



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

Fifth Submission: PP_CIV_2025_368

Maria Reyes Retana, Kanika Shokeen

reproducibility@worldbank.org

January 27, 2026

This review verifies the reproducibility of the exhibits included in the paper " *Do Workfare Programs Live Up to Their Promises? Experimental Evidence from Côte d'Ivoire*".

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit in the main section of the paper and 10 randomly selected exhibits in the appendix were reproduced accurately**
- The code was successfully executed on a new computer after:
 1. Change the path in the main do file and run step 1.
 2. Open the R project, restore the environment and run the master R script.
 3. Go back to the stata main do file and run step 3.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 48 hours (11 stata + 37 R) to run.
- We conducted our reproducibility analysis based on the paper shared by the authors on January 16th.
- We ran the code starting from two points: (i) generating the intermediate data using the Machine Learning (ML) model, and (ii) using the intermediate ML data provided directly by the authors to produce the final exhibits. This report presents the results from both runs. An exhibit is marked as reproduced if it matches the paper using the ML intermediate data provided by the authors. The differences observed when generating the ML data from scratch are minimal and likely attributable to differences in operating systems (e.g., Mac vs. Windows) or package versions. We note them below for transparency. See detailed results below.
- **Note:** The starting point of this package is the dataset published in the Microdata Library (see more details in the README). For transparency, the authors have included the code they used to generate

that dataset, and this code is provided within the reproducibility package. The replicators verified that the datasets produced by the included code match exactly the published data available in the Microdata Library, which serves as the starting point of the package.

- **Verification Process and Data Handling:**

- The reproducibility package relies on 1 type of data: accessible (public but not included in the reproducibility package).
- Reviewers used accessible data provided directly by the authors to conduct the reproducibility verification, and this is not included in the public reproducibility package. The data is published on the World Bank Microdata Library.
- Reviewers verified that the accessible data matches what was provided by the authors; results of this verification are documented in the file *comparison_report.csv*.
- *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks. Users who acquire the accessible data can use this file to verify that the data has not been altered.

- **Reproducibility Summary:**

- **Data:** All data sources are publicly available and included in the reproducibility package.
- **Code:** All code files (from cleaning to analysis) are included in the reproducibility package.
- **Outputs:** All outputs are generated by code included in the reproducibility package.
- **Reproducibility verification:** Reviewers had access to the same materials included in the public reproducibility package. They verified that the publicly available data matches the data used in the package. Hashes for the datasets are included in the reproducibility package for verification purposes.
- **Dependencies environment:** The reviewers reproduced an existing environment for dependencies using dependency files or an environment metadata file provided by the authors.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Table 1 Reproduced.** Verified against *Table_1_balance_weighted_prueba.tex*.
- **Table 2 Reproduced** Verified against *Table_2_ENSETE_comp.tex*
- **Table 3 Reproduced.** Verified against *Table_3_main_itt.tex*
- **Table 4 Reproduced.**
- **Table 5 Reproduced.** Verified against *Table_5_main_itt_behavior_wellbeing.tex*.
- **Table 6 Reproduced.** Verified against *Table_A5_itt_bhv_skills.tex*.
- **Table 7 Reproduced.** Verified against *Table_6_ml_v3.tex*. When using the ML created from scratch by the replicators multiple values differ, most of them minimal. However, when reproducing the exhibit using the intermediate ML results provided by the authors, the values match exactly, so this table is considered reproducible. For transparency, the different values are highlighted below.

Paper

Table 7: Heterogeneity in impacts on earnings *during* and *post* program, machine learning results

	(1) La total savings (Monthly)	(2) La total savings (Monthly)	(3) Total earnings (Monthly)	(4) Total earnings (Monthly)
	<i>Midline</i>	<i>Endline</i>	<i>Midline</i>	<i>Endline</i>
Panel A: Predicted average treatment effect and heterogeneity loading parameter				
ATE (β_1)	2.642 (2.595,2.947) [0.000]	-0.0297 (-0.466,0.397) [0.000]	2.4363.0 (17018.5,31150.3) [0.000]	3314.0 (-1842.7,8523.0) [0.414]
HET (β_2)	1.257 (0.883,1.635) [0.000]	0.111 (-0.515,0.721) [0.099]	0.389 (0.00173,0.774) [0.099]	0.416 (-1.149,5.384) [0.756]
Best ML method				
	Generalized Random Forest	Random Forest	Random Forest	R-Learner
Panel B: By quartile of predicted impacts on earnings <i>during</i> program (using <i>midline</i>) (GATES)				
Mean in quartile 1 (0 to 25%)	1.408 (0.870,2.063)	-0.294 (-1.324,0.760)	1.4953.4 (1891.4,28251.8)	469.9 (-11962.0,12389.1)
Mean in quartile 2 (25 to 50%)	2.277 (1.670,2.895)	-0.309 (-1.355,0.733)	23172.3 (9865.0,36605.8)	2793.9 (-9810.5,15283.1)
Mean in quartile 3 (50 to 75%)	2.835 (2.239,3.444)	-0.0515 (-1.082,0.978)	29754.2 (16232.9,42960.7)	5658.1 (-6700.1,17794.3)
Mean in quartile 4 (75 to 100%)	3.986 (3.382,4.585)	-0.271 (-1.314,0.769)	31495.4 (18088.7,44832.7)	4490.7 (-7835.4,16887.3)
P-value all coefficients are equal				
Best ML method	0.000 Generalized Random Forest	1.000 Generalized Random Forest	0.302 Random Forest	1.000 Random Forest
Panel C: By quartile of predicted impacts on earnings <i>post</i> program (using <i>endline</i>) (GATES)				
Mean in quartile 1 (0 to 25%)	2.860 (2.240,3.474)	-0.0834 (-0.963,0.803)	20120.2 (7275.6,32824.1)	2884.7 (-7540.5,12739.3)
Mean in quartile 2 (25 to 50%)	2.641 (2.036,3.253)	-0.0969 (-0.950,0.791)	26809.6 (10405.8,39322.7)	4020.6 (-6197.8,14214.9)
Mean in quartile 3 (50 to 75%)	2.426 (1.795,3.041)	-0.0609 (-0.915,0.796)	27443.4 (14798.2,40110.1)	3920.1 (-6313.2,14117.8)
Mean in quartile 4 (75 to 100%)	2.652 (2.020,3.277)	0.0360 (-0.823,0.886)	25493.3 (11190.1,36318.8)	2549.3 (-7586.6,13104.9)
P-value all coefficients are equal				
Best ML method	0.000 Random Forest	0.100 Random Forest	0.888 Random Forest	0.477 Random Forest

Heterogeneity analysis based on the approach in Chernozhukov et al. (forthcoming) (see discussion in Section 5.1.2 and Appendix C.1). Column (1) and (2) (respectively column (3) and (4)) show the estimated average treatment effect (respectively endline). Estimates in Panel A are based on equation 1. Panel B is based on equation 2, while the heterogeneity of the effect. P-values for the heterogeneity of the effect are reported in Panel C. The chosen algorithm is the one with the highest AUC (lambda). The larger AUC indicates the better the model is. The chosen algorithm is the one with the highest AUC (lambda). The larger AUC indicates the better the model is. The chosen algorithm is the one with the highest AUC (lambda). The larger AUC indicates the better the model is.

- **Table 8 Reproduced.** Verified against *Table_7_ml_charac_v3.tex*. When using the ML created from scratch by the replicators, multiple values differ; most of them are minimal. However, when reproducing the exhibit using the intermediate ML results provided by the authors, the values match exactly, so this table is considered reproducible. For transparency, the different values when running the ML process are highlighted below.

Code

	(1) Mean in 1st quartile	(2) Mean in 4th quartile	(3) Test (1)-(2) (p-value)
Individual characteristics			
Female	0.140	0.540	0
Live in urban area	0.940	0.960	0.560
Age	25.24	24.27	0
Nb of children	0.780	0.870	0.290
Education			
Years of education	11.03	9.790	0.0250
Primary education not completed	0.400	0.480	0.103
Has participated in vocational training	0.520	0.280	0
Household characteristics and assets			
Household size	5.910	6.440	0.0640
Is head of household	0.390	0.0900	0
Total Nb of assets	0.630	0.420	0
Nb of transportation assets	0.840	0.600	0.0220
Nb of agricultural assets	5.430	3.310	0.0410
Nb of household durables	1.950	1.620	0.0240
Nb of communication assets	7.570	6.640	0.00900
Employment			
Has an activity	1	0.440	0
Is Wage-Employed	0.550	0.0900	0
Is Self-Employed	0.510	0.150	0
Nb of activities	1.380	0.460	0
Total Earnings (monthly, CFA)	49110.9	702.4	0
Searched for a job (last month)	0.600	0.520	0.903
Savings, constraints and expenditures			
Has saved (last 3 months)	0.730	0.240	0
Of which: share of formal savings (cond. on savings)	0.330	0.170	0.00200
Has a savings account	0.200	0.0600	0
Savings Stock (CFA)	66498.1	6869.1	0
Self-reported constraints to repay loans	0.200	0.170	0.508
Self-reported constraints to access credit	0.430	0.580	0
Self-reported constraints for basic needs expenditures	0.650	0.710	0.266
Transportation expenditures (last 7 days, CFA)	3012.2	1012.8	0
Communication expenditures (last 7 days, CFA)	3228.0	657.5	0
Cognitive skills and risk preference			
Cognitive (deduction, Raven Test)	0.240	0.230	0.830
Dexterity (Nuts Test)	0.370	0.380	0.0780
Dexterity (Bolts Test)	0.340	0.330	0.436
Positive affect (CES-D scale, Nb positive days)	6.460	6.080	0.0650
Positive attitude towards the future (ZITPI scale)	29.33	29.12	0.702
Is Risk averse (based on hypothetical lotteries)	0.710	0.700	1
Patience (scale 0 to 10, 10=very patient)	3.420	3.280	1
Preference for present (actualization rate for 1 month)	0.580	0.580	1

Paper

Table 8: Baseline characteristics of the bottom and top quartiles of the distribution of predicted impacts on (ln) earnings *during* program

	(1) Mean in 1st quartile	(2) Mean in 4th quartile	(3) Test (1)-(2) (p-value)
Individual characteristics			
Female	0.13	0.53	0
Live in urban area	0.94	0.96	0.56
Age	25.22	24.26	0
Number of children	0.78	0.86	0.31
Education			
Years of education	11	9.69	0.026
Primary education not completed	0.40	0.48	0.103
Has participated in vocational training	0.52	0.28	0
Household characteristics and assets			
Household size	5.91	6.45	0.060
Is head of household	0.39	0.090	0
Total number of assets	0.63	0.43	0
Number of transportation assets	0.83	0.61	0.03
Number of agricultural assets	5.38	3.39	0.041
Number of household durables	1.96	1.63	0.021
Number of communication assets	7.57	6.68	0.011
Employment			
Has an activity	1	0.44	0
Is Wage-Employed	0.55	0.09	0
Is Self-Employed	0.52	0.14	0
Number of activities	1.38	0.46	0
Total Earnings (monthly, CFA)	49113.3	702.4	0
Searched for a job (last month)	0.61	0.52	0.9
Savings, constraints and expenditures			
Has saved (last 3 months)	0.72	0.25	0
Of which: share of formal savings (cond. on savings)	0.33	0.17	0.002
Has a savings account	0.20	0.06	0
Savings Stock (CFA)	66498.9	6795.1	0
Self-reported constraints to repay loans	0.20	0.18	0.513
Self-reported constraints to access credit	0.43	0.58	0
Self-reported constraints for basic needs expenditures	0.66	0.71	0.057
Transportation expenditures (last 7 days, CFA)	3011.8	1021.0	0
Communication expenditures (last 7 days, CFA)	3221.9	655.9	0
Cognitive skills and risk preference			
Cognitive (deduction, Raven Test)	0.24	0.23	0.738
Dexterity (Nuts Test)	0.37	0.39	0.128
Dexterity (Bolts Test)	0.34	0.33	0.436
Positive affect (CES-D scale, No. of positive days)	6.45	6.09	0.064
Positive attitude towards the future (ZITPI scale)	29.33	29.11	0.678
Is Risk averse (based on hypothetical lotteries)	0.71	0.70	1
Patience (scale 0 to 10, 10=very patient)	3.44	3.28	0.072
Preference for present (actualization rate for 1 month)	0.58	0.58	1

Heterogeneity analysis based on the approach in Chernozhukov et al. (forthcoming) (see discussion in Section 5.1.2 and Appendix C.1). Column (1) (respectively (2)) displays average characteristics of the bottom (respectively top) quartile of the distribution of predicted impacts on (ln) earnings during the program. Column (3) reports p-values for a test of equality between the top and bottom quartile. Reported results are based on the algorithm with the best prediction for *midline* : Generalized Random Forest. The chosen algorithm is the one with the highest AUC (lambda): the larger AUC gets, the stronger the correlation between $\hat{y}(X)$, $\hat{y}(X)$, and $y(X)$ (see appendix table A12 for comparisons across algorithms). Means by quartile are estimated for each sample split; reported values are the median across 100 sample splits. P-values are adjusted for partition uncertainty. *Household assets* and *savings stock* variables are winsorized at the 95th percentile.

- **Table 9 Reproduced.** Verified against *Table_8_during_post_v3.tex*. When using the ML created from scratch by the replicators, multiple values differ; most of them are minimal. However, when reproducing the exhibit using the intermediate ML results provided by the authors, the values match exactly, so this table is considered reproducible. For transparency, the different values are highlighted below.

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

	(1) Employed	(2) Wage-Empl. (ln at least 1 activity)	(3) Self-Empl. (ln at least 1 activity)	(4) Total Hours worked (weekly)	(5) Total earnings in CFA (monthly)	(6) Ln total earnings (monthly)	(7) Earnings in Wage-Empl. in CFA (monthly)	(8) Earnings in Self-Empl. in CFA (monthly)	(9) Total expenditures in CFA (monthly)	(10) Savings in CFA (stock)	(11) Well-being index (z-score)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)											
Mean in quartile 1 (0 to 25%)	0.039 (-0.0078, 0.087)	0.35 (0.28, 0.42)	-0.059 (-0.16, 0.046)	0.81 (-4.39, 6.08)	14805 (1639, 27760)	1.46 (0.87, 2.06)	28834 (21922, 35807)	-6859 (-12362, -1345)	15414 (6633, 24276)	35561 (23460, 47674)	0.23 (0.025, 0.44)
Mean in quartile 2 (25 to 50%)	0.093 (0.046, 0.14)	0.40 (0.33, 0.46)	-0.11 (-0.21, -0.0031)	1.12 (-4.09, 6.35)	21937 (8698, 35146)	2.30 (1.70, 2.89)	33479 (26487, 40457)	-6486 (-12019, -966)	10622 (1709, 19388)	32056 (20219, 44174)	0.21 (0.0014, 0.43)
Mean in quartile 3 (50 to 75%)	0.13 (0.082, 0.18)	0.44 (0.37, 0.50)	-0.10 (-0.20, -0.00068)	4.15 (-1.09, 9.28)	30085 (16876, 43013)	2.83 (2.22, 3.42)	36303 (29276, 43299)	-4536 (-9914, 738)	11175 (2444, 19858)	40038 (27897, 51787)	0.11 (-0.10, 0.31)
Mean in quartile 4 (75 to 100%)	0.23 (0.19, 0.28)	0.53 (0.46, 0.60)	-0.12 (-0.22, -0.013)	8.00 (2.63, 13.3)	31604 (18150, 44869)	3.97 (3.35, 4.56)	39464 (32525, 46318)	-4154 (-9726, 1312)	18011 (9471, 26864)	40551 (28577, 52610)	0.21 (0.0047, 0.42)
P-value all interactions equal	0.000	0.002	1.000	0.166	0.228	0.000	0.164	0.911	0.663	1.000	0.984
P-value treat*Q1=treat*Q4	0.000	0.001	0.869	0.104	0.114	0.000	0.068	0.771	1.000	1.000	1.000
Mean in control for Q(1)	0.97	0.85	0.34	48.3	74622.6	10.5	50610.5	144069.4	73954.4	53428.6	0.20
Mean in control for Q(2)	0.97	0.85	0.28	45.0	63875.5	10.3	47993.5	9482.7	59611.2	45068.4	0.16
Mean in control for Q(3)	0.94	0.83	0.25	41.9	60839.7	9.89	46146.6	7528.4	55768.8	42968.5	0.14
Mean in control for Q(4)	0.91	0.79	0.21	37.0	52419.4	9.37	42025.2	5264.7	47311.8	35408.5	0.07
Panel B: <i>Post</i> program impacts (pooled treatment) (12 to 15 months after program ends)											
Mean in quartile 1 (0 to 25%)	-0.020 (-0.098, 0.060)	0.0019 (-0.11, 0.12)	-0.0025 (-0.12, 0.11)	-1.36 (-8.53, 5.80)	1579 (-10755, 13912)	-0.31 (-1.32, 0.73)	3157 (-3737, 10005)	235 (-9751, 10546)	-3729 (-5357, 12897)	12399 (-7700, 32587)	0.16 (-0.077, 0.40)
Mean in quartile 2 (25 to 50%)	-0.018 (-0.096, 0.061)	-0.010 (-0.12, 0.10)	0.0047 (-0.11, 0.12)	-1.06 (-8.27, 6.30)	736 (-11458, 13060)	-0.31 (-1.34, 0.73)	593 (-6330, 7529)	836 (-9441, 11045)	1290 (-7596, 10055)	3689 (-16501, 23859)	0.071 (-0.17, 0.31)
Mean in quartile 3 (50 to 75%)	0.018 (-0.061, 0.096)	-0.0028 (-0.12, 0.11)	-0.018 (-0.13, 0.097)	2.32 (-4.89, 9.55)	6656 (-6066, 18844)	-0.037 (-1.08, 0.99)	830 (-6015, 7696)	5076 (-5057, 15254)	-368 (-9472, 8493)	6800 (-13642, 27063)	0.056 (-0.18, 0.30)
Mean in quartile 4 (75 to 100%)	0.012 (-0.067, 0.090)	-0.028 (-0.14, 0.087)	0.049 (-0.064, 0.16)	-0.065 (-7.30, 7.17)	2771 (-9307, 15356)	-0.25 (-1.29, 0.80)	2661 (-9488, 4389)	5650 (-4708, 16012)	1946 (-7036, 11018)	10040 (-9921, 30277)	0.068 (-0.17, 0.30)
P-value all interactions equal	1.000	1.000	1.000	1.000	1.000	1.000	0.800	1.000	1.000	1.000	1.000
P-value treat*Q1=treat*Q4	1.000	1.000	1.000	1.000	1.000	1.000	0.397	0.855	1.000	1.000	0.979
Mean in control for Q(1)	0.94	0.64	0.40	47.6	64297.7	9.76	30622.6	27727.4	66767.3	77089.8	0.14
Mean in control for Q(2)	0.90	0.56	0.34	44.5	48049.5	8.86	23910.2	19502.6	53638.7	58768.3	0.10
Mean in control for Q(3)	0.88	0.54	0.32	42.6	45313.4	8.37	21591.9	18006.8	49538.3	52302.9	0.10
Mean in control for Q(4)	0.80	0.48	0.32	35.7	34223.3	7.31	16394.0	15165.5	43487.1	42764.8	-0.04

Heterogeneity analysis based on the approach in Chernozhukov et al. (forthcoming) (see discussion in Section 5.1.2 and Appendix C.3). Estimates of treatment effects on the main outcomes during the program (Panel A), and after the program (Panel B) by quartile of the (predicted) impacts on earnings during the program. At the bottom of each panel, we present a test of equality of treatment effects across all quartiles, and a test of equality between the bottom (Q1) and top (Q4) quartiles. Estimation is based on a specification similar to equation 4 but replacing the dependent variable with the alternative (midline or endline) outcome variable. The best predictions during program are used: Generalized Random Forest. The chosen algorithm is the one with the highest Λ (lambda): the larger Λ gets, the stronger the correlation between $\hat{s}_0(Z)$, noted $S(Z)$, and $s_0(Z)$ (see appendix table A12 for comparisons across algorithms). Predictions are estimated for each sample split, reported values are the medians across 100 sample splits. P-values are adjusted for partition uncertainty. *Hours, earnings, expenditures, and savings* winsorized at the 97th percentile. For variables (y) in logarithms we take $\ln(y + 1)$.

- **Table 10 Reproduced.** Verified against *Table_9_ml_post_bis_v3.tex*. When using the ML created from scratch by the replicators, some values differ, most of them are minimal. When reproducing the exhibit using the intermediate ML results provided by the authors, the values match exactly, so this table is considered reproducible. For transparency, the different values are highlighted below.

Paper

	Investment in Self-Employed Activities				Search for Wage Jobs	
	(1) Total number of income generating activities	(2) Value of productive assets in CFA (stock)	(3) Start-up capital in CFA (main self-empl. act., stock)	(4) Value of investments in CFA (last 3 months)	(5) Searched for a job (last 3 months)	(6) Total spent in job search in CFA (last 12 months)
Mean in quartile 1 (0 to 25%)	0.026 (-0.12, 0.18)	11189 (261, 22223)	14723 (-310, 29499)	538 (-1431, 2560)	0.082 (-0.038, 0.20)	1806 (-3341, 6861)
Mean in quartile 2 (25 to 50%)	0.0093 (-0.14, 0.16)	3843 (-7105, 14666)	2899 (-11880, 18088)	322 (-1654, 2301)	0.0062 (-0.11, 0.13)	-580 (-5819, 4584)
Mean in quartile 3 (50 to 75%)	-0.0043 (-0.16, 0.15)	-1325 (-12576, 9517)	2086 (-13125, 17246)	-625 (-2558, 1338)	0.025 (-0.096, 0.14)	828 (-4278, 5934)
Mean in quartile 4 (75 to 100%)	0.038 (-0.11, 0.19)	-20 (-11026, 10771)	5612 (-9583, 20764)	-138 (-2121, 1916)	0.057 (-0.065, 0.18)	887 (-4198, 5942)
P-value all interactions equal	1.000	0.350	0.893	1.000	0.956	1.000
P-value treat*Q1=treat*Q4	1.000	0.282	0.710	1.000	1.000	1.000
Mean in control for Q(1)	0.55	23944.3	43103.8	3673.3	0.57	11914.3
Mean in control for Q(2)	0.50	18478.1	30054.0	2704.7	0.55	8466.7
Mean in control for Q(3)	0.47	14974.2	26206.3	2242.7	0.55	8269.3
Mean in control for Q(4)	0.40	10781.7	19455.5	974.5	0.56	8413.1

Robustness checks based on the approach in Pischke (2014) as implemented in the R package *lme4* (Bates & Kuznetsov, 2019). Estimates

Code

	Self-employment				Job search	
	(1) Total number of income generating activities	(2) Value of prod. assets in CFA (stock)	(3) Start-up capital in CFA (main self-empl. act., stock)	(4) Value of investments in CFA (last 3 months)	(5) Searched for a job during last 3 months	(6) Total exp. in job search in CFA (last 12 months)
Mean in quartile 1 (0 to 25%)	0.024 (-0.13, 0.18)	11462 (444, 22593)	14866 (-250, 30017)	554 (-1391, 2467)	0.085 (-0.035, 0.21)	1962 (-3348, 7044)
Mean in quartile 2 (25 to 50%)	0.011 (-0.14, 0.16)	3617 (-7474, 14587)	2969 (-12028, 17857)	249 (-1733, 2245)	0.0085 (-0.11, 0.13)	-687 (-5884, 4529)
Mean in quartile 3 (50 to 75%)	0.00084 (-0.15, 0.15)	-784 (-11572, 9932)	1784 (-13336, 16915)	-591 (-2554, 1355)	0.025 (-0.094, 0.15)	834 (-4282, 5958)
Mean in quartile 4 (75 to 100%)	0.032 (-0.12, 0.19)	-33 (-10999, 10882)	5610 (-9393, 20854)	-63 (-2065, 1969)	0.053 (-0.068, 0.17)	973 (-4197, 6067)
p-value: all interactions equal	1.000	0.342	0.814	1.000	0.902	1.000
p-value: treatXQ1=treatXQ4	1.000	0.253	0.681	1.000	1.000	1.000
Mean in control for Q(1)	0.55	24110.8	42964.8	3667.0	0.57	11906.1
Mean in control for Q(2)	0.50	18540.1	30232.9	2697.4	0.55	8388.2
Mean in control for Q(3)	0.47	14772.7	26030.6	2219.0	0.55	8170.9
Mean in control for Q(4)	0.40	10882.4	19404.2	957.6	0.56	8353.3

- **Table 11 Reproduced.** Verified against *Table_10_target_MLv3.tex*. When using the ML created from scratch by the replicators, some values differ, most of them are minimal. When reproducing the exhibit using the intermediate ML results provided by the authors, the values match exactly, so this table is considered reproducible. For transparency, the different values are highlighted below.

Paper

Table 11: Impacts on (ln) earnings and cost-benefit ratios under alternative targeting approaches

	Random selection	Machine learning pred. by quartile of pred. impacts <i>during</i> program		Selection on baseline characteristics		
	(1)	(2)	(3)	(4)	(5)	(6)
	Treated (ITT)	Mean in quartile 4 (75% to 100%)	Mean in quartile 1 (0 to 25%)	Women	Low baseline earnings (self-declared) (bottom 25 %)	Low baseline earnings (predicted) (bottom 25%)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)						
Treatment	2.948*** (0.194)	3.966*** (0.307)	1.465*** (0.305)	3.728*** (0.349)	4.035*** (0.384)	3.657*** (0.373)
Observations	2912	2877	2877	2912	2877	2877
Panel B: Post program impacts (pooled treatment) (12 to 15 months after program ends)						
Treatment	-0.037 (0.181)	-0.252 (0.529)	-0.306 (0.525)	-0.050 (0.345)	-0.147 (0.517)	-0.020 (0.405)
Observations	3934	3865	3865	3934	3736	3736
Panel C: Cost-benefit Analysis						
Effect <i>during</i> program in CFA [A]	27083	31671	14660	32097	33954	32824
Effect <i>post</i> program in CFA [B]	4361	4491	461	7168	7602	10648
Discounted sum of impacts (<i>during</i> + <i>post</i> program) [C]	253920.50	287958.58	109428.85	330429.84	349837.28	386817.87
Total cost per beneficiary [D]	768708.10	768708.10	768708.10	768708.10	768708.10	768708.10
Cost-benefit ratio (<i>during</i> + <i>post</i> program) [E]	3.03	2.67	7.02	2.33	2.20	1.99

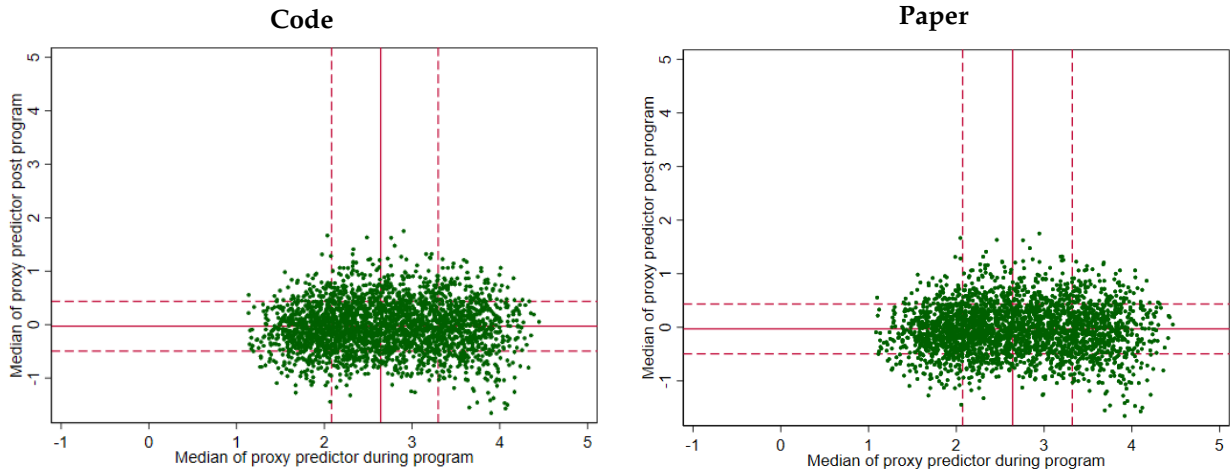
Column (1) is the ITT estimate based on specification in equation 1. Columns (4-6) show estimated treatment effects by subgroups (obtained from heterogeneity analysis using baseline covariates). Columns (2) and (3) document the effect of selecting the observations in the top and bottom quartile of the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

Code

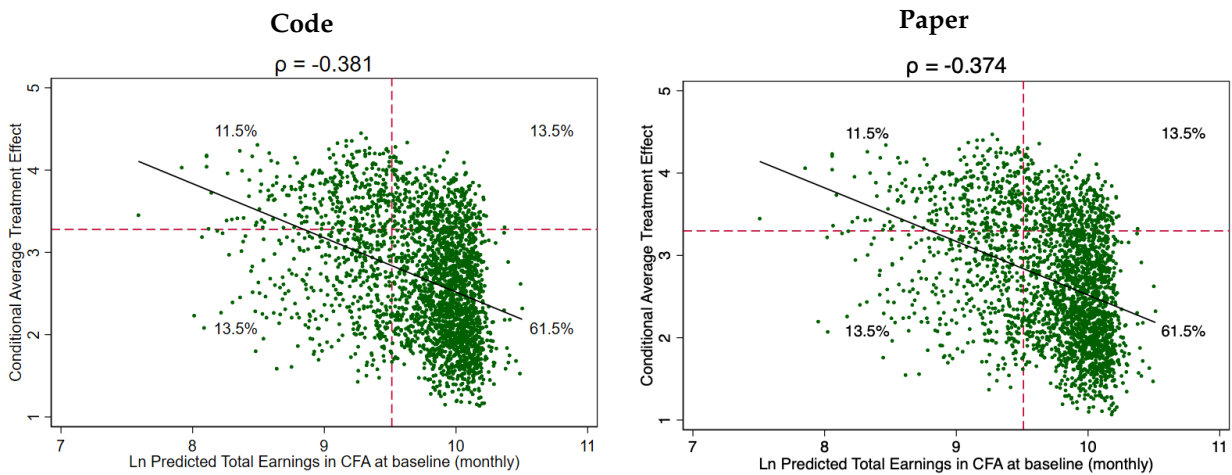
	Random selection	Machine Learning predictions		Selection on baseline characteristics		
	(1) Treated (ITT)	(2) Mean ($\hat{\rho}^{CT}$) in 4th quartile	(3) Mean ($\hat{\rho}^{CT}$) in 1st quartile	(4) Women	(5) Low baseline earnings (self-declared, bottom 25%)	(6) Low baseline earnings (predicted, bottom 25%)
Panel A: Impacts <i>during</i> the program (around 4.5 months after program starts)						
Treatment	2.948*** (0.194)	3.986*** (0.307)	1.468*** (0.306)	3.728*** (0.349)	4.035*** (0.384)	3.615*** (0.368)
Observations	2912	2877	2877	2912	2877	2877
Panel B: Post program impacts (pooled treatment) (12 to 15 months after program ends)						
Treatment	-0.037 (0.181)	-0.271 (0.529)	-0.294 (0.525)	-0.050 (0.345)	-0.147 (0.517)	0.101 (0.415)
Observations	3934	3865	3865	3934	3736	3736
Panel C: Cost-benefit Analysis						
Effect <i>during</i> program in CFA [A]	27083	31495	14953	32097	33954	32550
Effect <i>post</i> program in CFA [B]	4361	4491	420	7168	7602	10428
Discounted sum of impacts (<i>during</i> + <i>post</i> program) [C]	253920.50	286729.85	111618.38	330429.84	349837.28	381682.84
Total cost per beneficiary [D]	768708.10	768708.10	768708.10	768708.10	768708.10	768708.10
Cost-benefit ratio (<i>during</i> + <i>post</i> program) [E]	3.03	2.68	6.89	2.33	2.20	2.01

Notes: Column (1) reports the ITT estimate from Equation ???. Columns (4-6) show the β_i coefficient from Equation ???. Columns (2) and (3) document the effects for the top and bottom quartiles of predicted impacts. Column (3)'s post-program impact is assumed to be zero for [C]. Column (6) uses baseline earnings predicted via Elastic Net. [A] and [B] are winsorized at the 97th percentile. [C] is calculated as $\sum_{k=1}^T \rho^{(k-1)\beta^{dur+mid}} + \sum_{k=8}^T \rho^{(k-7)\beta^{post}}$, with $\rho = (1+\delta)^{-1/12}$ and $\delta = 10\%$. [E] is the cost-benefit ratio. Standard errors clustered at the brigade level. * $p < .1$, ** $p < .05$, *** $p < .01$.

- **Figure 1 Reproduced.** Verified against *figure_1_qte_all*
- **Figure 2 Reproduced.** Verified against *figure_B3_qte_Tarm.png*.
- **Figure 3 Reproduced.** Verified against *figure_2_gates_MLv3*
- **Figure 4 Reproduced with minor differences.** Verified against *figure_3_scatter_cate_v3.png* There are some visual differences, but as these are fairly minimal, this is considered reproducible.



- **Figure 5 Reproduced.** Verified against *graph_predbase_vs_predcate_mid*. When using the ML results provided by the authors this figure reproduces exactly, when running the ML model from scratch, there are minimal differences documented below.



Results in the Annex

For the Appendix, we did not review every exhibit. Instead, we randomly selected 10 exhibits to assess the appendix. Our review was based on those 10 exhibits. Since they were chosen randomly, we are operating under the assumption that if all randomly selected exhibits are reproducible, then the rest should be as well. The seed used to generate the random selection was 2646874, and the exhibits selected were: 1 (Table A1), 2 (Table A2), 5 (Table A5), 10 (Table A10), 13 (Table A13), 15 (Table A15), 19 (Figure B1), 21 (Figure B3), 24 (Table A20), 25 (Figure D1). Note: Tables initially numbered as A10, A19, and A20 were not included in the final manuscript, so these are not included in the final report.

- **Table A1 Reproduced.** Verified against *Table_A1_programs.tex*.
- **Table A2 Reproduced.** Verified against *Table_4_rob_itt.tex*.
- **Table A5 Reproduced.** Verified against *Table_A4.tex*.

- **Table A6 Reproduced.** Verified against *Table_A5.tex*
- **Table A12 Reproduced.** Verified against *Table_aspiration_itt*.
- **Table A14 Reproduced.** Verified against *Table_A7_ml_extcov_v3*
- **Figure B1 Reproduced.** Verified against figure *graph_cum_dist_income.png*.
- **Figure B2 Reproduced.** Verified against *graph_base_vs_pred.png*
- **Figure B3 Reproduced.** Verified against *figure_B5_CE_v3.png*

Reproduction Environment

- Paper exhibits were reproduced in a computer with the following specifications:
 - OS: MacOS Tahoe M4 pro
 - Memory available: 497 GB
 - Software version: Stata version 19.5 MP, R version 4.4