NBER WORKING PAPER SERIES

DOES POVERTY CHANGE LABOR SUPPLY? EVIDENCE FROM MULTIPLE INCOME EFFECTS AND 115,579 BAGS

Abhijit Banerjee
Dean Karlan
Hannah Trachtman
Christopher R. Udry

Working Paper 27314
http://www.nber.org/papers/w27314

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020, Revised November 2025

Approval from the Yale University Human Subjects Committee, IRB No. 0705002656, 1002006308, 1006007026, and 1011007628; and from the Innovations for Poverty Action Human Subjects Committee, IRB Protocol No. 19.08January-002, 09December-003, 59.10June-002, and 10November-003.494. Thanks to the Ford Foundation, and 3ie for funding. Thanks to Nathan Barker, Caton Brewster, Abubakari Bukari, David Bullon Patton, Alexis Devonish, Sébastien Fontenay, Angela Garcia, Yann Guy, Samantha Horn, Sana Khan, Hideto Koizumi, Matthew Lowes, Elizabeth Naah, Augustine Owusu, Michael Polansky, Elana Safran, Sneha Stephen, Rachel Strohm, and Stefan Vedder for outstanding research assistance and project management, and in particular Bram Thuysbaert for collaboration. The authors would like to thank the leadership and staff at Presbyterian Agricultural Services (PAS) for their partner-ship. Thanks to Frank DeGiovanni of the Ford Foundation, Syed Hashemi of BRAC University, and Aude de Montesquiou and Alexia Latortue of CGAP for their support and encouragement of the research. No authors have any real or apparent conflicts of interest, except Karlan was on the Board of Directors of Innovations for Poverty Action, which participated in oversight of the implementation. All data and code will be available upon publication at the IPA Dataverse (doi pending). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Does Poverty Change Labor Supply? Evidence from Multiple Income Effects and 115,579 Bags

Abhijit Banerjee, Dean Karlan, Hannah Trachtman, and Christopher R. Udry

NBER Working Paper No. 27314
June 2020, Revised November 2025
JEL No. H31, J22, O12

ABSTRACT

The income elasticity of labor supply is a central parameter of many economic models. We test the response of labor supply to exogenous changes in income using data from a randomized evaluation of a multi-faceted grant program in northern Ghana combined with a researcher-implemented bag-making operation. We find a non-negative "income effect" on labor supply. We argue that simple models with either labor or capital market frictions cannot explain the results, whereas a model that allows for positive physiological or psychological productivity effects from higher income fits with our findings.

Abhijit Banerjee
Massachusetts Institute of Technology
Department of Economics
and NBER
banerjee@mit.edu

Dean Karlan
Northwestern University
and CEPR
and also NBER
dean.karlan@gmail.com

Hannah Trachtman
The Hebrew University of Jerusalem
Department of Economics
hannah.lukzilberman@mail.huji.ac.il

Christopher R. Udry
Northwestern University
Department of Economics
and NBER
christopher.udry@northwestern.edu

1 Introduction

The income elasticity of labor supply is a central parameter of many economic models. The standard assumption that consumption and work are not strong complements drives the common prediction that higher income reduces labor supply. This affects optimal redistributive policy design because such labor supply reductions hamper net income gains.

The basic argument for why we should expect this negative labor supply response is well-known. If utility from consumption is $u(c)$, the disutility of labor supply is $v(l)$ and the relation between consumption and labor supply is $c = f(l) + t$, where f is income and is some increasing concave function of labor supply and t is a transfer, we immediately get a first order condition

$$u'(f(l) + t)f'(l) = v'(l)$$

from which it follows that any increase in t will reduce the marginal utility of income and therefore labor supply. Of course, several important assumptions are being made here. First, as pointed out by Benjamin (1992), we need that t does not directly raise the marginal product of labor (we cannot have $f(l, t)$ with $f_{lt}(l, t) > 0$). As Benjamin (1992) also points out, this is typically ruled out by the assumption of either perfect capital markets (in which case t should not enter $f(l, t)$) or perfectly substitutable household labor and market labor (in which case $f_l(l, t)$ equals the market wage). However neither assumption seems particularly plausible for low income families in developing countries (LaFave et al., 2025). A transfer may directly raise the marginal product of labor, making this kind of *investment productivity effect* quite relevant.

A second reason why the expected income effect may be absent for the very poor is that consumption (or income) and labor supply may be complements, so that the disutility of effort takes the form $v(l, c)$ with $v_{lc}(l, c) < 0$. The idea that a mechanical nutrition-productivity relationship generates complementarity between consumption and work lies at the heart of the earliest poverty trap models in which better-fed workers provide more effort (Leibenstein, 1957; Dasgupta and Ray, 1986). More generally, higher consumption may result in better health, which in turn reduces the disutility of work (Strauss and Thomas, 1998). We call this a *physiological productivity effect*.

More recently, psychological poverty trap models have made a similar case for why, at low levels of psychological well-being, positive income shocks may boost labor supply—what we call a *psychological productivity effect*. People living under scarcity may exhibit "tunnel vision," focusing intently to allocate scarce resources at the expense of other margins (Mullainathan and Shafir, 2013; Fehr et al., 2022), including,

in some cases, productivity (Fink et al., 2020; Pinto, 2021; Kaur et al., 2025). Other models propose that both individuals (Dalton et al., 2016) and economies (Genicot and Ray, 2017) may get trapped in poverty when aspirations and outcomes are jointly determined (Lybbert and Wydick, 2018; Bernard et al., 2023).

Consistent with this set of theories, syntheses and meta-analysis of cash transfer evaluations find income effects on labor supply are often non-negative (Baird et al., 2018; Banerjee et al., 2017; Crosta et al., 2024). Banerjee et al. (2015) and Bandiera et al. (2017) report on impact in seven sites of the Graduation program, a multi-faceted program built around an asset transfer to very poor households, and find higher incomes and labor supply. The positive impacts persisted three years after households received productive assets, and subsequent work shows persistence, of mixed magnitude, up to ten years later (Banerjee et al., 2021; Balboni et al., 2022; Barker et al., 2024). While the graduation program potentially changes persistent household circumstances and activities, other work shows that more temporary positive shocks also have a *positive* effect on labor supply of low-income households: e.g., in a field experiment with piece-rate workers in India, Kaur et al. (2025) finds higher levels of productivity on days when workers are cash-rich, and argues this is indicative of improved cognition and focus due to lower levels of financial stress.

Yet the above body of evidence has three potential limitations. First, hours worked does not measure the entirety of labor supply. For example, if much of the labor supply response is in the form of reduced (unmeasured) effort on a job, the person may eventually be fired, but we do not observe this long-term outcome.

Second, the Graduation experiments were not designed to distinguish between different possible mechanisms. The intervention involved both an asset transfer to plausibly credit constrained households (so an increase in t, which may shift $f(l,t)$) as well as encouragement and training intended to shift their $v(.)$ functions. The physiological or psychological effect of extra income and the non-monetary components of the Graduation program may contribute to non-negative income effects, but there is no way to tell from the data.

Finally, the productivity effects of temporary income shocks studied in Kaur et al. (2025) are indeed likely to be psychological: the immediacy of the impact makes it unlikely due to productive asset investments (unlike in Gertler et al. (2012)). We complement the Kaur et al. (2025) lab-in-the-field experiment, investigating whether sustained productivity effects may result from an intervention that substantially and durably increases incomes. It is precisely these larger and more prolonged transfers that policy-makers worry about.[1]

[1]Indeed, the position of the literature, as summarized by Baird et al. (2018) is that "the income effect underlying the labor-leisure trade-off appears most apparent when transfers are large and/or prolonged (as with lottery winnings and pensions), and appears to

With this context in mind, we make two contributions by building on our study of the Ghana Graduation program (also called "Graduating the Ultra Poor" or "GUP") reported on in Banerjee et al. (2015). First, we provide better measurement of labor supply in the context of a durable increase in the earnings of a household, and still find a non-negative income effect on labor supply over a non-trivial period of time (six months). Second, we provide evidence that the psychological or physiological productivity effect is driving the observed departure from the traditionally-hypothesized income effect.

A key to both contributions is a novel measurement exercise involving a bag-making operation. Treatment and control villages were randomly chosen to have bag production units. Those who were invited to work in these units were offered piece rate contracts to produce bags, and all inputs were provided. We counted, and carefully graded the quality of, the bags produced. The piece rate depended on the quality. Thus we consider bag output to be a reliable byproduct of effort, construed as a combination of hours worked and attention and diligence. Each bag-making unit was also randomly assigned to produce either simple or more complex bags, to test marginal productivity with respect to task complexity.

For those in the bags production sub-groups, the comparison of GUP and Control households tells us that GUP increases participation in bags, bags production, and earnings from bags by 20-30% (with p-values of 0.02, 0.03, and 0.06, respectively; and multiple-hypotheses q-values of 0.12 for each). If anything, the effects are stronger for those assigned complex bags: GUP households make 48% more complex bags than Control households (p=0.07, q=0.16).

Of course, this increased labor supply could reflect households cutting back on other activities. For hours worked, we estimate that GUP households supply only about two percent fewer hours to all forms of productive labor than do Control households (p=0.49, q=0.55). Survey data on agricultural outcomes, business outcomes, and wage income suggest GUP households neither make large labor-saving investments nor neglect their agricultural or non-agricultural businesses; in other words, our time-use measures are not likely to be missing important dimensions of effort.

Strikingly, GUP households supply more overall labor despite earning substantially more than Control households. Summing up across all the sources of earnings plus any cash transfers, during bag-making GUP households earned $9.4 more per month than Control households, excluding bags earnings (p<0.01, q<0.01). This implies that they earned more than double the Control monthly earnings of $8 while supplying more labor to bags production and no less labor elsewhere. In other words there is prima facie evidence of a *positive* GUP effect on labor supply.

be much weaker for one-time transfers." Recent evidence has substantiated this notion in the U.S. context (Vivalt et al., 2025) and in Poland (Gromadzki, 2024).

Turning to our second question, we argue that a physiological or psychological productivity effect, rather than an investment productivity effect, underlies our result. This is because bag-making offered no scope for additional investment by the households: all capital was provided by us, the researchers.

This still leaves the question of whether the GUP effect is merely an income effect, given the multifaceted program. The experimental design included several arms that allow us to investigate the role of access to savings as well as the role of pure income. Two points ensue: First, access to savings does not boost labor supply. Second, GUP participants with randomly larger cash transfers do not respond any less positively. Ultimately, the evidence for a pure income effect is not dispositive, and we cannot decisively rule out the GUP encouragement component contributing to higher labor supply.

We contribute to a large literature on labor markets in developing countries (e.g. Lewis (1954); Rosenzweig (1988); Foster and Rosenzweig (1996); Goldberg (2016); Guiteras and Jack (2018)). Our work builds on analysis of effects of credit constraints on labor supply (e.g. Kochar (1999); Rose (2001); Jayachandran (2006); Fink et al. (2020)); effects of positive income shocks on labor supply (e.g. Baird et al. (2018); Kaur et al. (2025)); and, determinants of effort (e.g. Breza et al. (2018); Brune et al. (2021); Kaur et al. (2015)), including the potential importance of psychological well-being and its effect on income (Shah et al., 2012; Mani et al., 2013; Haushofer and Fehr, 2014).

2 Experimental Design

We partnered with Presbyterian Agricultural Services (PAS), a local NGO in northern Ghana, whose field agents engaged in the direct field implementation of the Graduation program (GUP), the Savings-Only program (SOUP), and the employment program (Bags). PAS operates in three agricultural "stations" in Northern Ghana: Tamale, Langbensi, and Sandema.

2.1 GUP and SOUP

Table 1 Panel A shows the assignment of households and villages to GUP, SOUP and Control, and the cross-cutting bags measurement village assignments. Each village was assigned GUP, SOUP, or Control, and then within each treatment village, half of sample households actually received the treatment intervention, and half served as control households within treatment villages.[2]

In GUP villages, half of sample households were assigned to the GUP treatment. GUP included six

[2]Note that this is the same sample that was studied in Banerjee et al. (2015) and Banerjee et al. (2022).

components, all directed to the female household head: (1) productive asset transfer; (2) skills training for the productive asset, (3) life skills training and mentorship via two years of weekly household visits, (4) weekly lean-season cash (between $6 and $9 PPP depending on family size) for consumption support, (5) basic health services education, and (6) a local bank savings account and deposit collection. Additional details are described in Appendix A.1.

The sixth component of GUP, access to a savings account and weekly deposit collection by PAS field agents, was the entirety of the SOUP intervention. In SOUP villages, slightly more than half of sample households were assigned to the SOUP treatment. These households received a visit from the field officer to collect savings, but did not receive any other components of the program.

2.2 Bag-Making

We designed an employment program ("Bags") offering wages for the production of cloth bags, and implemented it such that it cross-cut GUP, SOUP, and control assignments. Half of the villages were randomly assigned to Bags, as shown in Table 1 Panel A. In GUP and SOUP villages assigned to Bags, all sample households assigned to GUP or SOUP were invited to participate. In Control villages assigned to Bags, half of sample households were invited to participate. For logistical reasons, we assigned all pure Control villages with fewer than 30 compounds to No-Bags; we thus restrict all of our analysis to villages with at least 30 compounds. This leaves 93 Bags villages and 72 No-Bags villages. Our sample includes 896 Bags households: 397 Control, 221 GUP, and 278 SOUP.

Table 1 Panel B presents the details of two sub-treatments within the bags measurement exercise. First, among Bags villages, half were assigned to produce a simple bag, and half were assigned to produce a complex bag. While the simple bag has basic "running stitches" on the hem and the strap, the complex bag alternates one "running stitch" with four more difficult "chain stitches," in a pattern that requires counting. (See Appendix Figure 1.) The purpose of this variation was to test whether income effects are more important for complex tasks. Second, we varied the amount of unconditional consumption support, in the form of a cash transfer, received by GUP-Bags households. This was varied at the village level, and was either $1.31 or $3.92 weekly.[3] The purpose of this variation was to test for a pure income effect, without the encouragement that came with GUP.

The bag-making employment program ran for six months, from June through November 2012, and was

[3]This amount was chosen to make expected weekly transfers, inclusive of bags earnings ($2.6 on average), roughly equivalent to the $6-$9 received by GUP-No-Bags households.

designed to coincide with the lean season. The GUP and SOUP programs, which launched in July 2011, had been running for almost a year by the time bag-making began, and continued to run until July 2013. (See Appendix Figure 2 for a full timeline.)

Immediately prior to the start of the employment program, bag-making field agents invited female household heads to participate in a community-level training, which lasted for four days. During production, GUP, SOUP, and Control field agents visited each community on a weekly basis. At each visit, they collected new bags, distributed replacement fabric, and paid wages for bags submitted two weeks prior. Households could submit a maximum of ten bags per week. In the two weeks between when bags were collected and when wages were paid, quality checks were carried out by bag-making field agents, and bags were classified as high, mid, or low quality (see Appendix A.2 for details).

Wages were paid with a two-week lag. Upon each visit, GUP, SOUP, and Control field agents informed households of the composition of high, mid, and low quality bags submitted two weeks prior, and distributed payment accordingly. Baseline piece-rates for mid-quality bags were randomly varied each month at the village level to be either \$0.40 or \$0.91. Bags judged to be high/low quality earned the baseline wage plus/minus \$0.13. The wage was not affected by whether the bag was simple or complex, except through quality scores. Every four weeks, bag-making field agents returned to communities to give feedback and remedial training.[4,5]

3 A Model of Labor Supply

We provide a simple model to clarify the interplay between the GUP, SOUP and Bags programs and to explain how we use them together to conclude that there was a non-negative (and perhaps positive) income effect on labor supply, driven by a psychological or physiological productivity effect.

The utility from income c is given by $\lambda u(\frac{c}{\lambda}))$, where λ is a shifter for the utility function. A higher λ is meant to capture the impact of the savings component of the GUP intervention, which makes it possible to spread the extra consumption over a longer future, hence raising the marginal utility of income. For results 1 and 2 we keep λ fixed. The household production function is $f(l,t)$, where the inclusion of t represents the possibility that the transfers raise the marginal product of labor. In other words we assume that $f_l(l,t) > 0$, $f_{ll}(l,t) < 0$, $f_t(l,t) \geq 0$ and $f_{lt}(l,t) \geq 0$. As noted, a necessary condition for this is that there are imperfections in both the capital market and the labor market. This framework captures the three

[4] We discuss potential experimenter demand effects in Appendix C.

[5] At the end of the program, the bags were sold to other research teams, who used them as participant gifts.

components of the GUP program (the productive asset transfer, the skills training, and the consumption support) that have the potential to raise the marginal product of labor.

The disutility of labor supply l is given by $v(l,T)$, where the inclusion of T is aimed to capture the relation between the various interventions and labor supply, via physiological or psychological channels. In other words it is possible that $T = t$, but we want to allow for possibility of interventions that shift labor supply without providing an income transfer (such as through encouragement). We assume that $v_l(l,T) > 0$, $v_{ll}(l,T) > 0$, $v_T(l,T) \leq 0$ and $v_{lT}(l,T) \leq 0$. One case where we might expect $v_T(l,T) < 0$ and $v_{lT}(l,T) < 0$, is when $T = t$, income transfers boost consumption and greater consumption raises labor supply. Another would be a coaching/encouragement treatment, where $t = 0$ but $T > 0$. Within this framework, two components of the GUP program (life skills training, and basic health services and health education) may have direct effects on the disutility of work via T. Finally we assume that $c = f(l,t) + t$. The asset transfer and consumption support components of GUP enter the budget constraint through t.

The first order condition for utility maximization is

$$u'\left(\frac{f(l,t)+t}{\lambda}\right) f_l(l,t) = v_l(l,T).$$

Suppose that $t = t(T)$ with $t'(T) > 0$. It is evident that $\frac{dl}{dT} < 0$ as long as $f_{lt}(l,t) = 0$ and $v_{lT}(l,T) = 0$. However $\frac{dl}{dT}$ can be positive if either $f_{lt}(l,t) > 0$ or $v_{lt}(l,T) < 0$. As before we call these two sources of a non-traditional income effects the investment productivity effect and the psychological/ physiological productivity effect.

Result 1: As long as λ is fixed, a necessary condition for the income effect on labor supply not to be negative is that there has to be either the investment productivity effect or the psychological/physiological productivity effect.

For our second result, we permit the household to have access to two production technologies, so that

$$c = f^a(l^a,t) + f^b(l^b,t) + t,$$

where $f^a(.)$ represents the bag making opportunity.

The household now maximizes

$$\lambda u\left(\frac{f^a\left(l^a,t\right) + f^b\left(l^b,t\right) + t}{\lambda}\right) - v(l^a + \gamma l^b, T)$$

by choosing l^a and l^b. γ represents the relative cost of effort in the two tasks. Now suppose $f^a_{l^a t}(l^a, t) = 0$. The first order condition with respect to l^a yields

$$u'\left(\frac{c}{\lambda}\right) f^a_{l^a}(l^a, t) = v_l(l, T)$$

We wish to compare $l^a(T)$ with $l^a(T')$ where $t(T) > t(T')$. Suppose $c(T) \geq c(T')$ and therefore $u'(c(T)) < u'(c(T'))$. Moreover let $l^b(T) \geq l^b(T')$. Then if it also true that $l^a(T) \geq l^a(T')$ then $l(T) \geq l(T')$. Now if $v_{lT}(l, T) = 0$, then $v_l(l(T), T) \geq v_l(l(T'), T')$. In this case the only way to satisfy the first order condition is for $f^a_{l^a t}(l^a, t) > 0$. Conversely, if $f^a_{l^a t}(l^a, t) = 0$ then it must be the case that $v_{lT}(l, T) < 0$. We summarize this as:

Result 2: As long as λ is fixed, if there is one activity where there is no investment productivity effect, and the labor supply to that activity is greater despite the fact the household is richer and is working no less, then there must be a psychological/physiological productivity effect on the disutility of labor.

The last observation is about λ. If λ goes up, say because of savings collection, the household's marginal utility of income goes up and therefore both its labor supply and its income must both go up.

Result 3: If λ goes up, the household's labor supply and its income must both go up.

4 Data and Empirical Methods

4.1 Data

We have three sources of data. First, we have weekly administrative data on labor supply (the number of bags submitted), the quality of each bag, and the resulting earnings. Of the 896 clients in the Bags sample, 88.7% chose to make bags at some point during the program, and on average, participants produced 3.9 bags per week. Among clients who participated in a given week, the average number of bags submitted was 6.7. Most people submitted either zero or 10 bags, as demonstrated in Appendix Figure 3a. Appendix Figure 3b shows the distribution of earnings, broken down by complex and simple bags. Both have a mode at zero (consistent with Figure 3a), and simple bags do earn more (undoubtedly because the task was easier, which led to more bag production and higher quality scores).

Second, we have time use surveys in which the female household head reported how she spent her time the previous day. We administered these surveys five times monthly during the Bags program, to 1,238 households, including almost all 896 Bags households and roughly 20% of No-Bags households. On

average, 80% of Bags households were found and surveyed each month. In our time use survey, rather than asking about time spent on bags directly, we asked only about "wage labor (including bags)" in order to maintain separation between the evaluation team and the team that was implementing the Bags program. We thus impute time on bags by taking reported time on wage labor, and subtracting average time on wage labor from the Control-No-Bags, GUP-No-Bags, and SOUP-No-Bags households for each Bags group, respectively. See Appendix B for details.

Third, we have a series of standard and comprehensive household surveys that were part of the larger program evaluation of the Graduation program (Banerjee et al., 2015), including a baseline survey, three shorter midline surveys (conducted with one third of the sample), a two-year follow-up survey and a three-year follow-up survey. These surveys included questions about income, consumption, agricultural outcomes, business outcomes, and welfare. The second midline survey is used heavily, as it took place during the Bags program. It includes 865 households: 288 Bags households and 577 No-Bags households.[6,7]

4.2 Orthogonality

Appendix Tables 1 and 2 show baseline indicators, including a wealth index that aggregates the seven variables related to overall wealth, across treatment groups. We fail to reject differences for 9 out of 14 variables examined, but do have imbalance on 5: average age, land area, monthly per capita consumption, monthly household income, and the food security index. We had intended to re-randomize, but due to a coding error, it did not happen. To be conservative, we control for these five baseline variables in all regressions, but our main results are robust to excluding them. Moreover, in each table, we report “baseline p-values" that correspond to treatment-control differentials, estimated using the same specification, for the baseline value of the outcome (when we have it).

4.3 Method of Analysis

We use two main specifications for our three types of data: one for the analysis of individual- or individual-month level outcomes measured in our surveys (Equation 1); and one for the analysis of individual-week level bag-making outcomes (Equation 2). Any deviations from these specifications are reported in table notes.

[6]We rely on these 288 observations to estimate treatment effects on income, expenditure, and revenue during the Bags program.
[7]We do not find differential attrition by treatment group for the key comparisons used in this paper; see Appendix D.

$$Y_{im}^{k} = \alpha^{k} + \beta^{k} T_i + \gamma^{k} Y_i^{k0} + \delta^{k} \mathbb{1}\{Y_i^{k0} \text{ missing}\} + W_i^{strata} + \theta^{k, interviewer_{im}} + \epsilon_{im}^{k} \quad (1)$$

$$Y_{iw}^{j} = \alpha^{j} + \beta^{j} T_i + W_i^{strata} + \rho^{station_i, w} + \epsilon_{iw}^{j} \quad (2)$$

In Equation 1, Y_{im}^{k} is outcome k measured for individual i in month m in the unbalanced panel of the midline surveys and time use surveys. T_i is an indicator for treatment. The treatment or set of treatments included varies across tables, and is described in the table notes. Y_i^{k0} is the baseline value of outcome k for individual i; we additionally include an indicator for Y_i^{k0} being missing for individual i. If the outcome k was not measured at all in the baseline survey, we exclude Y_i^{k0} and $\mathbb{1}\{Y_i^{k0} \text{ missing}\}$.[8] W_i^{strata} is a vector of baseline controls that consists of the variables we intended to use for re-randomization plus the five variables that were imbalanced at baseline, discussed in Section 4.2.[9] $\theta^{k, interviewer_{im}}$ denotes interviewer fixed effects. In Equation 2, Y_{iw}^{j} is outcome j measured for individual i in week w in the administrative bags-making data. $\rho^{station_i, w}$ denotes station-week fixed effects.[10] We cluster standard errors at the village level, since GUP/SOUP and Bags were assigned at the village level.

We use the Benjamini-Hochberg procedures (Benjamini and Hochberg, 1995) put forward in Anderson (2008) to compute q-values that correct for multiple hypotheses within each table or panel, excluding tests of linear combinations of coefficients which are dependent on individual coefficients. We do not extend these corrections beyond the boundary of an individual table (or panel) because the substantive aspects of the hypotheses we test change dramatically across tables.

In all analyses, we restrict the sample to villages with at least 30 compounds, as discussed in Section 2.2.

5 Results

5.1 Effects of GUP on income and labor supply

Table 2 reports the main results, the GUP and SOUP treatment effects for Bags households on both Bags and non-Bags labor supply. We focus on Bags households, so from now on we will refer to these households

[8] This is the case for time use measures and for bag production outcomes in Equation 2.

[9] The intended re-randomization variables include: household size, age of respondent, asset ownership index, whether the household owns a business, whether the respondent has savings, surface area of land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village.

[10] In Equation 1, enumerator fixed effects capture geographic differences (enumerators were assigned to different stations) as well as differences in interviewing style. In Equation 2, the outcome is objectively measured on a weekly basis, so we include station-week fixed effects to account for differences across geography and time.

simply as "GUP" "SOUP" or "Control."[11] We begin by discussing Panel A, which compares GUP, SOUP, and Control households.

In column 1 we look at monthly household income (excluding bags earnings) during the bag-making program. This is the sum of self-reported income at the midline that took place during Bags, plus monthly unconditional consumption support, which was received in varying amounts. GUP households earn more than double their Control counterparts (p<0.01, q=<0.01).

Columns 2 through 5 show impacts on bag-making. GUP participants produce 30% more bags (p=0.03, q=0.12), are 21% more likely to submit any bags (p=0.02, q=0.12), and earn 27% more from bags production than Control (p=0.06, q=0.12). On the other hand SOUP participants are actually less likely to produce bags than Control (p=0.06, q=0.15), and show lower overall production and earnings, though differences are not statistically significant. SOUP participants under-perform GUP participants on almost every measure (for example, there is a 24 percentage point gap in participation, with p<0.01).

Column 6 shows impacts on time spent on non-bags activities, including farming, businesses, animals, and home labor. GUP and SOUP households may spend a bit less time (9 minutes) on daily non-bags productive labor, but this difference constitutes only 1.6% of the control group mean and is not statistically different from zero (for GUP, p=0.49 and q=0.55).[12] In Table 3 columns 5-7 we see that this effect is composed of a reduction in time spent on agriculture combined with smaller increases in time spent on non-agricultural businesses and home labor. Column 9 indicates that GUP households are also spending slightly more time on leisure (13 minutes, p=0.08, q=0.22), which, given how many more bags they produce, may indicate improved productivity.

The differences in bag-making become more stark when we focus on complex bag production in Panel B of Table 2. GUP households produce 48% more complex bags than control households on a weekly basis (p=0.07, q=0.16).[13] SOUP households, on the other hand, produce 36% fewer complex bags than Control (p=0.04, q=0.14) and a fortiori than GUP.

5.2 Which components of GUP drive the result?

What is the source of the labor supply effect? We first consider the role of savings collection, since Result 3 from our model shows that access to savings can boost labor supply by raising the marginal utility of in-

[11] In Appendix E we discuss impacts on the full sample.

[12] We additionally replicate the basic finding that, considering self-reported labor supply among the full sample at the end of the program, GUP does not produce a classic negative income effect (see Appendix Table 3).

[13] In Appendix F we explore these effects further.

come. The dramatic gap in bag-making between GUP and SOUP households displayed in Table 2 suggests that savings collection is unlikely to be the source of GUP's positive effect on bag-making.

Second, we explore the role of income versus encouragement. Since the Graduation program aims to simultaneously shift income and encourage more labor through psychosocial improvements, we take advantage of experimental variation in the amount of the unconditional cash transfer to isolate the cash income effect.[14] We do not however have experimental variation in household visits, the component that aims to provide information and encouragement regarding income generating activities.[15]

Panel C of Table 2 compares the outcomes of GUP participants receiving transfers of $3.9 weekly with those receiving transfers of $1.3 weekly. Column 1 shows that the total monthly income of high UCT participants is 70% higher than their low UCT counterparts ($p<0.01$). Column 2 shows that the difference in the bags production index between GUP households receiving high versus low UCT is not significantly different from zero, and if anything, is positive ($p=0.43$). The same pattern is seen across the components of the index, suggesting that high UCT households are important drivers of the positive effect on bag production. Moreover, column 6 shows that any reduction in time on non-bags labor for GUP participants is likely to be driven by low UCT households. All together, it seems unlikely to be the case that this pure income shock generates a negative income effect.

The GUP program differs from a pure income transfer not only because it includes encouragement, but also because it enables households to generate their own income. Labor supply responses may depend on whether income is earned or granted, for traditional economic reasons like signaling and expectations-setting, or for other reasons such as positive (pride) or negative (stigma) image effects. Our results cannot speak to these mechanisms, but they do suggest that neither earned nor unearned income produce substantial reductions in labor supply.

5.3 Are we missing the effect on effort?

Empirically, we use bag output as a proxy for bag-making effort, since it encapsulates both labor supplied in hours and productivity. We now apply the same approach to other income-generating activities to verify that GUP households, who do not show detectable reductions in hours spent on these activities, do not reduce effort in other ways. We start from the fact that there is minimal wage labor in our context. Individ-

[14] This strategy enables us to test the effects of additional cash only on top of GUP, and its external validity may be limited if there is complementarity between GUP and the cash transfer.

[15] In other contexts, such visits have not produced an additive effect on income, although this remains under-researched (Blattman et al., 2016).

uals either work on their own farms or run their own businesses. In both cases the household is the residual claimant and the effective labor supply, including any differences in productivity, should be reflected in the income from the activity. We also look at labor-saving investments, which would allow earnings to go up even when effort has gone down.

Tables 3 and 4 report GUP-induced changes in agriculture (the dominant household enterprise) and non-farm enterprises for Bags households. There is no detectable difference in the amount of hired labor used by GUP households compared to Control households (Table 3 column 1, with p=0.47 and q=0.67), despite the fact that GUP recipients spend somewhat less time on their farms compared to Control households (Table 3 column 5, with p=0.04 and q=0.19). At the same time we see only minimal evidence of labor-saving expenditures, the most important of which would be herbicide. Column 2 shows an increase in expenditure on herbicide among GUP household (p=0.08, q=0.22), with a magnitude that is large relative to the control mean, but small in absolute terms, equal to about two percent of the average use by farmers in this region.[16] Column 3 shows that expenditure on fertilizer, which is labor-using rather than labor-saving because of its effects on weed growth and output, is not significantly different between GUP and control households. If anything, GUP households spend more, with the estimate amounting to 10 percent of the average use in the region (p=0.14, q=0.35). Table 4, column 1 shows that agricultural revenue is no lower for GUP households and the point estimate is positive (p=0.30, q=0.42). Moreover column 2 shows that GUP has no impact on residual productivity, which is the residual from regressing harvest value on input expenditure, acreage and labor time, and is an attempt to measure the effort the household is putting into agriculture (p=0.94, q=0.94). In other words there is no evidence that GUP households are neglecting their agricultural business, despite the fact that they are spending less time on them.

The same holds for their other businesses: GUP's effect on business revenue and earnings (Table 4 columns 3 and 4) is not significant, and if anything positive (p=0.21 and q=0.42 for revenue). The effect of GUP on time spent on the business is also positive (Table 3 column 6, with p=0.06 and q=0.22). We cannot measure labor substitution for these businesses, but given the (tiny) scale of the businesses, this seems unlikely.

Wage labor is uncommon in our sample.[17] Table 4 column 5 shows that for Bags households during the bags program, GUP households earned $1.11 less in monthly wage income relative to Control households, who only earn $1.52 in wage income monthly (p=0.06, q=0.34). Thus there may be substitution away from

[16] This is calculated from data from the same agroclimatic zone from a representative set of farmers in villages with fewer than 50 compounds (Udry, 2019). Households in our sample, who were classified as "ultra-poor," on average use much smaller amounts.

[17] See Appendix G for details.

wage labor, but this is small relative to the increases in earnings across the other sources.

Finally, it is possible that GUP households do less household work or shift toward more relaxing labor in ways that would be difficult to measure.[18] There is no difference in the time spent on household work among GUP, SOUP, and Control households assigned to Bags (Table 3). Although we cannot rule out the possibility that GUP households reduced effort on unmeasured margins, it seems equally likely that it could be the opposite: for example, GUP households might exert more effort in housework due to the presence of additional livestock. In net, these effects are likely to be second order.

Importantly, the high UCT GUP households do not drive any differences in time spent on non-bags labor (Table 2, Panel C, column 6) or expenditure on labor-saving substitutes (Table 3, Panel B, columns 1-3). Their harvest value is higher than the low UCT households (Table 4 Panel B columns 1-2, with p=0.02 and q=0.34), suggesting that if there is any crowd out of farming effort due to the GUP intervention, it is driven by the low UCT households.[19,20,21]

6 Conclusion

We find that GUP has a positive effect on income, but does not reduce labor supply, and in fact raises production of bags and especially production of complex bags. From Results 1 and 2 in our theoretical model, these results are consistent with a psychological/physiological effect on labor supply brought about by the GUP intervention. In Appendix Table 8, we check whether the labor supply effects of GUP are concentrated during the end of the lean season, when households are presumably under the greatest psychological and/or physiological duress. We interact GUP with a station-level indicator for the peak of the harvest, when food becomes widely available (see Appendix J for details). Indeed, the positive effect of GUP on labor supply is concentrated in the weeks before the harvest peak.

The idea that there may be positive rather than negative income effects on labor supply has a long pedigree. We provide support for this view based on a field experiment that features a long-term positive income shock and objective measurement of labor output.

[18] For example, GUP households might buy milled grain instead of grinding it themselves.

[19] In Appendix Table 4, we show that our main results are robust to the exclusion of the five baseline variables that were imbalanced at baseline.

[20] Appendix Tables 5, 6, and 7 report estimates from the fully saturated model including all GUP treatment arms (including GUP-savings, discussed in Appendix I.1, but excluding wage variations) and their interactions, following the recommendations of Muralidharan et al. (2025). While the coefficients are not directly comparable to those in the main pooled ("short") specifications, the qualitative patterns are broadly consistent.

[21] In Appendix H we describe analysis changes that have been made since our NBER Working Paper (Banerjee et al., 2020).

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. *Journal of the American Statistical Association*, 103(484):1481–1495.

Baird, S., McKenzie, D., and Özler, B. (2018). The effects of cash transfers on adult labor market outcomes. *IZA Journal of Development and Migration*, 8(1):22.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., and Heil, A. (2022). Why Do People Stay Poor? *The Quarterly Journal of Economics*, 137(2):785–844.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. (2017). Labor Markets and Poverty in Village Economies. *The Quarterly Journal of Economics*, 132(2):811–870.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., and Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236):1260799.

Banerjee, A., Duflo, E., and Sharma, G. (2021). Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights*, 3(4):471–86.

Banerjee, A., Karlan, D., Osei, R., Trachtman, H., and Udry, C. (2022). Unpacking a multi-faceted program to build sustainable income for the very poor. *Journal of Development Economics*, 155:102781.

Banerjee, A., Karlan, D., Trachtman, H., and Udry, C. R. (2020). Does Poverty Change Labor Supply? Evidence from Multiple Income Effects and 115,579 Bags. NBER Working Paper 27314, National Bureau of Economic Research.

Banerjee, A., Rema, H., Gabriel, K., and Benjamin, O. (2017). Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs. *World Bank Research Observer*, 32(2):155–184.

Barker, N., Karlan, D., Udry, C., and Wright, K. (2024). The Fading Treatment Effects of a Multi-Faceted Asset-Transfer Program in Ethiopia. *American Economic Review: Insights*, 6(2):277–294.

Benjamin, D. (1992). Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household. *Econometrica*, 60(2):287–322.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Bernard, T., Dercon, S., Orkin, K., and Seyoum Taffesse, A. (2023). The Future in Mind: Aspirations and Long-term Outcomes in Rural Ethiopia. Discussion Paper 18492, CEPR.

Blattman, C., Green, E. P., Jamison, J., Lehmann, M. C., and Annan, J. (2016). The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda. *American Economic Journal: Applied Economics*, 8(2):35–64.

Breza, E., Kaur, S., and Shamdasani, Y. (2018). The Morale Effects of Pay Inequality. *The Quarterly Journal of Economics*, 133(2):611–663.

Browning, M. and Chiappori, P.-A. (1998). Efficient Intra-Household Allocations: A General Characterization and Empirical Tests. *Econometrica*, 66(6):1241–1278.

Brune, L., Chyn, E., and Kerwin, J. (2021). Pay Me Later: Savings Constraints and the Demand for Deferred Payments. *American Economic Review*, 111(7):2179â2212.

Crosta, T., Karlan, D., Ong, F., Rüschenpöhler, J., and Udry, C. R. (2024). Unconditional Cash Transfers: A Bayesian Meta-Analysis of Randomized Evaluations in Low and Middle Income Countries. NBER Working Paper 32779, National Bureau of Economic Research.

Dalton, P. S., Ghosal, S., and Mani, A. (2016). Poverty and Aspirations Failure. *The Economic Journal*, 126(590):165–188.

Dasgupta, P. and Ray, D. (1986). Inequality as a Determinant of Malnutrition and Unemployment: Theory. *The Economic Journal*, 96(384):1011–1034.

Deaton, A. (2005). Measuring Poverty in a Growing World (Or Measuring Growth in a Poor World). *Review of Economics and Statistics*, 87(1):1–19.

Fehr, D., Fink, G., and Jack, B. K. (2022). Poor and Rational: Decision-Making under Scarcity. *Journal of Political Economy*, 130(11):2862–2897.

Fink, G., Jack, B. K., and Masiye, F. (2020). Seasonal Liquidity, Rural Labor Markets, and Agricultural Production. *American Economic Review*, 110(10):3351–3392.

Foster, A. D. and Rosenzweig, M. R. (1996). Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market. *The Review of Economic Studies*, 63(3):347–374.

Genicot, G. and Ray, D. (2017). Aspirations and Inequality. *Econometrica*, 85(2):489–519.

Gertler, P., Martinez, S., and Rubio-Codina, M. (2012). Investing Cash Transfers to Raise Long-Term Living Standards. *American Economic Journal: Applied Economics*, 4(1):164–192.

Goldberg, J. (2016). Kwacha Gonna Do? Experimental Evidence About Labor Supply in Rural Malawi. *American Economic Journal: Applied Economics*, 8(1):129–49.

Gromadzki, J. (2024). Labor Supply Effects of a Universal Cash Transfer. *Journal of Public Economics*, 239:105248.

Guiteras, R. P. and Jack, B. K. (2018). Productivity in piece-rate labor markets: Evidence from rural Malawi. *Journal of Development Economics*, 131:42–61.

Haushofer, J. and Fehr, E. (2014). On the psychology of poverty. *Science*, 344(6186):862–867.

Jayachandran, S. (2006). Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries. *Journal of Political Economy*, 114(3):538–575.

Karlan, D., Savonitto, B., Thuysbaert, B., and Udry, C. (2017). Impact of savings groups on the lives of the poor. *Proceedings of the National Academy of Sciences*, 114(12):3079–3084.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-Control at Work. *Journal of Political Economy*, 123(6):1227–1277.

Kaur, S., Mullainathan, S., Oh, S., and Schilbach, F. (2025). Do Financial Concerns Make Workers Less Productive? *The Quarterly Journal of Economics*, 140(1):635–689.

Kochar, A. (1999). Smoothing Consumption by Smoothing Income: Hours-of-Work Responses to Idiosyncratic Agricultural Shocks in Rural India. *Review of Economics and Statistics*, 81(1):50–61.

LaFave, D., Peet, E., and Thomas, D. (2025). Farm profits, prices and household behavior. *Journal of Development Economics*, 174:103423.

Leibenstein, H. (1957). *Economic Backwardness and Economic Growth*. Wiley, New York.

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, 22:139–191.

Lybbert, T. J. and Wydick, B. (2018). Poverty, Aspirations, and the Economics of Hope. *Economic Development and Cultural Change*, 66(4):709–753.

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty Impedes Cognitive Function. *Science*, 341(6149):976–980.

Mullainathan, S. and Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. Times Books/Henry Holt and Co.

Muralidharan, K., Romero, M., and Wüthrich, K. (2025). Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments. *The Review of Economics and Statistics*, 107(3):589–604.

Nakalembe, C., Becker-Reshef, I., Bonifacio, R., Hu, G., Humber, M. L., Justice, C. J., Keniston, J., Mwangi, K., Rembold, F., Shukla, S., et al. (2021). A Review of Satellite-Based Global Agricultural Monitoring Systems Available for Africa. *Global Food Security*, 29:100543.

Pinto, R. (2021). Financially stressed employees. SSRN Working Paper 3823267.

Rose, E. (2001). Ex ante and ex post labor supply response to risk in a low-income area. *Journal of Development Economics*, 64(2):371–388.

Rosenzweig, M. R. (1988). Chapter 15 Labor Markets in Low-Income Countries. *Handbook of Development Economics*, 1:713–762.

Sedda, L., Tatem, A. J., Morley, D. W., Atkinson, P. M., Wardrop, N. A., Pezzulo, C., Sorichetta, A., Kuleszo, J., and Rogers, D. J. (2015). Poverty, Health and Satellite-Derived Vegetation Indices: Their Inter-Spatial Relationship in West Africa. *International Health*, 7(2):99–106.

Shah, A. K., Mullainathan, S., and Shafir, E. (2012). Some Consequences of Having Too Little. *Science*, 338(6107):682–685.

Strauss, J. and Thomas, D. (1998). Health, Nutrition, and Economic Development. *Journal of Economic Literature*, 36(2):766–817.

Udry, C. (2019). Information, market access and risk: Addressing constraints to agricultural transformation in Northern Ghana. *Draft Report*.

Vivalt, E., Rhodes, E., Bartik, A. W., Broockman, D. E., Krause, P., and Miller, S. (2025). The Employment Effects of a Guaranteed Income: Experimental Evidence from Two U.S. States. NBER Working Paper 32719, National Bureau of Economic Research.

Table 1: Experimental Design

Panel A: Intervention and Bags Assignments

Intervention Village Assignment	Bags Village Assignment	# Villages	Household Assignment	# Households
Control	No Bags	15	Control-No-Bags	282
	Bags	42	Control-No-Bags	376
			Control-Bags	397
GUP	No Bags	29	Control-No-Bags	276
			GUP-No-Bags	294
	Bags	25	Control-No-Bags	215
			GUP-Bags	221
SOUP	No Bags	28	Control-No-Bags	208
			SOUP-No-Bags	283
	Bags	26	Control-No-Bags	203
			SOUP-Bags	278
TOTAL		165		3033

Panel B: Bags Sub-Treatment Assignment

Intervention Village Assignment - Bags	Bags Simple/Complex Sub-treatment	Bags UCT Sub-treatment	# Villages	# Households
Control-Bags	Simple	n/a	21	189
	Complex	n/a	21	208
GUP-Bags	Simple	High UCT	8	59
		Low UCT	7	71
	Complex	High UCT	7	61
		Low UCT	3	30
SOUP-Bags	Simple	n/a	13	141
	Complex	n/a	13	137
TOTAL			93	896

Panel A shows intervention treatment assignments (GUP, SOUP, and Control) and assignment to the Bags program. Both intervention treatment assignments and Bags assignments were at the village level. The village-level assignments are shown in columns 1-3 of Panel A and are defined as follows. Control-Bags = intervention Control villages assigned to Bags. Control-No-Bags = intervention Control villages not assigned to Bags. GUP-Bags = GUP intervention villages assigned to Bags. GUP-No-Bags = GUP intervention villages not assigned to Bags. SOUP-Bags = SOUP intervention villages assigned to Bags. SOUP-No-Bags = SOUP intervention villages not assigned to Bags. Within each village treated with an intervention (GUP/SOUP) and/or Bags, there are households who are randomly assigned to receive the corresponding set of treatments, and there are untreated households. Specifically, the household-level assignments, shown in columns 4-5 of Panel A, are defined as follows. Control-No-Bags = households who receive neither Bags nor GUP/SOUP. Control-Bags = households who receive Bags but not GUP/SOUP. GUP-Bags = households who receive GUP and Bags. GUP-No-Bags = households who receive GUP but not Bags. SOUP-Bags = households who receive SOUP and Bags. SOUP-No-Bags = households who receive SOUP but not Bags. Panel B shows sub-treatments within the Bags program. All sub-treatments were randomized at the village level such that all individuals within a village who received the Bags program received identical sub-treatment assignments. The village-level sub-treatments are defined as follows. Simple = assigned to sew the simple bag. Complex = assigned to sew the complex bag. High UCT = GUP intervention households (with Bags) who received an unconditional cash transfer of $3.92 each week. Low UCT = GUP intervention households (with Bags) who received an unconditional cash transfer of $1.31 each week.

Table 2: Effects of GUP, SOUP, and UCT on Income and Labor Supply Among Bags Households During Bags Program

Panel A: Effects of GUP and SOUP

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
GUP	ITT	9.36	0.27	1.14	0.12	0.66	-8.74
	SE	(1.34)	(0.12)	(0.50)	(0.05)	(0.35)	(12.75)
	p-val	0.00	0.03	0.03	0.02	0.06	0.49
	q-val	0.00	0.12	0.12	0.12	0.15	0.55
	Bsl p-val	0.30					
SOUP	ITT	-0.46	-0.18	-0.65	-0.11	-0.33	-8.64
	SE	(0.56)	(0.13)	(0.49)	(0.06)	(0.38)	(12.76)
	p-val	0.41	0.16	0.18	0.06	0.39	0.50
	q-val	0.48	0.28	0.29	0.15	0.47	0.55
	Bsl p-val	0.55					
Ctrl Mean		8.01	0.00	3.76	0.58	2.45	535.59
Ctrl SD		25.04	1.00	3.97	0.49	3.01	158.72
Obs		288	18816	18816	18816	18816	3442
GUP - SOUP	ITT	9.82	0.45	1.79	0.24	0.99	-0.10
GUP - SOUP	p-val	0.00	0.00	0.00	0.00	0.03	0.99

Panel B: Effects of GUP and SOUP for Households Assigned Complex Bags

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
GUP-Complex	ITT	11.12	0.29	1.43	0.12	0.59	-4.40
	SE	(0.97)	(0.19)	(0.77)	(0.10)	(0.43)	(18.39)
	p-val	0.00	0.13	0.07	0.23	0.17	0.81
	q-val	0.00	0.25	0.16	0.32	0.29	0.86
	Bsl p-val	0.29					
SOUP-Complex	ITT	-1.01	-0.28	-1.08	-0.15	-0.60	-0.48
	SE	(0.69)	(0.14)	(0.52)	(0.08)	(0.35)	(15.01)
	p-val	0.16	0.06	0.04	0.07	0.09	0.97
	q-val	0.28	0.15	0.14	0.16	0.19	0.98
	Bsl p-val	0.30					
Ctrl Mean		14.44	-0.16	3.01	0.54	1.88	541.23
Ctrl SD		34.83	0.91	3.54	0.50	2.59	171.77
Obs		160	9156	9156	9156	9156	1608
GUP - SOUP	ITT	12.14	0.56	2.50	0.27	1.20	-3.92
GUP - SOUP	p-val	0.00	0.00	0.00	0.01	0.01	0.81

Panel C: Effects of UCT

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
High UCT	ITT	13.05	0.35	1.46	0.16	0.86	2.69
	SE	(0.00)	(0.16)	(0.67)	(0.07)	(0.49)	(12.94)
	p-val	0.00	0.04	0.03	0.03	0.09	0.84
	q-val	0.00	0.13	0.13	0.12	0.19	0.86
	Bsl p-val	0.73					
Low UCT	ITT	4.35	0.18	0.77	0.08	0.44	-21.89
	SE	(0.00)	(0.15)	(0.66)	(0.07)	(0.41)	(19.90)
	p-val	0.00	0.24	0.24	0.22	0.28	0.27
	q-val	0.00	0.33	0.33	0.32	0.35	0.35
	Bsl p-val	0.21					
Ctrl Mean		8.01	0.00	3.76	0.58	2.45	535.59
Ctrl SD		25.04	1.00	3.97	0.49	3.01	158.72
Obs		288	18816	18816	18816	18816	3442
High - Low	ITT	8.70	0.17	0.70	0.07	0.42	24.58
High - Low	p-val	0.00	0.43	0.44	0.41	0.49	0.26

This table shows treatment effects among Bags households during the Bags program. In all panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP-Bags and SOUP-Bags. Panel B shows effects of GUP-Bags and SOUP-Bags for households assigned complex bags. Panel C shows effects of GUP-Bags separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with at least 30 compounds. In column 1, we estimate Equation 1 using household-level data from the midline survey during bag-making and including surveyor fixed effects as well as the baseline value of the outcome when possible. In columns 2-5, we estimate Equation 2 using household-week-level bag production data and including station-week fixed effects. In column 6, we estimate Equation 1 using household-month-level data from time use surveys during bag-making and including surveyor fixed effects. In all columns, we control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Daily time non-bags labor is minutes spent yesterday on farming, business operations, animal production and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Table 3: Effects of GUP, SOUP, and UCT on Inputs Among Bags Households During Bags Program

Panel A: Effects of GUP and SOUP

		Exp. Labor (1)	**Exp. Herbicide** (2)	**Exp. Fertilizer** (3)	**Time Bags** (4)	**Time Field** (5)	**Time Business** (6)	**Time Home** (7)	**Time All Labor** (8)	**Time Leisure** (9)
GUP	ITT	2.59	3.83	20.06	-2.12	-20.96	9.22	1.18	-12.63	12.89
	SE	(3.54)	(2.11)	(13.35)	(11.64)	(9.80)	(4.80)	(11.79)	(12.35)	(7.22)
	p-val	0.47	0.08	0.14	0.86	0.04	0.06	0.92	0.31	0.08
	q-val	0.67	0.22	0.35	0.88	0.19	0.22	0.92	0.49	0.22
	Bsl p-val	0.48	0.27	0.05	.	.	.	.	.	.
SOUP	ITT	-0.59	1.27	17.49	-21.85	-24.02	10.06	4.41	-23.30	16.24
	SE	(2.18)	(2.32)	(20.01)	(12.37)	(10.86)	(4.74)	(11.42)	(11.58)	(9.25)
	p-val	0.79	0.59	0.39	0.08	0.03	0.04	0.70	0.05	0.08
	q-val	0.86	0.71	0.58	0.22	0.19	0.19	0.79	0.21	0.22
	Bsl p-val	0.51	0.90	0.04	.	.	.	.	.	.
Ctrl Mean		4.02	3.83	57.58	112.94	206.61	28.48	299.27	603.49	127.17
Ctrl SD		15.67	10.82	88.11	106.61	185.65	78.23	155.09	168.19	140.45
Obs		272	272	272	1978	3442	3442	3442	3442	3442
GUP - SOUP	ITT	3.17	2.56	2.58	19.74	3.06	-0.84	-3.23	10.68	-3.35
GUP - SOUP	p-val	0.35	0.36	0.90	0.19	0.79	0.88	0.81	0.33	0.72

Panel B: Effects of UCT

		Exp. Labor (1)	**Exp. Herbicide** (2)	**Exp. Fertilizer** (3)	**Time Bags** (4)	**Time Field** (5)	**Time Business** (6)	**Time Home** (7)	**Time All Labor** (8)	**Time Leisure** (9)
High UCT	ITT	-2.62	-0.32	26.20	-28.90	-25.13	14.85	11.73	-13.91	12.51
	SE	(3.68)	(1.64)	(14.74)	(12.55)	(11.82)	(5.95)	(11.38)	(11.87)	(9.38)
	p-val	0.48	0.85	0.09	0.02	0.04	0.01	0.31	0.24	0.19
	q-val	0.67	0.88	0.22	0.19	0.19	0.19	0.49	0.42	0.38
	Bsl p-val	0.45	0.27	0.11	.	.	.	.	.	.
Low UCT	ITT	9.45	9.36	11.91	18.19	-16.15	2.75	-10.95	-11.14	13.33
	SE	(7.69)	(3.38)	(22.04)	(13.66)	(12.74)	(6.07)	(17.66)	(18.73)	(9.29)
	p-val	0.23	0.01	0.59	0.19	0.21	0.65	0.54	0.55	0.15
	q-val	0.41	0.19	0.71	0.38	0.40	0.76	0.71	0.71	0.35
	Bsl p-val	0.79	0.61	0.11	.	.	.	.	.	.
Ctrl Mean		4.02	3.83	57.58	112.94	206.61	28.48	299.27	603.49	127.17
Ctrl SD		15.67	10.82	88.11	106.61	185.65	78.23	155.09	168.19	140.45
Obs		272	272	272	1978	3442	3442	3442	3442	3442
High - Low	ITT	-12.07	-9.68	14.30	-47.08	-8.98	12.10	22.68	-2.77	-0.83
High - Low	p-val	0.21	0.03	0.56	0.01	0.55	0.11	0.22	0.88	0.94

This table shows treatment effects on inputs among Bags households during the Bags program. In both panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP-Bags and SOUP-Bags. Panel B shows effects of GUP-Bags separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with at least 30 compounds. In columns 1-3, we estimate Equation 1 using household-level data from the midline survey during bag-making and including surveyor fixed effects as well as the baseline value of the outcome when possible. In columns 4-9, we estimate Equation 1 using household-month-level data from time use surveys (minutes spent on each activity yesterday) during bag-making and including surveyor fixed effects. In all columns, we control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). Expenditure on labor, herbicide, fertilizer are expenditures in the last 12 months. Home labor includes childcare, cleaning, cooking, collecting firewood/water, and shopping. All labor includes bags and other wage labor, agriculture, businesses, animals (which is minimal, so we do not show separately), and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Table 4: Effects of GUP, SOUP, and UCT on Non-Bags Outputs Among Bags Households During Bags Program

Panel A: Effects of GUP and SOUP

		Monthly Harvest Value (1)	**Monthly Residual Prod.** (2)	**Monthly Business Revenue** (3)	**Monthly Business Income** (4)	**Monthly Wage Income** (5)	**Monthly Non-Bags Income** (6)
GUP	ITT	59.31	-4.52	14.97	8.76	-1.11	6.74
	SE	(55.62)	(64.00)	(11.66)	(6.05)	(0.56)	(5.88)
	p-val	0.30	0.94	0.21	0.16	0.06	0.26
	q-val	0.42	0.94	0.42	0.38	0.34	0.42
	Bsl p-val	0.27	.	0.13	$0.04^{(+)}$	0.44	0.87
SOUP	ITT	-22.20	-60.59	-9.47	-2.94	-0.73	-3.59
	SE	(70.42)	(54.77)	(6.16)	(2.56)	(0.91)	(2.78)
	p-val	0.75	0.28	0.13	0.26	0.43	0.21
	q-val	0.79	0.42	0.38	0.42	0.57	0.42
	Bsl p-val	0.12	.	0.44	$0.08^{(+)}$	0.01	0.44
Ctrl Mean		408.47	-13.50	13.94	6.54	1.52	8.01
Ctrl SD		461.91	349.39	55.61	24.83	4.66	25.04
Obs		272	266	287	287	288	288
GUP - SOUP	ITT	81.51	56.07	24.44	11.71	-0.38	10.34
GUP - SOUP	p-val	0.18	0.39	0.05	0.06	0.63	0.08

Panel B: Effects of UCT

		Monthly Harvest Value (1)	**Monthly Residual Prod.** (2)	**Monthly Business Revenue** (3)	**Monthly Business Income** (4)	**Monthly Wage Income** (5)	**Monthly Non-Bags Income** (6)
High UCT	ITT	138.60	94.92	10.58	6.42	-0.79	5.21
	SE	(54.79)	(59.36)	(18.88)	(8.52)	(0.53)	(8.56)
	p-val	0.02	0.12	0.58	0.46	0.15	0.55
	q-val	0.34	0.38	0.63	0.57	0.38	0.63
	Bsl p-val	0.13	.	0.73	$0.03^{(+)}$	0.71	0.78
Low UCT	ITT	-46.51	-136.67	21.01	12.00	-1.52	8.81
	SE	(64.83)	(68.40)	(11.59)	(8.23)	(0.69)	(8.28)
	p-val	0.48	0.06	0.08	0.16	0.03	0.30
	q-val	0.57	0.34	0.38	0.38	0.34	0.42
	Bsl p-val	0.82	.	$0.09^{(+)}$	0.34	0.00	0.98
Ctrl Mean		408.47	-13.50	13.94	6.54	1.52	8.01
Ctrl SD		461.91	349.39	55.61	24.83	4.66	25.04
Obs		272	266	287	287	288	288
High - Low	ITT	185.11	231.60	-10.43	-5.58	0.73	-3.60
High - Low	p-val	0.02	0.01	0.65	0.64	0.21	0.77

This table shows treatment effects on non-bags outputs among Bags households during the Bags program. In both panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP and SOUP. Panel B shows effects of GUP separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with at least 30 compounds. We estimate Equation 1 using household-level data from the midline survey during bag-making and including surveyor fixed effects as well as the baseline value of the outcome when possible. We control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). Harvest value is the total revenue from harvest, six months after the bag-making period. Residual productivity contains the residuals from a regression of harvest value on input expenditure, acreage, and average time on field. Business revenue and income pertain to all non-agricultural businesses. Wage income does not include earnings from bags. Non-Bags income includes income from agriculture, business, wages, and animals. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Online Appendix

A Program Implementation Details

A.1 GUP and SOUP

Presbyterian Agricultural Services (PAS), our implementation partner, has experience doing extension work and promotion of savings groups, including a prior randomized controlled trial with Innovations for Poverty Action (Karlan et al., 2017). PAS field agents engaged in the direct field implementation, while Innovations for Poverty Action coordinated the implementation with senior management of PAS.

Participants in the program come from three areas of Northern Ghana corresponding to three agricultural "stations" run by PAS: Tamale, Langbensi, and Sandema. PAS first identified poor communities, and in each identified community, staff members then facilitated a Participatory Wealth Ranking (PWR) in which members of the community ranked households by economic status. Finally, PAS staff members returned for a verification of the households judged to be the poorest. In order to be eligible, households were required to have a female between the ages of 18-24, who we will call the female household head, to be the direct recipient of all treatment activities and the primary respondent for surveys.

As mentioned in the text, the GUP program included six components, all of which were directed to the female household head: (1) the transfer of a productive asset; (2) skills training for the management of the asset, (3) life skills training and mentorship, via weekly household visits over two years, (4) a weekly cash stipend for consumption support, worth between $6 and $9 PPP depending on family size, during each lean season, (5) some basic health services and health education, and (6) access to a savings account at a local bank and deposit collection. The first component, the productive asset transfer, was provided at the beginning of the program, and households were permitted to choose a package of productive assets from a set list. Most households chose a package that included four goats.[22] The skills training, in which participants learned how to take care of the asset (e.g., when to vaccinate goats), took place at the start of the program, and then also as part of weekly household visits by the PAS field officer. The household visits also provided the backbone for delivering components three through six. The third component, a "hand-holding" or life-skills component, provided nudges to help the household focus on building productive assets to generate positive change in long-term outcomes, and more generally, to set aspirations and plans for coping with current problems and improving the future. The consumption support was explicitly intended to help this process in the short-run, by helping to absorb short-run shocks that could lead to households consuming the transferred assets. The fifth component included basic education on health and hygiene as well as enrollment in the national health insurance scheme (about $2 per month).

As mentioned in the text, the sixth component of GUP, access to a savings account and weekly deposit collection by PAS field agents, was the entirety of the SOUP intervention.

In Section I.1, we describe additional sub-treatments within GUP and SOUP that are not the main focus of this paper.

[22] Other assets included hens, pigs, and inputs for the production of shea butter, maize, and sorghum.

A.2 Bag-Making

Prior to the start of the program, we established 18 quality standards for simple bags, and 25 quality standards for complex bags, and reviewed them with Bags households during the training. Bags were assigned one point for meeting the quality standards at the "excellent" level, half a point for "satisfactory," and zero points for "unsatisfactory." At the end of the quality check, the final quality score was calculated and the bag was classified as high, mid, or low quality. Over the course of the study, 35% of bags collected were low quality, 34% were mid quality, and 31% were high quality. Every four weeks, bag-making field agents returned to communities to give feedback and remedial training. In Appendix Figure 4 we plot bag production and quality scores over time. Subfigure (a) demonstrates that bag production was relatively stable over the duration of the program, with the exception of a drop in weeks 15 and 17 for Control households, motivating our analysis in Appendix Table 8. Subfigure (b) indicates that quality scores were also relatively stable over the program, increasing slightly during the first half of the program and falling toward the end.

B Imputing Time Spent on Bags

In our time use survey, rather than asking about time spent on bags directly, we asked only about "wage labor (including bags)" in order to maintain separation between the evaluation team and the team that was implementing the bags program. In what follows, we describe how we impute time spent on bags.

In Appendix Table 9 we report levels of monthly wage income and time spent on bags and/or wage labor across treatment groups. In Column 1, we can see that within each treatment group—control, GUP, and SOUP—there is very little difference in wage income between Bags and No-Bags households within each treatment group. Since wage income does not include bags earnings, we infer from this that there were not significant differences in non-bags wage labor across Bags and No-Bags groups. Therefore, any differences in time spent on "time bags and/or wage labor" within each treatment group, between Bags and No-Bags households, can be attributed to time spent on bags. Column 2 shows that indeed, time spent on "time bags and/or wage labor" is much higher among households assigned to bag-making.

We thus impute time spent on bags by taking the time spent on "time bags and/or wage labor" for each Bags participant, and subtracting the mean time spent on "time bags and/or wage labor" from the corresponding No-Bags treatment group. For example, for a GUP-Bags participant, we take "time spent on bags and/or wage labor" and subtract the mean time spent on "bags and/or wage labor" in GUP-No-Bags.

An important caveat is that if any part of GUP causes participants to be more productive in non-bags wage labor, they may earn the same wage income with fewer hours. If this were true, then our imputed measure of time spent on bags for participants with GUP would be biased downward (since time spent on non-bags wage labor is biased upward). For this reason, we do not draw strong conclusions from any estimated differences in time on bags across treatments (between high and low UCT, for example).

C Experimenter Demand Effects

To the extent possible, activities connected to the bag-making program were conducted by separate bag-making field agents (including training, quality-checking, feedback, and announcements about wage changes).

However, the regular GUP/SOUP field agents were responsible for collecting the bags and delivering payment each week. As a result, the two programs were certainly not perceived as separate.

In general, we cannot rule out the possibility that the observed effects of GUP are in part due to demand effects. That said, we would not expect demand effects on bag-making labor supply to be concentrated during the weeks when households are most constrained, which is the case (see Appendix Table 8). Also, though GUP households were indeed 12 percentage points more likely to submit at least one bag each week than control households (see Table 2 column 4), it is still the case that 30% of GUP households chose not to make any bags each week.

D Attrition

In Appendix Table 10 we report attrition by treatment status for the second midline, the two-year follow-up survey, and the time use surveys. For the two-year survey, GUP-No-Bags had slightly higher participation than Control-No-Bags ($p = 0.10$), but participation was between 94% and 100% across all groups, and these are not our principal outcomes. In our time use midlines, we do find that Bags households were more likely to participate than No-Bags households across all treatment groups ($p < 0.01$). (Surveyors reported that they were easier to find, since they were also anticipating visits from field agents who collected bags on a weekly basis.) Fortunately, we do not focus on this comparison in our analysis, and we do not find significant differences between GUP-Bags and Control-Bags, between GUP-Bags and SOUP-Bags, or between high and low UCT. (We do not show attrition results by high versus low UCT in this table, but have checked that there are no significant differences in participation for any of the three surveys.)

E Overall Impacts of GUP and SOUP

In what follows, we discuss overall impacts of GUP and SOUP for the full sample of participants. (The analysis in the paper is restricted to Bags households). In Appendix Table 3 Panel A we report impacts of the basic treatments, GUP and SOUP, including both Bags and No-Bags households. Columns 1-5 use data collected at the end of the two-year program (7-9 months after the end of bag-making); columns 6 and 7 use time use data collected in the same survey. For each outcome, we estimate Equation 1 with one additional control: a binary indicator for being assigned to the bag-making program.

Columns 6-7 suggest that, consistent with our results from the bag-making sample in Table 2, GUP and SOUP households spend the same amount of time providing productive labor as do Control households, and report the same amount of leisure time (each of the estimated treatment effects is smaller than three percent of the control mean, and statistically indistinguishable from zero, with p=0.98 and q=1.00 for the effect of GUP on productive labor).

Columns 1-5 report economic impacts. The GUP treatment raised the value of livestock owned by the household by more than 30 percent relative to control (itt = \$73, with q<0.01). SOUP households also acquire more livestock (itt = \$32, with p=0.03 and q=0.17), but the net increase is significantly less than that for the GUP households. On the other hand, as column 2 reports, SOUP has as large an effect on total asset value as GUP (and both larger than control, with p<0.05 and q<0.22). The pattern for self-reported monthly income at the two-year mark (in column 3) is similar: both SOUP and GUP have positive point estimates,

but the GUP effect is almost twice as large as the SOUP and is the only one that is significant (p<0.01, q=0.05). SOUP households, like GUP households, do not report more time spent on leisure. There are no statistically significant effects on consumption or health (columns 4 and 5), although on consumption, the more complete analysis reported in Banerjee et al. (2022) does find statistically significant impacts.[23,24]

We next describe the results for the GUP-No-Bags and SOUP-No-Bags interventions, reported in Appendix Table 3 Panel B. For each outcome, we estimate Equation 1. Here, GUP-No-Bags is the classic "graduation" intervention. GUP-No-Bags households report similar amounts of time spent on both leisure and productive labor relative to control No-Bags households. SOUP-No-Bags households may spend more time on productive labor than control No-Bags households, but the estimate is noisy (p=0.24, q=0.46) and cannot be distinguished from the corresponding GUP effect. The effects of GUP-No-Bags and SOUP-No-Bags on livestock, total assets and income parallel those of GUP and SOUP overall: GUP-No-Bags has a stronger effect on livestock than SOUP-No-Bags, they have similar impacts on total assets, and GUP-No-Bags has the largest impact on income. Neither GUP-No-Bags nor SOUP-No-Bags has a noticeable impact on health, but SOUP-No-Bags does increase consumption.

These program impacts indicate, first, that self-reported income was higher among GUP households, both with bags (p=0.05, q=0.22) and without bags (p=0.01, q=0.05), at the end of the two-year program. Second, GUP households show no evidence of a reduction in labor supply.

F Exploring Effects on Complex Bags

In Appendix Tables 11-12, we take a closer look at the effects on complex bag-making displayed in Table 2 Panel B. In Appendix Table 11, we show that the additional complex bag production by GUP households does not come at the expense of quality. In Appendix Table 12 we show that this effect is consistent over the course of the bags program, suggesting that the difference in production between GUP and Control households is not due to differential learning rates in the early weeks of the program.

G Wage Labor

In Control-No-Bags, average monthly wage labor earnings (excluding bags) are $1.13, and only 16% of households have positive wage earnings in a month. In terms of time, in Control-No-Bags, average time spent on wage labor is 9.7 minutes daily, and only 7.1% of households spent any time on wage labor yesterday. Demand for wage labor is also low: in Control-No-Bags, yearly expenditure on wage labor is $4.06 and only 9.7% of households demand any labor from the market in a year.

[23]Note that there are three key differences between the specification in Equation 1 and that used in Banerjee et al. (2022). First, since the randomization of GUP among bag-making villages was at the village level, here we include interviewer fixed effects instead of village fixed effects, and we cluster at the village level. Second, here we exclude villages with fewer than 30 compounds due to the issue with the randomization of bag-making discussed in Section 2.2. Finally, here we only use data from the two-year follow-up household survey (rather than averaging over the two-year survey and the second and third midlines), given the use of interviewer fixed effects. We chose to estimate Equation 1 for consistency, but this specification was designed for the analysis of the bag-making program, and our more complete estimates for the effect of GUP on consumption are those reported in Banerjee et al. (2022).

[24]The discrepancy between household income and consumption that we find is common, see Deaton (2005).

H Updates from NBER Working Paper

There are several differences between the analysis in this version of the paper and the analysis in our NBER Working Paper (Banerjee et al., 2020). For the sake of transparency, below we list some of the analysis choices that were updated in this version of the paper.

1. The table depicting the experiment design (Table 1) includes only the observations that are used in the analysis (excluding participants from villages with fewer than 30 compounds.)
2. We report time use outcomes at two years for all sample households (Appendix Table 3 columns 6-7), as well time use outcomes during the bags program for the bags sample (Table 3 columns 8-9).
3. We cluster standard errors at the village level for every analysis.

I Additional Treatments

Unfortunately, a pre-analysis plan was not made before the field experiment began in 2011. In this section, we describe the full experiment as it was originally designed. Specifically, there were two additional treatment dimensions that were randomized, but are not the main focus of this paper.

I.1 GUP-Savings and SOUP-Match

As mentioned in Appendix A.1, the GUP and SOUP programs contained sub-treatments, which were designed to allow us to "unpack" the bundled GUP and SOUP interventions (Banerjee et al., 2022). Specifically, for half of the households assigned to GUP, the weekly collection of savings deposits was randomly withheld. For the GUP households who did receive deposit collection services, the treatment is equivalent to the combination of GUP and SOUP. We found no evidence that the presence or absence of savings collection made a difference for the impact of GUP on consumption or income (Banerjee et al., 2022). There are thus two GUP treatments: "GUP-Savings" and "GUP-No-Savings" which are pooled for the purpose of our main analyses.

In Appendix Table 13 we show separately the effects of GUP-No-Savings and GUP-Savings on bags labor supply. The effects are not significantly different for households with and without savings collection, further supporting the conclusion from Section 5.2 that savings is unlikely to drive the positive effect of GUP.

Among households assigned to SOUP, there was also an additional sub-treatment: half received savings accounts and deposit collection without a match ("SOUP without match") and half received savings accounts and deposit collection with a 50% match ("SOUP-match"). Specifically, for every GHC 1 deposited, households in this group received a matching contribution of GHC 0.50. At the onset of the program, there was a maximum match of GHC 1.50 GHC per week (for a GHC 3 deposit) but this cap was eventually removed. Thus, like GUP, there are two SOUP treatments: "SOUP-Match" and "SOUP-No-Match" which are pooled for the purpose of our main analyses.

I.2 Wage Elasticities

As mentioned in Section 2.2, we also introduced variation in the wage at the village level over time. The original intent was to estimate the effects of GUP not only on labor supply, but also on labor supply elasticities. Every four weeks, villages were assigned a different baseline wage: \$0.40 or \$0.91. At the start of that week, participants were informed of the payment per bag they would be receiving for bags made in the subsequent four weeks. As discussed in Section 2.2, bags produced in week one of a given wage rotation would be collected at the end of week one and inspected for quality over the course of weeks two and three. Payment for these bags would be made at the end of week three (two weeks from the date of collection). This means that only bags collected at the end of week four were made with actual experience of the new wage.

We explore wage elasticities in Appendix Table 14. In column 1, we regress the inverse hyperbolic sine of weekly bags submitted on the log wage, including individual and week fixed effects and clustering standard errors at the village level. We find an elasticity that is close to zero. In column 2, we provide evidence that participants were responsive to wages they were receiving for bags submitted previously, as opposed to the correct relevant wage for the bags they were making. We define "experience" to mean either the fourth week of the month, or for participants whose villages were randomly assigned two consecutive high wage months and two consecutive low wage months, any week in the second consecutive month with the same wage. Given this evidence, in column 3, we use the 3-week lagged wage, and find positive but small elasticity estimates. Given the fuzzy understanding of wage changes and when they would take effect, we do not focus on wage elasticities in the main part of the paper.

J NDVI Measure

We use the Normalized Difference Vegetation Index (NDVI) as a proxy for vegetation growth in each study area (station). NDVI is derived from satellite imagery, raging from -1 to +1, with values closer to 1 indicating dense, healthy vegetation, and values near 0 or negative indicating bare soil, water, or cloud cover. We use 16-day composite NDVI values from the MODIS MOD13Q1 dataset at 250-meter resolution, extracted for each of the three stations in northern Ghana where the study was conducted (Tamale, Sandema, Langbensi) for the bag-making period. NDVI increases during the growing season and typically peaks shortly before or during the main harvest period. It has been used for early warning systems to promote food security (Nakalembe et al., 2021) and has a negative correlation with poverty in West Africa (Sedda et al., 2015).

To approximate when households were likely to have begun eating from their own harvests and experiencing improved food security, we construct a binary indicator equal to 0 before the peak NDVI date (the date in the season when the index took the highest value) and 1 starting at the peak and onward. This approach is based on the assumption that the NDVI peak coincides with widespread harvest availability. Since food insecurity in the region is driven largely by seasonal crop cycles, and most households rely on their own production for staple consumption, the NDVI peak is likely to be a reasonable proxy for the transition out of the lean season.

K Intra-Household Allocation of Labor

Here, we address the possibility that our treatments shifted intra-household allocation of labor, which our simple model did not account for. However, a close analogue to Result 2 holds for any collective household (Browning and Chiappori, 1998). The primary beneficiary of all of the GUP interventions was the female head of household, and she was the direct recipient of the bag-sewing training. If the receipt of the GUP intervention does not decrease the bargaining power (Pareto weight) of the recipient, and the recipient supplies more labor to the bags activity despite the fact that the household is richer and she is working no less, then there must be a psychological/physiological productivity effect on the bags activity. We have shown in Table 2 that GUP increased household income and did not meaningfully reduce the recipient's overall labor supply. Overall, 95% of bags were made by the female head of household and 96% were made by an adult female (over age 15). We find no evidence that GUP, SOUP, or high consumption support affected the fraction of bags made by adult females within the household. Specifically, if we estimate Equation 2 with the share of bags made by the female household head as the outcome, we obtain a point estimate on GUP of -0.01, with $p = 0.58$ and $q = 0.77$.[25] Shifts in intrahousehold labor allocation, therefore, do not provide an alternative to a physiological or psychological effect of GUP on the supply of labor.[26]

L Inter-Temporal Substitution of Labor Supply

Any intertemporal substitution of labor supply induced by the Bags treatment is likely similar for GUP and Control, and thus are unlikely to drive our main results. It is possible that GUP interacted with the Bags treatment in ways that produced intertemporal changes in incentives and opportunities more subtle than the investment productivity effect we describe. Indeed, such effects are part of what we intend to capture, even if they do not fit perfectly under the labels of physiological or psychological productivity effects.

[25] In our replication package, we provide the relevant code for this estimate, but we do not show the table here due to restrictions on the appendix length.

[26] The bags program itself may have shifted the intra-household allocation of labor but our evidence focuses on differences between GUP-Bags and Control-Bags households.

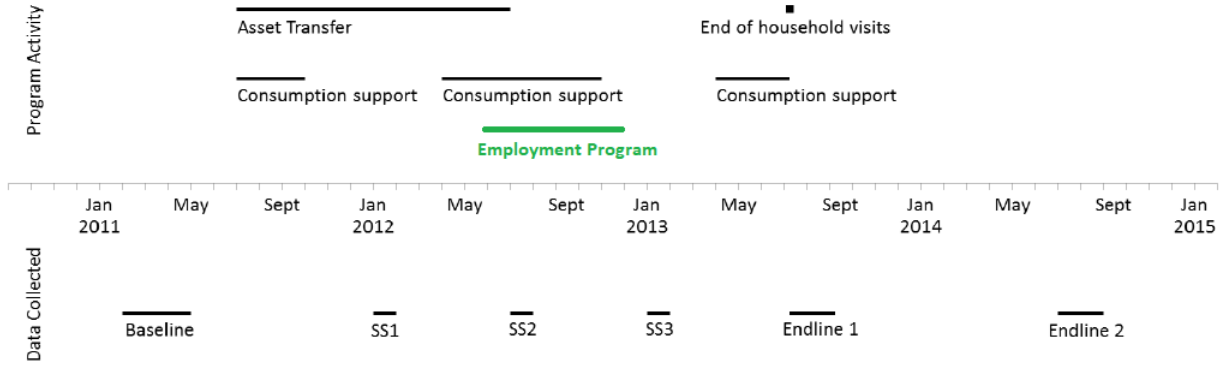
M Additional Figures and Tables

Appendix Figure 1: Simple Bag (left) and Complex Bag (right)

(a) Simple bag *(b) Complex bag*

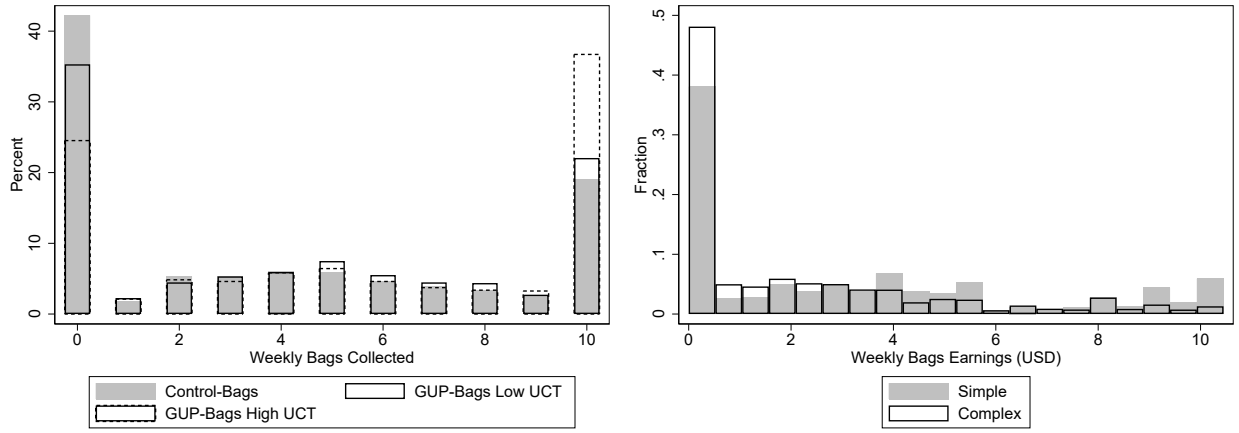
The simple bag shown in subfigure (a) has "running" stitches on the hem and strap. The complex bag shown in subfigure (b) has a more complicated pattern on the hem and strap: a sequence of four "chain" stitches alternating with one "running" stitch.

Appendix Figure 2: Timeline



This figure presents a timeline of program implementation and data collection. In the top part of the timeline we show program activities, and in the bottom part we show data collection. SS1, SS2, and SS3 refer to the three short midlines that were conducted for about one third of sample households. Throughout the paper, we rely heavily on SS2, which was conducted during the employment program. During the employment program we conducted additional time use surveys each month, over five months.

Appendix Figure 3: Weekly Bags Produced by Treatment and Weekly Bags Earnings by Complexity

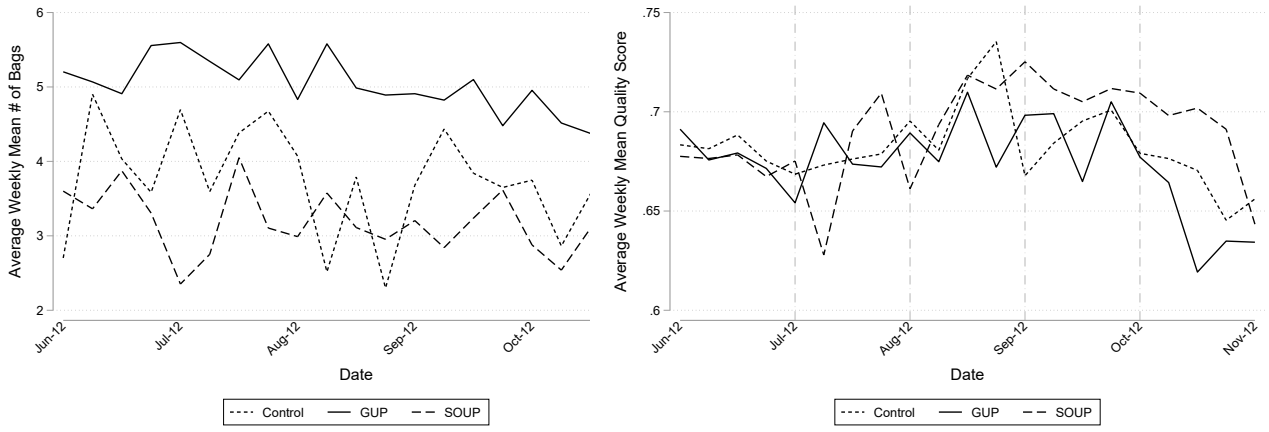


(a) Distribution of Bags by GUP & UCT

(b) Distribution of Bags Earnings by Complexity

Subfigure (a) shows the distribution of weekly bags produced by households assigned to Bags, shown separately for control households, GUP households with high UCT, and GUP households with low UCT. Subfigure (b) shows the distribution of weekly earnings from bags produced by households assigned to Bags, shown separately for households assigned simple bags and households assigned complex bags. Weekly earnings include natural variations in the piece-rate due to the quality of the bags submitted, as well as random variations in the piece-rate due to the wage assignment.

Appendix Figure 4: Bag Production & Bag Quality Over Time



(a) Average Weekly Bags

(b) Average Weekly Bag Quality Score

This figure shows bag-making outcomes by treatment group over the 21 weeks of the bag-making program. Subfigure (a) shows the average weekly bags and subfigure (b) shows the average weekly quality score. In subfigure (b), the vertical gray dashed lines indicate the weeks that immediately followed monthly feedback visits by bag-making field agents.

Appendix Table 1: Household Level Baseline Indicators, Intervention Treatments (Means and Standard Deviations)

	Ctrl No-Bags (1)	**GUP No-Bags** (2)	**SOUP No-Bags** (3)	**Ctrl Bags** (4)	**GUP Bags** (5)	**SOUP Bags** (6)	**p-value F-test Joint Sig** (7)
Household Size	7.16	7.61	7.28	7.18	7.29	7.48	0.82
	(3.88)	(4.16)	(3.72)	(3.50)	(3.62)	(3.52)	
Average Age, Household	25.57	25.02	25.13	24.77	24.37	23.04	0.02
	(10.90)	(10.55)	(10.69)	(9.69)	(9.13)	(8.12)	
Land Area (Acres)	4.54	4.50	4.84	4.65	5.03	4.82	0.81
	(3.92)	(4.02)	(4.12)	(3.77)	(3.88)	(3.94)	
Monthly Per Capita Cons (USD)	58.96	54.00	62.61	58.58	59.24	58.55	0.53
	(38.68)	(34.49)	(40.88)	(39.18)	(41.96)	(37.03)	
Monthly HH Income (USD)	42.34	42.60	48.74	41.12	45.07	45.49	0.87
	(56.76)	(58.84)	(56.77)	(53.13)	(50.23)	(56.72)	
Savings Balances (USD)	2.40	1.63	2.78	1.65	2.46	3.11	0.50
	(11.98)	(8.54)	(13.91)	(9.87)	(11.09)	(14.00)	
Food Security Index	0.00	0.03	0.16	-0.14	0.14	0.19	0.04
	(1.01)	(1.02)	(1.11)	(0.91)	(1.05)	(1.07)	
Asset Value Index	0.03	0.03	0.09	0.03	0.00	0.02	0.99
	(1.07)	(1.01)	(1.12)	(1.02)	(0.73)	(0.77)	
Financial Inclusion Index	0.00	0.00	0.05	-0.08	-0.08	0.19	0.24
	(0.96)	(0.94)	(1.07)	(0.84)	(0.76)	(1.36)	
Wealth Index	0.01	0.00	0.16	-0.06	0.07	0.15	0.50
	(1.01)	(0.94)	(1.13)	(0.93)	(0.89)	(1.04)	
Physical Health Index	-0.06	0.00	0.00	0.03	0.13	0.00	0.12
	(1.02)	(1.07)	(1.04)	(0.98)	(0.93)	(0.99)	
Mental Health Index	0.00	-0.04	0.14	-0.08	0.03	0.08	0.38
	(1.02)	(0.97)	(1.00)	(1.02)	(1.02)	(1.02)	
Political Involvement Index	0.00	0.09	-0.16	0.06	0.00	0.01	0.27
	(1.00)	(0.99)	(1.01)	(0.99)	(1.00)	(1.00)	
Female Empowerment Index	0.00	0.11	-0.09	0.02	-0.12	0.00	0.57
	(0.99)	(1.03)	(0.99)	(1.01)	(1.03)	(1.00)	

This table shows means and standard deviations for key indicators at baseline by intervention treatment assignment. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms. Indices are centered around mean baseline value. The wealth index aggregates the seven previous indicators (land area, consumption, income, savings, food security, asset value, and financial inclusion). The last column contains the p-value from an F-test of joint significance of all treatments. See Appendix Table 15 for components of all indices.

Appendix Table 2: Household Level Baseline Indicators, Bags Treatments (Means and Standard Deviations)

	Ctrl-Bags Simple (1)	**Ctrl-Bags Complex** (2)	**GUP-Bags Simple Low UCT** (3)	**GUP-Bags Simple High UCT** (4)	**GUP-Bags Complex Low UCT** (5)	**GUP-Bags Complex High UCT** (6)	**SOUP-Bags Simple** (7)	**SOUP-Bags Complex** (8)	**p-value F-test Joint Sig** (9)
Household Size	6.96 (3.38)	7.38 (3.61)	6.59 (3.21)	7.71 (4.33)	8.13 (4.31)	7.28 (2.83)	7.58 (3.38)	7.38 (3.66)	0.83
Average Age, Household	24.20 (8.53)	25.28 (10.63)	26.87 (11.44)	23.02 (7.43)	24.45 (9.27)	22.74 (6.78)	22.72 (7.45)	23.36 (8.76)	0.34
Land Area, Acres	4.76 (4.22)	4.54 (3.31)	3.45 (3.11)	5.01 (4.14)	6.50 (3.89)	6.17 (3.85)	5.01 (4.05)	4.63 (3.84)	0.04
Monthly Per Capita Cons. (USD)	51.98 (33.12)	64.57 (43.18)	46.03 (35.43)	51.17 (34.99)	73.18 (49.59)	75.58 (44.50)	63.31 (37.35)	53.64 (36.18)	0.00
Monthly HH Income (USD)	41.37 (58.35)	40.89 (48.04)	25.69 (30.27)	43.46 (44.11)	71.51 (79.01)	56.17 (48.62)	53.74 (67.07)	36.99 (42.18)	0.02
Savings Balances (USD)	1.46 (8.19)	1.82 (11.21)	4.27 (15.26)	1.43 (6.25)	1.76 (9.61)	1.73 (9.55)	2.54 (12.41)	3.70 (15.48)	0.69
Food Security Index	-0.28 (0.79)	0.00 (0.99)	-0.13 (0.86)	0.28 (1.16)	0.08 (1.03)	0.36 (1.11)	0.31 (1.05)	0.06 (1.08)	0.00
Asset Value Index	-0.10 (0.88)	0.15 (1.13)	-0.12 (0.67)	0.00 (0.71)	0.12 (0.98)	0.10 (0.68)	0.06 (0.78)	0.00 (0.75)	0.59
Financial Inclusion Index	0.00 (0.89)	-0.14 (0.79)	0.00 (0.91)	-0.05 (0.76)	-0.21 (0.58)	-0.10 (0.67)	0.08 (1.18)	0.30 (1.52)	0.14
Wealth Index	-0.15 (0.93)	0.02 (0.91)	-0.28 (0.89)	0.05 (0.70)	0.35 (1.06)	0.37 (0.84)	0.24 (1.01)	0.05 (1.07)	0.15
Physical Health Index	0.00 (0.98)	0.06 (0.99)	0.12 (1.02)	0.27 (0.62)	-0.09 (1.14)	0.11 (0.95)	0.00 (1.02)	0.03 (0.97)	0.30
Mental Health Index	-0.20 (0.96)	0.02 (1.06)	-0.10 (1.03)	0.18 (1.11)	0.00 (0.98)	0.07 (0.94)	0.14 (1.07)	0.01 (0.96)	0.35
Political Inv. Index	0.06 (0.99)	0.06 (0.99)	0.03 (1.00)	0.12 (0.99)	-0.14 (1.02)	-0.08 (1.01)	0.06 (1.00)	0.00 (1.01)	0.91
Female Emp. Index	0.05 (0.98)	0.01 (1.04)	0.10 (1.07)	-0.10 (0.96)	-0.39 (1.09)	-0.27 (0.97)	-0.06 (1.03)	0.00 (0.97)	0.50

This table shows means and standard deviations for key indicators at baseline by bag-making treatment assignment. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms. Indices are centered around mean baseline value. The wealth index aggregates the seven previous indicators (land area, consumption, income, savings, food security, asset value, and financial inclusion). The last column contains the p-value from an F-test of joint significance of all treatments. See Appendix Table 15 for components of all indices.

Appendix Table 3: Intervention Treatment Effects at Two Years

Panel A: Impacts of GUP and SOUP

		Livestock Value (1)	Asset Value (2)	Monthly HH Income (3)	Monthly Cons./Cap. (4)	Health Index (5)	Time All Labor (6)	Time Leisure (7)
GUP	ITT	72.92	74.85	7.49	1.13	0.04	-0.45	-1.63
	SE	(17.61)	(38.64)	(2.54)	(1.62)	(0.03)	(16.77)	(9.52)
	p-val	0.00	0.05	0.00	0.48	0.24	0.98	0.86
	q-val	0.00	0.22	0.05	0.74	0.46	1.00	0.96
	Bsl p-val	0.37	0.08	0.66	0.57	$0.04^{(+)}$	.	.
SOUP	ITT	32.09	83.33	4.00	2.91	-0.01	7.31	0.41
	SE	(14.65)	(39.53)	(2.49)	(1.64)	(0.05)	(23.04)	(10.90)
	p-val	0.03	0.04	0.11	0.08	0.85	0.75	0.97
	q-val	0.17	0.18	0.33	0.27	0.96	0.91	1.00
	Bsl p-val	0.68	0.29	0.97	0.18	0.50	.	.
Ctrl Mean		240.17	589.48	36.59	44.15	-0.16	330.83	84.63
Ctrl SD		348.59	764.13	43.08	30.15	0.84	269.48	127.04
Obs		2909	2900	2907	2880	2767	1130	1130
GUP - SOUP = 0	p-val	0.05	0.86	0.25	0.35	0.36	0.76	0.87

Panel B: Impacts by Bags Treatment

		Livestock Value (1)	Asset Value (2)	Monthly HH Income (3)	Monthly Cons./Cap. (4)	Health Index (5)	Time All Labor (6)	Time Leisure (7)
Control-Bags	ITT	-3.16	56.57	2.48	2.98	-0.02	7.45	11.39
	SE	(16.21)	(46.06)	(2.49)	(2.01)	(0.05)	(22.42)	(10.30)
	p-val	0.85	0.22	0.32	0.14	0.65	0.74	0.27
	q-val	0.96	0.45	0.56	0.38	0.91	0.91	0.49
	Bsl p-val	0.88	0.85	0.59	0.71	0.48	.	.
GUP-No-Bags	ITT	77.20	104.70	9.40	2.73	0.01	12.04	9.93
	SE	(23.75)	(47.92)	(3.33)	(2.07)	(0.04)	(38.38)	(22.98)
	p-val	0.00	0.03	0.01	0.19	0.76	0.75	0.67
	q-val	0.04	0.17	0.05	0.44	0.91	0.91	0.91
	Bsl p-val	0.25	0.12	0.56	0.13	0.22	.	.
GUP-Bags	ITT	64.73	72.67	6.44	0.63	0.05	0.34	6.02
	SE	(21.75)	(54.46)	(3.32)	(1.63)	(0.05)	(21.59)	(10.31)
	p-val	0.00	0.18	0.05	0.70	0.35	0.99	0.56
	q-val	0.05	0.44	0.22	0.91	0.57	1.00	0.83
	Bsl p-val	0.61	0.09	0.59	0.56	$0.01^{(+)}$	.	.
SOUP-No-Bags	ITT	26.77	107.61	4.97	5.09	0.03	48.41	0.10
	SE	(18.55)	(49.79)	(3.14)	(1.99)	(0.05)	(41.40)	(14.46)
	p-val	0.15	0.03	0.12	0.01	0.63	0.24	0.99
	q-val	0.39	0.17	0.33	0.09	0.90	0.46	1.00
	Bsl p-val	0.15	0.88	0.60	0.73	0.56	.	.
SOUP-Bags	ITT	34.92	91.45	4.22	1.99	-0.07	2.24	10.99
	SE	(20.01)	(50.65)	(3.36)	(1.61)	(0.08)	(27.95)	(11.74)
	p-val	0.08	0.07	0.21	0.22	0.37	0.94	0.35
	q-val	0.27	0.27	0.45	0.45	0.59	1.00	0.57
	Bsl p-val	0.19	0.01	0.25	0.06	0.25	.	.
Ctrl Mean		242.93	578.33	35.94	43.64	-0.17	325.83	83.81
Ctrl SD		356.19	760.35	42.14	29.43	0.85	267.81	128.06
Obs		2909	2900	2907	2880	2767	1130	1130

This table shows treatment effects among all sample households (Bags and No-Bags) two years after the start of the GUP and SOUP interventions. Panel A shows average effects of GUP and SOUP; the omitted group is Control households (Bags and No-Bags) in any village. Panel B shows effects by Bags sub-treatment; the omitted group is Control No-Bags households in any village. The sample is restricted to villages with at least 30 compounds. We estimate Equation 1 using household-level data from the two-year survey, including surveyor fixed effects and the baseline value of the outcome when possible, and adding an indicator for whether or not household was treated with bags (Panel A only). We control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village), imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). Livestock value is the total number of livestock owned times the median reported price for each animal. Asset value is the total number of assets (including livestock, household and productive assets, and stocks), valued using asset prices relative to the price of goats from other countries. Monthly household income is the sum of income from the household's business, farm, wage labor, and (revenue from) animals. (Income from wage labor does not include bags earnings.) Monthly consumption per capita includes both food and non-food expenditure. Health index includes two variables: average daily living score (mean of capacity bathing, lifting, walking, and working) and sick day (1 if the member did not miss a day of work due to illness in the last year, 0 otherwise). Time all labor includes time spent on bags and other wage labor, agriculture, businesses, animals, and home labor. Time leisure is minutes spent yesterday on religious activities, social activities, ceremonies, traveling, personal care, and resting. (These time use measures are taken at the two-year mark, not during the bags program.) Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 4: Effects of GUP, SOUP, and UCT on Income and Labor Supply Among Bags Households During Bags Program - No Controls for Imbalanced Variables

Panel A: Effects of GUP and SOUP

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
GUP	ITT	9.48	0.27	1.13	0.12	0.66	-8.76
	SE	(1.35)	(0.12)	(0.50)	(0.05)	(0.35)	(12.39)
	p-val	0.00	0.03	0.03	0.02	0.06	0.48
	q-val	0.00	0.11	0.11	0.11	0.16	0.56
SOUP	ITT	-0.37	-0.17	-0.61	-0.11	-0.29	-8.66
	SE	(0.56)	(0.13)	(0.50)	(0.06)	(0.38)	(12.75)
	p-val	0.52	0.20	0.22	0.07	0.45	0.50
	q-val	0.56	0.33	0.33	0.16	0.55	0.56
Ctrl Mean		8.01	0.00	3.76	0.58	2.45	535.59
Ctrl SD		25.04	1.00	3.97	0.49	3.01	158.72
Obs		288	18816	18816	18816	18816	3442
GUP - SOUP	ITT	9.85	0.44	1.74	0.23	0.95	-0.10
GUP - SOUP	p-val	0.00	0.01	0.01	0.00	0.04	0.99

Panel B: Effects of GUP and SOUP for Households Assigned Complex Bags

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
GUP-Complex	ITT	11.28	0.26	1.33	0.10	0.53	-6.72
	SE	(0.94)	(0.18)	(0.75)	(0.09)	(0.42)	(18.59)
	p-val	0.00	0.16	0.08	0.27	0.22	0.72
	q-val	0.00	0.30	0.17	0.34	0.33	0.76
SOUP-Complex	ITT	-0.91	-0.29	-1.12	-0.15	-0.64	-1.98
	SE	(0.67)	(0.14)	(0.51)	(0.08)	(0.35)	(14.67)
	p-val	0.19	0.05	0.03	0.06	0.07	0.89
	q-val	0.33	0.15	0.11	0.16	0.16	0.89
Ctrl Mean		14.44	-0.16	3.01	0.54	1.88	541.23
Ctrl SD		34.83	0.91	3.54	0.50	2.59	171.77
Obs		160	9156	9156	9156	9156	1608
GUP - SOUP	ITT	12.19	0.55	2.45	0.26	1.17	-4.74
GUP - SOUP	p-val	0.00	0.01	0.00	0.01	0.01	0.76

Panel C: Effects of UCT

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
High UCT	ITT	13.05	0.35	1.46	0.16	0.87	2.59
	SE	(0.00)	(0.16)	(0.67)	(0.07)	(0.49)	(12.61)
	p-val	0.00	0.03	0.03	0.03	0.08	0.84
	q-val	0.00	0.11	0.11	0.11	0.17	0.86
Low UCT	ITT	4.35	0.18	0.73	0.08	0.42	-22.41
	SE	(0.00)	(0.15)	(0.66)	(0.07)	(0.41)	(19.58)
	p-val	0.00	0.25	0.27	0.22	0.31	0.26
	q-val	0.00	0.34	0.34	0.33	0.39	0.34
Ctrl Mean		8.01	0.00	3.76	0.58	2.45	535.59
Ctrl SD		25.04	1.00	3.97	0.49	3.01	158.72
Obs		288	18816	18816	18816	18816	3442
High - Low	ITT	8.70	0.17	0.73	0.08	0.45	25.00
High - Low	p-val	0.00	0.41	0.41	0.40	0.46	0.24

This table shows treatment effects among Bags households during the Bags program, without controls for imbalanced variables. (In other words, we report the coefficients from Table 2, estimated in a specification without controls for imbalanced variables. In our replication package, we provide code to do the same for Tables 3 and 4, but we do not show the results here due to restrictions on the appendix length.) In all panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP-Bags and SOUP-Bags. Panel B shows effects of GUP-Bags and SOUP-Bags for households assigned complex bags. Panel C shows effects of GUP-Bags separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with at least 30 compounds. In column 1, we estimate Equation 1 using household-level data from the midline survey during bag-making, excluding controls for the five imbalanced variables. We include surveyor fixed effects as well as the baseline value of the outcome when possible. In columns 2-5, we estimate Equation 2 use household-week-level bag production data, excluding controls for the five imbalanced variables. We include station-week fixed effects. The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. In column 6, we estimate Equation 1 using household-month-level data from time use surveys during bag-making, excluding controls for the five imbalanced variables. We include surveyor fixed effects. In all columns, we control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village). Daily time non-bags labor is minutes spent yesterday on farming, business operations, animal production and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 5: Effects of GUP, SOUP, and UCT on Income and Labor Supply Among Bags Households During Bags Program - Long Specification

		Monthly HH Income + UCT	Weekly Bags Prod. Index	Weekly # of Bags	Weekly Partic. (0/1)	Weekly Bags Earnings	Daily Time Non-Bags Labor
		(1)	(2)	(3)	(4)	(5)	(6)
GUP	ITT	4.35	-0.04	-0.09	-0.02	-0.14	6.09
	SE	(0.00)	(0.20)	(0.85)	(0.08)	(0.59)	(22.70)
	p-val	0.00	0.85	0.91	0.83	0.81	0.79
	q-val	0.00	0.92	0.93	0.92	0.92	0.92
GUP-Savings	ITT	0.00	0.03	0.06	0.02	0.09	18.34
	SE	(0.00)	(0.13)	(0.48)	(0.07)	(0.33)	(13.59)
	p-val	0.00	0.81	0.90	0.77	0.79	0.18
	q-val	0.07	0.92	0.93	0.92	0.92	0.41
GUP High UCT	ITT	8.70	0.51	1.68	0.24	1.52	-3.88
	SE	(0.00)	(0.27)	(1.14)	(0.08)	(0.90)	(26.87)
	p-val	0.00	0.06	0.15	0.01	0.10	0.89
	q-val	0.00	0.30	0.38	0.07	0.30	0.93
GUP-Savings X High UCT	ITT	-0.00	-0.04	-0.03	-0.02	-0.19	12.56
	SE	(0.00)	(0.19)	(0.72)	(0.08)	(0.59)	(27.00)
	p-val	0.02	0.82	0.96	0.78	0.74	0.64
	q-val	0.23	0.92	0.96	0.92	0.92	0.92
Complex	ITT	-0.00	-0.24	-0.97	-0.08	-0.74	24.75
	SE	(0.00)	(0.14)	(0.55)	(0.06)	(0.38)	(15.29)
	p-val	0.22	0.09	0.08	0.19	0.05	0.11
	q-val	0.43	0.30	0.30	0.41	0.30	0.31
GUP X Complex	ITT	0.00	0.15	0.77	0.08	0.24	-57.03
	SE	(0.00)	(0.41)	(1.67)	(0.20)	(0.99)	(31.14)
	p-val	0.78	0.71	0.65	0.71	0.81	0.07
	q-val	0.92	0.92	0.92	0.92	0.92	0.30
GUP-Sav X Complex	ITT	-0.00	0.52	2.04	0.22	1.40	-51.72
	SE	(0.00)	(0.31)	(1.09)	(0.16)	(0.84)	(23.90)
	p-val	0.04	0.10	0.07	0.18	0.10	0.03
	q-val	0.26	0.30	0.30	0.41	0.30	0.26
GUP High UCT X Complex	ITT	-0.00	-0.36	-0.90	-0.20	-1.08	38.07
	SE	(0.00)	(0.49)	(2.04)	(0.24)	(1.26)	(36.05)
	p-val	0.27	0.47	0.66	0.39	0.39	0.29
	q-val	0.50	0.77	0.92	0.67	0.67	0.53
GUP-Sav X High UCT X Complex	ITT	0.00	-0.49	-2.16	-0.19	-1.30	55.80
	SE	(0.00)	(0.35)	(1.25)	(0.17)	(1.00)	(36.19)
	p-val	0.04	0.16	0.09	0.27	0.20	0.13
	q-val	0.26	0.40	0.30	0.50	0.41	0.35
Ctrl Mean		2.13	0.18	4.59	0.62	3.07	529.85
Ctrl SD		5.14	1.06	4.24	0.49	3.30	144.11
Obs		220	12978	12978	12978	12978	2409

This table shows estimates from a fully saturated model, incorporating all GUP treatments and their interactions, for labor supply outcomes among Bags households during the Bags program. SOUP households are excluded. The omitted group is Control-Bags households making simple bags. The sample is restricted to villages with at least 30 compounds. In column 1, we use household-level data from the midline survey during bag-making, and we include surveyor fixed effects as well as the baseline value of the outcome when possible. In columns 2-5, we use household-week-level bag production data, and include station-week fixed effects. In column 6, we use household-month-level data from time use surveys during bag-making, and include surveyor fixed effects. In all columns, we control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Daily time non-bags labor is minutes spent yesterday on farming, business operations, animal production and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in the table. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 6: Effects of GUP, SOUP, and UCT on Inputs Among Bags Households During Bags Program - Long Specification

		Exp. Labor (1)	**Exp. Herbicide** (2)	**Exp. Fertilizer** (3)	**Time Bags** (4)	**Time Field** (5)	**Time Business** (6)	**Time Home** (7)	**Time All Labor** (8)	**Time Leisure** (9)
GUP	ITT	16.54	22.74	-21.13	9.35	-20.73	8.89	14.78	7.84	26.01
	SE	(11.28)	(8.42)	(27.02)	(13.25)	(15.08)	(6.21)	(21.84)	(25.27)	(13.89)
	p-val	0.16	0.01	0.44	0.48	0.17	0.16	0.50	0.76	0.07
	q-val	0.54	0.20	0.76	0.76	0.54	0.54	0.76	0.89	0.34
GUP-Savings	ITT	-6.06	-18.40	-37.79	25.14	5.61	-1.61	13.43	34.66	-18.63
	SE	(3.05)	(7.52)	(12.05)	(22.70)	(9.49)	(6.50)	(19.38)	(20.70)	(17.31)
	p-val	0.06	0.02	0.00	0.27	0.56	0.81	0.49	0.10	0.29
	q-val	0.34	0.23	0.13	0.62	0.78	0.89	0.76	0.47	0.63
GUP High UCT	ITT	-15.99	-18.67	6.18	-10.05	-28.35	14.35	9.08	-11.63	-13.90
	SE	(13.42)	(8.59)	(29.66)	(28.30)	(22.90)	(11.66)	(29.52)	(28.69)	(15.67)
	p-val	0.25	0.04	0.84	0.72	0.22	0.22	0.76	0.69	0.38
	q-val	0.59	0.30	0.90	0.89	0.56	0.56	0.89	0.88	0.73
GUP-Sav X High UCT	ITT	-5.01	14.67	-12.44	-24.28	17.98	-5.61	4.34	-3.97	-13.50
	SE	(4.77)	(7.64)	(17.52)	(43.74)	(26.08)	(10.83)	(24.79)	(27.67)	(21.39)
	p-val	0.30	0.07	0.48	0.58	0.49	0.61	0.86	0.89	0.53
	q-val	0.65	0.34	0.76	0.80	0.76	0.81	0.91	0.92	0.76
Complex	ITT	-1.16	4.14	-103.86	13.01	-4.82	7.11	21.71	34.47	-7.03
	SE	(4.12)	(2.08)	(17.92)	(9.09)	(13.07)	(5.56)	(11.67)	(15.58)	(9.18)
	p-val	0.78	0.06	0.00	0.16	0.71	0.21	0.07	0.03	0.45
	q-val	0.89	0.34	0.00	0.54	0.89	0.56	0.34	0.25	0.76
GUP X Complex	ITT	-12.92	-13.44	83.98	-1.59	41.49	-0.07	-95.15	-52.01	-43.59
	SE	(11.39)	(8.89)	(33.81)	(41.34)	(26.96)	(17.74)	(37.20)	(52.24)	(19.81)
	p-val	0.27	0.14	0.02	0.97	0.13	1.00	0.01	0.32	0.03
	q-val	0.62	0.54	0.23	0.98	0.54	1.00	0.20	0.67	0.25
GUP-Sav X Complex	ITT	3.59	13.28	-16.86	-16.40	-41.69	-3.20	-5.99	-57.45	25.50
	SE	(4.82)	(8.44)	(19.05)	(30.82)	(32.64)	(13.36)	(20.47)	(40.93)	(19.90)
	p-val	0.46	0.13	0.39	0.60	0.21	0.81	0.77	0.17	0.20
	q-val	0.76	0.54	0.73	0.81	0.56	0.89	0.89	0.54	0.56
GUP High UCT X Complex	ITT	6.54	5.51	11.18	-35.26	-6.45	-13.01	58.20	12.53	68.39
	SE	(14.62)	(8.81)	(44.93)	(51.55)	(34.55)	(20.68)	(43.95)	(53.89)	(23.37)
	p-val	0.66	0.54	0.81	0.50	0.85	0.53	0.19	0.82	0.00
	q-val	0.86	0.76	0.89	0.76	0.91	0.76	0.56	0.89	0.13
GUP-Sav X High UCT X Complex	ITT	16.63	-7.92	80.76	2.14	13.33	18.92	19.49	55.05	-17.99
	SE	(11.32)	(8.94)	(30.58)	(54.09)	(44.19)	(19.22)	(29.40)	(45.94)	(25.66)
	p-val	0.16	0.38	0.01	0.97	0.76	0.33	0.51	0.24	0.49
	q-val	0.54	0.73	0.20	0.98	0.89	0.67	0.76	0.58	0.76
Ctrl Mean		3.60	1.62	80.68	107.77	245.28	19.45	263.99	578.30	164.14
Ctrl SD		14.27	9.08	98.87	96.82	175.36	58.73	132.31	152.18	148.78
Obs		206	206	206	1358	2409	2409	2409	2409	2409

This table shows estimates from a fully saturated model, incorporating all GUP treatments and their interactions, for inputs among Bags households during the Bags program. SOUP households are excluded. The omitted group is Control-Bags households making simple bags. The sample is restricted to villages with at least 30 compounds. In columns 1-3, we use household-level data from the midline survey during bag-making and include surveyor fixed effects as well as the baseline value of the outcome when possible. In columns 4-9, we use household-month-level data from time use surveys (minutes spent on each activity yesterday) during bag-making and include surveyor fixed effects. In all columns, we control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). Expenditure on labor, herbicide, fertilizer are expenditures in the last 12 months. Home labor includes childcare, cleaning, cooking, collecting firewood/water, and shopping. All labor includes bags and other wage labor, agriculture, businesses, animals (which is minimal, so we do not show separately), and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in the table. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 7: Effects of GUP, SOUP, and UCT on Non-Bags Outputs Among Bags Households During Bags Program - Long Specification

		Monthly Harvest Value (1)	**Monthly Residual Prod.** (2)	**Monthly Business Revenue** (3)	**Monthly Business Income** (4)	**Monthly Wage Income** (5)	**Monthly Non-Bags Income** (6)
GUP	ITT	-278.52	-398.20	1.52	3.83	-1.32	2.47
	SE	(154.21)	(135.13)	(11.57)	(5.54)	(0.97)	(6.07)
	p-val	0.08	0.01	0.90	0.50	0.19	0.69
	q-val	0.21	0.03	0.94	0.62	0.33	0.79
GUP-Savings	ITT	-122.50	-4.27	76.10	41.73	0.17	37.45
	SE	(45.13)	(35.02)	(16.10)	(4.57)	(0.31)	(3.35)
	p-val	0.01	0.90	0.00	0.00	0.59	0.00
	q-val	0.05	0.94	0.00	0.00	0.71	0.00
GUP High UCT	ITT	140.79	179.10	108.91	47.11	0.45	50.16
	SE	(111.51)	(112.60)	(55.95)	(24.97)	(0.84)	(27.13)
	p-val	0.22	0.13	0.06	0.07	0.59	0.08
	q-val	0.35	0.29	0.19	0.20	0.71	0.20
GUP-Savings X High UCT	ITT	78.35	121.23	-173.01	-82.48	-0.74	-80.05
	SE	(176.09)	(146.94)	(41.03)	(12.58)	(1.02)	(15.45)
	p-val	0.66	0.42	0.00	0.00	0.47	0.00
	q-val	0.78	0.55	0.00	0.00	0.61	0.00
Complex	ITT	-460.38	-360.54	10.35	7.77	-0.42	6.47
	SE	(146.47)	(130.16)	(9.85)	(3.98)	(1.18)	(4.49)
	p-val	0.00	0.01	0.30	0.06	0.73	0.16
	q-val	0.02	0.04	0.43	0.19	0.81	0.30
GUP X Complex	ITT	346.73	389.48	-1.00	-12.15	-1.72	-12.66
	SE	(275.25)	(254.56)	(25.43)	(9.05)	(2.09)	(10.07)
	p-val	0.22	0.14	0.97	0.19	0.42	0.22
	q-val	0.35	0.30	0.97	0.34	0.55	0.35
GUP-Sav X Complex	ITT	104.59	9.74	-89.60	-43.87	-0.18	-43.38
	SE	(72.34)	(96.29)	(37.67)	(7.30)	(0.79)	(6.07)
	p-val	0.16	0.92	0.03	0.00	0.83	0.00
	q-val	0.30	0.94	0.09	0.00	0.89	0.00
GUP High UCT X Complex	ITT	241.32	247.06	-114.86	-39.82	2.20	-43.33
	SE	(232.48)	(211.62)	(62.52)	(27.32)	(1.40)	(28.58)
	p-val	0.31	0.25	0.08	0.16	0.13	0.14
	q-val	0.43	0.38	0.20	0.30	0.29	0.30
GUP-Sav X High UCT X Complex	ITT	-80.15	-251.84	173.49	78.63	1.26	82.31
	SE	(232.78)	(213.58)	(54.55)	(14.75)	(1.09)	(15.83)
	p-val	0.73	0.25	0.00	0.00	0.26	0.00
	q-val	0.81	0.38	0.02	0.00	0.38	0.00
Ctrl Mean		476.02	48.85	3.34	0.57	1.57	2.13
Ctrl SD		526.14	345.38	12.46	2.65	3.29	5.14
Obs		206	203	219	219	220	220

This table shows estimates from a fully saturated model, incorporating all GUP treatments and their interactions, for non-bags outputs among Bags households during the Bags program. SOUP households are excluded. The omitted group is Control-Bags households making simple bags. The sample is restricted to villages with at least 30 compounds. We use household-level data from the midline survey during bag-making and include surveyor fixed effects as well as the baseline value of the outcome when possible. We control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). Harvest value is the total revenue from harvest, six months after the bag-making period. Residual productivity contains the residuals from a regression of harvest value on input expenditure, acreage, and average time on field. Business revenue and income pertain to all non-agricultural businesses. Wage income does not include earnings from bags. Non-Bags income includes income from agriculture, business, wages, and animals. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in the table. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 8: Interaction between GUP and NDVI on Labor Supply Among Bags Households During Bags Program

		Weekly Bags Prod. Index (1)	**Weekly # of Bags** (2)	**Weekly Partic. (0/1)** (3)	**Weekly Bags Earnings** (4)	**Daily Time Non-Bags Labor** (5)
GUP	ITT	0.43	1.72	0.21	1.01	0.19
	SE	(0.12)	(0.50)	(0.05)	(0.35)	(13.66)
	p-val	0.00	0.00	0.00	0.01	0.99
	q-val	0.00	0.00	0.00	0.02	0.99
past NDVI peak	ITT	0.04	0.08	0.00	0.27	19.14
	SE	(0.05)	(0.16)	(0.02)	(0.19)	(14.09)
	p-val	0.35	0.60	0.80	0.16	0.18
	q-val	0.47	0.69	0.86	0.30	0.30
GUP X past NDVI peak	ITT	-0.22	-0.82	-0.11	-0.55	-22.27
	SE	(0.10)	(0.33)	(0.04)	(0.46)	(26.68)
	p-val	0.03	0.01	0.01	0.23	0.41
	q-val	0.06	0.04	0.03	0.34	0.51
Ctrl Mean		-0.09	3.48	0.53	2.22	518.82
Ctrl SD		1.00	3.96	0.50	2.95	161.38
Obs		18816	18816	18816	18816	3442

This table shows treatment effects among Bags households during the Bags program, interacted with an indicator equal to 1 for weeks at or after peak vegetation, and 0 otherwise. We use the NDVI (Normalized Difference Vegetation Index) as a station-specific measure of vegetation; see Appendix J for details. The omitted group is Control-Bags households in any village during the weeks before the NDVI peak. The sample is restricted to villages with at least 30 compounds. In columns 1-4 we estimate Equation 2 using household-week-level bag production data, including station fixed effects instead of station-week fixed effects. In column 5, we estimate Equation 1 using household-month-level data from time use surveys during bag-making, including surveyor fixed effects. In all columns, we control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Daily time non-bags labor is minutes spent yesterday on farming, business operations, animal production and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in the table. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 9: Justifying Imputation of Time Spent on Bags

	(1) Monthly Wage Income (USD)	(2) Time Bags and/or Wage Labor
GUP-No-Bags	-0.78	-10.39
	(0.34)	(5.30)
SOUP-No-Bags	-0.36	-6.26
	(0.50)	(4.81)
Control-Bags	0.21	105.49
	(0.39)	(6.21)
GUP-Bags	-0.92	98.71
	(0.31)	(12.21)
SOUP-Bags	-0.25	75.71
	(0.69)	(10.84)
Ctrl Mean	1.13	9.73
Obs	864.000	789.000
p-val, GUP-Bags = GUP-No-Bags	0.651	0.000
p-val, SOUP-Bags = SOUP-No-Bags	0.890	0.000

This table shows treatment effects among all sample households (Bags and No-Bags) on wage income and wage labor during bag-making. The omitted group is Control No-Bags households in any village. The sample is restricted to villages with at least 30 compounds. In column 1, we estimate Equation 1 using household-level data from the midline survey during bag-making and including surveyor fixed effects. In column 2, we estimate Equation 1 using household-level averages over the five monthly time use surveys administered during the bags program, and we include surveyor fixed effects. In both columns, we control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). Standard errors are clustered at the village level. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 10: Attrition

		Midline During Bags (1)	Two-Year Survey (2)	Time-Use Midline (3)
Control-Bags	ITT	-0.01	-0.00	0.11
	SE	(0.01)	(0.01)	(0.03)
	p-val	0.31	0.81	0.00
GUP-No-Bags	ITT	0.00	0.01	0.01
	SE	(0.00)	(0.00)	(0.05)
	p-val	0.93	0.10	0.87
GUP-Bags	ITT	0.00	0.00	0.14
	SE	(0.00)	(0.01)	(0.03)
	p-val	0.45	0.45	0.00
SOUP-No-Bags	ITT	0.01	0.00	0.04
	SE	(0.01)	(0.01)	(0.05)
	p-val	0.67	0.81	0.46
SOUP-Bags	ITT	0.01	-0.00	0.10
	SE	(0.01)	(0.01)	(0.03)
	p-val	0.37	0.52	0.00
Ctrl Mean		0.99	0.95	0.67
Ctrl SD		0.10	0.21	0.28
Obs		870	2915	1238
GUP-Bags - Ctrl-Bags	p-val	0.20	0.47	0.36
GUP-Bags - SOUP-Bags	p-val	0.45	0.27	0.17

This table shows treatment effects on survey participation, or selective attrition. The omitted group is Control No-Bags households in any village. The sample is restricted to villages with at least 30 compounds. In each column, we estimate Equation 1, including surveyor fixed effects. We control for stratification variables and imbalanced variables. The outcome in column 1 is a binary indicator for whether the household was surveyed in the midline that took place during bags, conditional on being one of the 1085 households selected to participate. The outcome in column 2 is a binary indicator for whether the household was surveyed at the two-year mark. The outcome in column 3 is the household's average participation rate in the six time use surveys that took place during the bags program, conditional on being one of the 1542 households selected to participate. Standard errors are clustered at the village level.

Appendix Table 11: Effect of Complex vs. Simple on Bag Quality

		Weekly High Bags (1)	Mean Quality Score (out of 1) (2)	Quality Bag Index (3)
Control Complex	ITT	-0.68	-0.03	-0.33
	SE	(0.29)	(0.03)	(0.09)
	p-val	0.02	0.30	0.00
	q-val	0.05	0.44	0.00
GUP Simple	ITT	0.37	-0.01	-0.02
	SE	(0.52)	(0.03)	(0.12)
	p-val	0.48	0.68	0.84
	q-val	0.62	0.76	0.84
GUP Complex	ITT	-0.79	-0.06	-0.44
	SE	(0.30)	(0.04)	(0.09)
	p-val	0.01	0.12	0.00
	q-val	0.03	0.21	0.00
Ctrl Mean		1.84	0.73	1.20
Ctrl SD		3.20	0.19	0.70
Obs		18816	10854	10854
GUP Complex - Ctrl Complex	ITT	-0.11	-0.03	-0.11
GUP Complex - Ctrl Complex	p-val	0.65	0.43	0.19

This table shows treatment effects by complexity on bag quality among Bags households during the Bags program. The omitted group is Control-Bags households with simple bags. The sample is restricted to villages with at least 30 compounds. We estimate Equation 2 using household-week-level bag production data and including station-week fixed effects. We control for stratification variables and imbalanced variables. Column 1 shows the number of "high quality" bags submitted, column 2 shows the mean quality score (the percent of criteria fulfilled, which was used to determine whether the bag was "low," "mid," or "high" quality) over all of the bags submitted, and column 3 shows a composite measure of bag quality that awards 2 points for every "high" quality bag submitted and 1 point for every "mid" quality bag submitted. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in the table.

Appendix Table 12: Effect of Complex vs. Simple on Bags Labor Supply: Dynamics

		Bag Production Index, First Four Weeks (1)	**Bag Production Index, Final 17 Weeks** (2)
Control Complex	ITT	-0.16	-0.24
	SE	(0.15)	(0.13)
	p-val	0.26	0.08
	q-val	0.46	0.23
GUP Simple	ITT	0.44	0.14
	SE	(0.19)	(0.17)
	p-val	0.02	0.40
	q-val	0.13	0.48
GUP Complex	ITT	0.18	0.08
	SE	(0.18)	(0.21)
	p-val	0.31	0.70
	q-val	0.46	0.70
Ctrl Mean		0.15	0.19
Ctrl SD		1.03	1.06
Obs		3584	15232
GUP Complex - Ctrl Complex	ITT	0.35	0.32
GUP Complex - Ctrl Complex	p-val	0.06	0.11

This table shows treatment effects by complexity on bag-making labor supply for Bags households, looking separately at the first four weeks (after 5 weeks of initial piloting) versus the final 17 weeks. The omitted group is Control-Bags households with simple bags. The sample is restricted to villages with at least 30 compounds. We estimate Equation 2 using household-week-level bag production data and including station-week fixed effects. We control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in the table.

Appendix Table 13: Bags Labor Supply among Bags Households During Bags Program, GUP-No-Savings vs. GUP-Savings

		Weekly Bags Prod. Index (1)	**Weekly of Bags** (2)	**Weekly Partic. (0/1)** (3)	**Weekly Bags Earnings** (4)
GUP-No-Savings	ITT	0.22	0.97	0.09	0.53
	SE	(0.13)	(0.53)	(0.06)	(0.35)
	p-val	0.09	0.07	0.11	0.14
	q-val	0.12	0.12	0.13	0.14
GUP-Savings	ITT	0.33	1.32	0.15	0.80
	SE	(0.13)	(0.51)	(0.05)	(0.38)
	p-val	0.01	0.01	0.01	0.04
	q-val	0.03	0.03	0.03	0.07
Ctrl Mean		0.00	3.76	0.58	2.45
Ctrl SD		1.00	3.97	0.49	3.01
Obs		18816	18816	18816	18816
GUP-Sav - GUP-No-Sav = 0	p-val	0.18	0.25	0.12	0.22

This table shows treatment effects of GUP on bag-making labor supply for Bags households during the Bags program, separately for GUP households with savings collection and GUP households without. The omitted group is Control-Bags households. The sample is restricted to villages with at least 30 compounds. We estimate Equation 2 using household-week-level bag production data, and including station-week fixed effects. We control for stratification variables (household size, age of primary respondent, asset ownership index, whether the household owns a business, whether the primary respondent has any savings, total surface area of the household's land owned, livestock ownership index, the village's distance to the nearest market, and the number of compounds in the village) and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in the table.

Appendix Table 14: Wage Elasticity Results

	(1) IHS(bags)	(2) IHS(bags)	(3) IHS(bags)
log(wage)	0.01	-0.10**	
	(0.04)	(0.04)	
log(wage) X experience		0.26***	
		(0.08)	
log(wage(t-3))			0.18***
			(0.05)
Individual FE's	Yes	Yes	Yes
Week FE's	Yes	Yes	Yes
N	18816	18816	16128
R-Sq	0.49	0.49	0.50
Mean Bags (Level)	3.89	3.89	3.89

This table shows effects of the (log) wage on weekly bag production (inverse hyperbolic sine). We include individual and week fixed effects and cluster standard errors at the village level. In column 1 we use the effective, current wage. In column 2, we interact the wage with an indicator of experience with the wage. Specifically, "experience" indicates either the fourth week of the month (when participants had the wage for three weeks already) or, for participants whose village was randomly assigned two consecutive high wage months and two consecutive low wage months, any week in the second consecutive month with the same wage. In column 3, we use the 3-week lagged wage.

Appendix Table 15: Variable Definitions and Construction

Variable	Definition
Asset Value (USD)	The total number of assets (including livestock, household and productive assets, and stocks), valued using asset prices relative to the price of goats from other countries.
Business Income (and Revenue), Monthly (USD)	Monthly business revenues minus expenses. For each business within the household, we ask about how many months in the last year the business was operating, how many months were "normal" (neither higher nor lower than last month), how many months were "high profit," and how many months were "low profit." We ask about sales and profits in the last month, as well as profits in normal, high, and low months. We use the ratio of last month's profits to high and low profits to impute sales in high and low profit months, compute total sales in the last year by summing over sales in normal, high, and low months, and divide by the number of months the business was operating to get monthly revenue. We do the same exercise for expenses.
Consumption, Monthly per Capita (USD)	Total of all food consumption, temptation goods and fuel expenditures, school, clothing, festival expenditures, transportation, rent, medical expenditures, and home improvement spending. Food consumption is (quantity*market price) for each food item. Scaled to per capita monthly values.
Expenditure on Inputs	Expenditures on manure, fertilizer, labor, herbicide, insecticide, and other inputs in the last year.
Financial Inclusion Index	Standardized index of two variables, centered around the baseline mean. The first variable is the amount received in loans by the household in the last year. The second variable is the savings balances at the time of the survey.
Food Security Index	Standardized index of three variables, centered around baseline means. The first two variables equal 0 if the household answered "all year" or "during the lean season only" to the following questions, about adults and kids, respectively: "Did adults/kids ever reduce number of meals per day or reduce portions over the past year?" The third variable equals 0 if the household answered "all year" or "during the lean season only" to the question "Did adults ever skip entire days without eating?"
Female Empowerment Index	Standardized index of five variables, centered around baseline means. Each variable is the answer to the question, "To what extent do you believe yourself able to make your own decisions concerning X?" The categories X are food, school expenses, health expenses, visiting friends, and purchases. They are measured on a scale from 1 to 3.
Harvest Value	We ask about the quantity of each crop sold in the last year. If the units of harvested crops are the same as the units of sold crops and we have the sale price, then we use this price to compute the sale value of each crop, and then sum over crops. Otherwise, we use the median price for that crop.
Income, Monthly HH (USD)	The sum of monthly business income, monthly crop income, monthly wage income, and monthly animal revenue. Monthly crop income is harvest value minus expenditure on inputs (annual), divided by twelve.
Livestock Value (USD)	Sum of values of goats, fowl, pigs, sheep, and cows.
Mental Health Index	Standardized index of three variables, centered around baseline means. The first is economic satisfaction, measured on a scale from 1 to 5. The second is a standardized index of five measures: feeling sad, crying, not eating, not working, and feeling restless, measured on a scale from not at all, hardly ever, some of the time, or most of the time. The third is whether the individual was not worried in the last year (0 if the member experienced a period of worry in the last year, 1 otherwise).
Physical Health Index	Standardized index of two variables, centered around baseline means. The first is the average daily living score, which is the mean of four variables: capacity bathing, capacity lifting, capacity walking, and capacity working (each measured on a scale from 1 being easily done to 4 being unable to do). The second is whether or not they did not take a sick day (1 if the member did not miss a day of work due to illness in the last year, 0 otherwise).
Political Involvement Index	z-score of attendance at village meetings in the last year (1 if the person attended a village meeting in the last 12 months, 0 otherwise), centered around baseline mean.