The Fading Treatment Effects of a Multi-Faceted Asset-Transfer Program in Ethiopia

Nathan Barker, Dean Karlan, Christopher Udry, Kelsey Wright[1]

We study the long-run effects of a big-push "graduation" program in Ethiopia in which very poor households received a one-time transfer of productive assets (mainly livestock), technical training, and access to savings accounts. After seven years, treatment effects on wealth and consumption remain economically meaningful but dissipated relative to the two- and three-year results. Treatment effects on other outcomes attenuated further. Based on absolute wellbeing (e.g., food security) not dropping, we argue that the treatment effect dissipation is driven primarily by improved living standards for control households, rather than losses of the previously accrued benefits for the treatment households.

JEL Classification: I32, I38, J24, O12

[1] Nathan Barker: University of Chicago, barkern@uchicago.edu; Dean Karlan: Northwestern University, CEPR, NBER, IPA, and J-PAL, karlan@northwestern.edu; Christopher Udry: Northwestern University, CEPR, NBER, and J-PAL, christopher.udry@northwestern.edu; Kelsey Wright: Paris School of Economics, kelsey.wright@psemail.eu.
This study received approval from the Yale University Human Subjects Committee, IRB Protocol #1002006308, and from the Innovations for Poverty Action Human Subjects Committee, IRB Protocol #5.09December-003. The authors would like to thank the Relief Society of Tigray (REST) for partnering on the study, especially Teklewoini Assefa, Dr. Mulugeta Berhanu, Desta Gebremichael, and Zeratsion Fessha. Thanks to Sana Khan, Matthew Lowes, Hideto Koizumi, Noor Sethi, and Liam Frolund for outstanding research assistance. Thanks to Nathanael Goldberg for his expertise and work on project design and evaluation. JaRco Consulting assisted Innovations for Poverty Action with the first wave of data collection; Innovations for Poverty Action led subsequent waves. Teame Tesfay provided exceptional management of the follow-up surveys. Thanks to the Ford Foundation, 3ie, and U.S. Agency for International Development (USAID) [through the Financial Integration, Economic Leveraging, broad-based Dissemination and Support Leader with Associates (FIELD-Support LwA) managed by FHI 360] for funding. The contents of this paper are the responsibility of the authors and do not necessarily reflect the views of FHI 360, USAID, or the United States Government. We also sincerely thank the study participants for their time.

I. Introduction

A central hypothesis within development economics is that poverty traps can explain why some individuals remain poor. At sufficiently low levels of wealth, individuals face a limited set of occupations, thus limiting wealth accumulation potential and in turn limiting income growth potential (Galor and Zeira 1993; Banerjee and Newman 1993). Such theories imply that the same individuals could achieve a higher steady-state income if given a one-time infusion of capital. This theoretical literature, coupled with empirical evidence suggesting high short-run returns to capital in low-income countries (Blattman et al., 2016, 2014; De Mel et al., 2008; Udry and Anagol, 2006), has provided a basis for public policies that provide one-time capital transfers sufficiently large to start income-generating activities.

Consistent with this microeconomic theory and empirical results, several studies have found promising short- and medium-run evidence on the effectiveness of "graduation" programs at increasing the earnings and wealth of very low-income individuals in developing countries. In graduation programs, individuals receive a one-time transfer of productive assets (or cash to buy productive assets), coupled with training, consumption support (often), and improved access to savings; all together the aim is to push these individuals into a higher steady-state income. Randomized evaluations of graduation-style programs have found positive impacts on quantity and diversity of income-generating activities, earnings, and wealth (Bandiera et al. 2017; Banerjee et al. 2015; Bedoya et al. 2019; Brune et al. 2022). The theory inherent in these programs—that a one-off "big push" is sufficient to bring about higher long-run incomes—has been broadly adopted by policymakers. As of 2020, over 20 million households across 75 countries received some program centered around a one-off transfer (Andrews et al., 2021).

However, demonstration that these programs successfully overcome important poverty traps requires more long-run evidence. In particular, if these programs push individuals over some critical threshold, we should expect that the short-run gains of these programs are likely to persist, and moreover to see untreated households likely fail to clear the same threshold (Balboni et al., 2022). The available long-run evidence on graduation programs, while limited, is largely consistent with this hypothesis. In randomized evaluations of graduation programs in Bangladesh and India, treatment effects persist for at least seven and ten years, respectively (Bandiera et al. 2017; Banerjee, Duflo, and Sharma 2021). Showing that these programs lead to positive impacts over long time horizons across a broad variety of contexts is consistent with the hypothesis of prevalent poverty traps.

However, there are also theoretical reasons we might expect to see convergence in the longer run. If the threshold theory is right and a program fails to push individuals over a critical threshold, we should expect treated households to be likely to return rapidly to their pre-program equilibrium. Alternatively, if the constraints faced by the poor are not binding in the long-run, for example because individuals can borrow or save their way above a critical capital threshold, or because labor productivity in the occupations available to them is high, we might expect to see relatively high rates of convergence by control households (Buera, Kaboski, and Shin 2015). Distinguishing between these two candidate explanations for fading treatment effects thus depends on examining trends of both control and treatment households. Fading treatment effects coupled with control stagnation is consistent with (at least some share of) treated households

failing to clear a critical threshold; fading treatment effects due to strong control growth suggests that the transfers accelerated growth that was imminent in the absence of treatment.

While the India and Bangladesh graduation programs have persistent long-run treatment effects, similar programs elsewhere have found patterns more broadly consistent with capital transfers solely operating as an accelerant. For instance, in Uganda, cash transfers to young adults led to income and wealth gains at two and four years, but the control group caught up by year nine, driven by growth in control group earnings (Blattman et al., 2020). The similarity in short-run results between these graduation and cash transfer studies, coupled with differing degrees of persistent treatment effects, suggest further value in understanding the economic environments and program components in which capital transfer-centered programs are and are not consistent with the presence of and escaping from poverty traps.

We analyze the long-run effectiveness of a graduation program in Ethiopia, whose initial two-year results were reported in Banerjee et al (2015), and which was a local adaption of the NGO BRAC's graduation program. Our study consists of 925 individuals from the region of Tigray who were eligible for the Productive Safety Net Programme, Ethiopia's food-for-work program. Recruited subjects were enrolled in public lotteries in their local communities; winning households received a one-off transfer of either sheep and goats, oxen, bees, or inventory for petty trade, coupled with technical training and coaching, and were given access to local bank accounts (with the requirement that $270 must be saved before individuals could draw down their savings). The program yielded large and positive impacts on consumption, income and wealth after two and three years.

Our core result is that while there continue to be positive treatment effects on household welfare seven years after the initial program, long-run impacts have faded relative to the impacts present in earlier waves, driven by strong growth in control group living standards.

Two and three years after the transfer of assets, the graduation program led to large average increases in per capita consumption: treatment effects of 0.24 standard deviations (SDs) at two years, 0.25 SDs at three years; and large increases in asset wealth: 0.95 SDs at both two and three years. Concurrently, we see large gains in the frequency with which individuals engage in livestock market transactions. Two years after the asset transfer, treated have bought an average of $74 more livestock in the last 12 months (control mean of $13—these and all other values are in 2021 USD Purchasing Power Parity (PPP) terms, using an index of local prices, detailed in Appendix A), and have sold $79 more livestock (control mean of $18).

By year seven, treated households continue to have greater wealth and higher consumption, but the gap has closed meaningfully. Treatment group wealth continued to grow modestly between years three and seven (by 0.12 SDs), but control households grew much faster, by 0.64 SDs, causing the treatment effect to fall to 0.43 SDs. Similarly, while treatment per capita consumption grew by 0.50 SDs in this window, control per capita consumption grew by 0.58 SDs, causing the treatment effect to fall to 0.17 SDs. The frequency of livestock transactions has also narrowed between the groups; in the seventh post-treatment year, treated households bought just $6 more livestock and sold $38 more (in this case, reflecting both control group growth and treatment group decline).

Between years three and seven, control households caught up to treatment households in food security and livestock ownership, two domains where it is especially feasible to measure economic welfare in real terms. On our food security index, control households have improved by 0.71, 0.74 and 1.02 standard deviations in the two-, three- and seven-year surveys. For example, the share of control households reporting that all members get enough food everyday has risen from 34% at baseline to 83% at year seven. As a result, treatment effects fall from a peak of 0.15 SDs in year seven to 0.04 (and not significantly different from zero), despite treatment households also improving their food security in this period. Similarly, we find that livestock ownership among treated households held constant over time (from $2,308 to $2,469 to $2,449 at years two, three and seven, respectively); fading treatment effects reflect control household livestock accumulation.

Our results are consistent with the hypothesis that the graduation program accelerated treatment household consumption, income, and wealth growth, but not with the idea that it pushed households into an equilibrium they would have been unlikely to reach absent the program. Unlike the results found in other longer-run studies of similar programs, the control households closed the initial gap in earnings and wealth achieved by the program. They experienced meaningful gains in living standards during the study period, and engaged in many of the same economic behaviors as treatment households (specifically livestock ownership and sales).

II. Experimental Design, Intervention and Data

The study took place in ten *tabias* (wards) in Tigray Region, Ethiopia.[2] Subjects were recruited from the government's food-for-work program, the Productive Safety Net Programme (PSNP). In our study area, PSNP participants earn chickpeas, wheat, and cooking oil in exchange for manual labor (e.g., digging trenches and engaging in land rehabilitation).

A local Community Task Force identified 1,000 PSNP participants for the study, on the basis of their having below average landholdings, low levels of livestock, high dependency ratios, and limited external income. Individuals had to be capable of engaging in a physical activity, and not have an outstanding loan at any financial institution. After this initial screening, 925 individuals were contacted and deemed eligible.[3] A separate public lottery was conducted in each of the ten tabias to determine treatment status. To our knowledge, our study had full compliance—all treated households received the graduation program, and no control households did. Appendix Table A1 shows balance across the two groups at baseline, consistent with successful randomization.

The graduation program was implemented by Tigrigna NGO Relief Society of Tigray (REST), a large and well-known NGO in the region that offers a variety of services, including water and sanitation, livelihoods programs, and education programs.

[2] A tabia is the smallest administrative unit in Ethiopia. In our setting, a tabia comprises 1-6 villages (the median tabia contains four).

[3] 19 were not found, 52 were deemed ineligible on the basis of existing loans, and four were discovered to be duplicates.

Treatment households were offered one of four productive asset options, each meant to be worth roughly $270 USD in value at the time ($1,371 2021 USD PPP): (1) 16 sheep or goats for fattening, (2) two oxen for fattening, (3) two beehives and colonies, and (4) a comparable value of inputs for petty trade. Most households chose some type of livestock: 62% chose shoats, 24% chose oxen, 10% beekeeping, and 4% petty trade. Each of the asset choices was accompanied by technical training specific to their asset.

Households additionally received a bank account at the Dedebit Credit and Savings Institution (a microfinance institution) in the nearest market center. Individuals were required to accumulate savings equal to the value of the initial transfer prior to being able to withdraw their savings.[4] REST staff visited the program individuals for coaching on a regular basis over two years (initially weekly, then monthly). REST also offered quarterly refresher trainings of the core income generating principles. Unlike in some program contexts, treated households did not receive additional short-term cash or food transfers beyond what was already available to them (and to control households) via the PSNP program. Program activities ran from May 2010 to May 2012.

The research NGO Innovations for Poverty Action conducted four waves of surveys. The baseline survey took place two months prior to the lottery and follow-up surveys after two, three, and seven years. In each, we administered both a household survey and up to two adult surveys. The household survey included household member characteristics, housing, assets, income, consumption, food security, health spending and shocks. The adult survey was administered to the household head (and their spouse, if married); questions included time use, household decision-making, physical and mental health, and political participation. We report a timeline of the program and surveying in Appendix Table A2.

Attrition is low: we successfully interviewed 99%, 98%, and 96% of households in the two-, three, and seven- year surveys, respectively. Appendix Table A3 reports more details, including tests showing that attrition does not differ statistically significantly by treatment status.

III. Results

Empirical Strategy

We estimate average treatment effects of the graduation program on outcome *y* for household *h* at time *t* with the specification:

$$(1)\ y_{ht} = \alpha + \beta\ Treat_h + \delta\ y_{ho} + \ \mu_v + \ \epsilon_{ht}$$

Where $Treat_h$ is an indicator variable for whether the household was randomly chosen to receive the program, y_{ho} is the outcome at baseline (coded as 0 at missing if baseline, with an

[4] This was done because of the Ethiopian government's general opposition to handouts. This program requirement to save the transfer amount prior to withdrawing funds ensured that the program was "like a loan," though unlike a loan, households did not have to repay anyone.

additional indicator variable for baseline missingness), μ_v are tabia fixed-effects (reflecting that the randomization was stratified at the tabia level), and ϵ_{ht} is our error term. Standard errors are clustered at the household level (i.e., household-level outcomes have a cluster size of one; adult-level outcomes, where we survey 1-2 members per households, share a cluster).

In Figures 1-4, we show the evolution of treatment and control means over time, and present average treatment effects from these regressions.[5]

Indexed Family Outcomes

To provide an aggregate view of how the graduation program affects household and individual welfare over time, we report results on indexed family outcomes, following the grouping of families laid out in Banerjee et al. (2015). A full description of every variable component in each index is listed in Appendix B. We follow Kling, Liebman, and Katz (2007) and create z-score indices for each family. We standardize against the baseline whenever possible, or against the control group at year two (when any relevant variables were not collected at baseline).[6] Additionally, we correct for multiple hypothesis testing using the Benjamini-Hochberg step-up procedure to control for the false discovery rate (FDR) (Benjamini and Hochberg, 1995), following the procedure discussed in Anderson (2008). We report q-values for each indexed outcome, equal to the minimum significance level at which we can reject the null hypothesis, once adjusting for the FDR.

Results are reported in Figure 1. We report outcomes on political and women's empowerment in Appendix Table A4 because of space constraints.[7]

Three key patterns emerge from Figure 1.

First, the evaluation took place during a period of meaningfully improving living standards for control households. On our measure of asset ownership—equal to the sum of the value of durable goods, livestock, and other productive assets (e.g., farm tools), standardized to the baseline mean and standard deviation (SD)—the control household mean has improved by 1.16 SDs by year seven relative to baseline, and similarly, food security has improved by 1.02 SDs. Relative to the two-year mark (the first year for which we have comparable data), real per capita consumption—equal to the value of all food consumed by the household, other expenses occurred by the household, and durable good purchases, with the monthly sum standardized against the endline 1 control group—has increased by 0.49 SDs at year seven, and income and revenues by 1.25 SDs. These changes in control means suggest that the program's economic environment was not one in which households were stagnating in a low-level equilibrium, but rather one of solid economic growth.

---

[5] The outcomes in Figures 1-3 correspond to the outcomes reported in Tables 1-3 of Banerjee, Duflo, and Sharma (2021) ("BDS"). The results from our site and BDS are also reported in Appendix Tables A4-A6, to facilitate comparison across the two sites.

[6] Three measures—per capita consumption, asset wealth and productive time use—are aggregates. In these cases, we index against the aggregate rather than the mean of the component parts.

[7] We still adjust our q-values for these families, i.e., we adjust for the fact we are testing ten indexed family outcomes.

Second, treatment effects on economic outcomes (wealth, consumption, food security, income and revenues, financial inclusion, productive time use) all follow a similar pattern: large and positive effects at years two and three, and with positive coefficients, albeit of reduced magnitude by year seven, and in many cases, a loss of statistical significance. For example, the estimated treatment effect on asset ownership has fallen from 0.95 SDs at years two and three to 0.43 SDs at year seven. Treatment effects on our income and revenues and financial inclusion indices both exceeded 1 SD at their peak at year two (1.41 SDs in the case of income and revenues, 1.85 in the case of financial inclusion). By year seven, both point estimates are still moderately sized (0.24 and 0.30 SDs, respectively), but in neither case can we reject the null of no effect, though because variance in our sample is growing over time, we are less powered to detect effects of a given size by year seven.

Our consumption estimates paint an intermediate picture: the treatment effect decline from year three to year seven is relatively modest: 0.25 (p-value<0.01) to 0.17 SDs (p-value=0.036). However, given the large number of families we are testing (and the lower rate of statistical significance across the ten families at year seven, leading to a greater false discovery rate adjustment), the adjusted q-value on per capita consumption is 0.196 at year seven, suggesting ambiguity about the persistence of this effect.

In the case of asset ownership, consumption, food security, and income and revenues, fading treatment effects by year seven can be fully explained by growing control group means; treatment means have not fallen below year two levels. For financial inclusion, of the 1.55 SD decline in treatment effects, 47% can be explained by control growth improvements.

Finally, we observe limited evidence of treatment effects on downstream outcomes. In no wave can we reject the null that the program had no effect on mental or physical health, or women's decision-making (the last of which is reported in Appendix Table A4, given space constraints).

Consumption and Food Security

In Figure 2, we document the evolution of treatment and control households' consumption and food security, with a focus on the extent to which the program pushes individuals out of extreme poverty.

Control household living standards are improving considerably in the time horizon studied. At year seven, control household per capita consumption (the same outcome in Figure 1, reported here in 2021 USD PPP) is 33% higher than at the two-year mark. The share of households getting enough food everyday has grown from 34% at baseline to 83% by year seven. Treatment effects are initially positive on both consumption and food security—by year three, per capita consumption is 18% higher among treatment households; food security has improved by 0.15 SDs. By year seven, consumption gains fade but remain positive (9% higher), but we can no longer reject that food security treatment effects are zero.

While these accumulated gains are sizable, treated households are still very poor by most standards. At year seven, control household per capita consumption is equal to $2.23 per day,

and treatment consumption $2.43. Mean treatment household consumption has thus just exceeded the World Bank's extreme poverty threshold of $2.38 per day.[8]

Income and Revenues

The same pattern is present for most income-based outcomes: strong treatment effects in years two and three, that have faded but are generally still significant in year seven. Here however, fading treatment effects reflect a combination of strong control group growth and declining treatment values for certain outcomes. We report aggregate outcomes in Figure 3, and explore mechanisms in Figure 4.

We see large effects on livestock revenues at the two-year mark; treated individuals have $129 more per month (relative to a control mean of $33). This effect fades by years three and seven to $31 and $40 per month, respectively. At the seven-year mark, 65% of the declining treatment effects are due to control households increasing the extent to which they engage in livestock market transactions—their mean increases from $33 per month at year three to $90 per month at year seven.

We see some evidence that treatment households have diversified their income sources beyond those directly promoted by the program, and towards agricultural activity. Treatment effects on net agricultural profits are positive at years two and three, and smaller and statistically insignificant at year seven. Figure 4 shows that treatment effects on total household agriculture expenditures and land cultivated are both positive in all three follow-up waves. In years two, three and seven, the treatment effects on acres cultivated by the household are 0.15, 0.22, and 0.25, respectively (representing 11%, 14%, and 18% increases). In years two and three, these increases translate to increases in agricultural revenue; by year seven we can no longer reject that treatment and control means are equal.

Are spillovers biasing the results?

Our randomization was at the household level; control households reside in the same tabias as treated households. Moreover, given the public nature of the lottery, control households would have been aware of which individuals received the program. Our identification strategy rests on the assumption that control households represent the counterfactual outcome for treatment households in the absence of the program, an assumption violated in the presence of spillovers.

Other studies of capital transfers offer evidence of possible margins through which spillovers might operate. Egger et al. (2022) shows that cash transfers in western Kenya equal to 15% of annual village GDP generated local multiplier effects of approximately 2.5 in treated villages. In rural Mexico, Angelucci and De Giorgi (2009) documents that the cash transfer program PROGRESA led to increased ineligible household consumption (driven by increased lending and transfers to these households) and reduced precautionary savings (driven by improved risk-sharing networks).

[8] Inflation-adjusted

The evidence of spillovers from graduation programs is more scarce. In three of the six sites in Banerjee et al (2015) (Ghana, Honduras, Peru), a two-stage randomization allowed for a direct test of spillovers—there is some evidence of negative spillovers on mental health, but overall evidence of spillovers is limited. In Bangladesh, Bandiera et al., (2017) examines the effect of the graduation program on ineligible households. They find that the graduation program led to increased wages paid for low-skill work (for maids and in agriculture, due to reduced supply as treated households shifted away from this work), but find limited other evidence of spillovers.

We examine spillovers and general equilibrium effects in multiple ways in order to explore whether they may explain the dissipating treatment effect. First, we consider the magnitude of the transfers, and conclude they are small relative to the local economy and livestock markets. There are approximately 55,000 individuals in our study tabias, and approximately 11,000 households. 458 households (approximately 4%) received the program. Asset transfers were equal to approximately $1,371 USD PPP, or 50% of the annual consumption value of these households. If we assume that sample household consumption is representative of the community at large (a conservative assumption, given that eligibility in the program was determined on the basis of poverty), asset transfers would have represented approximately 2% of year two intervention area tabia GDP. This transfer size is therefore small as compared to the consumption growth experienced by control households, whose consumption grew by $1,087, or 33%, between years two and seven. For spillovers to be a key driver of our results, they would need to operate through a more concentrated mechanism than a margin like the broad multiplier effects observed by Egger et al. (2022). Similarly, the livestock transferred represents a modest increase in the total supply of animals in the economy—we estimate no more than a 4.3% increase in the value of livestock.[9]

Next, we examine the limited data we have on gifts and loans between households. Although not an exhaustive survey module, we find little support for a change in gifts and loans across treatment and control households. Appendix Table A8 shows a statistically significant relative decline in the likelihood of receiving remittances for treatment versus control households in year three, but the amount received shows no significant change, and the volume of informal borrowing moves the other way.

We also examine participation in government programs. Although we find some evidence that the government food-for-work program is taken up differentially more by control households in years three and seven, we cannot reject the null that the groups have (i) equal labor supply in the program or (ii) equal benefits received. Our (not statistically significant) point estimates suggest that the growing gap in PSNP benefits can explain roughly 15% of the decline in per capita consumption treatment effects between years three and seven.

Last, we examine spillovers by examining control households more physically proximate to treatment households compared to those more distant. We lack data on social networks to do a similar analysis for social proximity. Appendix Tables 10-14 presents the effect on control households of living within 200 meters of a treatment household, once controlling for all neighbors within 200 meters. We do not find evidence that control households benefit (in terms of resource-sharing, consumption, or income) from spatial proximity to treatment households. In

[9] A fuller description of our calculations in this section is reported in Appendix C.

contrast, the most reliable effect we observe is that control households proximate to treated households have *lower* asset wealth, driven by fewer livestock purchases, and *less* total livestock ownership. One possible mechanism here is that finite pasture for grazing made livestock-rearing less optimal for these households.

Thus, spillovers may actually be causing an overestimate, rather than explaining the dissipation. To examine this, we return to our core treatment effect analysis and estimate:

$$(2)\ y_{ht} = \alpha + \theta\ treat_h + \beta\ neighbors_{h0} + \gamma\ neighbors\_treated_{h0} + \delta\ y_{ho} + \mu_v + \epsilon_{ht}$$

in Appendix Tables 15-19, for the full sample (where $neighbors_{h0}$ is the number of neighbors within 200m, and $neighbors\_treated_{h0}$ the number within 200m who received the program). Reassuringly, we find that our estimated treatment effects are similar when controlling for proximity to treatment households among our full sample.

Cost-Benefit Analysis

While the benefits of the graduation program to treatment households were sizable, intervention costs are also substantial: in 2021 USD PPP terms, the total cost of the program was $4,011 per participant, reported in Panel A of Appendix Table A5. Of this, $1,371 (34%) constituted a direct transfer to the household in the form of livestock or business inputs; the other 66% included a combination of staff salaries, training, supervision costs, and indirect expenses. Of the six sites reported in Banerjee et al. (2015), Ethiopia is the third-cheapest, albeit three times more expensive than in India, the cheapest of the six. While in practice, some of the costs were incurred over the two years the training and coaching continued, we assume that all costs were incurred immediately at time zero when discounting.[10]

We calculate benefits on the basis of per capita consumption, (from Appendix Table A5). We assume that the full benefit to households is equal to the sum of all consumption treatment effects, and make three key additional assumptions. First, we assume that the impacts on consumption at year one are equal to those in year two. Second, we assume that the decline in consumption treatment effects from year three to year seven followed a linear decline. Third, we assume that this linear rate of decline from year three to seven continues up to the final year at which treatment effects would be greater than zero under this decline (year 16). We annualize these consumption benefits, and apply a 5% annual discount rate, following the social discount rate used by the World Bank and International Monetary Fund (International Monetary Fund 2013).

Our estimated benefits are restricted to consumption effects. We are therefore neglecting any insurance benefit associated with increased wealth (including via the directly transferred assets), and any non-pecuniary benefits, for example, through feeling more politically empowered.

[10] Assuming half of non-transfer costs are incurred in year two causes our (present-discounted) estimate of program costs to fall by 1.5%.

Under these assumptions, we find that the total costs and total benefits are very similar—total benefits to the household equal 97% of costs. This result depends on the fact that while program costs are incurred immediately, benefits are gradual and extend over a long-time horizon. As shown in Figure 5, and Panel C of Appendix Table A20, with no discount rate applied, the total projected benefits are well in excess of program costs. Benefits equal costs at a social discount rate of 4.4%.

We quantify the uncertainty of our benefit-cost ratio estimates using bootstrap simulations. For each of 10,000 repetitions, we construct a bootstrap sample, re-estimate consumption treatment effects on this sample, and calculate our benefit-cost ratio on these estimated values. (This procedure is further described in Appendix D). We find that benefits exceed costs in 52% of our simulations; our 95% confidence interval of the benefit cost-ratio is [0.42, 2.36]. This sizable uncertainty about the relative ratio of costs and benefits emphasizes the further importance of accumulating additional evidence on the effectiveness and costs of these programs.

Comparison with India

Our findings diverge from those in other longer-run studies of graduation studies, including Banerjee et al., (2021), (henceforth "BDS"), the study most similar to ours. In BDS, very poor households received either cattle, goats, or business inputs, and other support—both our site and BDS' were part of the same six-site pooled study of graduation programs (Banerjee et al., 2015). We show in Appendix Tables A4-A6 that BDS and our setting have similar trajectories in the short run, with strong treatment effects on consumption, income, and wealth at 18 months/two years and three years. In contrast, BDS find that treatment effects increased between years three and seven and persisted through year ten, driven in part by income diversification into migration. Of note, as in Ethiopia, control living group standards were also rising meaningfully between years three and seven in India—growing treatment effects took place in spite of this.

While there are many possible explanations as to why effects might differ between the two sites, there are at least two substantive differences worth highlighting. First, BDS hypothesize that a key mechanism through which treatment households compounded their gains in India is via migration—remittances and migration grew strongly by year seven for both groups; by year ten, treatment households were sending more remittances, and migrating to more distant locations for longer periods of time. In contrast, this mechanism is not a key driver of differences in Ethiopia. Migration out of Tigray, Ethiopia is very low, and we do not observe positive treatment effects on remittances. Second, a major difference between the two sites is the degree of take-up. We have 100% take-up in our study, while in BDS, only 52% of households took the offer (their estimates are intention-to-treat). It is possible that our fading treatment effects are partially driven by negatively selected households with respect to returns to capital (who plausibly may have been among the share most likely to decline the offer in India).

IV. Conclusion

In Ethiopia, a transfer of productive assets, coupled with training, coaching and access to savings accounts induced large treatment effects on consumption, income and wealth two and three years later, and led to positive but diminished treatment effects seven years after the initial transfer.

These fading treatment effects took place in a context in which control household living standards were improving consistently. We do not find evidence of dissipation of any of the previously accrued absolute improvements for treatment households. Consumption and related welfare measures continued to grow for both treatment and control households throughout the study period, with the gap across groups closing over time.

Our results are largely inconsistent with the hypothesis that the program unlocked a poverty trap for poor households in Ethiopia, given the high rate of growth on the part of control households between years three and seven. This result therefore stands in contrast to what has been observed in India and Bangladesh, in which results are consistent with, albeit perhaps not dispositive of, the hypothesis that the programs pushed individuals over some critical threshold.

This difference also matters from a policy perspective, as NGOs and governments consider the scale-up of graduation programs. In a public finance sense, the question is whether from a cost-effectiveness perspective they should be compared to short-run fixes of humanitarian poverty-based crises or compared to attempts to address long-run income and inclusive market development approaches. In a practical sense, this has implications for the number of years over which one would reasonably project benefits in a cost-benefit analysis. Of the six graduation pilots studied in Banerjee et al. (2015), for none did the discounted benefits from the first three years exceed the costs. The five that did have benefits in excess of costs only did so with assumptions that the benefits continued past the three years, which seemed reasonable to assume some continuation given the similar results at two and three years. The predicted continued benefits have been realized in India, but have been positive but smaller in Ethiopia. The fading of results here, compared to the persistence elsewhere, suggest continued long-run study of these programs is essential in order to understand how and when such results persist.

Moreover, the results in India are important because they suggest the possibility of large treatment effects with a much cheaper version of the program. In India, the direct transfers to households equaled roughly half of total program costs, compared to roughly a third in Ethiopia. Work on a graduation program in Ghana has found that neither simply providing access to savings, nor providing a one-off transfer of goats with no additional training, is sufficient to achieve the same impacts as those observed with the full graduation program (Banerjee et al. 2022), suggesting that the solution is not as straightforward as simply eliminating core components. We view the question of how to either pare down the cost of these programs (perhaps through reduced costs associated with scale up), or how to increase the effectiveness (e.g., by strengthening the socioemotional skills of individuals receiving the assets, as in Barker et al. (2022) and Bossuroy et al. (2022)), as important areas for future study.

The divergence between India and Ethiopia also speaks to the value of studying the long-run impacts of public policies and programs, especially those whose cost-benefit calculus depends on benefits persisting in the long run. These two studies offer proof-of-concept for the idea that similar short-term trajectories do not necessarily imply similar trajectories in the longer-run.

Coda: Civil Conflict in Tigray

As in any empirical study, the effects of programs depend on how they interact with the external conditions, including both positive and negative shocks (Rosenzweig and Udry, 2020). Unfortunately, in the time since our most recent data collection, our study area experienced a severe, negative shock in the form of violent conflict and displacement.

In the time since the end of the Civil War in 1992 through 2018, the Tigray People's Liberation Front (TPLF), the political party sharing an ethnicity with our sample, had been a central power in Ethiopia's ruling coalition. This changed in 2018, when a new ruling coalition was formed that excluded the TPLF from power, leading to tensions between the TPLF and the newly ruling Prosperity Party. These tensions escalated to violent conflict in late 2020 (year ten of our study), with the proximate cause of disputes over the legitimacy of Tigrayan elections held during the COVID pandemic. As of October 2022, this conflict has led to an estimated 385,000-600,000 deaths and more than 2 million displacements (York, 2022); reporting suggests the study area was heavily affected.

As of November 2022, a peace agreement was signed between the two parties; while conflict had not fully ceased, the parties agreed to a coordinated disarmament.

Thus, the extent to which treatment effects persisted or faded in our study site will depend on how the program interacted with this violent conflict and its immediate resolution, rather than with a context of macroeconomic growth and stability, as was the case from baseline through our three follow-ups. Previous work finds positive impacts of the graduation program in the face of conflict (Bedoya et al., 2019; Brune et al., 2022 report strong impacts in Afghanistan and modestly positive impacts in Yemen, respectively). Clearly, meaningful uncertainty remains about the relationship between graduation programs and resilience, and how treatment effects will evolve for participants in this program in Tigray, Ethiopia.

References

Anderson, M.L., 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. J. Am. Stat. Assoc. 103, 1481–1495.

Andrews, C., de Montesquiou, A., Sánchez, I.A., Dutta, P.V., Samaranayake, S., Heisey, J., Clay, T., Chaudhary, S., 2021. The state of economic inclusion report 2021: The potential to scale. World Bank Publications.

Angelucci, M., De Giorgi, G., 2009. Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption? Am. Econ. Rev. 99, 486–508.

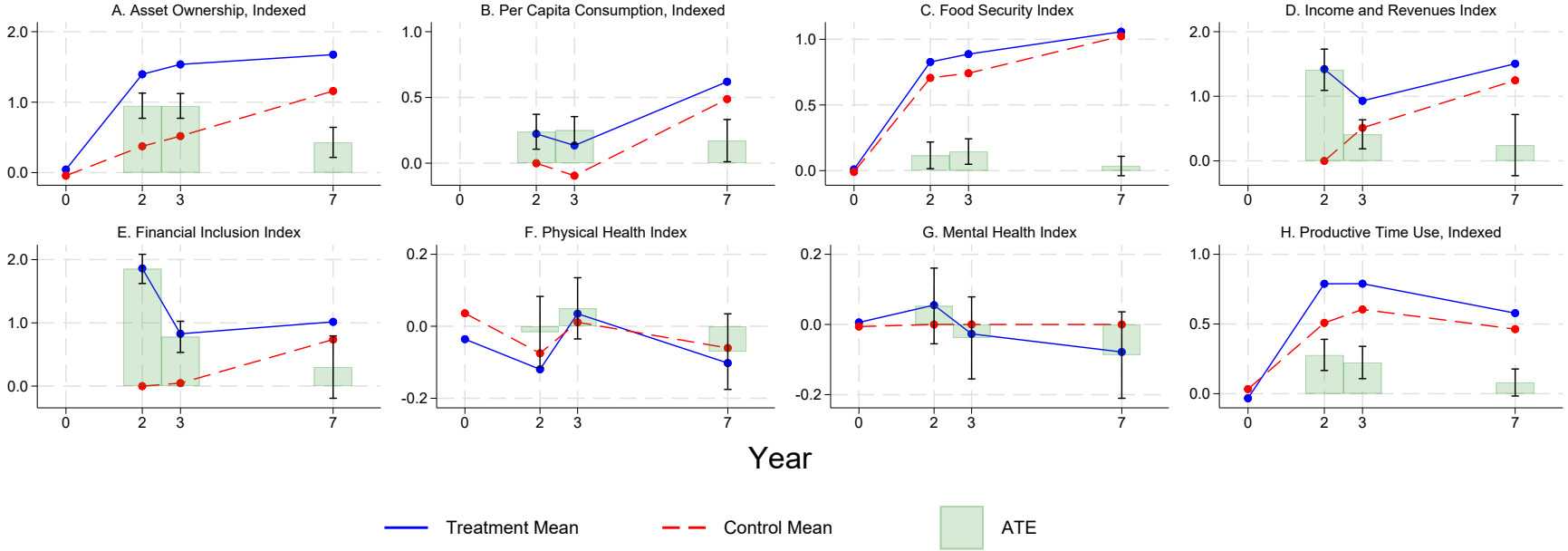
Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., Heil, A., 2022. Why do people stay poor? Q. J. Econ. 137, 785–844.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., Sulaiman, M., 2017. Labor markets and poverty in village economies. Q. J. Econ. 132, 811–870.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., Udry, C., 2015a. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. Science 348.
Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., Udry, C., 2015b. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. Science 348, 1260799.
Banerjee, A., Duflo, E., Sharma, G., 2021. Long-Term Effects of the Targeting the Ultra Poor Program. Am. Econ. Rev. Insights 3, 471–486. https://doi.org/10.1257/aeri.20200667
Banerjee, A., Karlan, D., Osei, R., Trachtman, H., Udry, C., 2022. Unpacking a multi-faceted program to build sustainable income for the very poor. J. Dev. Econ. 155, 102781.
Banerjee, A.V., Newman, A.F., 1993. Occupational choice and the process of development. J. Polit. Econ. 101, 274–298.
Barker, N., Bryan, G., Karlan, D., Ofori-Atta, A., Udry, C., 2022. Cognitive behavioral therapy among Ghana's rural poor is effective regardless of baseline mental distress. Am. Econ. Rev. Insights.
Bedoya, G., Coville, A., Haushofer, J., Isaqzadeh, M., Shapiro, J.P., 2019. No household left behind: Afghanistan targeting the ultra poor impact evaluation. National Bureau of Economic Research.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.
Blattman, C., Fiala, N., Martinez, S., 2020. The long-term impacts of grants on poverty: Nine-year evidence from Uganda's youth opportunities program. Am. Econ. Rev. Insights 2, 287–304.
Blattman, C., Fiala, N., Martinez, S., 2014. Generating skilled self-employment in developing countries: Experimental evidence from Uganda. Q. J. Econ. 129, 697–752.
Blattman, C., Green, E.P., Jamison, J., Lehmann, M.C., Annan, J., 2016. The returns to microenterprise support among the ultrapoor: A field experiment in postwar Uganda. Am. Econ. J. Appl. Econ. 8, 35–64.
Bossuroy, T., Goldstein, M., Karimou, B., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C.C., Udry, C., Vaillant, J., Wright, Kelsey, 2022. Tackling psychosocial and capital constraints to alleviate poverty. Nature 605, 291–297.
Brune, L., Karlan, D., Kurdi, S., Udry, C., 2022. Social protection amidst social upheaval: Examining the impact of a multi-faceted program for ultra-poor households in Yemen. J. Dev. Econ. 155, 102780.
Buera, F.J., Kaboski, J.P., Shin, Y., 2015. Entrepreneurship and financial frictions: A macrodevelopment perspective. economics 7, 409–436.
De Mel, S., McKenzie, D., Woodruff, C., 2008. Returns to capital in microenterprises: evidence from a field experiment. Q. J. Econ. 123, 1329–1372.
Egger, D., Haushofer, J., Miguel, E., Niehaus, P., Walker, M., 2022. General equilibrium effects of cash transfers: experimental evidence from Kenya. Econometrica 90, 2603–2643.
Galor, O., Zeira, J., 1993. Income distribution and macroeconomics. Rev. Econ. Stud. 60, 35–52.
Kling, J.R., Liebman, J.B., Katz, L.F., 2007. Experimental analysis of neighborhood effects. Econometrica 75, 83–119.
Rosenzweig, M.R., Udry, C., 2020. External validity in a stochastic world: Evidence from low-income countries. Rev. Econ. Stud. 87, 343–381.
Udry, C., Anagol, S., 2006. The return to capital in Ghana. Am. Econ. Rev. 96, 388–393.

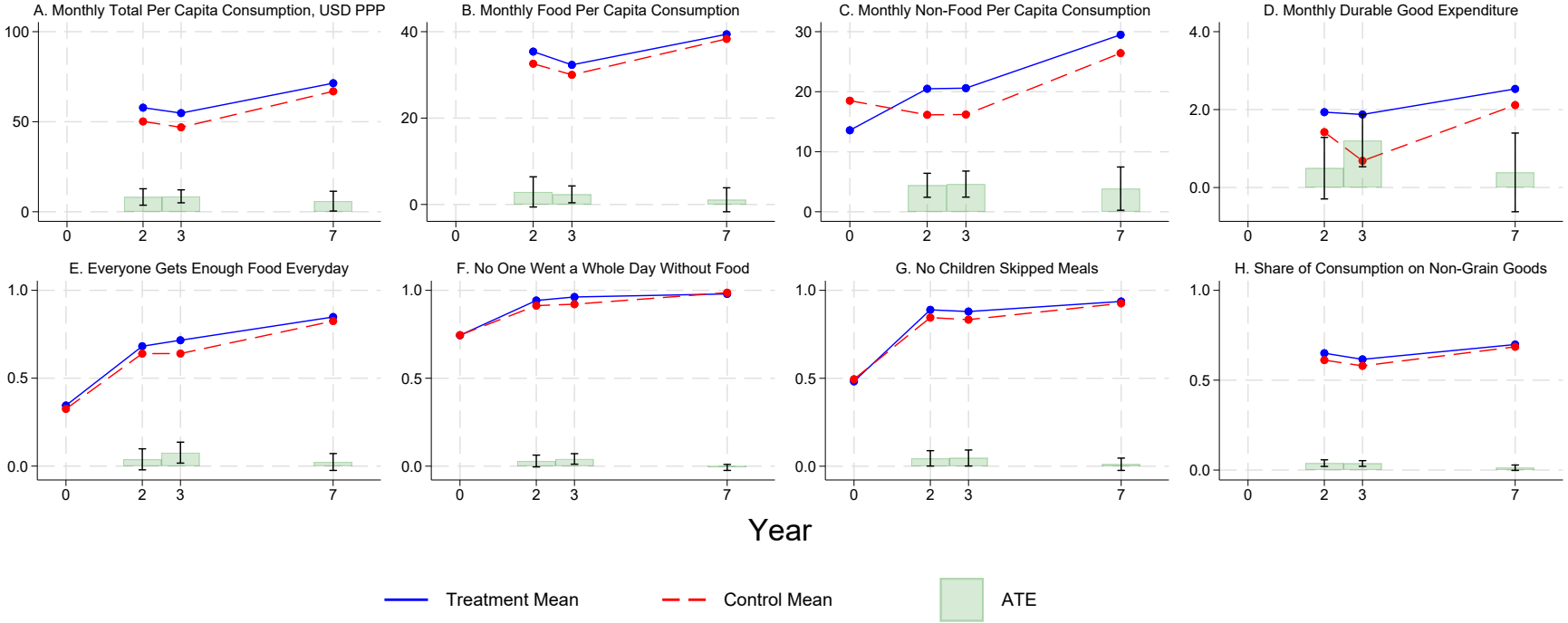
York, G., 2022. Surge of dehumanizing hate speech points to mounting risk of mass atrocities in northern Ethiopia, experts say. Globe Mail.

Figure 1: Indexed Family Outcomes



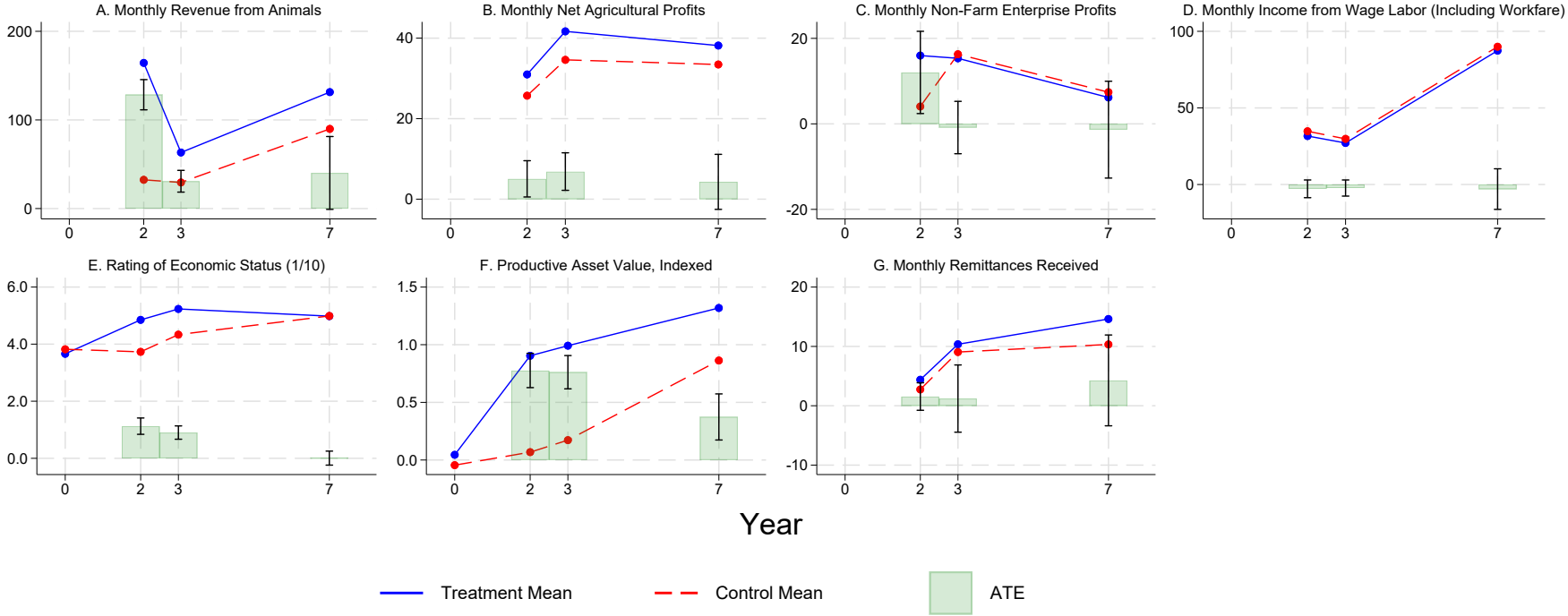
Notes: Reported outcomes are indices, such that the reference group has mean zero and standard deviation one. Outcomes are standardized against the baseline sample when possible, or against the year two control group when comparable data does not exist at baseline. Three outcomes are aggregates that have been indexed (asset ownership, per capita consumption, minutes working), the others are z-score indices of multiplecomponent outcomes. The full construction of each variable is described in Appendix B. The lines are equal to (unadjusted) means for treatment and control households; the bars are estimated treatment effects from equation (1), which controls for baseline values, where available, and tabia-level strata fixed effects. Baseline values are plotted when comparable data is available. The error bars report 95% confidence intervals. Standard errors are clustered at the household level. That is, household-level outcomes (asset wealth, consumption, food security, income and revenues, and financial inclusion) have a cluster size of one; adult-level outcomes (physical health, mental health and time use), where we survey 1-2 members per households, share a cluster.

Figure 2: Consumption and Food Security



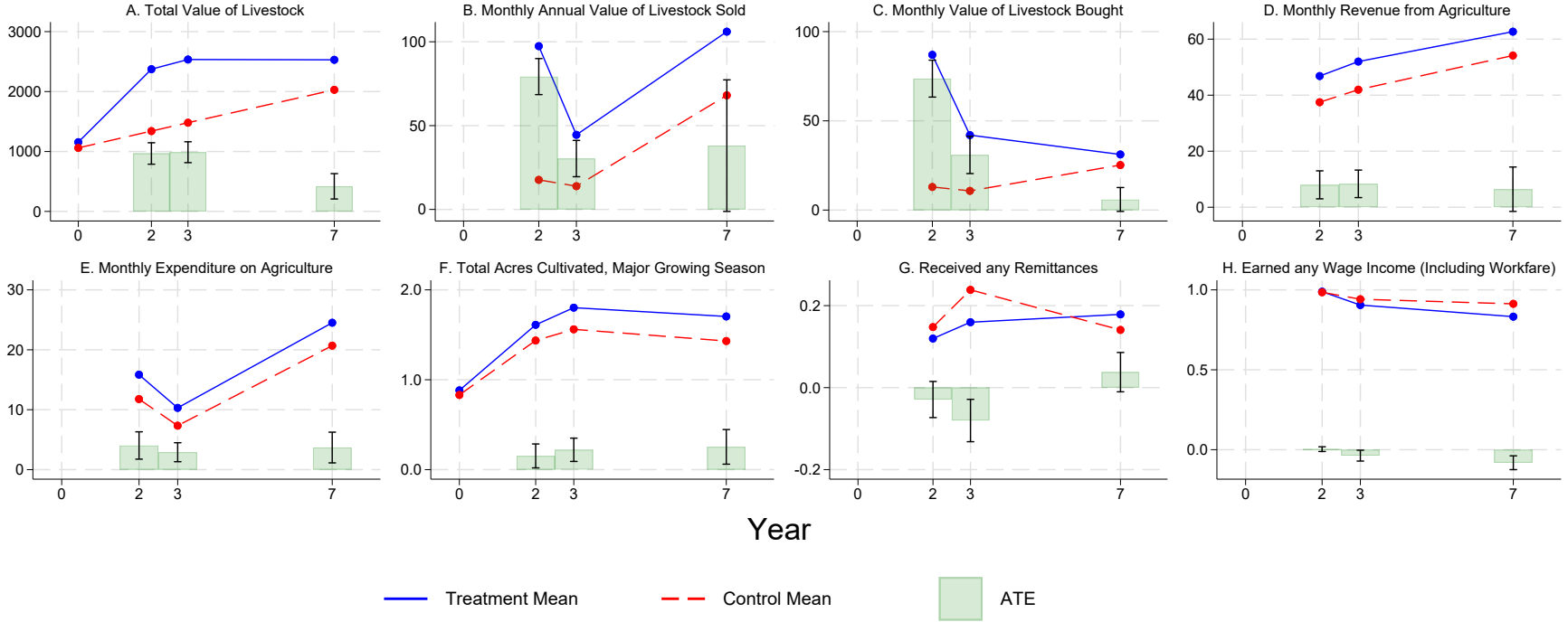
Notes: All financial variables are reported in 2021 USD in Purchasing Power Parity (PPP) terms, using a local price index described in Appendix A. The full construction of each variable is described in Appendix B. The lines are equal to (unadjusted) means for treatment and control households; the bars are estimated treatment effects from equation (1), which controls for baseline values, where available, and tabia-level strata fixed effects. Baseline values are plotted when comparable data is available. The error bars report 95% confidence intervals. Standard errors are Huber-White heteroskedastic.

Figure 3: Income Aggregates



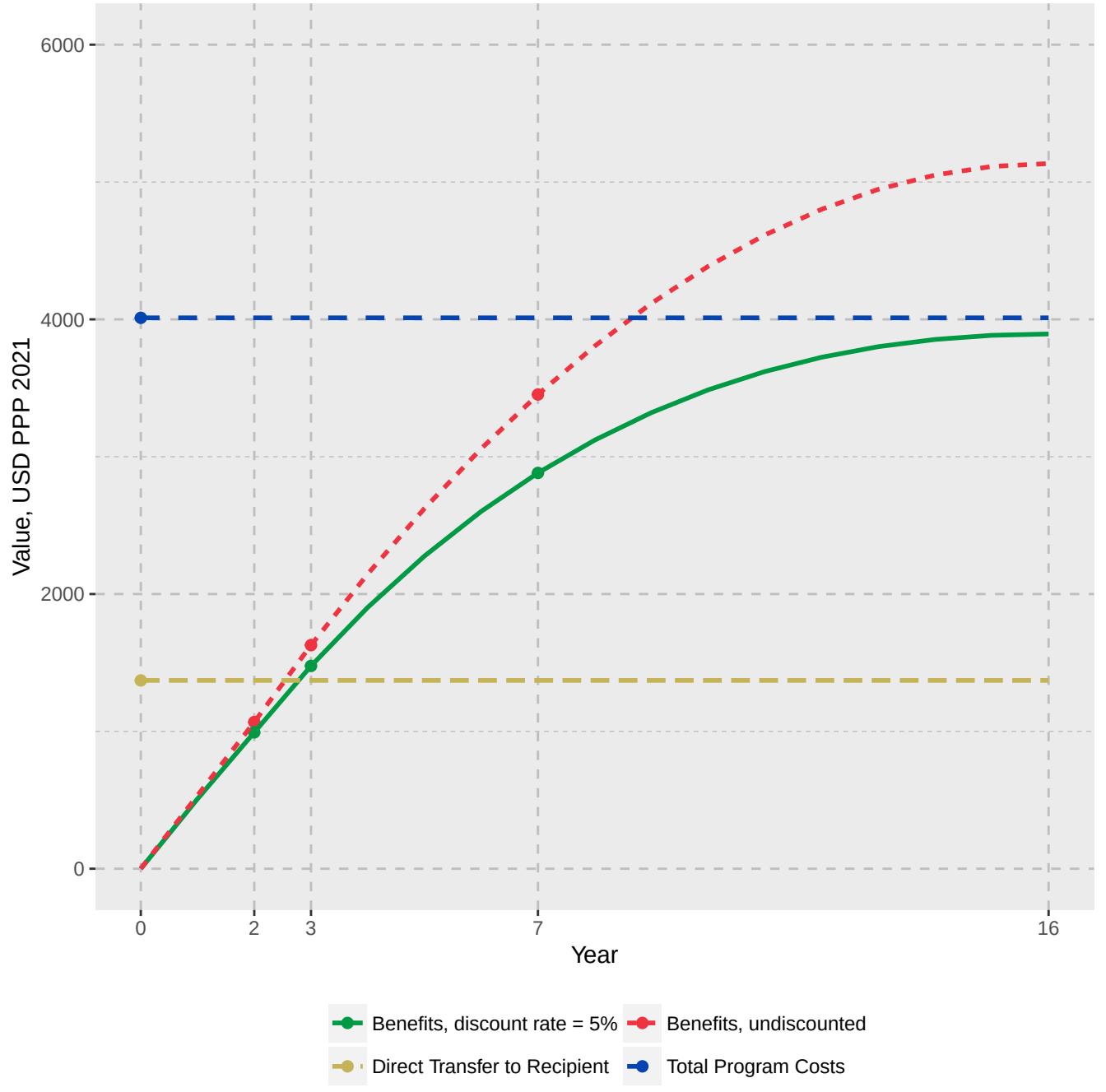
Notes: All financial variables are reported in 2021 USD in Purchasing Power Parity (PPP) terms, using a local price index described in Appendix A. The full construction of each variable is described in Appendix B. The lines are equal to (unadjusted) means for treatment and control households; the bars are estimated treatment effects from equation (1), which controls for baseline values, where available, and tabia-level strata fixed effects. Baseline values are plotted when comparable data is available. The error bars report 95% confidence intervals. Standard errors are Huber-White heteroskedastic.

Figure 4: Income Mechanisms



Notes: All financial variables are reported in 2021 USD in Purchasing Power Parity (PPP) terms, using a local price index described in Appendix A. The full construction of each variable is described in Appendix B. The lines are equal to (unadjusted) means for treatment and control households; the bars are estimated treatment effects from equation (1), which controls for baseline values, where available, and tabia-level strata fixed effects. Baseline values are plotted when comparable data is available. The error bars report 95% confidence intervals. Standard errors are Huber-White heteroskedastic.

Figure 5: Cost Benefit Trajectory



Notes: Figure 5 compares program costs and benefits. Program costs are directly reported by the implementing partner and scaled to 2021 USD in Purchasing Power Parity Terms, with a full breakdown given in Appendix Table A20. The Direct Transfer to the Recipient is the value of the asset (livestock, petty trade inputs, or beehives and bee colonies). Program benefits are calculated as monthly per capita consumption treatment effects times household size times 12. The dots represent the data points we use to estimate the trajectory (we assume a linear decay between years three and seven, and that the same trend continues until benefits reach zero). The green solid line assumes a 5% annual discount rate; the red dotted line does not apply a discount rate.

Online Appendix

The Fading Treatment Effects of a Multi-Faceted Asset-Transfer Program in Ethiopia

Nathan Barker, Dean Karlan, Christopher Udry, and Kelsey Wright

Contents:

A: Price Conversions
B: Variable Construction
C: Spillover Calculations and Estimation
D: Appendix D: Cost-Benefit Simulations
E: Appendix Tables
F: Appendix Figures

Appendix A: Price Conversions

We convert prices across waves using a local price index, based on the prices that respondents report spending on goods of food in the consumption module of the household surveys.

Specifically, we build our price index using all good-unit pairs (e.g., beer bottles of cooking oil) purchased by at least five households in each of the four waves. This restriction leaves us with 22 good-unit pairs. For each good-unit pair, we calculate the median purchase price per unit (which we use as our price in the price index), and the mean amount purchased (which we use as our quantity).[1] With these values, we construct a single Fisher price index for each wave. We estimate price indices of 151, 164 and 186 for our two-, three- and seven-year surveys, respectively (with our baseline survey standardized to 100).

As a comparison, the Consumer Price Index in Ethiopia for the three follow-up years (standardizing 2010 to 100), are 156, 168, and 240, respectively. These numbers suggest that national inflation is in line with locally observed inflation in our two- and three-year surveys, but that national inflation is substantially higher by 2017.

We first use our constructed indices to convert all values from ETB in the survey year to 2010 ETB. We then convert these values to 2021 USD PPP by multiplying by 4.28, the 2010 ETB-2010 USD PPP exchange rate (per the World Bank). Finally, we multiply this value by (271.0/218.1), the ratio of the US consumer price index in 2021 over the US consumer price index in 2010.

The divergence between local and national price indices matters for the interpretation of treatment and control group trends for financial outcomes by year seven (albeit not the ratio of treatment to control group means in each wave). A higher CPI by year seven (as implied by national prices) implies lower living standards for both treatment and control households than estimated in our main results.

In assessing the relative appropriateness of local prices versus the national CPI to adjust for prices, our evaluation is that using local prices leads to a more accurate measure than do national prices. There are three reasons for this assessment.

First, there are two measures that we are able to measure especially well in real terms: livestock ownership and asset wealth. For these two measures, we can therefore compare the trajectory of these measures in real terms for treatment and control households against estimates using local and national prices, and assess the relative performance of these two price indices against the real measures.

For livestock ownership, we measure the total stock held by individuals using Tropical Livestock Units, a measure that imposes a relative value of animals based on their metabolic weight (which in turn closely correlates with their value in livestock markets. For example, a cattle is valued at 0.7 TLU, a goat at 0.1). This measure therefore captures the evolution of livestock wealth in real

[1] We use medians for prices, to limit the potential influence of outliers. However, we use means rather than medians for quantities, because for most of these goods, the median amount purchased is equal to zero.

(metabolic weight-based) terms, and offers a chance to interrogate whether local food prices or the national price index better align with this measure. We present the TLU evolution for treatment and control households in Panel A of Appendix Figure 1, the same evolution in local price terms in Panel B, and with the national CPI in Panel C. Panels A and B show very strong agreement regarding how livestock ownership has evolved over the study years, while Panel C in contrast would imply livestock wealth has significantly declined.

Similarly, for (non-livestock) asset ownership, we observe the actual quantities of each asset type owned by each individual, and their estimate of its value if they were to sell it today. We therefore create a single price for each good (taking median prices across waves), and multiply the number of each good a given household owns by the (cross-wave) median valuation. This approach applies the same unit price of each good across waves (e.g., we assign all beds the value of $153.91 in every wave, and every bicycle a value of $182.77). If anything, this "real value" might understate gains in wealth over time, because it doesn't allow for quality improvements (despite the fact that we might expect e.g., one mobile phone in 2017 to be more valuable than one mobile phone in 2010), and thus potentially advantages the national CPI when validating the two measures.

We plot the growth of "Real Asset Wealth" in Panel A of Appendix Figure 2, and our estimates of asset wealth using local and national prices in Panels B and C (all values are standardized such that the two-year mean is equal to 1). Here, the pattern is less striking than for livestock, albeit still more consistent with local prices better capturing real wealth than national prices do. For control households, the trajectory for real wealth and wealth using local prices is very similar. For treatment households, the real wealth is directionally similar, although treatment wealth in year seven grows faster when measured with local prices than with constant prices. Again, national CPI implies a very different trajectory from what is found with real wealth—it suggests declining wealth in years three and seven for both groups, steeply so for treatment households. This comparison again suggests to us that local prices better correspond to real living standards than the national CPI does.

Second, on non-price-based measures of food security-based wellbeing, we observe results consistent with household living standards increasing rather than decreasing between years three and seven. On our food security index, control households have improved by 0.28 standard deviations (measured in terms of baseline variance of food security) between years three and seven. For example, while 17% of control households reported that children skipped meals due to a lack of food in the last 12 months at the three-year mark, that number had fallen to 7% by year seven. Similarly, the total share of consumption allocated to the consumption of grains—a measure of proximity to subsistence (Subramanian and Deaton 1996)—fell from 42% to 31% between years three and seven. Treatment households also experienced gains on these measures between years three and seven: treatment food security rose by 0.17 standard deviations, the share of households with children who skipped meals fell from 12% to 6%, and the share of expenditure on grains fell from 38% to 30%.

In Appendix Figure 3, we plot our food security index in Panel A, the share of consumption on non-grain goods in Panel B, per capita consumption using our local price index in Panel C, and per capita consumption using the national CPI in Panel D. Again, the use of the local price index

seems to capture the evolution of the real outcomes well, while the national prices suggest a meaningful deviation from the real outcomes in year 7.

Finally, food itself represents a sizable share of total consumption by households. For control households, food consumption represents 65%, 64% and 57% of total consumption at years two, three and seven, respectively. These numbers thus offer a de facto bound of sorts on the degree to which local prices are failing to capture the basket of prices faced by these poor households—it seems very likely that the prices *actually paid* by these individuals for food is a better measure to them of the cost of food than a measure that uses the national price index, and which suggests a very different evolution of living standards.

These three assessments thus lead us to believe that local prices better capture the living standards of households in our sample than do national prices, and thus that our local price index is appropriate.

For completeness, for key financial outcomes (asset wealth, per capita consumption, livestock revenues, agricultural profits, livestock purchases, livestock sales, and financial savings), we also report estimates using the national CPI in Appendix Table A27, to allow readers to assess how our patterns differ when using the national CPI.

For our comparisons of the results in Ethiopia to those in India in Appendix Tables A4 to A6, we are unable to build a similar price index across waves in India. In the household survey data in India, very few expenditures were measured in terms of physical quantities—respondents were instead asked to aggregate and report the value of consumption goods at the group level (e.g., "total value of cereals and cereal products consumed in the last 30 days"). Given this, we instead take the values in BDS (reported in 2017 USD PPP) and multiply by the ratio of the US Consumer Price Index in 2021 over the US Consumer Price Index in 2017.

Appendix B: Variable Construction

This section details the construction of the variables used in the paper.

i. Outcomes forming part of Indexed Family Outcomes

Figure 1 and Appendix Table A4 report means and treatment effects of indexed family outcomes. Eight outcomes are reported in Figure 1; these eight plus the additional two indices (that we did not present in the figure, due to space constraints) are reported in Appendix Table A4.

Indexed family outcomes are listed in bold.

Asset Ownership is equal to the aggregate value of all assets owned by the household. This includes livestock, other productive assets (e.g. farm tools, sewing machines), and durable goods (such as phones, televisions, and furniture). In the two-, three-, and seven-year surveys, individuals are asked to estimate the price they could receive if they sold the item today. We sum up the value across all of these assets. At baseline, individuals report their ownership of goods, but do not estimate the value. We therefore recover estimates of their baseline value by taking the (inflation-adjusted) median value of each asset type across the three waves with data and multiplying this median price by quantities owned.

In Figure 1 and Appendix Table A4, we standardize the values, such that the baseline sample has mean 0 and standard deviation 1. In Appendix Table A6, we report the value (and its component parts) in USD PPP.

Our measure of **per capita consumption** is the aggregate of: (1) the value of all food consumed, equal to the quantities consumed multiplied by the median purchase price in the sample (asked about in the last month), (2) other expenditures incurred, such as school fees, soap, and home repairs (asked about in both the last month and last year, depending on the good in question), and (3) the total purchase value of durable goods in the last 12 months. All values are scaled to be monthly values.

Our baseline measure of consumption differs in two ways from post-intervention measures. At baseline, individuals were asked to estimate both (a) the value of food consumed from home production, and (b) total purchases of the food in the last 30 days, whereas in our follow-up surveys, they are just asked about total consumption of each food type (and what amounts were purchased). Also, at baseline, we do not ask about durable good purchases. We therefore aggregate across (a) food consumption, and (b) non-food expenditure to produce a baseline estimate that we control for in our baseline specifications (both total consumption, and the food aggregate for food consumption), but do not include the values in the figures because they are not directly comparable.

The index in Figure 1 and Appendix Table A4 is the aggregate value of consumption, standardized against the year two control mean. The value in USD PPP is shown in Figure 2 and Appendix Table A5, as are its component parts.

The **food security index** is a z-score aggregate of the following five measures: everyone gets enough food every day, no adults skipped meals, no household member went a whole day without food, no children skipped meals, everyone eats at least two meals every day. We follow BDS and report a subset of these outcomes in Figure 2; the full set is reported in Appendix Table A22. It is standardized against the baseline sample.

The **income and revenues index** is a z-score index with the following measures: livestock revenues: equal to revenues from livestock sales and the value of livestock products, such as eggs, whey or honey, agricultural profits: equal to yields (using sample median sale prices in cases when individuals did not sell their crops) and rental income minus expenses, not including family labor, microenterprise profits: equal to revenues minus costs, again not counting family labor, wage earnings (including via the Productive Safety Net Programme), and economic self-rating, on a scale of 1/10. Component parts are reported in Figure 3 and Appendix Table A6. All income measures in Appendix Table A6 are scaled to be in monthly terms.

The measures collected at baseline were not directly comparable. At baseline, individuals gave a single estimate for the value of livestock sales, whereas in the follow up surveys sales were asked about for each animal. For agricultural income, individuals were asked about harvests and expenses in the last twelve months at baseline, and in the last harvest in the follow-up surveys (there are two harvests in this part of Ethiopia, the follow-ups all asked about the major harvest). For non-farm enterprise income, at baseline, individuals were asked about specific types of products/services, the sale price, and total sales for each product. In follow-up surveys, individuals were instead asked to report total sales and expenses for months with high, medium, and low sales, and to report how sales were (on a low/medium/high scale) in each of the last 12 months.

We therefore include these measures (and their indexed value) as baseline controls, but standardize against the year two control group, and do not report these (not directly comparable) measures in our figures and tables.

The **financial inclusion** index is a z-score index, with the following measures: amount borrowed in the last 12 months from formal sources, amount borrowed in the last 12 months from informal sources, savings balance, and savings deposits in the last three months. It is standardized against the year two control mean, given that savings questions were not asked about at baseline. Individual outcomes are reported in Appendix Table A23.

The **productive time use index** is an aggregate of minutes spent working per day on average in the last 48 hours (individuals report their time use in each of the last two days). Productive activities include time spent working on agriculture, tending to livestock, managing a business, and working for a wage. This value is standardized against the baseline sample, and reported as an index in Figure 1 and Appendix Table A4. Component parts in minutes are reported in Appendix Table A24.

The **physical health index** is a z-score index that includes: no days missed due to poor physical health in the last month, a mean score of activities of daily living (ability to lift a 10 kg bag,

ability to walk for four hours without resting, ability to work in a field all day), and a 1/5 rating of an individual's physical health. This index is standardized against the baseline sample, with component parts reported in columns 1-3 of Appendix Table A25.

The **mental health index** is a z-score index that includes: a 1/5 overall life satisfaction rating, not having an extended period of time with worry in the last 12 months, and a stress index (with component parts including how often the individual felt sad, cried a lot, didn't feel like eating, didn't feel like doing their work, and had restless sleep). This index is standardized against the year two control sample (given the stress index methodology was different at baseline); component parts are reported in columns 4-6 of Appendix Table A25.

The **political empowerment index** is a z-score index that includes: whether the individual attended a meeting with a local leader of politician, whether they asked a question of a leader or politician at a local meeting, and whether they are a member of a political party. The index is standardized against the baseline sample and reported in Appendix Table A4; component parts are reported in columns 1-3 of Appendix Table A26.

The **women's empowerment index** is a z-score index that includes whether or not adult women in the household report having a major say in: food-related spending, education-related spending, healthcare-related spending, home improvement spending, and household management decisions. The index is standardized against the baseline sample and reported in Appendix Table A4; component parts are reported in columns 4-8 of Appendix Table A26.

ii. Other Outcome Variables

Figure 2 and Appendix Table A5: Consumption and Food Security

The share of total consumption on non-grain items uses total consumption minus the value of grain consumption (calculated using the total quantity consumed times median prices) as the numerator, and total consumption as the denominator. Because the method of measuring consumption differed at baseline, we include the share of consumption on non-grains as a control, but do not report it in the Figure/Table.

Figure 3 and Appendix Table A6

The productive asset value is equal to the sum of the reported value of livestock, and non-livestock productive assets (e.g., farm tools). For baseline productive asset ownership, we use the median (inflation-adjusted) price from the three follow-up waves to estimate the value (we observe quantities only). It is standardized such that the baseline mean is equal to 0, with standard deviation 1.

Remittances received is equal to the sum of all remittances that individuals reported receiving in the past 12 months, scaled to a monthly value. (This variable is also reported in Appendix Tables A8, A14, and A19, i.e. our tables on resource-sharing).

Figure 4 and Appendix Table A7

Livestock values is equal to individuals' total estimate of the value of each type of animal owned. For years two, three and seven, individuals reported an estimate. For baseline, we use median (inflation-adjusted) prices from the three follow-up waves to infer value, based on the reported quantity owned.

For total value of livestock sold, and total value of livestock bought, we ask individuals to report how many of each animal type they sold/bought in the last twelve months, and the total money they received/paid for them.

Revenues from agriculture is equal to the total value of individuals' agricultural yields in their last harvest (using median sales prices per crop in cases where individuals did not sell their output) and rents. (As noted above, at baseline, this measure asked about the last 12 months rather than last harvest, so we include the measure as a control, but do not present the measure in our figure and table, as it is not directly comparable).

Expenditure from agriculture is equal to the total amount paid to rent land and agricultural inputs in households' last harvest (e.g., seeds, fertilizer, pesticides). (As noted above, at baseline, this measure asked about the last 12 months rather than last harvest, so we include the measure as a control, but do not present the measure in our figure and table, as it is not directly comparable).

Total acres cultivated is equal to the sum of the size of all agricultural plots that individuals report cultivating. Most units are reported in local units (timads or gemads, the amount of land an ox can plow in a day), we convert to acres using median reported conversion rates from our village survey.

Received any remittances is an indicator variable equal to one if the household received any remittances in the last 12 months, and zero otherwise.

Earned any wage income is an indicator variable equal to one if individuals worked for others for pay (cash or in-kind), including in the Productive Safety Net Programme. We asked about work for others at baseline, but did not specifically ask about PSNP work. We therefore include this measure as a control, but do not report the mean in the figure and table, as it is not directly comparable.

Appendix Tables Only, Not in Indices

Received any informal loans is an indicator variable for whether an individual reported receiving a loan in the last 12 months from a neighbor, friend, shopkeeper, family member (not in the household), moneylender, co-op, or iqub (savings group).

Cash value of informal loans is equal to the sum of the value of all cash loans received by household members from the aforementioned informal sources in the last 12 months.

For each food item in our consumption module, we asked if the household received any of this food type as a gift from others. Received any food as gifts is an indicator variable equal to one if the household received any such gifts, and zero otherwise.

Number of types of food received as gift is equal to the number of distinct types of food individuals report receiving as gifts.

Household enrolled in PSNP is an indicator variable for whether the household reports having any member enrolled in the program.

Days of PSNP in the last month is equal to the sum of all days worked by household members in PSNP in the last month.

Monthly PSNP earnings is equal to the value of cash benefits from PSNP and food benefits (valued at median prices from the consumption food survey) in a typical month, multiplied by the number of months the household reported receiving benefits in the last 12 months, divided by 12.

Appendix C: Spillover Calculations and Estimations

In this section, we expand upon the calculations and methodological steps taken to evaluate the extent to which our data supports the hypothesis that spillovers are partially responsible for the evolution of control group living standards that we observe.

i. Magnitude of transfers relative to local economy

We estimate that approximately 4% of households in the study tabias received the transfers, that the livestock transfers constitute no more than 2% of annual tabia GDP. The transfers are likely equal to an increase in supply of livestock in the communities of no more than 4.3%.

We have two (informal) estimates of total population in the ten tabias where the program operated. First, at year-seven, we conducted a village survey in which asked local leaders to estimate the population of the tabias where they reside. Aggregating their estimates suggests a total population of 55,169 in study tabias. Second, the US Census Bureau estimates annual Ethiopia population at the *woreda* (district) level. They estimate a total population of 111,950 in Kilte Awlaelo (the district in which this study took place) in 2010. Kilte Awlaelo comprises 19 tabia—our study includes ten of these 19. Assuming an equal population across tabias implies a total population in our sample tabias of 58,921—the two methods thus produce similar estimates of the total population.[2]

If we assume an average household size of 5.05 (from our baseline sample), these population estimates would suggest there are a total of 10,925 to 11,668 households in our sample tabias. Because we lack data from ineligible households, we are unable to directly test whether sample household size is greater than or less than the population at large.

As a result of this study, 458 households received capital transfers, equal to 3.9-4.1% of households in these tabias. Direct transfers to the household were equal to $1,371 2021 USD PPP, or 50% of average annual household consumption for control households at year two. We lack data on ineligible households' consumption. Because having low income is an inclusion criterion for being in PSNP (and thus our sample), it is likely that the average household in the tabia has higher consumption than do control households. However, in a conservative case where our sample's consumption is representative of consumption at the tabia level, direct livestock transfers would constitute 2% of local annual GDP in the year they took place.

At year two, the mean control value for livestock ownership is $1,341 2021 USD PPP. Our estimate is therefore that the livestock transfers represent an approximately 102% increase in livestock value for treatment households. Again, it is likely the case that because sample households are identified on the basis of poverty, their livestock ownership is lower than that of the mean household in the community. If we assume however that the households are example representative of the population at large, it would suggest that livestock transfers led to a 4.3% increase in total livestock value in the communities, likely suggesting that price effects on livestock sales are not large.

[2] While crude, this is not an unreasonable assumption. Villages and neighborhoods are designated to tabias in a way that ensures each has a roughly equivalent population.

ii. Value of PSNP relative to consumption treatment effects

We report in our main text that growing differences between treatment and control households in PSNP receipts can explain roughly 15% of the declining per capita consumption treatment effects between years three and seven. This estimate depends on the following calculations:

We estimate in year three that treatment households receive $1.92 less per month in PSNP benefits, and that by year seven, this gap is equal to $4.21. In year three, the mean household size is 5.45, by year seven it is 5.57. If we assume a marginal propensity to consume of one from these benefits, this equates to an additional $0.40 of monthly per capita consumption to control households by year seven. Our decline in treatment effects between years three and seven is equal to $2.67 in per capita consumption, suggesting our point estimate of monthly PSNP benefit declines corresponds to 15% of the total per capita consumption decline.

iii. Spatial Spillovers

We report in the main text that control households do not appear to have received more transfers (in the form of remittances, lending or food donations) as a result of the program. One possibility though is that spillovers (in the form of transfers, or some other mechanism) might operate specifically through social networks, rather than in a diffused way to all control households.

We did not collect data information about (a) self-reported social networks, (b) participation in groups at a sufficiently fine level to link individuals, or (c) extended families (surnames are not sufficiently distinctive to decisively link families, as in e.g., Angelucci et al. (2009)).

However, we did collect GPS coordinates of households at year-seven (i.e., for 889 of the 925 households in the original sample). There is some movement by households within our study's geographic area—5% are now more than 6km from the median GPS point of other households in the same baseline village, suggesting our GPS measure likely does not capture where these households resided prior to the roll-out of the program. However, of the (non-random) 91% of control households who did not attrit and who appear not to have moved, we can examine how proximity to treatment households (once controlling for total proximity to sample households) is associated with economic outcomes.

Specifically, on the subsample of control households, we estimate:

$$(A1)\ y_{ht} = \alpha + \beta\ neighbors_{h0} + \gamma\ neighbors\_treated_{h0} + \delta\ y_{ho} + \mu_v + \epsilon_{ht}$$

Where $neighbors_{h0}$ is equal to the number of households within 200 meters of household h at baseline, and $neighbors\_treated_{h0}$ is the number of households within the same bandwidth who were randomly selected to receive the program. We report the degree of spillovers for five families of outcomes: (1) our indexed family outcomes, (2) consumption and food security, (3) income aggregates, (4) income mechanisms, and (5) resource-sharing (i.e., the outcomes in our main figures plus resource-sharing).

We report results in Appendix Tables A10-A14. γ therefore measures the marginal impact on a control household of living proximate to a treatment household, once controlling for the total number of proximate sample households.

Given our negative point estimate for many outcomes, a relevant question is whether (and the degree to which) negative spillovers are potentially overstating the treatment effects of the program. We therefore estimate:

$$(A2)\ y_{ht} = \alpha + \theta\ treat_h + \beta\ neighbors_{h0} + \gamma\ neighbors\_treated_{h0} + \delta\ y_{ho} + \mu_v + \epsilon_{ht}$$

Where $treatment_h$ is equal to the household's treatment status. We report results in Appendix Tables A15-A19. Reassuringly, we find when re-estimating our main effects with these spillovers that our estimates of the average treatment effect of the program are qualitatively quite similar.

Appendix D: Cost-Benefit Simulations

We estimate the uncertainty of our cost-benefit estimates using bootstraps. In particular, we construct bootstrapped samples of our study population, re-calculate consumption benefits on the bootstrapped sample, and use these calculations to re-estimate the benefit-cost ratio.

We take our full sample of 925 households ever in the sample, and in each of 10,000 simulations, randomly draw 925 households from this sample with replacement. (In each bootstrap, we sample at the household, rather than household-survey wave level. This preserves any autocorrelation that might exist between waves). We then re-estimate household benefits (equal to monthly per capita consumption average treatment effects * household size * 12) for our two-, three-, and seven-year results, and store these estimates in a matrix.

For each of these 10,000 estimates, we perform the same calculations done from our main estimates on these bootstrapped estimates to re-calculate our benefit-cost ratio. In particular, we assume that year one consumption benefits equal year two benefits, and that the benefits from years three to seven evolve linearly. In cases where the benefits decline from years three to seven, we assume that they follow a linear decline, until the point they are equal to zero.

In 19% of simulations, we find that benefits grew from years three to seven. For these simulations, we assume that benefits remain constant beyond year seven (rather than allowing these benefits to continue to grow in perpetuity). We additionally impose a cap that benefits extend for no more than 30 years.[3]

We discount all benefits with an annual rate of 5%.

Finally, for each of these 10,000 estimates, we calculate a benefit-cost ratio, by dividing the accumulated benefits over $4,011, our estimated program cost (we assume our costs are measured without error). We find that benefits exceed costs in 51.7% of our simulations; our median benefit-cost ratio is 1.02. Our 95% interval is [0.42, 2.36].

[3] This simplification is done because there are households for whom consumption estimates decline between years three and seven, albeit very slowly (our 99th percentile is an additional 363 years of benefits). Applying a perpetuity formula slightly overstates the benefits for these individuals, but there is no clear cut-off between "benefits extend forever" and "benefits ultimately reach zero." In practice, results are not very sensitive to the time horizon used in the very long run. Increasing our time horizon to 72 years increases the share for whom benefits exceed costs from 51.72% to 51.90%. Extending the benefits beyond year 72 is not pivotal for whether any additional households have benefits in excess of program costs.

Appendix References

Angelucci, Manuela, Giacomo De Giorgi, Marcos Rangel, and Imran Rasul. 2009. "Village Economies and the Structure of Extended Family Networks." *The BE Journal of Economic Analysis & Policy* 9 (1).

Subramanian, Shankar, and Angus Deaton. 1996. "The Demand for Food and Calories." *Journal of Political Economy* 104 (1): 133–62.

E. Appendix Tables

List of Tables

A1 – Baseline Balance Tests
A2 – Program Timeline
A3 – Attrition
A4 – Indexed Family Outcomes, Ethiopia and India
A5 – Consumption and Food Security, Ethiopia and India
A6 – Income Aggregates, Ethiopia and India
A7 – Income Mechanisms
A8 – Resource-Sharing
A9 – PSNP Participation
A10 – Spatial Spillovers, Control Households Only, Indexed Family Outcomes
A11 – Spatial Spillovers, Control Households Only, Consumption and Food Security
A12 – Spatial Spillovers, Control Households Only, Income Aggregates
A13 – Spatial Spillovers, Control Households Only, Income Mechanisms
A14 – Spatial Spillovers, Control Households Only, Resource-Sharing
A15 – Spatial Spillovers, All Households, Indexed Family Outcomes
A16 – Spatial Spillovers, All Households, Consumption and Food Security
A17 – Spatial Spillovers, All Households, Income Aggregates
A18 – Spatial Spillovers, All Households, Income Mechanisms
A19 – Spatial Spillovers, All Households, Resource-Sharing
A20 – Cost-Benefit Calculations
A21 – Asset Ownership
A22 – Food Security
A23 – Financial Inclusion
A24 – Time Use
A25 – Physical and Mental Health
A26 – Political Involvement and Women's Decision-Making
A27 – Financial Outcomes, with National CPI used as Inflator

Appendix Table A1: Baseline Balance Tests

	(1)	(2)	(3)	(4)	(5)
Panel A. Household-Level Outcomes					
	Asset Ownership	Per Capita Consumption	Food Security Index	Income and Revenues Index	Financial Inclusion Index
Treatment	0.095	-0.092	0.026	0.012	-0.134
	(0.063)	(0.064)	(0.064)	(0.064)	(0.063)
Observations	925	925	925	925	925
Control mean	0.000	0.000	0.000	0.000	0.000
Panel B. Adult-Level Outcomes					
	Physical Health Index	Mental Health Index	Time Working	Political Involvement Index	Women's Empowerment Index
Treatment (ITT): Three Year	-0.081	0.026	-0.051	0.003	-0.099
	(0.059)	(0.059)	(0.058)	(0.055)	(0.069)
Observations	1305	1304	1305	1305	773
Control Mean	0.000	0.000	0.000	0.000	0.000

Appendix Table A1 tests for balance on baseline outcomes between treatment and control. Each regression controls for tabia-level strata. Standard errors are clustered at the household level (i.e., household-level outcomes have a cluster size of one; adult-level outcomes, where we survey 1-2 members per households, share a cluster). Each outcome is standardized to have mean 1, standard deviation 0.

Appendix Table A2: Program Timeline

Activity	Date
Baseline Survey	April 2010
Public Lottery	May 2010
Asset Transfer	June-August 2010, December 2010*
Training and Coaching	June 2010-May 2012
Endline Survey 1	July 2012
Endline Survey 2	July-August 2013
Endline Survey 3	September 2017

Initital transfers took place from June-August. Due to concerns livestock purchases would drive up the price of livestock (and reduce the number of livestock that could be transferred), households receiving sheep and goats received half their livestock in June-August, and the other half in December

Appendix Table A3: Attrition

	(1) 2 Years	(2) 3 Years	(3) 7 Years
Panel A: Attrition by Treatment Status			
Treatment	0.000	-0.006	-0.008
	(0.007)	(0.009)	(0.013)
Control Mean	0.0107	0.0214	0.0428
Panel B: Correlates of Attrition			
Asset Ownership Index	0.006	0.002	0.000
	(0.004)	(0.005)	(0.007)
Per Capita Consumption Index	0.003	0.001	-0.001
	(0.003)	(0.004)	(0.006)
Food Security Index	-0.004	-0.005	-0.011
	(0.003)	(0.005)	(0.007)
Income and Revenues Index	0.001	0.002	-0.005
	(0.003)	(0.005)	(0.007)
Financial Inclusion Index	-0.002	-0.004	-0.002
	(0.003)	(0.004)	(0.006)
Physical Health Index, HH-Level Average	0.000	0.000	0.007
	(0.004)	(0.005)	(0.008)
Mental Health Index, HH-Level Average	0.000	-0.005	-0.007
	(0.004)	(0.005)	(0.008)
Productive Time Use Index, HH-Level Average	-0.001	0.002	0.006
	(0.004)	(0.005)	(0.007)
Political Involvement Index, HH-Level Average	0.000	-0.001	0.001
	(0.004)	(0.005)	(0.008)
Women's Empowerment Index	0.001	0.001	0.011
	(0.004)	(0.005)	(0.007)
Panel C: Test for Differential Composition of Attriters by Treatment			
Joint F-Test: Treatment and Indices Interacted with Treatment Status	0.315	0.348	0.770
p-value	0.99	0.984	0.693

Panel A reports regression results of whether or not an individual attrited from the sample on treatment status, with attrition as the dependent variable, with adult-level outcomes averaged at the household level. Panel B regresses attrition on the baseline values of our indexed family outcomes. Panel C reports the joint F-Test from a regression of attrition on the correlates in Panel B interacted with treatment. In all cases, we include tabia-level strata indicator variables; standard errors are Huber-White heteroskedastic.

Appendix Table A4: Indexed Family Outcomes

	Ethiopia (this paper)			India (BDS)		
	Average Treatment Effects (Full Compliance)			*Intention-to-Treat Effects (52% Take-Up)*		
	(1)	(2)	(3)	(4)	(5)	(6)
	2 Years	3 Years	7 Years	1.5 Years	3 Years	7 Years
Asset Ownership	0.950	0.947	0.429	0.217	0.389	0.814
	(0.091)	(0.090)	(0.109)	(0.111)	(0.103)	(0.132)
q-value	0.001	0.001	0.001	0.041	0.001	0.001
Control Mean	0.374	0.518	1.159	-0.20	-0.25	-0.46
Baseline Mean		0.000			0.000	
Per Capita Consumption	0.239	0.25	0.172	0.311	0.292	0.717
	(0.068)	(0.054)	(0.082)	(0.076)	(0.079)	(0.125)
q-value	0.001	0.001	0.196	0.001	0.001	0.001
Control Mean	0.000	-0.095	0.487	0.35	0.85	1.09
Baseline Mean		-			0.00	
Food Security Index	0.116	0.145	0.035	0.184	0.251	0.431
	(0.052)	(0.050)	(0.038)	(0.05)	(0.06)	(0.06)
q-value	0.022	0.003	0.467	0.00	0.00	0.00
Control Mean	0.706	0.742	1.024	0.35	0.94	1.09
Baseline Mean		0.000			0.00	
Income and Revenues Index	1.411	0.411	0.244	0.145	0.172	0.334
	(0.163)	(0.115)	(0.241)	(0.08)	(0.07)	(0.07)
q-value	0.001	0.001	0.467	0.04	0.02	0.00
Control Mean	0.00	0.512	1.247	0	0	0
Baseline Mean		-			-	
Financial Inclusion Index	1.853	0.779	0.301	-0.004	0.192	0.181
	(0.118)	(0.126)	(0.252)	0.04	(0.06)	(0.14)
q-value	0.001	0.001	0.449	0.26	0.00	0.05
Control Mean	0.00	0.0473	0.736	0.14	0.3	0.67
Baseline Mean		-			0.000	
Physical Health Index	-0.0158	0.0499	-0.0704	0.061	0.027	0.13
	(0.050)	(0.043)	(0.054)	(0.028)	(0.027)	(0.031)
q-value	0.293	0.104	0.434	0.028	0.160	0.001
Control Mean	-0.0749	0.0112	-0.0603	0.913	0.921	0.987
Baseline Mean		0.005			0.000	
Mental Health Index	0.053	-0.038	-0.087	0.115	0.012	0.011
	(0.055)	(0.060)	(0.063)	(0.029)	(0.037)	(0.018)
q-value	0.148	0.210	0.434	0.00	0.33	0.00
Control Mean	0.000	0.000	0.000	0.32	0.75	1.09
Baseline Mean		-			0.000	
Productive Time Use	0.278	0.224	0.0805	0.285	0.102	0.165
	(0.057)	(0.059)	(0.049)	(0.05)	(0.04)	(0.04)
q-value	0.001	0.001	0.385	0.001	0.018	0.000
Control Mean	0.508	0.604	0.463	0.23	0.28	-0.04
Baseline Mean		0.002			0.000	
Political Empowerment Index	0.097	0.115	0.007	0.009	0.021	0.031
	(0.057)	(0.060)	(0.060)	(0.34)	(0.03)	(0.03)
q-value	0.055	0.032	0.632	0.248	0.232	0.060
Control Mean	0.442	0.235	0.214	-0.05	0.13	0.27
Baseline Mean		0.001			0.000	
Women's Decision-Making Index	-0.016	-0.027	-0.054			
	(0.047)	(0.049)	(0.059)			
q-value	0.293	0.214	0.467			
Control Mean	0.188	0.241	0.572			
Baseline Mean		-0.005				

Appendix Table A4 reports results from intention-to-treat effects of the graduation program on indexed family outcomes. Columns (1) to (3) report results from Ethiopia, and Columns (4) to (6) from India, as presented in Banerjee et al. 2021. Each cell reports results from a separate regression. Observations range from 889 to 915 for the household level outcomes in Ethiopia and from 679 to 875 in India, and from 723 to 1307 for the adult-level outcomes in Ethiopia and from 1,229 to 1,950 in India. For asset ownership, consumption and time use, outcomes are aggregated and then rescaled; other outcomes are a z-score index of outcomes variables in the given category. Outcomes are either (a) standardized so the baseline has mean zero and standard deviation one (done whenever possible), or (b) standardized so the endline 1 control group has mean zero and standard deviation one (done in cases where we do not have baseline data, or the baseline components differed from the components in subsequent waves). Reported q-values are sharpened using the false discovery rate procedure detailed in Anderson (2008). They reflect a correction for 10 family outcomes in Ethiopia, and 9 family outcomes in India. The full list of variables used to construct each index is reported in Appendix B. Each regression controls for baseline values, and for tabia-level strata. Standard errors are clustered at the household level (i.e., household-level outcomes have a cluster size of one; adult-level outcomes, where we survey 1-2 members per households, share a cluster).

Appendix Table A5: Monthly Consumption and Food Security

	Ethiopia (this paper) *Average Treatment Effects (Full Compliance)*			India (BDS) *Intention-to-Treat Effects (52% Take-Up)*		
	(1)	(2)	(3)	(4)	(5)	(6)
	2 Years	3 Years	7 Years	1.5 Years	3 Years	7 Years
Monthly Per Capita Consumption	8.21	8.56	5.90	8.35	7.83	19.22
	(2.32)	(1.83)	(2.81)	(2.03)	(2.12)	(3.35)
Control Mean	50.2	46.9	66.9	54.5	67.9	74.2
Baseline Mean		-			45.0	
Monthly Per Capita Food Consumption	2.91	2.35	1.08	5.93	3.34	10.81
	(1.78)	(1.00)	(1.42)	(1.32)	(1.27)	(1.91)
Control Mean	32.6	30.0	38.3	36.8	41.5	41.8
Baseline Mean		-			29.6	
Monthly Per Capita Nonfood Consumption	4.41	4.61	3.86	2.36	4.43	8.34
	(1.02)	(1.11)	(1.84)	(1.20)	(1.26)	(1.86)
Control Mean	16.2	16.2	26.4	17.7	26.3	32.5
Baseline Mean		16.1			15.5	
Monthly Per Capita Durable Good Expenditure	0.50	1.21	0.39	-0.39	0.97	2.73
	(0.40)	(0.34)	(0.52)	(0.41)	(0.43)	(0.56)
Control Mean	1.42	0.685	2.12	2.52	2.13	2.51
Baseline Mean		-			1.07	
Everyone in HH gets enough food every day	0.039	0.077	0.023	0.074	0.141	0.205
	(0.031)	(0.030)	(0.025)	(0.025)	(0.034)	(0.032)
Control Mean	0.640	0.640	0.826	0.110	0.420	0.590
Baseline Mean		0.335			0.11	
No one HH went whole day without food	0.029	0.041	-0.007	0.128	0.038	0.095
	(0.017)	(0.015)	(0.009)	(0.03)	(0.02)	(0.02)
Control Mean	0.913	0.921	0.987	0.68	0.85	0.83
Baseline Mean		0.745			0.28	
No Children Skipped Meals	0.045	0.047	0.011	0.032	0.085	0.045
	(0.022)	(0.023)	(0.018)	(0.034)	(0.025)	(0.026)
Control Mean	0.845	0.833	0.926	0.750	0.860	0.870
Baseline Mean		0.489			0.510	
Share of total consumption on non-grain items	0.038	0.037	0.013	0.006	0.021	0.024
	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.006)
Control Mean	0.612	0.580	0.686	0.708	0.714	0.773
Baseline Mean		0.525			0.237	

Appendix Table A5 reports results from intention-to-treat effects of the graduation program on consumption and food security-related outcomes. Columns (1) to (3) report results from Ethiopia, and Columns (4) to (6) from India, as presented in Banerjee et al. 2021. Each cell reports results from a separate regression. Observations range from 889 to 915 for Ethiopia and from 679 to 875 in India. All outcomes in Ethiopia are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic. Additional food security outcomes are reported in Appendix Table A22.

Appendix Table A6: Monthly Income and Revenue

	Ethiopia (this paper) *Average Treatment Effects (Full Compliance)*			India (BDS) *Intention-to-Treat Effects (52% Take-Up)*		
	(1)	(2)	(3)	(4)	(5)	(6)
	2 Years	3 Years	7 Years	1.5 Years	3 Years	7 Years
Monthly Livestock Revenues	128.5	30.9	40.2	11.3	8.5	30.1
	(8.7)	(6.2)	(21.0)	(2.6)	(6.9)	(5.7)
Control Mean	32.5	29.5	89.9	3.7	8.8	10.7
Baseline Mean		-			0.0	
Monthly Agricultural Profits (ETH) / Fishing & Horticulture (IND)	5.05	6.85	4.30	20.64	34.34	119.81
	(2.30)	(2.38)	(3.49)	(6.408)	(7.626)	(16.750)
Control Mean	25.72	34.62	33.46	50.96	66.89	114.05
Baseline Mean		-			18.02	
Monthly Non-Farm Enterprise Profits	12.03	-0.86	-1.35	8.77	27.77	74.74
	(4.91)	(3.13)	(5.78)	(5.018)	(6.918)	(15.771)
Control Mean	4.07	16.30	7.42	40.07	54.70	100.05
Baseline Mean		-			14.64	
Monthly Income from Wage Labor (including workfare)	-2.80	-2.29	-2.98	5.62	5.12	98.43
	(2.95)	(2.68)	(6.76)	(9.382)	(15.119)	(28.449)
Control Mean	34.78	29.83	89.97	117.70	240.91	333.15
Baseline Mean		-			-	
Rating of Economic Status (1/10)	1.13	0.90	0.01	0.20	0.30	1.58
	(0.15)	(0.12)	(0.13)	(0.07)	(0.08)	(0.14)
Control Mean	3.73	4.34	4.99	2.77	3.36	4.73
Baseline Mean		3.742			1.97	
Productive Asset Value, Indexed	0.776	0.761	0.374	0.444	0.571	0.795
	(0.076)	(0.074)	(0.102)	(0.086)	(0.072)	(0.083)
Control Mean	0.068	0.173	0.863	-0.230	-0.300	-0.400
Baseline Mean		0.000			0.000	
Monthly Remittances Received	1.55	1.21	4.27	-	4.09	9.81
	(1.18)	(2.89)	(3.90)	-	(2.62)	(7.14)
Control Mean	2.75	9.06	10.34	-	14.24	38.55
Baseline Mean		-			-	

Appendix Table A6 reports results from intention-to-treat effects of the graduation program on income and revenue outcomes. Columns (1) to (3) report results from Ethiopia, and Columns (4) to (6) from India, as presented in Banerjee et al. 2021. Each cell reports results from a separate regression. Observations range from 889 to 915 for Ethiopia and from 679 to 875 in India. All financial outcomes in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A7: Income Mechanisms, Ethiopia

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Value of Livestock	Total Value of Livestock Sold per Month, last 12 months	Total Value of Livestock Bought per Month, last 12 months	Monthly Revenue from Agriculture	Monthly Expenditure on Agriculture	Total Acres Cultivated, Major Growing Season	Received any remittances	Earned any wage income (including from workfare program)
Treatment (ITT): Two Year	967	79.2	73.7	8.00	4.03	0.153	-0.029	0.004
	(91.2)	(5.49)	(5.28)	(2.55)	(1.17)	(0.068)	(0.023)	(0.007)
Observations	914	914	914	915	915	915	911	915
Control mean	1341	17.66	12.97	37.5	11.8	1.44	0.148	0.99
Treatment (ITT): Three Year	988	30.3	30.9	8.36	2.91	0.221	-0.080	-0.037
	(89.0)	(5.52)	(5.33)	(2.51)	(0.81)	(0.066)	(0.026)	(0.017)
Observations	908	908	908	908	908	908	908	915
Control mean	1481	13.9	10.8	42.0	7.35	1.56	0.239	0.94
Treatment (ITT): Seven Year	419	38.1	5.93	6.43	3.69	0.254	0.038	-0.082
	(108)	(20.0)	(3.41)	(4.04)	(1.31)	(0.098)	(0.025)	(0.022)
Observations	889	889	889	889	889	889	889	889
Control mean	2030	68.1	25.2	54.2	20.7	1.43	0.141	0.91
Baseline Mean	1107	-	-	-	-	0.856	-	-

Appendix Table A7 reports average treatment effects of the graduation program on mechanisms through which household-level income and revenues have evolved. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A8: Transfers and Lending In

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Remittances Received, Last 12 Months	Remittance Income per Month	Received any Informal Loans	Cash Value of Informal Loans Received	Received any Food as Gifts	Number of Types of Food Received as Gift
Treatment (ITT): Two Year	-0.029	1.55	-0.021	19.6	0.017	-0.116
	(0.023)	(1.18)	(0.028)	(16.5)	(0.030)	(0.076)
Observations	911	915	915	915	915	915
Control mean	0.148	2.75	0.729	117	0.294	0.606
Treatment (ITT): Three Year	-0.080	1.21	-0.003	28.9	0.006	-0.013
	(0.026)	(2.89)	(0.026)	(11.3)	(0.025)	(0.051)
Observations	908	908	908	908	908	908
Control mean	0.239	9.06	0.799	135	0.171	0.271
Treatment (ITT): Seven Year	0.038	4.27	0.004	26.7	-0.032	-0.115
	(0.025)	(3.90)	(0.019)	(45.9)	(0.033)	(0.080)
Observations	889	889	889	889	889	889
Control mean	0.141	10.3	0.085	55.6	0.593	1.24
Baseline Mean	-	-	0.014	4.70	-	-

Appendix Table A8 reports average treatment effects of the graduation program on mechanisms on transfer receipt. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A9 Productive Safety Net Participation

	(1)	(2)	(3)
	Household enrolled in Productive Safety Net Programme (PSNP)	Days of PSNP by Household in Last Month	Monthly Earnings from PSNP
Treatment (ITT): Two Year	-0.003	1.30	2.94
	(0.009)	(0.573)	(1.839)
Observations	913	911	913
Control mean	0.98	15.1	46.8
Treatment (ITT): Three Year	-0.059	-0.624	-1.92
	(0.021)	(0.600)	(2.13)
Observations	908	908	908
Control mean	0.91	11.0	50.5
Treatment (ITT): Seven Year	-0.052	-0.370	-4.21
	(0.027)	(0.645)	(2.61)
Observations	889	889	889
Control mean	0.81	8.19	56.9

Appendix Table A9 reports average treatment effects of the graduation program on participation and benefits from the Productive Safety Net Programme. Monthly PSNP earnings are reported in 2021 USD in PPP terms. Each regression controls tabia-level strata (we do have data for these outcomes at baseline). Standard errors are Huber-White heteroskedastic.

Appendix Tables A10: Spatial Spillovers - Indices - Control Households Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Asset Ownership	Per Capita Consumption	Food Security Index	Income and Revenues Index	Financial Inclusion Index	Physical Health Index	Mental Health Index	Time Working	Political Involvement Index	Women's Empowerment Index
Two Year										
Neighbors within 200m	0.101	0.045	0.030	0.008	-0.012	0.047	-0.060	-0.029	-0.027	-0.035
	(0.046)	(0.051)	(0.034)	(0.038)	(0.027)	(0.026)	(0.038)	(0.035)	(0.029)	(0.028)
Treated Neighbors within 200m	-0.163	-0.028	-0.086	-0.105	-0.056	-0.048	-0.004	-0.002	0.065	0.043
	(0.070)	(0.107)	(0.052)	(0.056)	(0.059)	(0.048)	(0.004)	(0.002)	0.065	0.043
Observations	426	426	426	426	426	615	615	388	615	352
Mean: No neighbors within 200m	0.442	-0.080	0.810	0.082	0.065	0.004	0.072	0.635	0.517	0.138
Three Year										
Neighbors within 200m	0.054	0.067	0.087	0.135	-0.015	0.018	-0.003	0.041	-0.032	0.004
	(0.039)	(0.031)	(0.031)	(0.111)	(0.032)	(0.026)	(0.027)	(0.041)	(0.030)	(0.031)
Treated Neighbors within 200m	-0.097	-0.061	-0.116	-0.128	-0.021	-0.019	-0.021	-0.117	0.047	0.013
	(0.052)	(0.045)	(0.046)	(0.072)	(0.039)	(0.019)	(0.021)	(0.117)	0.047	0.013
Observations	426	426	426	426	426	596	595	386	595	340
Mean: No neighbors within 200m	0.580	-0.218	0.723	0.464	0.256	0.040	-0.013	0.659	0.402	0.167
Seven Year										
Neighbors within 200m	0.006	0.049	0.012	0.219	-0.023	0.041	0.020	-0.020	0.011	0.030
	(0.051)	(0.037)	(0.020)	(0.091)	(0.170)	(0.048)	(0.042)	(0.024)	(0.034)	(0.036)
Treated Neighbors within 200m	-0.180	-0.009	-0.042	-0.336	0.138	-0.079	-0.020	-0.006	0.029	-0.051
	(0.071)	(0.053)	(0.030)	(0.115)	(0.245)	(0.079)	(0.020)	(0.006)	0.029	(0.051)
Observations	426	426	426	426	426	593	593	593	593	352
Mean: No neighbors within 200m	1.53	0.473	1.029	1.101	0.842	0.020	0.098	0.588	0.259	0.463

Appendix Table A10 reports spillover effects of the graduation program on control households. The sample includes the 426 control households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. For asset ownership, consumption and time use, outcomes are aggregated and then rescaled; other outcomes are a z-score index of outcomes variables in the given category. Outcomes are either (a) standardized so the baseline has mean zero and standard deviation one (done whenever possible), or (b) standardized so the endline 1 control group has mean zero and standard deviation one (done in cases where we do not have baseline data, or the baseline components differed from the components in subsequent waves). Reported q-values are sharpened using the false discovery rate procedure detailed in Anderson (2008). They reflect a correction for 10 family outcomes. The full list of variables used to construct each index is reported in Appendix B. Each regression controls for baseline values, and for tabia-level strata. Standard errors are clustered at the household level (i.e., household-level outcomes have a cluster size of one; adult-level outcomes, where we survey 1-2 members per households, share a cluster).

Appendix Table A11: Spatial Spillovers - Consumption - Control Households Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Monthly Per Capita Consumption	Monthly Per Capita Food Consumption	Monthly Per Capita Nonfood Consumption	Monthly Per Capita Durable Good Expenditure	Everyone in HH gets enough food every day	No one HH went whole day without food	No Children Skipped Meals	Share of total consumption on non-grain items
Two Year								
Neighbors within 200m	1.54	0.247	0.721	0.537	0.011	0.012	0.001	0.019
	(1.74)	(1.55)	(0.497)	(0.238)	(0.019)	(0.011)	(0.015)	(0.006)
Treated Neighbors within 200m	-0.944	0.781	-1.383	-0.400	-0.031	-0.026	-0.020	-0.020
	(3.66)	(3.418)	(0.787)	(0.365)	(0.029)	(0.018)	(0.022)	(0.010)
Observations	426	426	426	426	425	423	423	426
Mean: No neighbors within 200m	52.9	33.6	17.8	1.42	0.685	0.945	0.890	0.632
Three Year								
Neighbors within 200m	2.29	0.723	1.289	0.241	0.051	0.002	0.023	0.013
	(1.06)	(0.471)	(0.765)	(0.179)	(0.017)	(0.011)	(0.013)	(0.005)
Treated Neighbors within 200m	-2.08	-0.789	-1.004	-0.329	-0.076	-0.019	-0.023	-0.010
	(1.53)	(0.816)	(1.16)	(0.236)	(0.027)	(0.017)	(0.020)	(0.007)
Observations	426	426	426	426	425	425	425	426
Mean: No neighbors within 200m	47.7	29.6	16.8	1.31	0.660	0.945	0.817	0.600
Seven Year								
Neighbors within 200m	1.68	1.01	0.501	0.151	0.013	-0.002	0.008	0.005
	(1.26)	(0.643)	(0.771)	(0.131)	(0.016)	(0.005)	(0.009)	(0.004)
Treated Neighbors within 200m	-0.307	-0.261	0.296	-0.403	-0.021	-0.002	-0.024	-0.004
	(1.80)	(0.939)	(1.16)	(0.149)	(0.022)	(0.005)	(0.019)	(0.006)
Observations	426	426	426	426	426	426	392	426
Mean: No neighbors within 200m	69.1	38.9	27.7	2.56	0.835	0.980	0.935	0.702

Appendix Table A11 reports spillover effects of the graduation program on control households. The sample includes the 426 control households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A12: Spatial Spillovers - Income Aggregates - Control Households Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Monthly Livestock Revenues	Monthly Agricultural Profits	Monthly Non-Farm Enterprise Profits	Monthly Income from Wage Labor (including workfare)	Rating of Economic Status (1/10)	Productive Asset Value, Indexed	Monthly Remittances Received
Two Year							
Neighbors within 200m	-0.515	0.018	-1.21	0.583	0.109	0.067	0.322
	(2.18)	(0.433)	(1.27)	(2.19)	(0.087)	(0.037)	(0.555)
Treated Neighbors within 200m	-1.75	-0.776	0.434	-2.75	-0.297	-0.129	-0.015
	(3.30)	(0.489)	(1.99)	(3.19)	(0.133)	(0.056)	(0.596)
Observations	426	426	426	426	426	426	426
Mean: No neighbors within 200m	110	13.4	27.5	34.4	4.43	0.74	2.05
Three Year							
Neighbors within 200m	-1.37	4.60	-0.904	2.52	-0.003	0.027	-0.171
	(1.87)	(3.51)	(1.43)	(1.936)	(0.070)	(0.032)	(1.63)
Treated Neighbors within 200m	-1.21	-2.61	-1.01	-3.63	-0.024	-0.066	0.570
	(2.53)	(1.91)	(2.27)	(2.610)	(0.099)	(0.044)	(1.92)
Observations	426	426	426	426	426	426	426
Mean: No neighbors within 200m	52.1	14.6	33.2	25.4	4.81	0.78	5.70
Seven Year							
Neighbors within 200m	5.46	3.82	-0.738	7.85	0.092	-0.032	0.647
	(8.34)	(2.20)	(1.73)	(4.02)	(0.077)	(0.050)	(1.35)
Treated Neighbors within 200m	-3.63	-5.69	-5.18	-7.95	-0.125	-0.141	-0.103
	(8.13)	(2.64)	(2.89)	(5.39)	(0.120)	(0.068)	(2.52)
Observations	426	426	426	426	425	426	426
Mean: No neighbors within 200m	120	11.3	36.1	86.4	5.00	1.58	6.58

Appendix Table A12 reports spillover effects of the graduation program on control households. The sample includes the 426 control households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A13: Spatial Spillovers - Income Mechanisms - Control Households Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Value of Livestock	Total Value of Livestock Sold per Month, last 12 months	Total Value of Livestock Bought per Month, last 12 months	Monthly Revenue from Agriculture	Monthly Expenditure on Agriculture	Total Acres Cultivated, Major Growing Season	Received any remittances	Earned any wage income (including from workfare program)
Two Year								
Neighbors within 200m	95.0	-0.724	-0.302	-1.51	0.156	0.023	0.024	-0.005
	(49.4)	(1.47)	(1.18)	(1.24)	(0.689)	(0.040)	(0.017)	(0.003)
Treated Neighbors within 200m	-170	0.877	1.60	0.873	-0.305	-0.046	-0.003	-0.003
	(74.890)	(2.07)	(1.73)	(2.06)	(1.08)	(0.053)	(0.024)	(0.004)
Observations	425	425	425	426	426	426	424	426
Mean: No neighbors within 200m	2141.00	63.20	53.26	42.33	14.87	1.36	0.09	1.00
Three Year								
Neighbors within 200m	41.1	0.199	2.29	-1.17	0.334	0.006	0.014	-0.001
	(42.7)	(1.33)	(1.20)	(1.46)	(0.463)	(0.037)	(0.019)	(0.007)
Treated Neighbors within 200m	-92.5	-2.552	-3.13	-0.761	-0.785	-0.083	0.005	-0.002
	(57.9)	(1.59)	(1.41)	(2.30)	(0.673)	(0.050)	(0.029)	(0.014)
Observations	426	426	426	426	426	426	426	426
Mean: No neighbors within 200m	2190.00	32.04	25.04	42.94	9.76	1.53	0.19	0.94
Seven Year								
Neighbors within 200m	-22.5	8.18	5.07	-1.18	0.542	0.022	0.018	0.019
	(54.9)	(8.27)	(2.33)	(2.03)	(0.825)	(0.051)	(0.013)	(0.012)
Treated Neighbors within 200m	-155	-3.92	-8.44	-5.47	-1.99	-0.127	0.004	-0.010
	(74.7)	(7.87)	(2.87)	(3.29)	(1.12)	(0.066)	(0.019)	(0.015)
Observations	426	426	426	426	426	426	426	426
Mean: No neighbors within 200m	2765	90.3	31.2	57.5	21.3	1.44	0.13	0.89

Appendix Table A13 reports spillover effects of the graduation program on control households. The sample includes the 426 control households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A14: Spatial Spillovers - Resource Sharing - Control Households Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Remittances Received, Last 12 Months	Remittance Income per Month	Received any Informal Loans	Cash Value of Informal Loans Received	Received any Food as Gifts	Number of Types of Food Received as Gift
Two Year						
Neighbors within 200m	0.024	0.322	0.019	-4.63	0.025	0.050
	(0.017)	(0.555)	(0.016)	(7.07)	(0.019)	(0.040)
Treated Neighbors within 200m	-0.003	-0.015	-0.046	-3.20	0.010	0.010
	(0.024)	(0.596)	(0.025)	(10.8)	(0.027)	(0.061)
Observations	424	426	426	426	426	426
Mean: No neighbors within 200m	0.095	2.05	0.705	123	0.287	0.453
Three Year						
Neighbors within 200m	0.014	-0.171	-0.010	-6.52	0.009	-0.011
	(0.019)	(1.63)	(0.014)	(5.70)	(0.016)	(0.029)
Treated Neighbors within 200m	0.005	0.570	-0.010	1.77	0.007	0.049
	(0.029)	(1.92)	(0.021)	(8.43)	(0.025)	(0.055)
Observations	426	426	426	426	426	426
Mean: No neighbors within 200m	0.185	5.70	0.776	158	0.193	0.311
Seven Year						
Neighbors within 200m	0.018	0.647	-0.007	6.99	-0.018	0.014
	(0.013)	(1.35)	(0.010)	(10.8)	(0.019)	(0.060)
Treated Neighbors within 200m	0.004	-0.103	-0.001	-27.8	0.008	-0.066
	(0.019)	(2.52)	(0.013)	(22.3)	(0.029)	(0.082)
Observations	426	426	426	426	426	426
Mean: No neighbors within 200m	0.126	6.58	0.098	111	0.602	1.21

Appendix Table A14 reports spillover effects of the graduation program on control households. The sample includes the 426 control households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A15: Spatial Spillovers - Indices - All Households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Asset Ownership	Per Capita Consumption	Food Security Index	Income and Revenues Index	Financial Inclusion Index	Physical Health Index	Mental Health Index	Time Working	Political Involvement Index	Women's Empowerment Index
Two Year										
Treatment	0.983	0.235	0.124	1.48	1.90	-0.033	0.079	0.274	0.078	-0.002
	(0.098)	(0.072)	(0.054)	(0.181)	(0.132)	(0.049)	(0.053)	(0.059)	(0.060)	(0.049)
Neighbors within 200m	0.034	0.023	0.062	-0.001	-0.030	0.042	-0.018	-0.018	-0.004	-0.020
	(0.041)	(0.037)	(0.023)	(0.056)	(0.060)	(0.019)	(0.023)	(0.029)	(0.025)	(0.021)
Treated Neighbors within 200m	-0.114	-0.029	-0.097	-0.034	-0.063	-0.022	-0.023	0.001	0.047	0.028
	(0.058)	(0.071)	(0.034)	(0.067)	(0.055)	(0.027)	(0.033)	(0.040)	(0.038)	(0.031)
Observations	844	844	844	844	844	1227	1227	774	1227	704
Mean: No neighbors within 200m	0.442	-0.080	0.810	0.082	0.065	0.004	0.072	0.635	0.517	0.138
Three Year										
Treatment	0.978	0.263	0.134	0.451	0.755	0.036	-0.050	0.189	0.116	-0.001
	(0.095)	(0.055)	(0.051)	(0.124)	(0.133)	(0.043)	(0.057)	(0.061)	(0.061)	(0.051)
Neighbors within 200m	0.046	0.063	0.067	0.122	-0.006	0.014	0.007	0.062	-0.043	-0.001
	(0.039)	(0.025)	(0.021)	(0.065)	(0.054)	(0.018)	(0.026)	(0.030)	(0.024)	(0.024)
Treated Neighbors within 200m	-0.079	-0.046	-0.070	-0.118	-0.054	0.005	-0.039	-0.092	0.106	0.006
	(0.055)	(0.035)	(0.031)	(0.062)	(0.059)	(0.026)	(0.036)	(0.041)	(0.037)	(0.033)
Observations	844	844	844	844	844	1198	1196	772	1196	676
Mean: No neighbors within 200m	0.580	-0.218	0.723	0.464	0.256	0.040	-0.013	0.659	0.402	0.167
Seven Year										
Treatment	0.444	0.197	0.030	0.369	0.311	-0.130	-0.089	0.082	-0.001	-0.050
	(0.112)	(0.081)	(0.039)	(0.212)	(0.263)	(0.072)	(0.059)	(0.051)	(0.062)	(0.061)
Neighbors within 200m	-0.009	0.047	0.006	0.155	0.081	0.041	-0.002	-0.027	-0.007	0.033
	(0.051)	(0.035)	(0.016)	(0.082)	(0.116)	(0.030)	(0.026)	(0.021)	(0.026)	(0.028)
Treated Neighbors within 200m	-0.082	-0.008	-0.016	-0.264	-0.074	-0.034	0.024	0.006	0.026	-0.050
	(0.066)	(0.051)	(0.022)	(0.107)	(0.145)	(0.045)	(0.037)	(0.031)	(0.038)	(0.039)
Observations	844	844	844	844	844	1179	1179	1179	1179	690
Mean: No neighbors within 200m	1.525	0.473	1.029	1.101	0.842	0.020	0.098	0.588	0.259	0.463

Appendix Table A15 reports spillover effects of the graduation program on all households. The sample includes the 844 households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. For asset ownership, consumption and time use, outcomes are aggregated and then rescaled; other outcomes are a z-score index of outcomes variables in the given category. Outcomes are either (a) standardized so the baseline has mean zero and standard deviation one (done whenever possible), or (b) standardized so the endline 1 control group has mean zero and standard deviation one (done in cases where we do not have baseline data, or the baseline components differed from the components in subsequent waves). Reported q-values are sharpened using the false discovery rate procedure detailed in Anderson (2008). They reflect a correction for 10 family outcomes. The full list of variables used to construct each index is reported in Appendix B. Each regression controls for baseline values, and for tabia-level strata. Standard errors are clustered at the household level (i.e., household-level outcomes have a cluster size of one; adult-level outcomes, where we

Appendix Table A16: Spatial Spillovers - Consumption - All Households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Per Capita Consumption	Per Capita Food Consumption	Per Capita Nonfood Consumption	Per capita Durable Good Expenditure	Everyone in HH gets enough food every day	No one HH went whole day without food	No Children Skipped Meals	Share of total consumption on non-grain items
Two Year								
Treatment	8.05	3.00	4.18	0.531	0.045	-0.033	0.048	0.035
	(2.48)	(1.93)	(1.06)	(0.448)	(0.032)	(0.049)	(0.024)	(0.010)
Neighbors within 200m	0.781	-0.204	0.631	0.342	0.029	0.042	0.014	0.014
	(1.27)	(1.03)	(0.485)	(0.190)	(0.014)	(0.019)	(0.010)	(0.004)
Treated Neighbors within 200m	-0.988	0.508	-0.982	-0.536	-0.034	-0.022	-0.026	-0.018
	(2.42)	(2.01)	(0.879)	(0.263)	(0.020)	(0.027)	(0.014)	(0.007)
Observations	844	844	844	844	843	839	839	844
Mean: No neighbors within 200m, Control	47.4	30.5	15.9	0.970	0.679	0.939	0.885	0.613
Three Year								
Treatment	9.01	2.58	4.80	1.24	0.061	0.036	0.038	0.035
	(1.89)	(1.04)	(1.13)	(0.364)	(0.032)	(0.043)	(0.024)	(0.008)
Neighbors within 200m	2.17	0.715	1.24	0.189	0.041	0.014	0.022	0.009
	(0.854)	(0.399)	(0.563)	(0.178)	(0.013)	(0.018)	(0.009)	(0.004)
Treated Neighbors within 200m	-1.57	-0.631	-0.721	-0.238	-0.048	0.005	-0.016	-0.005
	(1.21)	(0.630)	(0.837)	(0.245)	(0.020)	(0.026)	(0.013)	(0.005)
Observations	844	844	844	844	843	842	840	844
Mean: No neighbors within 200m, Control	42.7	27.6	14.5	0.504	0.646	0.939	0.794	0.586
Seven Year								
Treatment	6.76	1.96	4.10	0.453	0.030	-0.130	0.004	0.011
	(2.77)	(1.27)	(1.84)	(0.403)	(0.025)	(0.072)	(0.018)	(0.008)
Neighbors within 200m	1.60	0.397	0.956	0.255	0.006	0.041	0.000	0.006
	(1.19)	(0.542)	(0.808)	(0.151)	(0.011)	(0.030)	(0.008)	(0.003)
Treated Neighbors within 200m	-0.271	-0.139	0.141	-0.299	-0.006	-0.034	-0.011	-0.004
	(1.74)	(0.789)	(1.26)	(0.217)	(0.015)	(0.045)	(0.012)	(0.005)
Observations	844	844	844	844	844	844	765	844
Mean: No neighbors within 200m, Control	66.4	37.3	26.2	2.91	0.824	0.985	0.917	0.696

Appendix Table A16 reports spillover effects of the graduation program on all households. The sample includes the 844 households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A17: Spatial Spillovers - Income Aggregates - All Households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Monthly Livestock Revenues	Monthly Agricultural Profits	Monthly Non-Farm Enterprise Profits	Monthly Income from Wage Labor (including workfare)	Rating of Economic Status (1/10)	Productive Asset Value, Indexed	Monthly Remittances Received
Two Year							
Treatment	129	13.6	5.88	-3.32	1.20	0.804	1.09
	(9.00)	(5.55)	(2.42)	(3.00)	(0.153)	(0.081)	(1.238)
Neighbors within 200m	4.66	-0.972	-2.66	0.730	0.125	0.018	-0.146
	(4.76)	(1.49)	(0.992)	(1.39)	(0.070)	(0.033)	(0.646)
Treated Neighbors within 200m	-4.95	0.368	2.17	-0.913	-0.160	-0.082	-0.054
	(6.71)	(1.35)	(1.61)	(2.32)	(0.104)	(0.046)	(0.552)
Observations	844	844	844	844	842	844	844
Mean: No neighbors within 200m, control	37.6	5.28	23.5	37.2	3.85	0.14	1.09
Three Year							
Treatment	32.4	-0.153	8.14	-3.72	0.948	0.794	1.72
	(6.42)	(3.45)	(2.47)	(2.79)	(0.123)	(0.078)	(3.08)
Neighbors within 200m	3.92	2.73	-1.36	2.34	0.071	0.023	0.720
	(3.01)	(1.91)	(1.07)	(1.26)	(0.052)	(0.033)	(1.519)
Treated Neighbors within 200m	-3.44	-1.66	-1.63	-2.30	-0.043	-0.060	-1.51
	(2.96)	(1.64)	(1.61)	(1.78)	(0.077)	(0.046)	(1.56)
Observations	844	844	844	844	843	844	844
Mean: No neighbors within 200m, control	36.5	14.7	29.1	28.6	4.47	0.23	6.51
Seven Year							
Treatment	41.4	1.50	5.72	-2.25	0.036	0.382	4.79
	(21.5)	(4.49)	(3.59)	(6.72)	(0.128)	(0.104)	(4.00)
Neighbors within 200m	13.4	1.02	-1.51	5.45	0.060	-0.037	0.238
	(8.79)	(1.60)	(1.35)	(3.03)	(0.057)	(0.047)	(1.75)
Treated Neighbors within 200m	-18.9	-1.97	-4.35	-2.61	-0.047	-0.062	2.46
	(10.1)	(2.03)	(1.91)	(4.76)	(0.086)	(0.062)	(3.72)
Observations	844	844	844	844	843	844	844
Mean: No neighbors within 200m, control	72.7	7.72	31.5	89.4	4.89	1.23	6.25

Appendix Table A17 reports spillover effects of the graduation program on all households. The sample includes the 844 households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A18: Spatial Spillovers - Income Mechanims - All Households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total Value of Livestock	Total Value of Livestock Sold per Month, last 12 months	Total Value of Livestock Bought per Month, last 12 months	Monthly Revenue from Agriculture	Monthly Expenditure on Agriculture	Total Acres Cultivated, Major Growing Season	Received any remittances	Earned any wage income (including from workfare program)
Two Year								
Treatment	992	79.7	74.1	9.07	4.24	0.177	-0.028	-0.003
	(97.2)	(5.71)	(5.54)	(2.70)	(1.26)	(0.071)	(0.023)	(0.007)
Neighbors within 200m	41.7	2.11	1.669	-3.34	-0.248	-0.012	0.020	-0.006
	(41.2)	(2.65)	(2.53)	(1.05)	(0.494)	(0.034)	(0.012)	(0.004)
Treated Neighbors within 200m	-100	-2.31	-0.387	2.10	-0.904	-0.030	-0.012	0.004
	(58.7)	(3.69)	(3.59)	(1.78)	(0.753)	(0.042)	(0.016)	(0.005)
Observations	843	843	843	844	844	844	840	844
Mean: No neighbors within 200m, control	1401	20.4	13.8	34.4	10.9	1.23	0.092	1.00
Three Year								
Treatment	1,026	31.8	32.0	9.86	3.09	0.231	-0.074	-0.045
	(93.5)	(5.69)	(5.49)	(2.62)	(0.858)	(0.069)	(0.028)	(0.017)
Neighbors within 200m	37.8	4.51	3.628	-1.98	-0.138	-0.026	0.009	-0.002
	(40.2)	(2.66)	(2.54)	(1.19)	(0.388)	(0.030)	(0.013)	(0.008)
Treated Neighbors within 200m	-77.0	-3.65	-0.797	-0.861	-0.259	-0.030	-0.012	0.004
	(56.8)	(2.15)	(2.26)	(1.74)	(0.568)	(0.039)	(0.018)	(0.011)
Observations	844	844	844	844	844	844	844	844
Mean: No neighbors within 200m, control	1498	17.5	13.1	36.0	6.89	1.39	0.198	0.962
Seven Year								
Treatment	428	39.7	5.294	8.13	4.05	0.266	0.046	-0.088
	(111)	(20.4)	(3.49)	(4.14)	(1.33)	(0.101)	(0.025)	(0.022)
Neighbors within 200m	-32.9	15.3	3.35	-1.55	0.570	-0.005	0.013	0.012
	(50.3)	(8.63)	(1.71)	(1.63)	(0.663)	(0.037)	(0.011)	(0.010)
Treated Neighbors within 200m	-72.2	-18.5	-3.97	-4.43	-1.34	-0.094	-0.009	0.007
	(68.1)	(9.85)	(2.58)	(2.30)	(0.926)	(0.049)	(0.016)	(0.013)
Observations	844	844	844	844	844	844	844	844
Mean: No neighbors within 200m, control	2362	46.1	28.5	49.5	18.0	1.25	0.107	0.939

Appendix Table A18 reports spillover effects of the graduation program on all households. The sample includes the 844 households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A19: Spatial Spillovers - Resource Sharing - All Households

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Remittances Received, Last 12 Months	Remittance Income per Month	Received any Informal Loans	Cash Value of Informal Loans Received	Received any Food as Gifts	Number of Types of Food Received as Gift
Two Year						
Treatment	-0.028	1.09	-0.030	19.4	0.010	-0.033
	(0.023)	(1.24)	(0.029)	(18.0)	(0.031)	(0.049)
Neighbors within 200m	0.020	-0.146	0.002	-13.5	0.028	0.042
	(0.012)	(0.646)	(0.012)	(6.24)	(0.014)	(0.019)
Treated Neighbors within 200m	-0.012	-0.054	-0.024	8.91	-0.014	-0.022
	(0.016)	(0.552)	(0.019)	(8.23)	(0.021)	(0.027)
Observations	840	844	844	844	844	844
Mean: No neighbors within 200m, Control	0.092	1.09	0.718	103	0.244	0.443
Three Year						
Treatment	-0.074	1.72	0.006	30.1	0.008	0.036
	(0.028)	(3.08)	(0.027)	(11.9)	(0.026)	(0.043)
Neighbors within 200m	0.009	0.720	-0.007	-6.59	0.008	0.014
	(0.013)	(1.52)	(0.011)	(4.578)	(0.012)	(0.018)
Treated Neighbors within 200m	-0.012	-1.51	-0.003	3.62	-0.003	0.005
	(0.018)	(1.560)	(0.016)	(6.829)	(0.017)	(0.026)
Observations	844	844	844	844	844	844
Mean: No neighbors within 200m, Control	0.198	6.51	0.771	150	0.176	0.282
Seven Year						
Treatment	0.046	4.79	0.011	29.5	-0.030	-0.130
	(0.025)	(4.00)	(0.019)	(49.280)	(0.033)	(0.072)
Neighbors within 200m	0.013	0.238	-0.005	2.30	-0.019	0.041
	(0.011)	(1.75)	(0.008)	(9.964)	(0.015)	(0.030)
Treated Neighbors within 200m	-0.009	2.46	-0.005	-6.61	0.045	-0.034
	(0.016)	(3.72)	(0.010)	(13.790)	(0.022)	(0.045)
Observations	844	844	844	844	844	844
Mean: No neighbors within 200m, Control	0.107	6.25	0.084	56.0	0.641	1.37

Appendix Table A14 reports spillover effects of the graduation program on sample households. The sample includes the844 households for whom we both have their GPS coordinates (from year seven), and who at year seven, are not more than 6 kilometers from the median GPS point of individuals they shared a baseline village with. All financial outcomes are reported in 2021 USD in PPP terms. Each regression controls for baseline values (when available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A20: Cost-benefit analysis

Panel A: Program Costs per Household, USD PPP 2021		
(1)	Direct Transfer Costs	1371
	Asset Cost	1371
	Food stipend	0
	Total Supervision Costs	2122
	Salaries of Implementing Organization Staff	387
	Materials	37
	Training	949
	Travel Costs	194
	Other Supervision Expenses	554
	Total Direct Costs	3493
	Start-up expenses	48
	Indirect Costs	470
(2)	Total Costs, calculated as if all incurred immediately at beginning of Year 0	4011
Panel B: Benefits per Household, USD PPP, All Values Deflated to Baseline at 5% annual social discount rate		
(3)	Year 1 Annual Consumption ITT, assuming treatment effect equal to Year 2	509
(4)	Year 2 Annual Consumption ITT Treatment Effect	484
(5)	Year 3 Annual Consumption ITT Treatment Effect	484
(6)	Year 7 Annual Consumption ITT Treatment Effect	280
	Estimated Benefits, Years 4-6, assuming linear decay from Year 3 to 7	1125
(7)	Projected Future Benefits, Years 8-16, assuming linear decay from Year 3 to Year 7 continues	1012
(8)	Total Estimated and Projected Consumption Benefits, (3) + (4) + (5) + (6) + (7)	3894
Panel C: Benefit-Cost Ratio		
	Total Consumption Benefits divided by Costs, (8) / (2)	0.971
	Total Consumption Benefits divided by Direct Transfer to Recipients, (8) / (1)	2.839
	Consumption Benefits Realized by Year 7, Divided by Costs, ((8) - (7)) / (1)	0.718
	Total Consumption Benefits divided by Costs, if not discounting future	1.233
	Share of bootstraps in which benefits exceed costs	0.519
	Median benefit-cost ratio from bootstrapped estimates	1.024
	95% Confidence Interval for benefit-cost ratio from bootstrapped estimates	[0.42, 2.36]

Appendix Table A20 presents cost-benefit estimates. Costs are reported by the implementing partner, and converted to 2021 USD in PPP terms. Benefits are calculated as equal to the sum of accumulated and projected future consumption benefits. We calculate the benefits in years 2, 3, and 7 from our consumption modules (and scaled to annual values). We assume that the decay in consumption benefits from Year 3 to 7 is linear, and that this decay continues linearly until the benefits reach 0, by Year 16.

Appendix Table A21: Asset Ownership

	(1)	(2)	(3)
	Total Asset Value	Productive Asset Value	Durable Good Value
Treatment (ITT): Two Year	1,162	931	61.8
	(112)	(91.0)	(20.0)
Control mean	1659	1239	207
Observations	915	915	910
Treatment (ITT): Three Year	1,158	914	72.5
	(109)	(88.4)	(17.8)
Control mean	1835	1364	206
Observations	908	908	908
Treatment (ITT): Seven Year	525	448	84.1
	(133)	(122)	(39.4)
Control mean	2618	2193	424
Observations	889	889	889
Baseline Mean	1201	1157	44.1

Appendix Table A21 reports average treatment effects of the graduation program on asset ownership (and its component parts), the indexed version of which is reported in Figure 1 and Appendix Table A4. Each regression controls for baseline values (which we construct using relative prices in subsequent waves--we observe baseline quantities of the same asset types, but not baseline values.). We additionally control for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A22: Food Security

	(1)	(2)	(3)	(4)	(5)
	Everyone gets enough food every day	No Adults skipped meals	No HH member went a whole day without food	No children skipped meals	Everyone eats at least two meals every day
Treatment (ITT): Two Year	0.039	0.063	0.029	0.045	0.018
	(0.031)	(0.030)	(0.017)	(0.022)	(0.018)
Control mean	0.640	0.656	0.913	0.845	0.910
Observations	914	910	910	910	909
Treatment (ITT): Three Yea	0.077	0.051	0.041	0.047	0.025
	(0.030)	(0.030)	(0.015)	(0.023)	(0.016)
Control mean	0.640	0.687	0.921	0.833	0.923
Observations	907	907	906	904	906
Treatment (ITT): Seven Yea	0.023	0.021	-0.007	0.011	0.012
	(0.025)	(0.023)	(0.009)	(0.018)	(0.019)
Control mean	0.826	0.850	0.987	0.926	0.908
Observations	889	889	889	802	889
Baseline Mean	0.335	0.370	0.745	0.489	0.827

Appendix Table A22 reports average treatment effects of the graduation program on food security. A subset of these variables is reported in Figure 2 and in Appendix Table A5, replicating the procedure of Banerjee et al. 2021. Each of these variables is used to construct the food security index in Figure 1 and Appendix Table A4. Each regression controls for baseline values (where available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A23: Financial Inclusion

	(1)	(2)	(3)	(4)
	Amount borrowed, last 12 months, formal sources	Amount borrowed, last 12 months, informal sources	Savings balance	Savings deposits, last 3 months
Treatment (ITT): Two Year	37.3	19.6	849.7	43.6
	(14.9)	(16.5)	(38.6)	(4.44)
Control mean	22.3	117	91.1	10.1
Observations	915	915	915	915
Treatment (ITT): Three Year	42.0	28.9	315.9	9.47
	(19.7)	(11.3)	(41.9)	(4.95)
Control mean	34.61	135	84.9	7.55
Observations	908	908	908	908
Treatment (ITT): Seven Year	-8.961	26.7	75.9	17.1
	(36.5)	(45.9)	(55.3)	(14.4)
Control mean	102	55.6	244	43.9
Observations	889	889	889	889
Baseline Mean	42.3	4.70	-	-

Appendix Table A23 reports average treatment effects of the graduation program on financial inclusion. Each of these variables is used to construct the financial inclusion index in Figure 2 and Appendix Table A4. Each regression controls for baseline values (where available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

Appendix Table A24: Productive Time Use

	(1)	(2)	(3)	(4)	(5)
	Minutes working per day on average, last 48 hours	Minutes working in agriculture on average, last 48 hours	Minutes working with livestock on average, last 48 hours	Minutes working on non-farm enterprise on average, last 48 hours	Minutes working in wage labor on average, last 48 hours
Treatment (ITT): Two Year	58.5	21.3	37.8	6.44	-6.14
	(12.0)	(9.34)	(7.58)	(3.54)	(6.35)
Control mean	225	96.4	80.3	2.86	45.0
Observations	834	834	834	834	834
Treatment (ITT): Three Year	47.1	22.6	30.9	4.41	-9.80
	(12.4)	(10.55)	(7.88)	(3.03)	(6.78)
Control mean	245	125	75.7	3.59	40.4
Observations	825	825	825	825	825
Treatment (ITT): Seven Year	16.9	8.2	6.52	3.40	0.176
	(10.4)	(9.04)	(6.90)	(2.65)	(4.43)
Control mean	215	103	88.1	3.96	20.4
Observations	1,228	1,228	1,228	1,228	1,228
Baseline Mean	118	18.5	39.6	2.67	57.5

Appendix Table A24 reports average treatment effects of the graduation program on productive time use. The aggregate (in column 1) is reported in Figure 1 (rescaled to have baseline mean 0 and standard deviation 1) and Appendix Table A4. Each regression controls for baseline values, and for tabia-level strata. Standard errors are clustered at the household level.

Appendix Table A25: Physical and Mental Health

	Physical Health Variables			*Mental Health Variables*		
	(1)	(2)	(3)	(4)	(5)	(6)
	No days of work missed due to poor physical health	Mean Score, Activites of Daily Living (0/1)	Perception of Physical Health (1/5)	Overall Satisfaction with Life (1/5)	No extended period of time with worry	Stress Index
Treatment (ITT): Two Year	0.004	-0.007	-0.029	0.146	0.004	-0.007
	(0.013)	(0.015)	(0.066)	(0.070)	(0.020)	(0.054)
Control mean	0.949	0.861	3.620	3.52	0.860	0.000
Observations	1,301	1,292	1,292	1,307	1,307	1,307
Treatment (ITT): Three Year	0.005	0.028	-0.028	0.083	-0.025	-0.059
	(0.010)	(0.014)	(0.060)	(0.061)	(0.015)	(0.059)
Control mean	0.959	0.879	3.731	3.55	0.939	0.000
Observations	1,263	1,256	1,256	1,265	1,264	1,265
Treatment (ITT): Seven Year	-0.002	-0.018	-0.102	-0.075	-0.009	-0.066
	(0.012)	(0.017)	(0.063)	(0.060)	(0.019)	(0.066)
Control mean	0.954	0.848	3.705	3.80	0.896	0.000
Observations	1,228	1,227	1,228	1,228	1,227	1,228
Baseline Mean	0.907	0.874	3.938	3.52	0.854	0.009

Appendix Table A25 reports average treatment effects of the graduation program on physical and mental health outcomes. The physical health index in Figure 1 and Appendix Table A4 is comprised of the variables in columns 1-3, while the mental health index is comprised of the variables in columns 4-6. Each regression controls for baseline values, and for tabia-level strata. Standard errors are clustered at the household level.

Appendix Table A26: Political and Women's Empowerment

	Political Involvement			***Women's Decision-Making***				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Individual has attended meeting with local leader or politician	Individual has asked a question of local leader or politician at meeting	Individual is member of political party	Women has major say in food-related spending decisions in household	Women has major say in education-related spending decisions in household	Women has major say in healthcare-related spending decisions in household	Women has major say in home improvement spending decisions in household	Women has major say in household management decisions in household
Treatment (ITT): Two Year	0.047	-0.009	0.061	-0.045	-0.007	-0.013	0.025	-0.020
	(0.026)	(0.025)	(0.026)	(0.030)	(0.030)	(0.030)	(0.027)	(0.028)
Control mean	0.613	0.296	0.349	0.735	0.491	0.556	0.406	0.402
Observations	1,307	1,307	1,306	758	742	757	755	745
Treatment (ITT): Three Year	0.047	0.004	0.061	0.010	-0.009	-0.015	-0.023	-0.030
	(0.028)	(0.023)	(0.027)	(0.029)	(0.031)	(0.032)	(0.030)	(0.031)
Control mean	0.522	0.216	0.328	0.780	0.492	0.563	0.447	0.393
Observations	1,262	1,265	1,265	711	717	725	722	655
Treatment (ITT): Seven Year	0.012	-0.003	-0.001	0.003	-0.025	-0.025	-0.046	-0.041
	(0.028)	(0.022)	(0.028)	(0.027)	(0.034)	(0.034)	(0.033)	(0.035)
Control mean	0.510	0.192	0.349	0.836	0.652	0.661	0.707	0.615
Observations	1,227	1,227	1,228	723	723	723	723	723
Baseline Mean	0.350	0.184	0.303	0.473	0.411	0.446	0.409	0.383

Appendix Table A26 reports average treatment effects of the graduation program on physical and mental health outcomes. The political empowerment index in Appendix Table A4 is comprised of the variables in columns 1-3, while the women's empowerment index is comprised of the variables in columns 4-8. Each regression controls for baseline values, and for tabia-level strata. Standard errors are clustered at the household level.

Appendix Table A27: Key Financial Variables, Inflated using National CPI

	(1)	(2)	(3)	(4)	(5)
	Total Asset Value	Monthly Per Capita Consumption	Livestock Revenues, Monthly	Total Value of Livestock	Total Savings Balance
Treatment (ITT): Two Year	1,274	9.01	141	883	932
	(122.40)	(2.55)	(9.55)	(83.2)	(42.3)
Control mean	1820	55.0	35.6	1223	100.0
Observations	915	915	915	914	915
Treatment (ITT): Three Year	1,231	9.10	32.8	905	336
	(116.30)	(1.95)	(6.64)	(81.5)	(44.5)
Control mean	1951	49.9	31.4	1356	90.3
Observations	908	908	908	908	908
Treatment (ITT): Seven Year	391	4.39	30.0	312	56.58
	(99.3)	(2.09)	(15.6)	(80.5)	(41.2)
Control mean	1951	49.8	67.0	1512	182.1
Observations	889	889	889	889	889
Baseline Mean	1342	-	-	1237	-

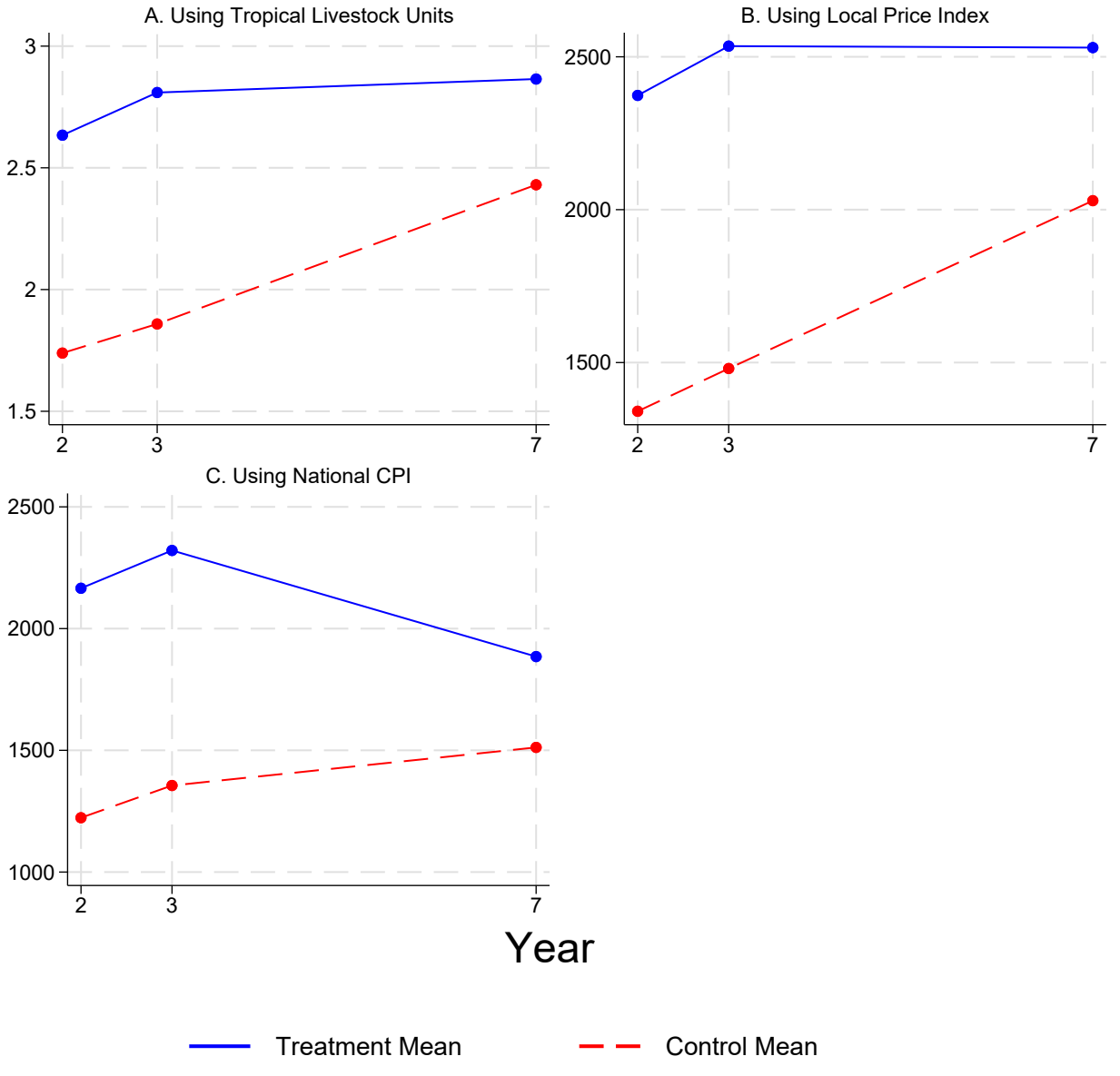
Appendix Table A27 reports average treatment effects of the graduation program on key financial outcomes. In contrast to our base case, in which we use a price index based on the prices of food faced by sample households, here we use Ethiopia's national consumer price index to convert waves to be in constant (USD PPP 2021) prices. A discussion of both methods can be found in Appendix A. Each regression controls for baseline values (where available), and for tabia-level strata. Standard errors are Huber-White heteroskedastic.

F. Appendix Figures

List of Figures

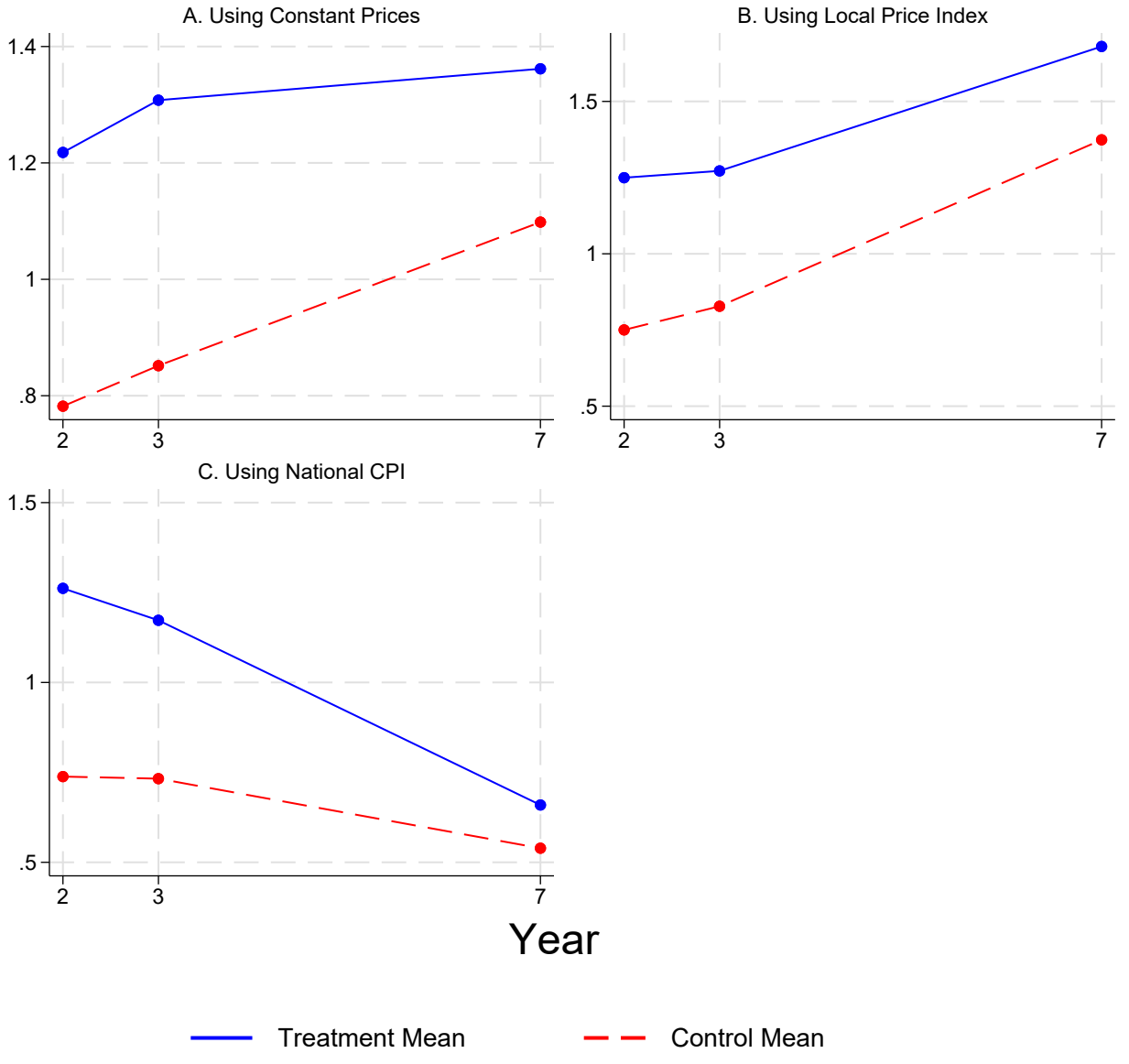
A1 – Validating Local vs. National Prices: Livestock Ownership
A2 – Validating Local vs. National Prices: Non-Livestock Asset Ownership
A3 – Validating Local vs. National Prices: Welfare Measures

Appendix Fig. 1: Validating Local vs. National Prices: Livestock Ownership



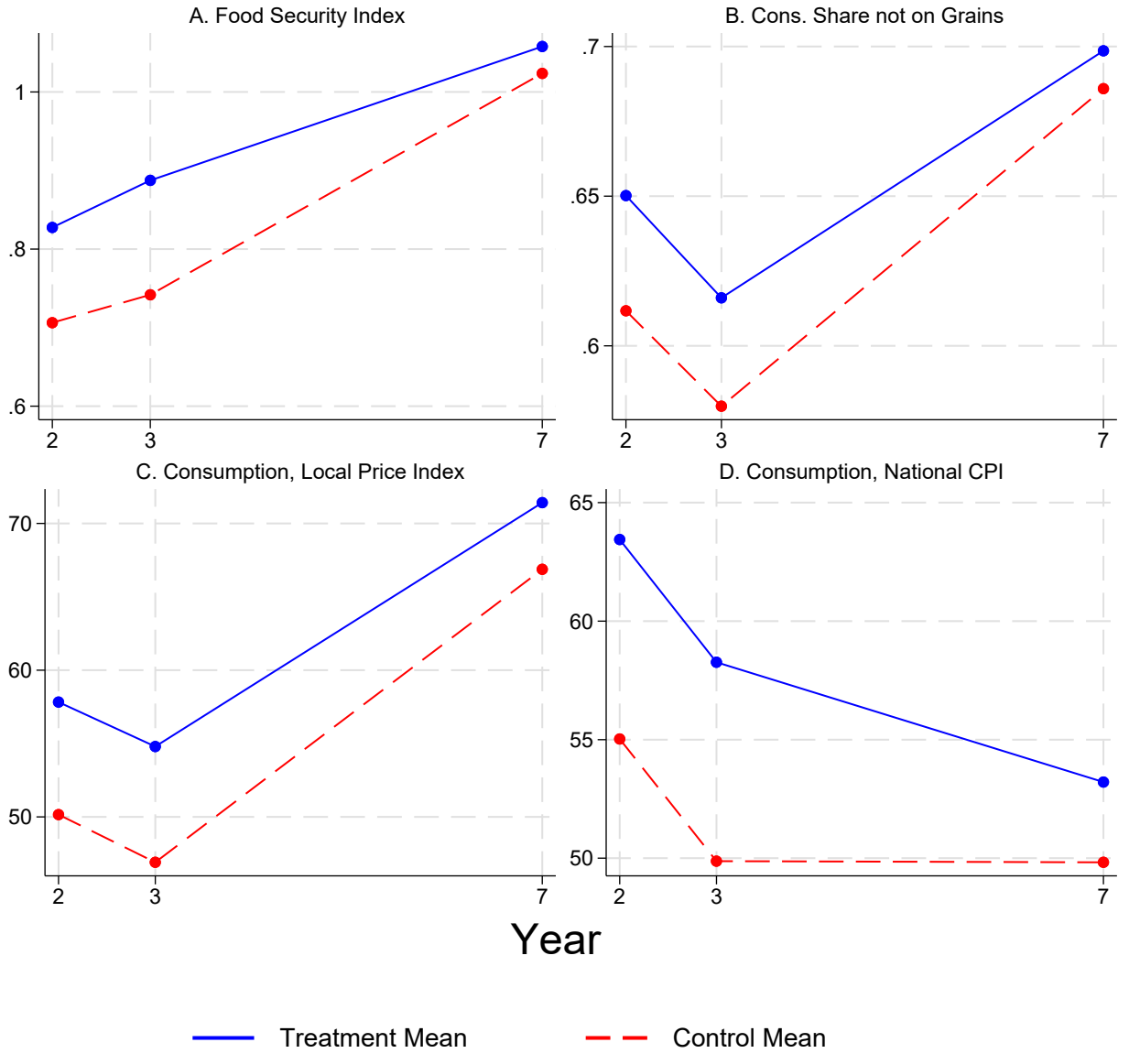
Notes: Appendix Figure 1 compares the evolution of livestock ownership for treatment and control groups over time under different measures of cross-wave prices, to assess the relative performance of using a local price index to measure inflation (the method used in our paper) versus the national consumer price index. Panel A reports the evolution in real terms (using Tropical Livestock Units), Panel B reports the evolution using our local price index, and Panel C reports the evolution using the national CPI.

Appendix Fig. 2: Validating Local vs. National Prices: Non-Livestock Asset Ownership



Notes: Appendix Figure 2 compares the evolution of non-asset livestock wealth for treatment and control groups over time under different measures of cross-wave prices, to assess the relative performance of using a local price index to measure inflation (the method used in our paper) versus the national consumer price index. Panel A reports the evolution in real terms (multiplying quantities owned by constant relative asset prices), Panel B reports the evolution using our local price index, and Panel C reports the evolution using the national CPI.

Appendix Fig. 3: Validating Local vs. National Prices: Welfare



Notes: Appendix Figure 3 compares the evolution of welfare for treatment and control groups over time under different measures of cross-wave prices, to assess the relative performance of using a local price index to measure inflation (the method used in our paper) versus the national consumer price index. Panels A and B report the evolution in real terms (using our food security index in Panel A, and the share of consumption on non-grain goods in Panel B), Panel C reports the evolution of per capita consumption using our local price index, and Panel D reports the evolution using the national CPI.