Supplemental Appendix

Do Workfare Programs Live Up to Their Promises?
Experimental Evidence from Côte d'Ivoire Dual Apprenticeships

Marianne Bertrand, Bruno Crépon, Alicia Marguerie, Patrick Premand

Appendix A Additional Tables

Table A1: Overview of Public Works Programs

(1) Country	(2) Program	(3) Public works program funding ($m)	(4) # of beneficiaries	(5) % of country labor force	(6) Max # of times the same beneficiary can participate	(7) # of work days per year	(8) Uses self-targeting (individuals apply)	(9) Uses other targeting methods (1=no, universal, 2=community targeting, 3=proxy means test, 4=Lottery)	(10) Daily wage ($, PPP)	(11) Total transfer per beneficiary per year ($, PPP)	(12) Has an objective to improve post-program outcomes	(13) Delivers complementary training
Panel A. Active Public Works Programs in sub-Saharan Africa (with World Bank support, 2020)												
Benin	Community and Local Government Basic Social Services Project	1	16,200	0.35	2	48	0	3	4.7	225	1	1
Cameroon	Social Safety Nets for Crisis Response	6	30,000	0.32	1	60	1	3	4.1	249	1	1
Congo, Dem. Rep.	Eastern Recovery Project	19	12,000	0.05	1	80	1	2, 4	5.2	414	1	1
Comoros	Social Safety Net Project	3	5,890	2.80	3	60	1	2	4.9	295	1	1
Côte d'Ivoire	Emergency Youth Employment & Skills Development Project	18	12,500	0.18	1	132	1	4	10.1	1,365	1	1
Ghana	Productive Safety Net Project	28	30,000	0.24	2	90	1	2, 3	7.2	646	1	1
Ethiopia	Urban Productive Safety Net	171	1,564,416	3.24	3	40	1	2, 3	7.4	298	1	1
Liberia	Youth, Employment, Skills Project	9	45,000	0.43	1	40	1	2	6.0	239	1	1
Liberia	Youth Opportunities Project	4	10,000	0.08	1	100	1	2	5.8	579	1	1
Madagascar	Productive Safety Net Project	15	32,500	0.27	3	80	1	2, 3	3.4	271	1	0
Malawi	Social Action Fund	20	600,000	10.54	1	15	0	2	3.4	51	1	0
Niger	Safety Net Project	11	60,000	0.88	1	60	1	3	4.4	265	0	0
Niger	Adaptive Safety Net Project 2	19	66,000	0.78	2	60	1	2	5.1	307	1	1
Nigeria	Youth Employment and Social Support Operation	200	250,000	0.47	2	240	0	2, 3	3.5	831	1	1
Tanzania	Productive Social Safety Net	25	230,000	1.04	1	60	0	2, 3	4.6	277	1	1
Tanzania	Productive Social Safety Net II	75	837,573	3.08	1	60	0	2, 3	3.3	201	1	0
Uganda	Third Northern Uganda Social Action Fund	49	499,000	3.54	2	60	0	2	3.6	213	1	1
	Average for projects in Panel A:	**39**	**253,005**	**1.66**	**1.65**	**76**	**0.65**		**5.23**	**396**	**0.94**	**0.76**
	Total for projects in Panel A:	**671**	**4,301,079**									
Panel B. Most Studied Programs in Quasi-Experimental Literature												
India	NREGA National Rural Employment Guarantee	10,533	111,900,000	23.72	unlimited	35	1	1	9.1	317	0	0
Ethiopia	PSNP Public Works sub program	282	6,803,770	12.89	unlimited	60	0	2, 3	4.9	292	0	1

Panel A provides an overview of public works programs supported by the World Bank in Sub-Saharan Africa (in 2020). This provides a lower bound for investments in public works programs as it excludes top-up financing for these projects, domestic financing, and investments in cash-for-works or labor-intensive public works from other agencies and donors.

Table A2: Impacts *during* and *post* program, robustness to alternative definition of earnings

	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%	Levels, winsorized at 99%	Ln, winsorized at 97%	Ln, winsorized at 99%	Levels, winsorized at 99%	Ln, winsorized at 97%	Ln, winsorized at 99%	Levels, winsorized at 99%
Panel A: Impacts *during* the program (around 4,5 months after program starts)									
Public Works Treatment (ITT)	2.95***	2.95***	24380.2***	-1.04***	-1.06***	-11303.4***	5.93***	5.93***	37181.6***
	(0.19)	(0.20)	(5998.7)	(0.20)	(0.20)	(4091.6)	(0.25)	(0.25)	(4033.5)
Strata f.e.	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	0.010	0.000	0.000	0.000
Panel B: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)									
Public Works Treatment (ITT)	-0.037	-0.028	7597.5***	0.22	0.23	7802.7***	-0.20	-0.20	-542.8
	(0.18)	(0.18)	(2380.5)	(0.23)	(0.23)	(2250.6)	(0.19)	(0.19)	(1145.3)
Strata f.e.	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	0.849	0.877	0.002	0.347	0.322	0.002	0.300	0.323	0.607
Panel C: *Post* program impacts (by treatment arm) (12 to 15 months after program ends)									
Public Works Treatment (PW)	-0.18	-0.18	5558.1**	0.13	0.14	5098.3**	-0.22	-0.23	-21.1
	(0.22)	(0.22)	(2665.5)	(0.27)	(0.28)	(2288.8)	(0.24)	(0.24)	(1486.7)
Self-Empl. training (SET)	0.22	0.23	5171.9	0.28	0.29	7999.6**	-0.065	-0.064	-1327.7
	(0.26)	(0.26)	(3709.3)	(0.34)	(0.35)	(3763.3)	(0.26)	(0.26)	(1531.2)
Wage-Empl. training (WET)	0.24	0.24	1188.7	-0.0027	-0.0025	439.3	0.15	0.15	-299.5
	(0.21)	(0.21)	(3179.3)	(0.32)	(0.32)	(3699.5)	(0.27)	(0.28)	(1544.4)
Strata f.e.	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	0.898	0.832	0.005	0.185	0.169	0.000	0.249	0.250	0.349
Perm. p-value PW+WET=0	0.822	0.778	0.054	0.727	0.698	0.176	0.771	0.787	0.818
Perm. p-value SET=WET	0.955	0.961	0.405	0.486	0.474	0.169	0.459	0.438	0.502

ITT estimates in panels A and B based on specification in equation 1. Estimates in panel C based on specification in equation 2. Robust standard errors clustered at (broad) brigade level in parentheses. For variables (y) in logarithms, we take $ln(y+1)$. Permutation tests use 1000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A3: Estimated impacts *during* and *post* program on economic outcomes, with baseline controls

	(1) Employed	(2) Wage employed (in at least 1 activity)	(3) Self-employed (in at least 1 activity)	(4) Total hours worked (weekly)	(5) Hours worked in wage empl. (weekly)	(6) Hours worked in self-Empl. (weekly)	(7) Total earnings in CFA (monthly)	(8) Ln total earnings (monthly)	(9) Earnings in wage empl. in CFA (monthly)	(10) Earnings in self-empl. in CFA (monthly)	(11) Total expenditures in CFA (monthly)	(12) Savings in CFA (stock)	(13) Well-being index (z-score)
Panel A: Impacts *during* the program (around 4.5 months after program starts)													
Public Works Treatment (ITT)	0.14***	0.48***	-0.09***	5.34***	15.89***	-5.60***	27485.77***	2.92***	36799.02***	-5567.13***	14417.38***	40162.30***	0.19***
	(0.01)	(0.02)	(0.02)	(1.23)	(1.24)	(0.93)	(2608.25)	(0.19)	(1463.69)	(1196.97)	(1314.87)	(2305.56)	(0.05)
Strata f.e.	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	39.18	20.79	11.28	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: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)													
Public Works Treatment (ITT)	0.01	0.00	0.02	0.52	-1.42	2.02*	5155.88***	-0.01	-665.49	4783.59***	2387.97	10143.10***	0.12***
	(0.02)	(0.02)	(0.02)	(1.33)	(1.22)	(1.14)	(1902.00)	(0.19)	(1079.42)	(1852.73)	(1466.46)	(3316.16)	(0.04)
Strata f.e.	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	40.49	22.06	13.26	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: *Post* program impacts (by treatment arms) (12 to 15 months after program ends)													
Public Works Treatment (PW)	0.01	0.00	0.02	-1.61	-1.55	0.20	3406.18	-0.15	69.13	3071.78	2131.18	8598.92**	0.15***
	(0.02)	(0.03)	(0.03)	(1.72)	(1.60)	(1.31)	(2116.14)	(0.22)	(1319.73)	(1901.47)	(1596.09)	(3644.89)	(0.05)
Self-Empl. training (SET)	0.01	-0.02	0.02	3.36*	0.49	2.62	4226.69	0.20	-1668.02	5203.42*	-207.36	8143.34*	-0.01
	(0.02)	(0.03)	(0.03)	(1.95)	(1.77)	(1.71)	(3044.51)	(0.24)	(1287.18)	(2826.96)	(1422.57)	(4479.69)	(0.05)
Wage-Empl. training (WET)	0.00	0.01	0.00	3.05	-0.16	2.93	947.80	0.24	-796.06	210.73	881.90	-3206.44	-0.08*
	(0.02)	(0.03)	(0.03)	(2.17)	(1.61)	(1.82)	(1963.81)	(0.19)	(1313.08)	(2173.59)	(1599.02)	(4242.97)	(0.05)
Strata f.e.	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	40.49	22.06	13.26	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 1. Estimates in panel C based on specification in equation 2. The set of baseline controls differs for each outcome. Variables are selected from a pool of 1257 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 A4: Estimated impacts *during* and *post* program on economic outcomes, without clustering standard errors

	(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 expenditures in CFA (monthly)	Savings in CFA (stock)
Panel A: Impacts *during* the program (around 4,5 months after program starts)												
Public Works Treatment (ITT)	0.14***	0.48***	-0.096***	4.89***	15.5***	-5.69***	27082.9***	2.95***	36799.0***	-5715.4***	14529.3***	39785.7***
	(0.013)	(0.017)	(0.019)	(1.16)	(1.00)	(0.80)	(2575.7)	(0.16)	(1302.7)	(1100.6)	(1490.0)	(2017.9)
Strata f.e.	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: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)												
Public Works Treatment (ITT)	0.015	0.0068	0.010	1.34	-0.61	1.70*	4360.6**	-0.037	-452.7	4005.2**	1361.7	11505.2***
	(0.015)	(0.019)	(0.019)	(1.20)	(1.07)	(1.01)	(2038.5)	(0.18)	(1007.9)	(1763.0)	(1394.0)	(3022.1)
Strata f.e.	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	0.310	0.755	0.646	0.293	0.610	0.139	0.027	0.844	0.652	0.028	0.349	0.001
Panel C: *Post* program impacts (by treatment arm) (12 to 15 months after program ends)												
Public Works Treatment (PW)	0.011	0.0081	0.0035	-0.76	-0.71	-0.12	2800.5	-0.18	312.2	2168.7	925.7	10429.5***
	(0.018)	(0.023)	(0.023)	(1.41)	(1.30)	(1.17)	(2308.5)	(0.22)	(1289.6)	(1964.9)	(1701.6)	(3639.5)
Self-Empl. training (SET)	0.011	-0.018	0.021	3.42**	0.46	2.77*	4229.3	0.22	-1591.8	5595.5*	278.1	7169.5*
	(0.017)	(0.023)	(0.025)	(1.55)	(1.31)	(1.44)	(3664.3)	(0.24)	(1266.1)	(3051.3)	(1774.0)	(4079.4)
Wage-Empl. training (WET)	0.0018	0.014	0.000048	3.12**	-0.14	2.89**	637.5	0.24	-792.6	135.8	1077.6	-3798.3
	(0.018)	(0.024)	(0.024)	(1.49)	(1.33)	(1.30)	(2641.8)	(0.22)	(1302.2)	(2223.9)	(1775.7)	(3736.3)
Strata f.e.	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	0.231	0.666	0.322	0.092	0.849	0.066	0.048	0.868	0.285	0.010	0.492	0.000
p-value PW+WET=0	0.485	0.344	0.883	0.121	0.522	0.034	0.168	0.813	0.698	0.286	0.254	0.074
p-value SET=WET	0.617	0.175	0.408	0.854	0.651	0.940	0.342	0.956	0.510	0.086	0.661	0.008
Observations	3934	3934	3934	3934	3934	3934	3934	3934	3934	3934	3814	3934
Perm. p-value PW+SET=0	0.273	0.665	0.391	0.159	0.859	0.168	0.015	0.865	0.303	0.004	0.457	0.001
Perm. p-value PW+WET=0	0.474	0.375	0.914	0.211	0.568	0.126	0.246	0.815	0.710	0.427	0.297	0.170
Perm. p-value SET=WET	0.675	0.224	0.511	0.914	0.732	0.965	0.366	0.959	0.562	0.180	0.634	0.050

ITT estimates in panels A and B based on specification in equation 1. Estimates in panel C based on specification in equation 2. Robust standard errors are in parentheses. Standard errors are not clustered by brigade. *Hours*, *earnings*, *expenditures*, and *savings* winsorized at the 97th percentile. For variables (y) in logarithms, we take $ln(y+1)$. Permutation tests use 1000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A5: Impacts *during* and *post* program, with full baseline sample at follow-up (ITT)

	(1) Employed	(2) Wage employed (in at least 1 activity)	(3) Self employed (in at least 1 activity)	(4) Total hours worked (weekly)	(5) Hours worked in wage empl. (weekly)	(6) Hours worked in self-empl. (weekly)	(7) Total earnings in CFA (monthly)	(8) Ln total earnings (monthly)	(9) Earnings in wage empl. in CFA (monthly)	(10) Earnings in self-empl. in CFA (monthly)	(11) Total expenditures in CFA (monthly)	(12) Savings in CFA (stock)
Panel A: Impacts *during* the program (around 4,5 months after program starts)												
Public Works Treatment (ITT)	0.14***	0.48***	-0.096***	4.89***	15.5***	-5.69***	27082.9***	2.95***	36799.0***	-5715.4***	14529.3***	39785.7***
	(0.015)	(0.024)	(0.020)	(1.25)	(1.29)	(0.94)	(2824.9)	(0.19)	(1472.5)	(1214.6)	(1441.4)	(2389.2)
Strata f.e.	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: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)												
Public Works Treatment (ITT)	0.0079	0.0069	0.020	0.73	-0.59	1.59	5030.5**	-0.072	0.93	4504.1**	1950.4	8953.4***
	(0.014)	(0.019)	(0.024)	(1.28)	(1.15)	(1.06)	(1996.1)	(0.19)	(978.4)	(1907.9)	(1452.2)	(3390.6)
Strata f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.86	0.51	0.35	41.02	21.97	13.34	42690.67	8.44	20132.39	18357.50	50010.79	48816.10
Observations	3910	3910	3910	3910	3910	3910	3910	3910	3910	3910	3788	3910
Perm. p-value: No effects	0.590	0.726	0.392	0.578	0.610	0.131	0.013	0.726	0.999	0.017	0.184	0.016
Panel C: *Post* program impacts (by treatment arm) (12 to 15 months after program ends)												
Public Works Treatment (PW)	0.0037	0.0085	0.013	-1.39	-0.69	-0.24	3461.2	-0.22	770.0	2655.2	1498.0	7875.4**
	(0.019)	(0.025)	(0.028)	(1.68)	(1.53)	(1.19)	(2236.8)	(0.23)	(1229.9)	(1955.1)	(1627.1)	(3354.3)
Self-Empl. training (SET)	0.011	-0.018	0.021	3.43*	0.46	2.78*	4220.4	0.22	-1604.8	5597.6**	272.2	7103.4
	(0.017)	(0.026)	(0.031)	(1.83)	(1.77)	(1.59)	(3135.6)	(0.25)	(1239.2)	(2782.6)	(1483.8)	(4515.6)
Wage-Empl. training (WET)	0.0022	0.013	0.00081	3.17	-0.14	2.92	674.6	0.24	-792.5	172.0	1134.1	-3725.1
	(0.020)	(0.026)	(0.031)	(2.18)	(1.61)	(1.85)	(2169.2)	(0.21)	(1247.4)	(2372.0)	(1779.3)	(4493.3)
Strata f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.86	0.51	0.35	41.02	21.97	13.34	42690.67	8.44	20132.39	18357.50	50010.79	48816.10
p-value PW+SET=0	0.388	0.677	0.218	0.245	0.884	0.121	0.009	0.998	0.474	0.002	0.280	0.001
p-value PW+WET=0	0.760	0.348	0.675	0.357	0.543	0.148	0.150	0.930	0.985	0.369	0.170	0.403
p-value SET=WET	0.682	0.208	0.530	0.914	0.699	0.955	0.359	0.935	0.518	0.174	0.586	0.034
Observations	3910	3910	3910	3910	3910	3910	3910	3910	3910	3910	3788	3910
Perm. p-value PW+SET=0	0.409	0.687	0.218	0.261	0.885	0.141	0.008	0.998	0.456	0.004	0.301	0.001
Perm. p-value PW+WET=0	0.761	0.351	0.686	0.345	0.571	0.148	0.151	0.935	0.990	0.357	0.185	0.386
Perm. p-value SET=WET	0.696	0.240	0.555	0.915	0.712	0.952	0.379	0.945	0.490	0.196	0.613	0.046

ITT estimates in panels A and B based on specification in equation 1. Individuals from the control group that were able to apply and participate in the third or fourth wave of the program are kept in the sample, and control individuals who applied to those waves but were not selected are not given larger weights. Robust standard errors clustered at 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 1000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A6: Impacts *post* program, LATE accounting for non-compliance in control between midline and endline

	(1) Employed	(2) Wage employed (in at least 1 activity)	(3) Self-employed (in at least 1 activity)	(4) Total hours worked (weekly)	(5) Hours worked in wage empl. (weekly)	(6) Hours worked in self-empl. (weekly)	(7) Total earnings in CFA (monthly)	(8) Ln total earnings (monthly)	(9) Earnings in wage empl. in CFA (monthly)	(10) Earnings in self-empl. in CFA (monthly)	(11) Total expenditures in CFA (monthly)	(12) Savings in CFA (stock)
Panel A: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)												
Public Works Treatment (LATE)	0.0089	0.0078	0.023	0.83	-0.66	1.80	5684.9**	-0.082	1.05	5090.0**	2202.8	10118.0***
	(0.016)	(0.022)	(0.027)	(1.45)	(1.29)	(1.19)	(2252.2)	(0.21)	(1100.7)	(2153.1)	(1634.7)	(3820.0)
Strata f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.86	0.51	0.35	41.02	21.97	13.34	42690.67	8.44	20132.39	18357.50	50010.79	48816.10
Observations	3910	3910	3910	3910	3910	3910	3910	3910	3910	3910	3788	3910
Perm. p-value: No effects	0.128	0.780	0.543	0.383	0.273	0.064	0.186	0.629	0.333	0.090	0.496	0.000

Estimates of Local Average Treatment Effects (LATE) accounting for one-sided non-compliance, whereby some control individuals were able to participate in the program between midline and endline. All treatment individuals are considered as participating in the program. Participation in the program is instrumented by the randomized treatment assignment. Robust standard errors clustered at 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 1000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A7: Estimated impacts *during* program on economic outcomes for household members other than the beneficiary

	(1) Employed	(2) Wage-employed (in at least 1 activity)	(3) Self-employed (in at least 1 activity)	(4) Total hours worked (weekly)	(5) Hours worked in wage empl. (weekly)	(6) Hours worked in self-empl. (weekly)	(7) Total earning in CFA (monthly)	(8) Earnings in wage empl. in CFA (monthly)	(9) Earnings in self-empl. in CFA (monthly)
Panel A: Impacts *during* the program (around 4,5 months after program starts)									
Public Works Treatment (ITT)	-0.0079	-0.010	0.0091	0.0059	-0.20	0.39	-28.1	-340.2	663.7
	(0.015)	(0.013)	(0.012)	(0.88)	(0.65)	(0.58)	(1483.6)	(628.4)	(757.9)
Strata f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.59	0.27	0.25	25.03	11.67	9.88	23673.33	8493.04	9578.04
Observations	9929	9895	9895	9929	9895	9895	9772	9739	9739
Perm. p-value: no effects	0.634	0.404	0.423	0.993	0.759	0.488	0.984	0.586	0.377

ITT estimates in panel A based on specification in equation 1. Based on information of households that were included at baseline, even if the beneficiary moved at midline. Collective households and attritors for the household survey at midline were excluded. Robust standard errors clustered at (broad) brigade level in parentheses. *Hours* and *earnings* winsorized at the 97th percentile. Permutation tests use 1000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A8: Estimated impacts *during* and *post* program on well-being index components

	(1) Well-being index (z-score)	(2) Self-esteem (Rosenberg scale) (z-score)	(3) Positive affect (CES-D sub scale) (z-score)	(4) Positive attitude towards the future (ZTPI sub scale) (z-score)	(5) Present fatalism (ZTPI sub scale) (z-score)	(6) Happiness in daily activities (z-score)	(7) Pride in daily activities (z-score)
Panel A: Impacts *during* the program (around 4.5 months after program starts)							
Public Works Treatment (ITT)	0.20***	0.14***	0.18***	0.086**	0.021	0.15***	0.15***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Strata f.e.	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	0.000	0.000	0.032	0.605	0.002	0.000
Panel B: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)							
Public Works Treatment (ITT)	0.11***	0.10**	0.041	0.061	-0.093**	0.076*	0.053
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Strata f.e.	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	0.005	0.024	0.332	0.196	0.030	0.078	0.182
Panel C: *Post* program impacts (by treatment arm) (12 to 15 months after program ends)							
Public Works Treatment (PW)	0.14***	0.12**	0.063	0.094*	-0.052	0.11**	0.089*
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Self-Empl. training (SET)	-0.0068	-0.044	-0.0078	-0.11*	-0.11**	-0.018	-0.025
	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Wage-Empl. training (WET)	-0.075	-0.019	-0.062	0.0034	-0.020	-0.082*	-0.087*
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Strata f.e.	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	0.012	0.167	0.285	0.819	0.000	0.066	0.184
Perm. p-value PW+WET=0	0.204	0.076	0.971	0.042	0.222	0.593	0.963
Perm. p-value SET=WET	0.191	0.695	0.227	0.044	0.092	0.145	0.235

ITT estimates in panels A and B based on specification in equation 1. Estimates in panel C based on specification in equation 2. The definition of the *well-being index* and variables entering the index is detailed in Appendix D. *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 A9: 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 *during* the program (around 4.5 months after program starts)					
Public Works Treatment (ITT)	0.12***	-0.031	0.023	-0.095**	-0.13***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Strata f.e.	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	0.008	0.444	0.573	0.040	0.004
Panel B: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)					
Public Works Treatment (ITT)	-0.012	0.013	-0.0032	0.0050	0.014
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Strata f.e.	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	0.775	0.769	0.925	0.904	0.728
Panel C: *Post* program impacts (by treatment arm) (12 to 15 months after program ends)					
Public Works Treatment (PW)	0.025	0.034	0.0066	-0.051	-0.013
	(0.05)	(0.05)	(0.06)	(0.04)	(0.05)
Self-Empl. training (SET)	-0.039	-0.062*	-0.054	0.054	0.012
	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)
Wage-Empl. training (WET)	-0.077	-0.0041	0.024	0.12***	0.072
	(0.05)	(0.05)	(0.07)	(0.04)	(0.05)
Strata f.e.	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	0.763	0.516	0.326	0.943	0.971
Perm. p-value PW+WET=0	0.288	0.564	0.556	0.157	0.228
Perm. p-value SET=WET	0.401	0.215	0.065	0.159	0.242

ITT estimates in panels A and B based on specification in equation 1. Estimates in panel C based on specification in equation 2. The definition of the *behavior index* and variables entering the index is detailed in Appendix D. *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 A10: Estimated impacts *during* and *post* program on risky behaviors

	(1) Stealing	(2) Assaulting someone	(3) Believing smuggling is necessary (to earn a living)	(4) Prostituting	(5) Threatening someone	(6) Taking illicit drugs	(7) Smuggling stolen objects	(8) Ties with a smuggling network	(9) Keeping fire arms at home
Panel A: Impacts *during* 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.	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: *Post* 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.	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: *Post* program impacts (by treatment arm) (12 to 15 months after program ends)									
Public Works Treatment (PW)	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.	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 D). Difference-in-means estimation (Miller, 1984) was used to estimate the mean in control and treatment effects. In Panels A and B, the specification $Y_i = \alpha + \gamma_1 L_i + \gamma_2 W_i + \gamma_3 (L_i \times W_i) + \delta_l X_{i,l} + \varepsilon_i$ 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 $Y_i = \alpha + \gamma_1 L_i + \gamma_2 W_i + \gamma_3 (W_i \times T1_i) + \gamma_4 (W_i \times T2_i) + \gamma_5 (L_i \times W_i) + \gamma_6 (L_i \times W_i \times T1_i) + \gamma_7 (L_i \times W_i \times T2_i) + \delta_l X_{i,l} + \varepsilon_i$; 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 job search 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 A11: Impacts *during* and *post* program on time use

	(1) Rest at 6 am of prev. day	(2) Leisure at 6 am of prev. day	(3) Work at 6 am of prev. day	(4) Rest at 10 pm of prev. day	(5) Leisure at 10 pm of prev. day	(6) Work at 10 pm of prev. day
Panel A: Impacts *during* the program (around 4,5 months after program starts)						
Public Works Treatment (ITT)	-0.14***	-0.044***	0.33***	0.071***	-0.054***	-0.028**
	(0.018)	(0.011)	(0.032)	(0.019)	(0.015)	(0.012)
LocXGender control	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.24	0.09	0.20	0.68	0.18	0.08
Observations	2955	2955	2955	2953	2953	2953
Perm. p-value: no effects	0.000	0.001	0.000	0.001	0.001	0.014
Panel B: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)						
Public Works Treatment (ITT)	0.025	-0.0029	0.0080	-0.000063	0.0011	0.0061
	(0.019)	(0.012)	(0.016)	(0.018)	(0.015)	(0.011)
LocXGender control	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.26	0.09	0.24	0.68	0.19	0.09
Observations	3933	3933	3933	3933	3933	3933
Perm. p-value: No effects	0.206	0.804	0.624	0.997	0.932	0.561
Panel C: *Post* program impacts (by treatment arms) (12 to 15 months after program ends)						
Public Works Treatment (ITT)	0.021	0.0036	0.021	-0.00080	0.0037	0.0041
	(0.022)	(0.014)	(0.021)	(0.023)	(0.019)	(0.014)
Self-empl.training (SET)	-0.0075	-0.0036	-0.016	0.0027	-0.0085	0.0030
	(0.032)	(0.013)	(0.023)	(0.023)	(0.019)	(0.012)
Wage-empl. training (WET)	0.022	-0.017	-0.025	-0.00038	0.00041	0.0033
	(0.026)	(0.014)	(0.022)	(0.022)	(0.020)	(0.015)
LocXGender control	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.26	0.09	0.24	0.68	0.19	0.09
p-value PW+SET=0	0.648	0.995	0.798	0.929	0.790	0.579
p-value PW+WET=0	0.076	0.383	0.834	0.957	0.839	0.583
p-value SET=WET	0.230	0.385	0.652	0.895	0.680	0.981
Observations	3933	3933	3933	3933	3933	3933
Perm. p-value PW+SET=0	0.649	0.999	0.795	0.938	0.793	0.616
Perm. p-value PW+WET=0	0.079	0.375	0.838	0.951	0.840	0.585
Perm. p-value SET=WET	0.236	0.401	0.644	0.894	0.692	0.985

ITT estimates in panels A and B based on specification in equation 1. Estimates in panel C based on specification in equation 2. Robust standard errors clustered at (broad) brigade level in parentheses. Permutation tests use 1000 permutations for each hypothesis. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A12: Estimated impacts *post* program on aspirations and reservation wage

	(1) Aspires to wage job in 20 years	(2) Aspires to self-employment in 20 years	(3) Expected earnings in wage empl. in CFA (monthly)	(4) Expected earnings in self-empl. in CFA (monthly)
Panel A: *Post* program impacts (pooled treatment) (12 to 15 months after program ends)				
Public Works Treatment (ITT)	0.0013	-0.0017	-3090.0	-5702.2
	(0.01)	(0.01)	(2159.72)	(3529.50)
Strata f.e.	Yes	Yes	Yes	Yes
Mean in Control	0.17	0.83	107539.41	118806.05
Observations	3934	3934	3934	3934
Perm. p-value: no effects	0.923	0.913	0.151	0.115
Panel B: *Post* program impacts (by treatment arm) (12 to 15 months after program ends)				
Public Works Treatment (PW)	0.011	-0.012	-1398.7	-1975.7
	(0.02)	(0.02)	(2540.13)	(4293.28)
Self-Empl. training (SET)	-0.036**	0.035**	-2422.2	-5451.8
	(0.02)	(0.02)	(2903.83)	(4423.61)
Wage-Empl. training (WET)	0.0044	-0.0039	-2846.1	-6156.1
	(0.01)	(0.01)	(2271.34)	(4346.66)
Strata f.e.	Yes	Yes	Yes	Yes
Mean in Control	0.17	0.83	107539.41	118806.05
p-value PW+SET=0	0.190	0.209	0.177	0.107
p-value PW+WET=0	0.339	0.346	0.104	0.047
p-value SET=WET	0.022	0.028	0.882	0.871
Observations	3934	3934	3934	3934
Perm. p-value PW+SET=0	0.203	0.224	0.193	0.112
Perm. p-value PW+WET=0	0.343	0.375	0.130	0.061
Perm. p-value SET=WET	0.035	0.033	0.886	0.880

ITT estimates in panels A and B based on specification in equation 1. Estimates in panel C based on specification in equation 2. 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 A13: 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 *during* program										
ATE (β_1)	2.631	2.642	2.643	2.635	2.634	24577.6	24351.6	24238.5	24549.6	24363.0
	(2.9,2.3)	(2.9,2.3)	(2.9,2.3)	(2.9,2.3)	(2.9,2.3)	(31323.2,17831.5)	(31214.6,17669.0)	(30976.0,17471.1)	(31423.5,17789.9)	(31150.3,17618.5)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
HET (β_2)	0.980	1.231	0.420	0.920	0.849	0.405	1.070	0.288	0.429	0.390
	(1.3,0.6)	(1.6,0.9)	(0.6,0.2)	(1.2,0.6)	(1.1,0.6)	(0.9,-0.07)	(2.1,-0.05)	(0.7,-0.03)	(1.3,-0.2)	(0.8,0.009)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.157]	[0.121]	[0.160]	[0.353]	[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 *post* program										
ATE (β_1)	-0.0551	-0.0508	-0.0484	-0.0570	-0.0297	3217.5	3527.3	3546.3	3314.0	3474.5
	(0.4,-0.5)	(0.4,-0.5)	(0.4,-0.5)	(0.4,-0.5)	(0.4,-0.5)	(8428.4,-1918.3)	(8737.3,-1706.6)	(8685.8,-1630.9)	(8523.0,-1842.7)	(8642.7,-1641.7)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.439]	[0.374]	[0.358]	[0.414]	[0.370]
HET (β_2)	-0.0581	0.156	0.0376	-0.00219	0.111	0.0837	0.391	0.0220	0.416	0.0800
	(0.6,-0.7)	(1.6,-1.3)	(0.3,-0.2)	(3.1,-2.8)	(0.7,-0.5)	(0.6,-0.5)	(1.8,-1.0)	(0.4,-0.3)	(5.4,-1.1)	(0.5,-0.4)
	[1.000]	[1.000]	[1.000]	[1.000]	[0.970]	[1.000]	[1.000]	[1.000]	[0.756]	[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. (forthcoming) (see discussion in Section 5.1.2 and Appendix E). Estimates are based on equation 3. 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% are provided in parentheses and adjusted p-values for partition uncertainty are provided in brackets. Earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms, we take $ln(y+1)$.

Table A14: 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)
	Midline	*Endline*	*Midline*	*Endline*
Panel A: Predicted average treatment effect and heterogeneity loading parameter				
ATE (β_1)	2.311	-0.136	27252.0	4011.4
	(1.918,2.714)	(-0.723,0.449)	(17338.2,37013.8)	(-4014.3,11850.4)
	[0.000]	[0.928]	[0.000]	[0.631]
HET (β_2)	2.336	0.272	2.297	0.136
	(1.156,3.577)	(-0.306,0.814)	(0.387,4.180)	(-1.855,2.883)
	[0.000]	[0.637]	[0.032]	[0.888]
Best ML method	Generalized Random forest	Random forest	Generalized Random forest	R-Learner
Panel B: By quartile of predicted impacts on earnings *during* program (using *midline*) (GATES)				
Mean in quartile 1 (0 to 25%)	1.430	-0.695	9616.7	1309.1
	(0.596,2.233)	(-2.047,0.682)	(-9485.5,28647.1)	(-17605.5,20036.9)
Mean in quartile 2 (25 to 50%)	2.063	-0.589	31539.2	3483.3
	(1.260,2.854)	(-1.953,0.810)	(12393.6,51035.5)	(-15126.9,22326.7)
Mean in quartile 3 (50 to 75%)	2.602	-0.126	32238.6	6594.5
	(1.763,3.397)	(-1.515,1.306)	(12793.6,51515.6)	(-11856.0,24930.0)
Mean in quartile 4 (75 to 100%)	3.312	-0.0495	33266.5	8086.5
	(2.521,4.094)	(-1.417,1.326)	(13826.0,52380.5)	(-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 *post* program (*using endline*) (GATES)				
Mean in quartile 1 (0 to 25%)	1.935	-0.338	21497.8	2309.3
	(1.102,2.761)	(-1.512,0.819)	(3457.5,39527.0)	(-12952.7,18286.3)
Mean in quartile 2 (25 to 50%)	1.989	-0.359	27897.3	5194.6
	(1.164,2.868)	(-1.520,0.828)	(10231.8,45734.2)	(-10461.8,20933.5)
Mean in quartile 3 (50 to 75%)	2.246	-0.432	28716.5	3064.9
	(1.456,3.038)	(-1.562,0.736)	(10596.6,46321.6)	(-12565.0,18074.0)
Mean in quartile 4 (75 to 100%)	2.555	0.368	29256.8	5016.5
	(1.768,3.370)	(-0.800,1.488)	(11482.7,46980.5)	(-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. (forthcoming) (see discussion in Section 5.1.2 and Appendix E). All baseline variables in the balance 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 3. 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 4. 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 $\widehat{s_0}(Z)$, 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% are provided in parentheses and adjusted p-values for partition uncertainty are provided in brackets. Earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms, we take $ln(y+1)$.

Table A15: Estimated impacts on (ln) earnings *post* program, by treatment arm

	(1) Public Works only (PW)	(2) PW and Self-Employment Training (SET)	(3) PW and Wage-Employment Training (WET)
	Endline	*Endline*	*Endline*
Panel A: Predicted average treatment effect and heterogeneity loading parameter			
ATE (β_1)	-0.0753	-0.0116	0.120
	(-0.601,0.447)	(-0.545,0.529)	(-0.406,0.648)
	[1.000]	[1.000]	[0.930]
HET (β_2)	0.0866	-0.572	0.450
	(-0.211,0.382)	(-1.431,0.299)	(-0.595,1.935)
	[0.818]	[0.425]	[0.722]
Best ML method	Gradient boosting	Elastic net	R-Learner
Panel B: By quartile of predicted impacts on earnings *post* program (*using endline*) (GATES)			
Mean in quartile 1 (0 to 25%)	-0.220	0.408	-0.0540
	(-1.270,0.826)	(-0.663,1.487)	(-1.112,1.004)
Mean in quartile 2 (25 to 50%)	-0.141	0.132	0.138
	(-1.188,0.923)	(-0.958,1.206)	(-0.913,1.187)
Mean in quartile 3 (50 to 75%)	-0.172	-0.116	0.238
	(-1.217,0.873)	(-1.193,0.962)	(-0.805,1.279)
Mean in quartile 4 (75 to 100%)	0.148	-0.522	0.168
	(-0.909,1.214)	(-1.587,0.543)	(-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. (forthcoming) (see discussion in Section 5.1.2 and Appendix E). Columns (1-3) show estimated impacts for each treatment arm compared to the control group. Estimates in Panel A are based on equation 3. 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 4. 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 $\widehat{s_0}(Z)$, 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% are provided in parentheses and adjusted p-values for partition uncertainty are provided in brackets. Earnings variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms, we take $ln(y+1)$.

Table A16: Estimated impacts on earnings (in levels) *post* program, by treatment arm

	(1) Public Works only (PW)	(2) PW and Self-Employment Training (SET)	(3) PW and Wage-Employment Training (WET)
	Endline	*Endline*	*Endline*
Panel A: Predicted average treatment effect and heterogeneity loading parameter			
ATE (β_1)	2155.1	4166.5	3970.3
	(-3648.0,8097.0)	(-2085.9,10463.6)	(-2233.2,10007.5)
	[0.799]	[0.397]	[0.412]
HET (β_2)	0.711	0.167	0.859
	(-0.793,2.217)	(-2.294,3.175)	(-0.383,2.666)
	[0.645]	[0.822]	[0.359]
Best ML method	Generalized Random forest	R-Learner	R-Learner
Panel B: By quartile of predicted impacts on earnings *post* program (*using endline*) (GATES)			
Mean in quartile 1 (0 to 25%)	-804.5	4252.3	954.7
	(-12141.7,10585.9)	(-8217.4,16697.1)	(-11358.2,13202.6)
Mean in quartile 2 (25 to 50%)	314.7	5075.0	3213.2
	(-11327.4,12027.1)	(-7721.7,17162.5)	(-8929.3,15509.9)
Mean in quartile 3 (50 to 75%)	2580.6	4686.6	3843.2
	(-9034.2,14329.5)	(-8013.1,17365.6)	(-8184.3,15983.8)
Mean in quartile 4 (75 to 100%)	6306.2	3651.1	5448.1
	(-5642.9,17819.6)	(-8919.7,15948.0)	(-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. (forthcoming) (see discussion in Section 5.1.2 and Appendix E). Columns (1-3) show estimated impacts for each treatment arm compared to the control group. Estimates in Panel A are based on equation 3. 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 4. 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 $\widehat{s}_0(Z)$, 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% are provided in parentheses and adjusted p-values for partition uncertainty are provided in brackets. Earnings variable is in CFA, winsorized at the 97th percentile.

Table A17: 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
Lives in urban area	0.91	0.97	0
Age	24.91	24.41	0.012
Number 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 number of assets	0.59	0.51	0
Number of transportation assets	1.22	0.42	0
Number of agricultural assets	7.74	2.42	0
Number of household durables	2.31	1.26	0
Number 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
Number 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	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 (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, No. of 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. (forthcoming) (see discussion in Section 5.1.2 and Appendix E). 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 $\widehat{s_0}(Z)$, noted $S(Z)$, and $s_0(Z)$ (see appendix Table A13 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 are provided. *Household assets* and *savings stock* variables winsorized at the 99th percentile.

Table A18: Estimated impacts *during* and *post* program on savings

	(1) Ln Savings (Monthly)	(2) Ln Savings (Monthly)	(3) Savings in CFA (Monthly)	(4) Savings in CFA (Monthly)
	Midline	*Endline*	*Midline*	*Endline*
Panel A: Predicted average treatment effect and heterogeneity loading parameter				
ATE (β_1)	3.543	0.472	36768.4	9492.9
	(2.981,4.100)	(-0.0260,0.977)	(30901.6,42641.1)	(1192.6,17761.4)
	[0.000]	[0.127]	[0.000]	[0.050]
HET (β_2)	1.303	0.247	0.440	0.193
	(0.454,2.169)	(-1.230,2.565)	(-0.0100,0.889)	(-0.101,0.505)
	[0.005]	[0.809]	[0.110]	[0.388]
Best ML method	Generalized Random forest	R-Learner	Random forest	Gradient boosting
Panel B: By quartile of predicted impacts on savings *during* program (GATES)				
Mean in quartile 1 (0 to 25%)	2.481	0.748	29852.0	10275.5
	(1.355,3.611)	(-0.440,1.920)	(18249.3,41593.9)	(-9226.9,29636.8)
Mean in quartile 2 (25 to 50%)	3.174	0.302	33873.9	7074.7
	(2.052,4.286)	(-0.870,1.471)	(21920.6,45466.9)	(-12459.8,27015.8)
Mean in quartile 3 (50 to 75%)	3.785	0.563	38335.9	10085.7
	(2.671,4.889)	(-0.619,1.746)	(26568.8,49862.4)	(-9870.0,29882.8)
Mean in quartile 4 (75 to 100%)	4.778	0.316	45983.1	9897.8
	(3.659,5.896)	(-0.885,1.497)	(34465.3,57655.4)	(-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 *post* program (GATES)				
Mean in quartile 1 (0 to 25%)	3.615	0.316	36489.5	3831.7
	(2.557,4.647)	(-0.703,1.346)	(26070.4,47031.0)	(-12411.8,19889.7)
Mean in quartile 2 (25 to 50%)	3.551	0.336	37213.5	7044.9
	(2.488,4.582)	(-0.705,1.367)	(26671.1,47659.7)	(-9401.7,23694.9)
Mean in quartile 3 (50 to 75%)	3.574	0.483	37366.7	9436.6
	(2.503,4.652)	(-0.527,1.506)	(26941.2,47912.0)	(-7461.4,25802.8)
Mean in quartile 4 (75 to 100%)	3.585	0.731	39460.8	16964.8
	(2.532,4.646)	(-0.286,1.740)	(28965.9,50086.8)	(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. (forthcoming) (see discussion in Section 5.1.2 and Appendix E). Columns (1) and (3) (respectively columns (2) and (4)) focus on outcomes at midline (respectively endline). Estimates in Panel A are based on equation 3. 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 4. 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 $\widehat{s_0}(Z)$, 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% are provided in parentheses and adjusted p-values for partition uncertainty are provided in brackets for Panel A. *Savings* variable is in CFA, winsorized at the 97th percentile. For variables (y) in logarithms, we take $ln(y+1)$.

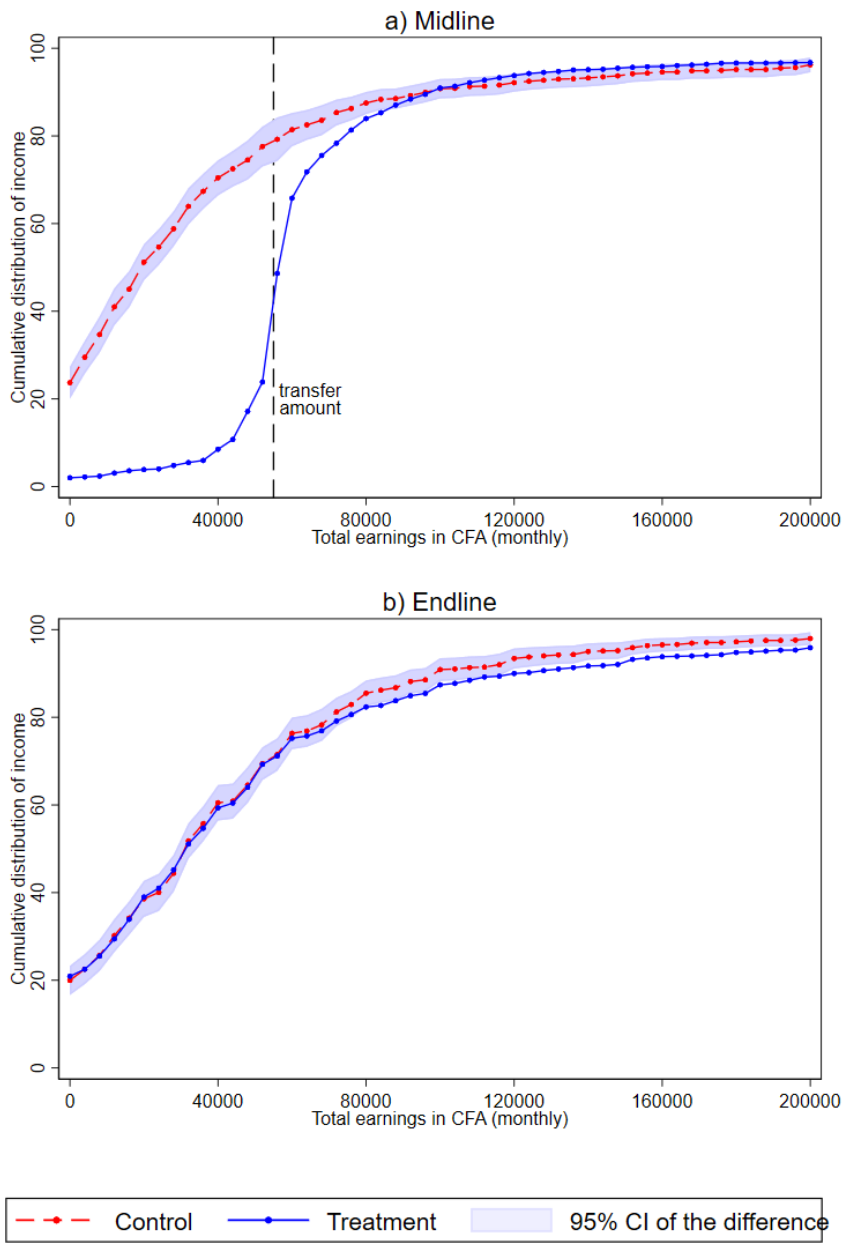
Table A19: Baseline characteristics of the bottom and top quartiles of predicted impacts on (ln) savings *during* program

	(1) Mean in 1st quartile	(2) Mean in 4th quartile	(3) Test (1)-(2) (p-value)
Individual characteristics			
Female	0.14	0.46	0
Lives in urban area	0.95	0.95	0.598
Age	24.65	24.29	0.080
Number 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 number of assets	0.65	0.44	0
Number of transportation assets	0.68	0.70	0.331
Number of agricultural assets	3.88	4.44	0.402
Number of household durables	1.75	1.70	0.125
Number 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
Number 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	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 (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, No. of 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. (forthcoming) (see discussion in Section 5.1.2 and Appendix E.3). 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 $\widehat{s_0}(Z)$, 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 are provided. *Household assets* and *savings stock* variables winsorized at the 99th percentile.

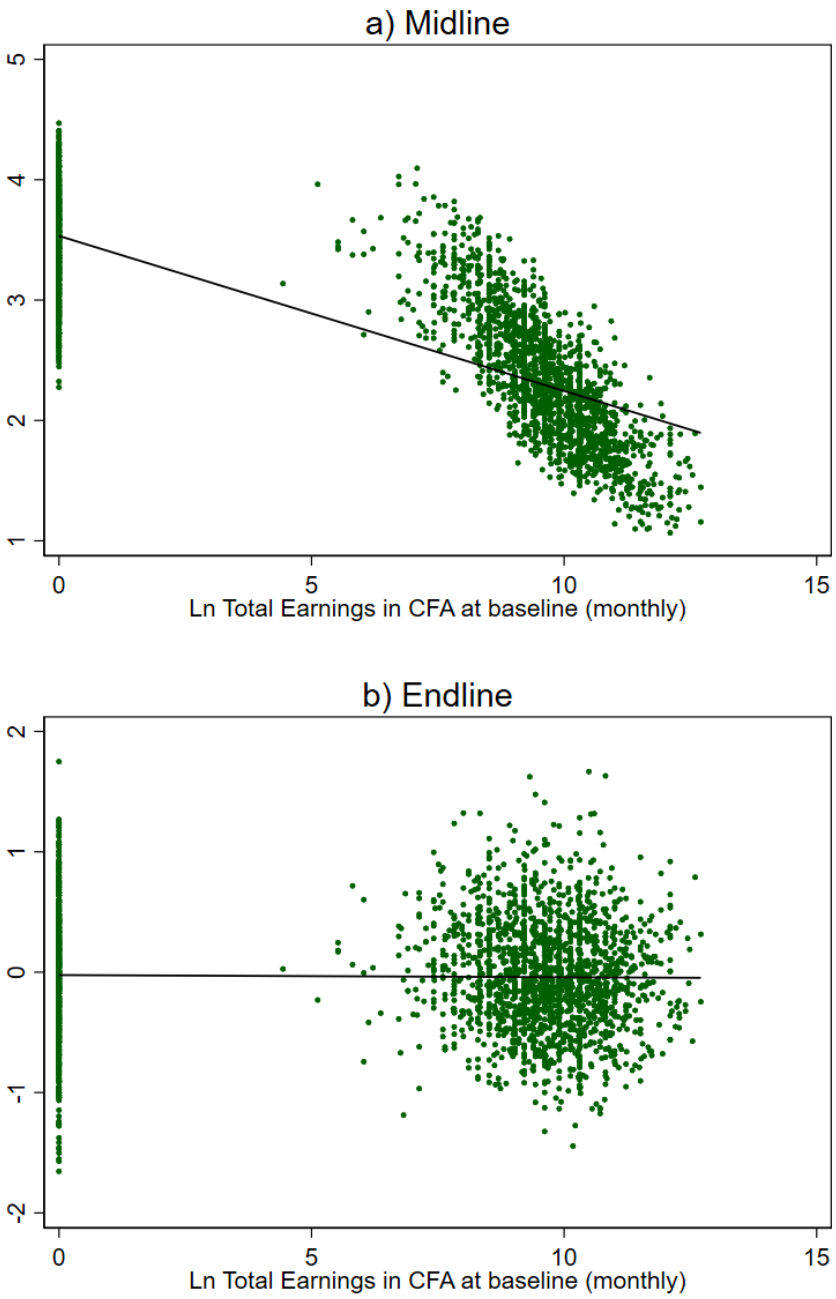
Appendix B Additional Figures

Figure B1: Cumulative distribution of earnings *during* and *post* program



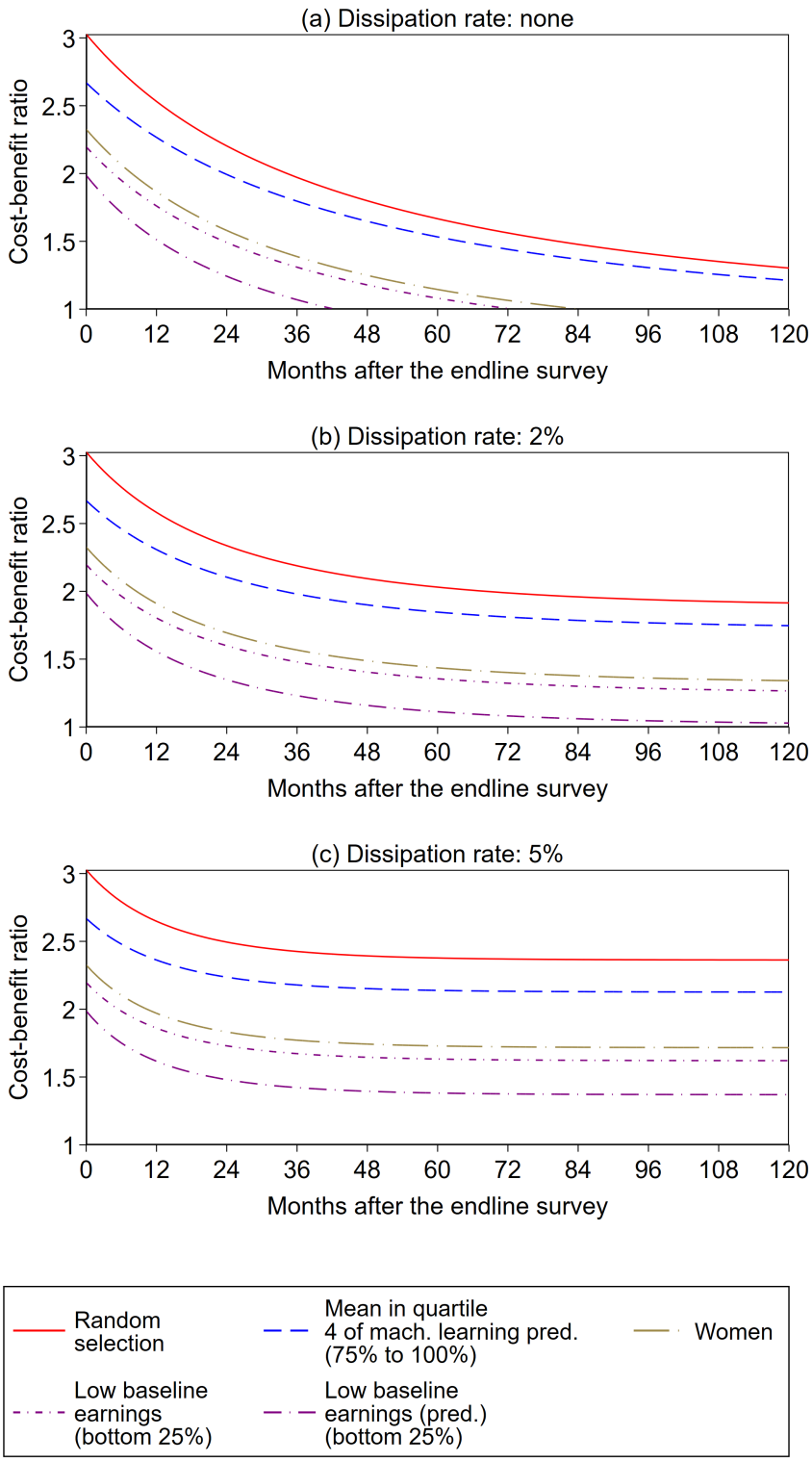
Note: Total monthly earnings variable winsorized at the 97th percentile. The figures show the results of the estimation of the cumulative distribution of potential outcomes in the treatment and control groups. They are based on the estimation of equation 1, with variables defined as $1(y < t)$ for t varying over the support of y. The red curve shows the average of those variables in the control group. The blue curve adds to this average the estimated treatment coefficient. The grey area around the red curve represents a band of +/- 1.96 times the standard error of the estimated treatment effect.

Figure B2: Reported baseline earnings vs predicted impacts *during* and *post* program



Note: Panel (a) compares baseline labor income to best predictions at midline, using Generalized Random Forest. Panel (b) compares baseline labor income to best predictions at endline using Random Forest. The correlation coefficient is -0.805 at midline and -0.018 at endline. In both panels, linear regression lines are added (corresponding to a R^2 of 0.648 in Panel (a)). For variables (y) in logarithms, we take $ln(y+1)$

Figure B3: Cost-benefit 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 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.[66] 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 and visuals.

The basic entrepreneurship training aimed to build skills to help youth set-up and manage a small 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) started with basic considerations on how to choose and set up and activity. It then covered business skills on topics such as simple accounting practices, managing stocks or setting up prices. 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 a basic business plan, and reviewed related topics from the curriculum with an operational focus. The final phase (20 hours) included post-training follow-up. The curriculum and training material were specifically developed for youths with limited education, relying on practical exercises and visuals.

The training on wage job search skills provided information on wage jobs opportunities, skills on jobs search techniques. It also offered 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 prepare a job interview. It included a self-evaluation of strength and weaknesses, and guidance on identifying the types of wage jobs suitable to participants' profiles. The second phase included field exercises to collect information on potential opportunities, identify

[66] Some youth were offered the second half of the training shortly 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 visit potential employers or professional networks, etc. During this phase, beneficiaries started implementing a "job search action plan" they had previously developed with their trainer. The third phase (40 hours over two weeks) provided individual and group feedback on field exercises, reviewed part of the curriculum (search, job applications and interviews) and provided additional practical guidance on how to search for a job.

In addition, and prior to the training on job search, supervisors of the brigades who were offered the wage job search training were also trained (over one week) 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. Supervisors then periodically rated youths on a range of skills to identify their strength and weaknesses. (This occurred 3 times per participant on average, over the remaining 2.5 months of public works), and these evaluations were later used to issue a work certificate that signaled between one and five competencies identified as strengths for each participant. It was recommended to participants to use those certificates in their job search, to better signal their skills. However, certificates were handed to them with a delay, a few months after the end of the public works.[67]

[67]The impact evaluation policy report contains additional details on these two types of training, including on the curricula (Bertrand et al., 2016).

Appendix D 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 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 media), 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 ac-

tivities from a time-use module,[68] the Rosenberg *self-esteem* scale,[69] the *positive affect sub-scale* from the CESD scale,[70] the sub-scale of *(positive) attitude towards the future* and the inverted sub-scale of *present fatalism* from the ZTPI scale.[71] 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.[72]

The *behavior index* aggregates 4 measures: an inverted measure of *anger* in daily activities taken from the time-use module,[73] an inverted measure of *impulsiveness* from the DERS scale,[74] the *conduct problems sub-scale* (inverted) and the *pro-social behavior sub-scale* from the Strengths and Difficulties Questionnaire (SDQ).[75] 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.[76]

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

[68]The 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.

[69]The Rosenberg self-esteem scale includes 10 items that measure self-esteem or self-worth. We use a validated version of the instrument in French (Vallieres and Vallerand, 1990).

[70]The 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).

[71]The 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).

[72]The 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.

[73]This was built as in footnote 68.

[74]The 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).

[75]The 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.

[76]The 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.

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 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 with 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 E Applying Machine Learning to Study Heterogeneity in Treatment Effects

This appendix complements section 5.2 by providing details on the application of machine learning methods to analyze treatment effect heterogeneity. The application is based on Chernozhukov et al. (forthcoming). Section E.1 provides an overview of the approach. Section E.2 describes the sample used to train the models and make predictions.

Section E.3 presents the machine learning algorithms and their parameters. Finally, Section E.4 describes how we adapted the procedure in Chernozhukov et al. (forthcoming) to our experimental setting.

We use similar notations as in Chernozhukov et al. (forthcoming). The Baseline Conditional Average (BCA) writes:

$$b_0(Z) := \mathbb{E}\left[Y(0)|Z\right]$$

The Conditional Average Treatment Effect (CATE) is defined as :

$$s_0(Z) := \mathbb{E}\left[Y(1)|Z\right] - \mathbb{E}\left[Y(0)|Z\right]$$

E.1 Overview

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 $s_0(Z) = E[Y(1) - Y(0)|Z^K = z]$ 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 ($k = 0$). We use ML methods to obtain an estimate $S(Z)$.

A key challenge when using machine learning techniques to study heterogeneity is to derive confidence intervals and perform inference. In this paper, we use the inference framework developed by Chernozhukov et al. (forthcoming), 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).[77]

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

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 distribution of $S(Z)$), and (iii) compare the characteristics of the population who benefit the most or the least from the program. The approach in Chernozhukov et al. (forthcoming) 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 E.3). We present the results for the best performing algorithm in the main tables of the paper, and provide robustness checks in Table A13.

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.[78] Chernozhukov et al. (forthcoming) offer a procedure to aggregate results across simulations and construct valid confidence intervals and p-values.[79]

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

$$Y = \alpha_1 + \alpha_2 B(Z) + \beta_1(T - P(Z)) + \beta_2(T - P(Z))(S(Z) - \widehat{E}(S(Z)) + \varepsilon \tag{F1}$$

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 $P(Z) = \widehat{E}(T|Z)$. 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

[78] Iterations 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.

[79] Their procedure takes into account the uncertainty coming from both the estimation of the parameters and 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. 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%.

(HET).[80] We are particularly interested in β_2, which offers a test for heterogeneity in the 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.[81]

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:

$$Y = \alpha_1 + \alpha_2 B(Z) + \sum_{k=1}^{4} \gamma_k (T - p(Z)) 1(G_k) + \nu \tag{F2}$$

The projection coefficients γ_k are the GATES parameters. The groups are defined as $G_k = \{S(Z) \in I_k\}$ 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 identified by the ML procedure as those who benefit the *most* from the program. Similarly, γ_1 can be interpreted as the average treatment effect among the 25% of individuals identified by the ML procedure as those who benefit the *least* from the program (group G_1).

Note that the question of targeting individuals for whom the effect of a program is the highest is only imperfectly resolved by using GATES for at least two reasons. The first reason is that

[80] In this framework, the quantity $BLP(Z) = \beta_1 + \beta_2(S(Z) - \widehat{E}(S(Z))$ can be interpreted as the best linear predictor of $s_0(Z)$ based on $S(Z)$. Also $\beta_1 = \mathbb{E}[s_0(Z)]$ 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).

[81] Chernozhukov et al. (forthcoming) 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%.

there is heterogeneity in the treatment effect beyond what is explained by the covariates $s_0(z)$. Buhl-Wiggers et al. (2022) shows that the proportion of the variance captured by estimators of treatment effect heterogeneity does not exceed a few percent of the lower bound of the treatment effect variance, even when resorting to ML estimators. The second reason is that the estimated parameters γ_k identify the expectation of the treatment effect conditionally on belonging to the group G_k. However, the group G_4, for example, identifies individuals in the upper quartile of the estimated treatment effect $S(Z)$. Nothing describes how these individuals compare to those in the upper quartile of the true heterogeneous treatment effect $s_0(Z)$. For these two populations to be considered as approaching each other when the sample size becomes large, a property of convergence of the point estimators $S(Z)$ is needed, which is precisely what the analysis of Chernozhukov et al. (forthcoming) accomplishes.

A natural next step is to study the characteristics of the groups of interest (i.e. $\mathbb{E}[g(Z)|G_k]$, 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 empirical section, we also assess how belonging to a particular group ("heterogeneity group") for a given outcome Y affects treatment effect on another outcome $\widetilde{Y}$. In other words, we seek to identify $\mathbb{E}[S_{\widetilde{Y}}(Z)|G_{Y,k}]$, where $S_{\widetilde{Y}}(Z)$ is the treatment effect on variable $\widetilde{Y}$ conditional on Z and $G_{Y,k}$ is the kth 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 for 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 (4) to perform this analysis, replacing Y as a dependent variable with the alternative outcome variable $\widetilde{Y}$.

E.2 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^K, Y^{During}, T) (respectively 'endline' (Z^K, Y^{Post}, T)) sample can be used to build and apply the model to predict 'during' (respectively 'post') conditional treatment effects, where T 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-attritors and non-missing outcomes for both surveys.[82] 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).[83]

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 F1). 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).

E.3 Machine Learning Algorithms

We consider five alternative machine learning algorithms to estimate the proxy predictors and apply the procedure in Chernozhukov et al. (forthcoming): elastic net, boosted trees, random forest, Rlearner (based on elastic net and proposed by Nie and Wager (2020)) and generalized

[82] There is also some attrition between survey rounds, and some missing values in baseline covariates. We exclude from each sample the attritors 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 nonattritors) are dropped from the sample.

[83] Recall 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.

random forest (proposed by Wager and Athey (2018)). All algorithms are implemented in **R**, and we adapted the codes provided by the authors.[84] These machine learning methods can be divided in two groups based on the way they approach the CATE function (Künzel et al. (2017)).

1. **Two learners**

   The first group of machine learning methods includes Elastic Net, Random Forest and Boosted Trees. They predict separately $\mathbb{E}[Y(1)]$ in the treatment group and $\mathbb{E}[Y(0)]$ 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.[85] The two fitted models are then used to predict potential outcomes $\hat{Y}(1)$ and $\hat{Y}(0)$ 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**

[84] https://github.com/demirermert/MLInference

[85] The 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.

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 $\mathbb{E}[Y(1)]$ and $\mathbb{E}[Y(0)]$, they directly fit a model to estimate $\mathbb{E}[Y(1)]$ -$\mathbb{E}[Y(0)]$. 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**[86] 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.

E.4 Adaptation to the Experimental Setting

In our application, we repeat $S = 100$ times the procedure developed by Chernozhukov et al. (forthcoming).[87]

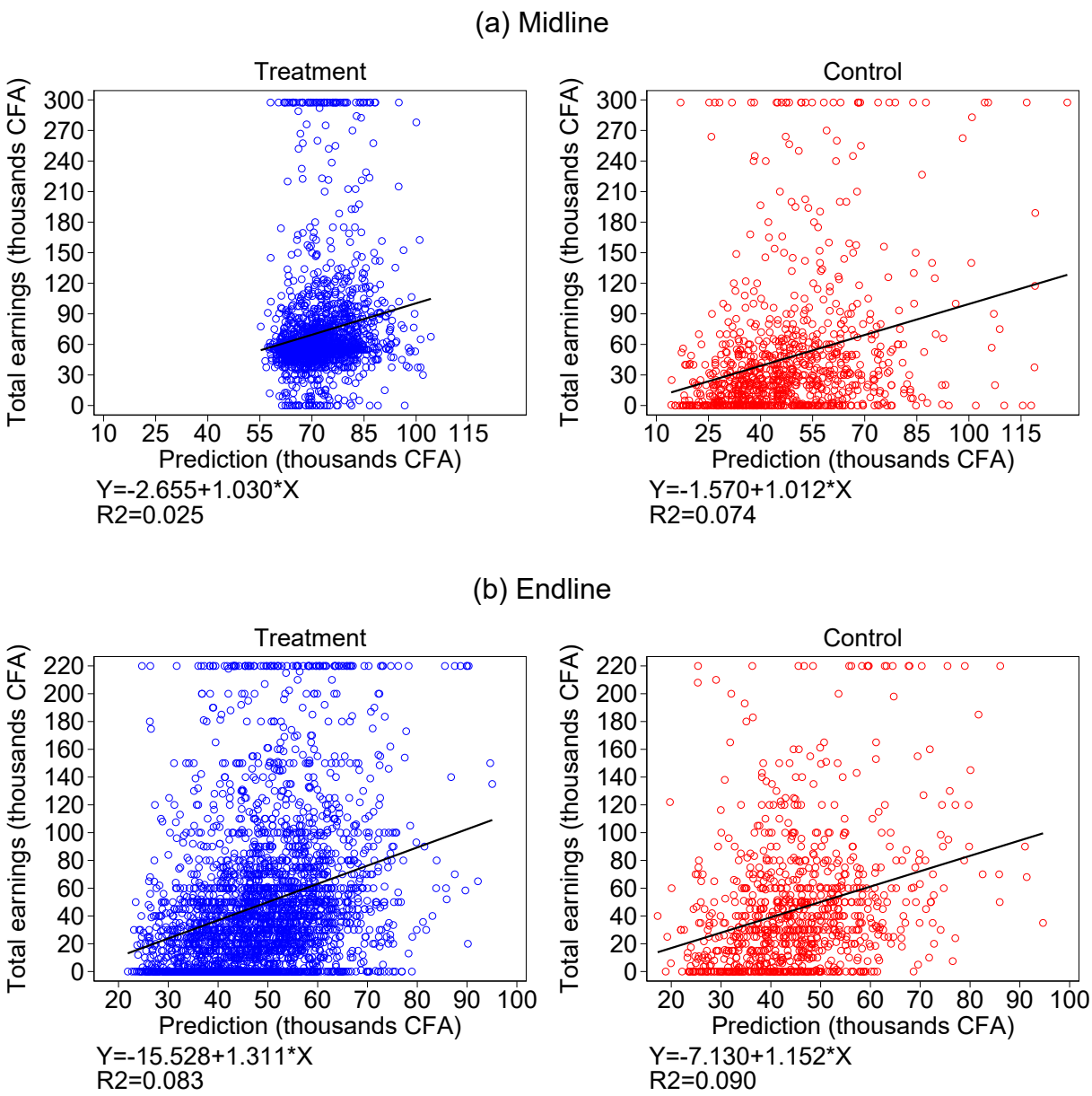
The first step of the method requires partitioning our dataset into an auxiliary and a main sample. We adapted the algorithm so that the sample splits are stratified by our randomization blocks ($locality * gender$), which represents 32 strata. This is important to preserve the identification strategy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

[86] https://github.com/xnie/rlearner

[87] Figure F1 shows the scatter plot of earnings y and $\widehat{y}$ for midline and endline as well as the regression line of y on $\widehat{y}$. 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.

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

Figure F1: 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.

Table F1: Baseline variables used in Machine Learning algorithms

Variable description	**Type**
Individual characteristics	
Female	Binary
Age	Continuous
Number of children	Continuous
Live in urban area	Binary
Education	
Education (total number of years)	Continuous
Has participated in vocational training	Binary
Household characteristics	
Household size	Continuous
Is head of household	Binary
Household assets	
Total number of assets[1]	Continuous
Employment	
Total number of activities	Continuous
Total number of wage-employment activities	Continuous
Total number of self-employment activities	Continuous
Is engaged in (at least one) casual activity	Binary
Total Earnings (monthly)	Continuous
Savings, Expenditures and Constraints	
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.

Appendix F Weights

This appendix describes the weights used in the analysis. Table G1 summarizes the weights used with midline data, and Table G2 with endline data. In general results are robust to weights not being 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 $w_{k1} = N_k \big/ N_{k1} \times N_1 \big/ N_P$ and $w_{k0s} = N_k \big/ N_{k0s} \times N_0 \big/ N_P$. 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:

- $w_{j,k} = N_{k1} \big/ N_{j,k} \times N_j \big/ N_1$ with $j = a, b, c$ [88]
- $w_{k0s} = N_{k1} \big/ N_{k0s} \times N_0 \big/ N_1$

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 $P_l = N_l^S / N_l$.

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

To take sub-sampling into account, original weights are multiplied by S/P_l . Therefore, 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.[89] 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 ($\widetilde{w}_{k0s,j}$). 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 $\bar{C}_3C_4T_4$.[90] 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:

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

[90] Hence $\widetilde{w}_{k0s,C_3T_3\bar{C}_4} = 0$; $\widetilde{w}_{k0s,\bar{C}_3C_4T_4} = 1 \times 0 = 0$; $\widetilde{w}_{k0s,C_3\bar{T}_3C_4T_4} = \frac{N_{k0s,C3}}{N_{k0s,C_3\bar{T}_3}} \times 0 = 0$.

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

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.[91] $E_{b,j}$ from which a random sub-sample is drawn in proportion $\pi_j = NE^S_{b,j}/NE_{b,j}$ (j is an index for treatment status * locality). Individuals interviewed during the tracking phase take a different weight than those interviewed during the main survey phase. Tracking respondents are weighted by $\omega^T_j = R^S_{a,j} + \lambda_j s_j R^S_{b_j}(1 - R^S_{a,j})E^S_{b,j}$, with λ_j, so that the final weight is $\omega^{S,f}_j = \omega^S_j \times \omega^T_j$.

The sum of the weights on population j is therefore : $\omega_j \times (N^S_{a,j} + \lambda_j NER^S_{s,b_j})$, with NER^S_{s,b_j} the number of individuals from the tracking sample who responded during tracking phase. We make the hypothesis that residual non-response $R^S_{b,j}$ is random. We seek to be representative of the respondent population of phases a and b. Therefore, we take $\lambda_j = NE^S_{b,j}/NER^S_{s,b,j}$

In group j, weights will be set such as:[92]

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

[91] Among 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.

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

Table G1: Summary of weights used with midline data

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

Table G2: Summary of weights used with endline data

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

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., N. Buehren, M. Goldstein, and J. Hoogeveen (2023): "Welfare impacts of public works in fragile and conflict affected economies: The Londö public works in the Central African Republic," *Labour Economics*, 81, 102293.

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. Fieulaine (2004): "Validation française de l'échelle de temporalité: The Zimbardo Time Perspective Inventory (ZTPI)," *European Review of Applied Psychology*, 54, 207–2017.

Athey, S. and G. Imbens (2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 113, 7353–7360.

Athey, S., J. Tibshirani, and S. Wager (2019): "Generalized random forests," *Annals of Statistics*, 47, 1148–1178.

Athey, S. and S. Wager (2021): "Policy Learning With Observational Data," *Econometrica*, 89, 133–161.

Bagga, A., M. Holmlund, N. Khan, S. Mani, E. Mvukiyehe, and P. Premand (2024): "Do Public Works Programs Have Sustained Impacts?: A Review of Experimental Studies from LMICs," *World Bank Research Observer*.

Banerjee, A., R. Hanna, B. A. Olken, and D. Sverdlin Lisker (2024): "Social Protection in the Developing World," *Journal of Economic Literature*, 62, 1349–1421.

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.

Berhane, G., D. O. Gilligan, J. Hoddinott, N. Kumar, and A. S. Taffesse (2014): "Can Social Protection Work in Africa? The Impact of Ethiopia's Productive Safety Net Programme," *Economic Development and Cultural Change*, 63, 1–26.

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.

——— (2025a): "Côte d'Ivoire - Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Public Works Impact Evaluation - Baseline Survey, 2013 (PEJEDEC PWP BL 2013)." Downloaded from - https://doi.org/10.48529/stxm-dw72.

——— (2025b): "Côte d'Ivoire - Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Public Works Impact Evaluation - Endline Survey, 2015 (PEJEDEC PWP EL 2015)." Downloaded from - https://doi.org/10.48529/6zxm-x259.

——— (2025c): "Côte d'Ivoire - Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Public Works Impact Evaluation - Midline Survey, 2013 (PEJEDEC PWP ML 2013)." Downloaded from - https://doi.org/10.48529/p1ec-xe78.

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 (2008): "Distributional impacts of the Self-Sufficiency Project," *Journal of Public Economics*, 92, 748–765.

Blair, G. and K. Imai (2012): "Statistical analysis of list experiments," *Political Analysis*, 20, 47–77.

Blattman, C. and S. Dercon (2018): "The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 10, 1–38.

Breda, T., J. Grenet, M. Monnet, and C. Van Effenterre (2023): "How Effective are Female Role Models in Steering Girls Towards STEM? Evidence from French High Schools," *The Economic Journal*, 133, 1773–1809.

Bryan, G., D. Karlan, and A. Osman (2024): "Big Loans to Small Businesses: Predicting Winners and Losers in an Entrepreneurial Lending Experiment," *American Economic Review*, 114, 2825–60.

Buhl-Wiggers, J., J. T. Kerwin, J. Muñoz-Morales, J. Smith, and R. Thornton (2022): "Some children left behind: Variation in the effects of an educational intervention," *Journal of Econometrics*.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (forthcoming): "Fisher-Schultz Lecture: Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments, with An Application to Immunization in India," *Econometrica*.

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 (2024): "Direct and Indirect Effects of Subsidized Dual Apprenticeships," *The Review of Economic Studies*, rdae094.

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.

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.

Franklin, S., C. Imbert, G. Abebe, and C. Mejia-Mantilla (2024): "Urban Public Works in Spatial Equilibrium: Experimental Evidence from Ethiopia," *American Economic Review*, 114, 1382–1414.

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.

Gazeaud, J., E. Mvukiyehe, and O. Sterck (2023): "Cash Transfers and Migration: Theory and Evidence from a Randomized Controlled Trial," *The Review of Economics and Statistics*, 105, 143–157.

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 Questionnaire: A pilot study on the validity of the self-report version," *European child & 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.

Haushofer, J., P. Niehaus, C. Paramo, E. Miguel, and M. Walker (2025): "Targeting Impact versus Deprivation," *American Economic Review*, 115, 1936–74.

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.

Hoddinott, J., G. Berhane, D. O. Gilligan, N. Kumar, and A. Seyoum Taffesse (2012): "The Impact of Ethiopia's Productive Safety Net Programme and Related Transfers on Agricultural Productivity," *Journal of African Economies*, 21, 761–786.

Hussam, R., E. M. Kelley, G. Lane, and F. Zahra (2022): "The Psychosocial Value of Employment: Evidence from a Refugee Camp," *American Economic Review*, 112, 3694–3724.

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.

Ivaschenko, O., D. Naidoo, D. Newhouse, and S. Sultan (2017): "Can public works programs reduce youth crime? Evidence from Papua New Guinea's Urban Youth Employment Project." *IZA Journal of Development and Migration*, 7.

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 (2020): "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.

Manski, C. F. (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246.

McKenzie, D. (2021): "Small business training to improve management practices in developing countries: re-assessing the evidence for 'training doesn't work'," *Oxford Review of Economic Policy*, 37, 276–301.

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.

——— (2023): "General equilibrium effects of (improving) public employment programs: Experimental evidence from India," *American Economic Review*, 91, 1261–1295.

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 (2020): "Quasi-oracle estimation of heterogeneous treatment effects," *Biometrika*, asaa076.

Perova, E., E. Johnson, A. Mannava, S. Reynolds, and A. Teman (2021): "Public Work Programs and Gender-Based Violence: Evidence from Lao PDR," Policy Research Working Paper No. 9691, World Bank, Washington, DC.

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.

Quinn, S. and C. Woodruff (2019): "Experiments and Entrepreneurship in Developing Countries," *Annual Review of Economics*, 11, 225–248.

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.

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.