b respondents

⁸⁴Among the attritors of phase (a) some individuals were considered ‘ineligible’ for tracking as they were (quasi) impossible to reach: dead individuals, individuals who migrated to another country, (for endline) individuals who were already impossible to find at baseline.

⁸⁵In theory, ω_j should be adjusted so that it does not use correction N_j / N_j^S 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 were randomly drawn. Finally, it means that we estimate the unknown amount $N_{a,j} + NE_{b,j}$ by $N_{a,j}^S + NE_{b,j}^S \times N_j / N_j^S$

Table A14: Summary of weights used with midline data

Randomization weights w_k		Sub Sampling weights ω_k^S		Tracking weights ω_j^T	
Treated	$w_{k1} = N_k / N_{k1} \times N_1 / N_P$	Treated	$w_{k,1}^S = N_{k1} / N_{k1}^{S_{k1}}, k=\text{locality}$	Respondents main phase ($R_a = 1$)	$\omega^T = 1$
Control	$w_{k0s} = N_k / N_{k0s} \times N_0 / N_P, k=\text{locality} \times \text{gender}$	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 $R_b = 1$), $j=\text{locality} \times \text{treatment status}$
					$\omega^T = 0$ 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), $k \in \llbracket 1, 32 \rrbracket$ (locality x gender)					

Table A15: Summary of weights used with endline data

Randomization weights $w_{j,k}$		Post-enrollment weights $\tilde{\omega}_{k,j}$		Tracking weights ω_j^T	
Treatment arm T_a, T_b or T_c	$w_{j,k} = N_k / N_{j,k} \times N_j / N_P, j = a, b, c$	Selected to participate to wave 3 or 4 (groups $C_3 T_3 \bar{C}_4, C_3 \bar{T}_3 C_4 T_4$ et $\bar{C}_3 C_4 T_4$)	0	Respondents main phase ($R_a = 1$)	$\omega^T = 1$
Control	$w_{k0s} = N_k / N_{k0s} \times N_0 / N_P, k=\text{locality x gender}$	Group $C_3 \bar{T}_3 C_4 \bar{T}_4$	$\frac{N_{k0s,C3}}{N_{k0s,C3\bar{T}_3}} \times \frac{N_{k0s,C3\bar{T}_3C4}}{N_{k0s,C3\bar{T}_3C4\bar{T}_4}}$	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 $R_b = 1$), $j=\text{locality x treatment status}$
		Group $C_3 \bar{T}_3 \bar{C}_4$	$\frac{N_{k0s,C3}}{N_{k0s,C3\bar{T}_3}}$		$\omega^T = 0$ if non respondent (but sampled) in tracking phase ($E_b = 1$ et $R_b = 0$)
		Group $\bar{C}_3 C_4 \bar{T}_4$	$\frac{N_{k0s,\bar{C}_3C4}}{N_{k0s,\bar{C}_3C4\bar{T}_4}}$		
		Group $\bar{C}_3 \bar{C}_4$	1		$\omega^T = 0$ if not sampled for tracking phase ($E_b = 0$)
Final weight: $w_{k,i}^F = w_{j,k} \times \tilde{\omega}_{i,l} \times \omega_{k,i}^T, j = 0, a, b, c$ (treatment status), $i = 1, 0s, l$ post-enrollment group, $k \in \llbracket 1, 32 \rrbracket$ (locality x gender)					