CENTER FOR GLOBAL DEVELOPMENT

A Firm of One's Own: Experimental Evidence on Credit Constraints and Occupational Choice

Andrew Brudevold-Newman, Maddalena Honorati, Gerald Ipapa, Pamela Jakiela, Owen Ozier

ABSTRACT

We evaluate two labor market interventions targeting young women in Nairobi, Kenya. The first was a multifaceted program involving vocational training, in-kind transfers of physical capital, and ongoing mentoring. The second was an unrestricted cash grant. Both interventions shift women into self-employment, impacts which persist after six years. Both programs also increase income in the short-term, but those effects disappear over time. Though the two treatments have similar impacts on labor market outcomes, women in the multifaceted program report significantly higher wellbeing six years after treatment relative to both women in the control group and those who received the grants.

KEYWORDS

youth unemployment, vocational training, cash grants, microenterprises, entrepreneurship, occupational choice, credit constraints, Africa, gende

JEL CODES

J24, M53, O12

WORKING PAPER 646 · MAY 2023

A Firm of One's Own: Experimental Evidence on Credit Constraints and Occupational Choice

Andrew Brudevold-Newman
World Bank
(abrudevoldnewman@worldbank.org)

Maddalena Honorati
World Bank
(mhonorati@worldbank.org)

Gerald Ipapa
University of Delaware
(gipapa@udel.edu)

Pamela Jakiela
Williams College, BREAD, Center for Global Development, IZA, and J-PAL
(pj5@williams.edu)

Owen Ozier
Williams College, BREAD, IZA, and J-PAL
(owen.ozier@williams.edu)

Andrew Brudevold-Newman, Maddalena Honorati, Gerald Ipapa, Pamela Jakiela, and Owen Ozier . 2023. "A Firm of One's Own: Experimental Evidence on Credit Constraints and Occupational Choice." CGD Working Paper 646. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/firm-ones-own-experimental-evidence-credit-constraints-and-occupational-choice.

We are grateful to Suleiman Asman, Sarah Baird, Esther Duflo, Maya Eden, David Evans, Deon Filmer, Jessica Goldberg, Markus Goldstein, Joan Hamory, Eeshani Kandpal, Jason Kerwin, David Lam, Isaac Mbiti, David McKenzie, Ted Miguel, Patrick Premand, Jeff Smith, Andy Zeitlin, numerous conference and seminar attendees, and four anonymous referees for their constructive comments. Rohit Chhabra, Emily Cook-Lundgren, Julian Duggan, and Laura Kincaide provided excellent research assistance. This research was funded by the IZA/DFID Growth and Labour Markets in Low Income Countries Programme, the CEPR/DFID Private Enterprise Development in Low-Income Countries Research Initiative, the ILO's Youth Employment Network, the National Science Foundation (award number 1357332), and the World Bank (SRP, RSB, i2i, Gender Innovation Lab). The study was registered at the AEA RCT registry under ID number AEARCTR-0000459. We are indebted to staff at the International Rescue Committee (the implementing organization) and Innovations for Poverty Action for their help and support. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent. All errors are our own.

CENTER FOR GLOBAL DEVELOPMENT
2055 L Street, NW Fifth Floor
Washington, DC 20036
202.416.4000

1 Abbey Gardens
Great College Street
London
SW1P 3SE

www.cgdev.org

1 Introduction

Millions of African youth cannot find good jobs. Young people entering the labor market in Africa often spend years transitioning between periods of short-term or informal employment, unpaid work, training, and job search (Filmer and Fox 2014, Bandiera, Elsayed, Smurra, and Zipfel 2022). During the extended period when out-of-school youth are searching for work and gaining relevant experience, they are heavily reliant on their families and social networks for support (Filmer and Fox 2014); this makes young women particularly vulnerable (Dupas 2011, Alfonsi, Bandiera, Bassi, Burgess, Rasul, Sulaiman, and Vitali 2020). In addition, credit and savings constraints can create a poverty trap for unemployed youth who would like to start a microenterprise. In this context, active labor market programs that promote female entrepreneurship may have the dual benefit of helping young people out of poverty traps while providing social protection for vulnerable young women.

We evaluate two labor market interventions intended to promote microentrepreneurship among young women in Nairobi, Kenya. One intervention is a "microfranchising" program designed to help women launch branded franchise businesses that were (loosely) linked to well-known Kenyan brands. The intervention is a multifaceted program that involves life skills training, vocational education, asset transfers, and ongoing mentoring. We evaluate the program through a randomized trial based on over-subscription. Eligible applicants were randomly assigned to either the franchise treatment, a control group, or an alternative grant treatment that offered women an unrestricted, one-off cash transfer (of just over 200 US dollars) without any additional monitoring, training, or mentoring. We analyze the impacts of these two labor market interventions – the **franchise** treatment and the **grant** treatment – using follow-up data collected in the first, second, and sixth years after treatment. Attrition rates are below ten percent in all three rounds of follow-up, and attrition is never correlated with either treatment, allowing us to interpret differences in outcomes between the treatment arms and the comparison group as the average causal impacts of the interventions.

Both treatments increase self-employment, and impacts persist for more than five years. In the sixth year after treatment, women assigned to the franchise treatment are 12 percentage points more likely to be self-employed than those in the control group, while women in the grant treatment are 10 percentage points more likely to be self-employed than control women. Year 6 impacts on self-employment are virtually identical to effects observed in Year 1 and Year 2, suggesting that treatment causes approximately one in ten women to make a permanent shift into microentrepreneurship. Yet, neither intervention leads to long-term increases in income. In the first year after treatment, both the franchise and grant treatments lead to statistically significant and economically meaningful increases in earned income, but these impacts mostly disappear by Year 2 and are completely absent in Year 6. Since incomes in the control group increased threefold between Year 1 and Year 6, our results are consistent with Blattman, Fiala, and Martinez (2020), who find that the short-term impacts on income from a youth entrepreneurship program in Uganda disappear over time as the control group enters the labor market. However, we do not observe any evidence that assignment to treatment eventually leads to lower incomes relative to the control group. Instead, over 90 percent of women in all three treatment arms state that they would prefer self-employment over paid work if the hours and wages were the same, suggesting that the persistence of the observed treatments effects on self-employment reflects women's preference for own account work.

The women in our sample are young – still in adolescence at baseline, when more than 80 percent still lived with a parent or other adult relative. This makes it difficult to use household consumption as a proxy for welfare, since the children of household heads often have limited information about household expenditures and the consumption of other household members. Instead, we measure individual wellbeing by aggregating indicators of living conditions (e.g. whether a women lives in a home with electricity or piped water), food security (which includes both objective and subjective elements), and subjective life satisfaction. Combining these into an aggregate wellbeing index, we find that the franchise treatment has positive and statistically significant impacts on wellbeing in the sixth year after treatment, increasing overall wellbeing by 0.19 standard deviations. In contrast, cash grants do not increase wellbeing, and women assigned to the franchise arm are significantly better off (in terms of overall wellbeing, not income or wealth) than those assigned to the grant treatment. Impacts are primarily driven by improvements in food security and subjective life satisfaction, and impacts on purely objective measures of living conditions are not statistically significant.

It may initially seem puzzling that two interventions that have remarkably similar impacts on young women's labor market outcomes have different effects on (relatively) long-run wellbeing. We present four pieces of evidence which, taken together, explain this finding by showing how the bundled franchise treatment shifted participants' perceptions in ways that the grant alone did not. First, causal mediation analysis suggests that the impacts of the franchise treatment on wellbeing are not primarily explained by self-employment; this is consistent with the possibility that life skills training, mentoring, and other holistic aspects of the franchise program impacted women's aspirations or other aspects of their psychosocial welfare.[1] Second, consistent with this possibility, we find the most pronounced positive impacts of the franchise treatment on the more subjective measures of wellbeing – food security (which captures both objective measures of access to food and subjective measures of anxiety related to food scarcity) and subjective life satisfaction. Again, this makes sense if impacts are partially explained by life skills training or mentoring, or potentially the specific combination of these elements within the franchise treatment. Third, measures of participants' beliefs about their own counterfactual outcomes suggest that those assigned to the franchise treatment perceive themselves as destined to have become entrepreneurs (consistent with psychological theories of hindsight bias); these tendencies are far less pronounced among those assigned to the grant treatment. Finally, using evidence from vignettes, we find that women assigned to the franchise treatment are significantly more likely to believe that a program that helps young women become self-employed is

[1]These conclusions are subject to the strong assumptions necessary for causal mediation analysis.

more beneficial than a program that helps young women find work.

Our estimates of the impacts of the franchise and grant treatments point to two broad conclusions. First, active labor market interventions targeting youth can have persistent impacts on occupational trajectories, even when they do not lead to measurable improvements in income or wellbeing. A one-off, unrestricted cash grant nudged young women into microentrepreneurship, and this effect does not appear to be diminishing over time in spite of the absence of any measurable benefit from self-employment (in terms of either earned income or subjective wellbeing). This highlights the importance of ethical considerations in policy evaluation, since even programs that fail may have long-term impacts on the lives of participants. Second, programs that look similar in terms of their initial impacts can lead to meaningfully different long-run outcomes. This highlights the importance of evaluating impacts on a broad range of life outcomes and time scales, and also suggests that policymakers should be cautious about assuming that cash grants can achieve the same developmental objectives as more (multifaceted) programs.

We contribute to active literatures on youth underemployment in Africa (Bandiera, Elsayed, Smurra, and Zipfel 2022), microenterprises (De Mel, McKenzie, and Woodruff 2008, Blattman, Fiala, and Martinez 2014, Jayachandran 2021), cash transfers (Haushofer and Shapiro 2016, Banerjee, Niehaus, and Suri 2019), vocational training and mentoring (Hamory, Kremer, Mbiti, and Miguel 2016, Brooks, Donovan, and Johnson 2018, Alfonsi, Bandiera, Bassi, Burgess, Rasul, Sulaiman, and Vitali 2020), the returns to capital for self-employed women (De Mel, McKenzie, and Woodruff 2009, Fafchamps, McKenzie, Quinn, and Woodruff 2011, Blattman, Green, Jamison, Lehmann, and Annan 2016, Bernhardt, Field, Pande, and Rigol 2019, Riley 2022), and poverty traps (Kraay and McKenzie 2014, Balboni, Bandiera, Burgess, Ghatak, and Heil 2022). The specific microfranchising model that we study has not been evaluated, but both individual components (e.g. vocational training) and multifaceted programs combining these elements have been evaluated in a number of settings. Our work is closely related to recent evaluations of multifaceted poverty alleviation programs that combing training, asset transfers, and ongoing support (Banerjee, Duflo, Goldberg, Karlan, Osei, Parienté, Shapiro, Thuysbaert, and Udry 2015, Banerjee, Duflo, and Sharma 2021), particularly those targeting women (Adoho, Chakravarty, Korkoyah Jr., Lundberg, and Tasneem 2014, Bandiera, Buehren, Burgess, Goldstein, Gulesci, Rasul, and Sulaiman 2020, Bossuroy, Goldstein, Karimou, Karlan, Kazianga, Parienté, Premand, Thomas, Udry, Vaillant, et al. 2022). Our findings also speak to recent work comparing cash grants to alternative poverty alleviation interventions (cf. McIntosh and Zeitlin 2021). In particular, our finding that a holistic program impacts wellbeing while cash grants do not resonates with recent work by Hussam, Kelley, Lane, and Zahra (2022), who find that employment raises refugees' psychosocial wellbeing while cash grants of comparable magnitude do not.

We also introduce a new method for eliciting program participants' beliefs about the treatment effects of development interventions. Our approach builds on work by Smith, Whalley, and Wilcox (2012), who suggest a range of approaches for measuring participant's assessments of program impacts, and McKenzie (2016), who shows that participants do a poor job of estimating their own counterfactual outcomes. We extend this work by offering a behavioral economic explanation for McKenzie's (2016) findings (hindsight bias), developing a survey-based approach to belief elicitation that circumvents these issues, and showing that our method suggests that program participants hold remarkably accurate beliefs about the treatment effects of the interventions they have participated in.

The rest of this paper is organized as follows. We describe our setting, the interventions we evaluate, and the design of our randomized evaluation in Section 2. We present our impact evaluation results in Section 3. We discuss our new measure of participant beliefs about program impacts in Section 4. Section 5 concludes.

2 Experimental Design and Data

We evaluate two labor market interventions targeting young women in three poor neighborhoods in Nairobi, Kenya. The implementing partner (the International Rescue Committee, or IRC) worked with community organizations to advertise an entrepreneurship program open to women under 21 years old in the Baba Dogo, Dandora, and Lunga Lunga neighborhoods in the eastern periphery of Nairobi. Our research design is based on oversubscription, since more women applied for the program than could be accommodated by the implementing organization. A total of 905 eligible applicants were stratified by neighborhood and application month, and then randomly assigned to one of three treatment arms: forty percent to the **franchise** treatment implemented by the IRC, twenty percent to an unrestricted **grant** treatment described below, and forty percent to the **control** group.[2]

2.1 Treatments

The first study arm was a multifaceted "microfranchising" program implemented by the IRC. The program helped young women launch branded microenterprises: either salons or food carts, each linked to a nationally-recognized brand. All participants in the franchise treatment completed a two-week business and life-skills training course before being matched with one of the two franchise partners. Women matched to the salon franchise completed six weeks of classroom training in a hairstyling school followed by a two-week internship at a local salon. After completing their training, participants identified a suitable premises and received necessary capital and business inputs such as branded aprons, a hair washing sink, a hair dryer, and a variety of hair styling products. Women matched to the food cart franchise partner attended a one-day training course and then received a mobile food cart, an apron or t-shirt displaying the franchise partner's logo, and an initial stock of prepared food to sell. Each microenterprise launched through the program was assigned a mentor who visited the business every few weeks. Mentors helped the young women in the program get their businesses off the ground — for example, by coordinating additional training with the franchise partners, helping women set up bank accounts for their busi-

[2] The trial is registered at `https://www.socialscienceregistry.org/trials/459`. The decision to use unequally-sized treatment arms reflects two considerations: the implementing organization's obligation to place as many women as possible in the microfranchising program and the limited amount of donor funding available for cash grants.

nesses, and assisting with financial management and record keeping. Additional details about the implementation of the franchise treatment are provided in Online Appendix B.

Women who were randomly assigned to the grant treatment were offered an unrestricted transfer of 20,000 Kenyan shillings (or 239 US dollars at the prevailing exchange rate). The value of the grant was selected to make it roughly comparable to the implementing cost of the microfranchising program.[3] Grant recipients were told that there were no restrictions on how the funds could be used and that the grant did not need to be paid back. Disbursements to grant recipients were timed to coincide with the launch of the microfranchise businesses.

2.2 Data Collection

Our analysis draws on four main sources of data. First, we administered a brief baseline survey to all eligible applicants prior to randomization. The baseline survey was conducted in the first half of 2013, and treatments were implemented between August and December of that year. We conducted our first follow-up survey between July and September of 2014, which was between seven and 11 months after microfranchises were launched. That Year 1 follow-up survey was conducted by phone; to keep the survey as brief as possible, we only collected information on labor force participation. Our second follow-up survey was conducted in-person in 2015, between 14 and 22 months after treatment.[4] In the Year 2 follow-up, we collected information about a broad range of labor market and wellbeing outcomes. Our final round of follow-up data collection took place between August of 2018 and August of 2019, the sixth year after treatment. To increase statistical power for noisy outcomes such as earned income, we conducted three waves of phone surveys over the course of a year, as suggested by McKenzie (2012).[5] For our analysis, we construct individual-level averages of labor market outcomes that were measured multiple times in our Year 6

[3] See Online Appendix B for a discussion of implementation costs.

[4] Over 98 percent of Year 2 follow-up interviews took place between February and July of 2015. Eight respondents were interviewed in August of 2015, and seven additional respondents were interviewed between September and December of 2015.

[5] The first wave of Year 6 phone surveys took place between August and November of 2018; the second wave between November of 2018 and February of 2019; and the third wave between March and August of 2019.

follow-up survey.

2.3 Baseline Characteristics

Online Appendix Table A1 describes the baseline characteristics of the 905 women in our sample. As expected, they are young: aged 17–20 at the time of the baseline survey. 92 percent completed primary school, but only 41 percent completed secondary school (though all women in the sample had left school prior to baseline). The average level of educational attainment was 10.3 years, which is very similar to the average of 10.6 years among women aged 18–20 living in Nairobi who were interviewed for the 2014 Demographic and Health Survey (Kenya DHS 2014). 48 percent of women in our sample were born outside of Nairobi, and 12 percent had lost both parents. Relative to the women living in Nairobi who were interviewed in the 2014 Kenya DHS, the women in our sample are less likely to be married (16 percent in our sample vs. 28 percent of DHS respondents from Nairobi) and more likely to have given birth (41 percent in our sample vs. 26 percent of DHS respondents). This likely reflects the targeting of the program to women living in some of Nairobi's poorest neighborhoods.

At baseline, 55 percent of women in our sample had any paid work experience, and 34 percent had done some form of vocational training. Only 15 percent were involved in any type of income-generating activity at baseline. However, many did substantial amounts of unpaid domestic work. For example, 63 percent report doing more than 20 hours of unpaid housework in the week prior to the baseline survey.

Online Appendix Table A1 reports summary statistics disaggregated by treatment. Most covariates are well-balanced. To adjust for any imbalances, we use lasso to select a set of baseline covariates that predict treatment status after controlling for stratum fixed effects. These variables are included in our main empirical specifications (though we also report specifications excluding baseline covariates in the Online Appendix).

2.4 Compliance with Treatment

Not all women assigned to the franchise and grant arms took up treatment. As is typical in training programs (McKenzie and Woodruff 2014), many of those assigned to the franchise treatment chose not to participate in the program. Just over 61 percent of those assigned to the franchise treatment attended at least one day of business training (which was the first component of the franchise treatment), and 44 percent completed the program and launched a business. We observe very little contamination of the grant and control groups: one woman in the grant treatment arm and zero women in the control group participated in the business training sessions, and a total of five women who were not assigned to the franchise treatment launched a microfranchise business. Among those assigned to the grant treatment, 95 percent accepted and received the grant. One woman assigned to the control group and zero women assigned to the franchise treatment received the grant. Online Appendix Table A2 summarizes the impacts of the randomly assigned treatments on take-up of the interventions.

2.5 Attrition

Attrition rates are low in all of our follow-up surveys: we successfully surveyed 94 percent of the baseline sample in the Year 1 follow-up, 93 percent in the Year 2 follow-up, and 91 percent in the Year 6 follow-up. Regressions testing for differential attrition across treatment arms are reported in Online Appendix Table A2. Attrition is not associated with either treatment in any of our three rounds of data collection.

3 Results

3.1 Estimation Strategy

We report intent-to-treat estimates of the impacts of being invited to participate in the franchise and grant treatments on a range of outcomes related to participation in the labor market and individual wellbeing. We report OLS regressions of the form

$$Y_i = \alpha + \beta_F F_i + \beta_G G_i + \eta_s + X_i + \varepsilon_i \tag{1}$$

where Y_i is an outcome of interest, F_i is an indicator for random assignment to the franchise treatment, G_i is an indicator for random assignment to the grant treatment, η_s is a vector of randomization stratum fixed effects, X_i is a vector of baseline controls chosen by lasso, and ε_i is a conditionally-mean-zero error term.[6] In addition to our coefficient estimates, we report standard errors, unadjusted p-values, and Benjamini-Hochberg q-values that control the false discovery rate (Benjamini and Hochberg 1995, Anderson 2008).

3.2 Impacts on Labor Market Outcomes

Impacts on labor market outcomes in the first, second, and sixth years after treatment are reported in Table 1.[7] We consider five main labor market outcomes: an indicator for self-employment, an indicator for paid work other than self-employment (either formal employment or informal work for which a respondent receives payment in cash or in-kind), an indicator for involvement in any income-generating activity (either self-employment or paid work), hours worked, and earned income. Additional information on the construction of our outcome variables is included in Online Appendix C.

Both the franchise treatment and the grant treatment shift women into self-employment, seemingly permanently since estimated impacts are nearly identical in Years 1, 2, and 6. Both treatments increase the likelihood of self-employment by between 10 and 12 percentage points, and all coefficients are statistically significant, even after adjusting for multiple inference (q-values across the three rounds of follow-up range from 0.001 to 0.054). 24 percent of women in the control group report being self-employed in the Year 1 follow-up,

[6]Online Appendix Table A1 indicates the set of baseline covariates selected by lasso. As a robustness check, we also report specifications that only control for randomization strata.

[7]Online Appendix Table A3 replicates Table 1 but omits the baseline covariates selected by lasso, including only randomization stratum fixed effects as covariates. All results are similar in magnitude and significance with and without baseline covariates.

so the treatment effect reflects a 41–45 percent increase in self-employment. By the Year 6 follow-up, 38 percent of women in the control group are self-employed, but both treatments still increase the likelihood of self-employment by more than 25 percent. We can never reject the hypothesis that the treatment effects of the franchise and grant treatments are equal, suggesting that liquidity alone is enough to shift women into microentrepreneurship.

Neither the franchise treatment nor the grant treatment has consistent, statistically significant impacts on overall labor force participation. In Year 1, both treatments lead to statistically significant reductions in the likelihood of doing paid work (q-values 0.030 and 0.022), but these impacts disappear by the second year. In Year 2, the franchise treatment causes an overall increase in the likelihood of doing any income-generating activity (q-value 0.025), but this is absent in Year 1 and does not persist into Year 6. We also do not observe statistically significant impacts on hours worked except in Year 1, when the grant treatment increases total hours by more than a third (q-value 0.046). Thus, the treatments seem to impact the likelihood of self-employment, but do not have consistent impacts on the overall likelihood of working or the probability of working for others.

We observe statistically significant and economically meaningful positive impacts on earned income in the first year after treatment, but these effects disappear over time. In Year 1, the franchise treatment increases weekly earned income by 168 Kenyan shillings (q-value 0.069), while the grant treatment increases earned income by 298 shillings (q-value 0.022). Since the average income in the control group is only 496 shillings per week, these effects represent substantial increases in income. In contrast, estimated impacts in Year 2 and Year 6 are smaller in magnitude and not statistically significant.

The distribution of income is positively skewed, with many zeros and a long right tail. Because the variance is large relative to the mean, statistical power is a concern.[8]

[8] In the analysis of treatment effects on earned income reported above, we winsorize the top and bottom 0.5 percent of the data to limit the influence of a small number of outliers. However, even in the winsorized data, the standard deviation of income is well above the mean. Moreover, since incomes in the control group increased over time (as did their standard deviations), in our Year 6 data we are not powered to detect positive impacts on income similar in magnitude to those observed in Year 1. The standard errors in Table 1 suggest that we would have a power of 0.8 to detect a franchise treatment effect of 481 shillings and a grant treatment effect of 580 shillings. These are more than double the magnitudes of the (statistically

Moreover, the distribution of treatment effects is of intrinsic interest: a program that shifts a large number of people from zero income to a low but positive income is quite different, from a policy perspective, than one that has substantial positive impacts on a very small number of treated individuals. To explore the incidence treatment effects on income, Figure 1 presents the CDFs of income in Year 1, Year 2, and Year 6 separately by treatment. We also estimate distribution regression models testing the impact of treatment across the distribution of income levels observed in the control group (Foresi and Peracchi 1995, Chernozhukov, Fernández-Val, and Melly 2013). Distribution regression models provide a framework for estimating treatment effects across the entirety of the distribution, which also allows us to test the hypothesis that treatment effects are constant. We plot our distribution regression results in Figure 2; each panel shows the impact of a treatment on the probability of having income below a range of different income thresholds (Chernozhukov, Fernández-Val, and Melly 2020).

In Year 1, the CDFs of income for the franchise and grant treatments are clearly to the right of the CDFs of income for the control group (Figure 1), and distribution regressions suggest that both treatments have impacts across much of the distribution of income levels observed in the control group (Figure 2). The colored lines and shaded areas in the top two panels of Figure 2 are generally below the horizontal axis, because positive treatment effects across the distribution correspond to a negative difference in CDFs. A Kolmogorov-Smirnov test rejects the hypothesis that the distribution of incomes is the same in the control and franchise treatment groups (p-value 0.021), but does not reject the hypothesis that the distributions are the same in the control and grant treatment groups (p-value 0.115) – possibly in part because the grant treatment arm is half the size of the franchise arm, so statistical power is reduced. However, the distribution regression models suggest that both the franchise and grant treatments increased incomes across the income distribution and fail to reject the hypothesis of constant treatment effects. That is, as shown in the top two panels of Figure 2, in Year 1, distribution regression leads to rejection of the null hypothesis

significant) effects observed in the first year after treatment.

of no effect with Cramer-von-Mises p-values of 0.017 for the franchise treatment, and 0.012 for the grant treatment; while the p-values associated with constant effects are 0.193 and 0.582 respectively.

Though the Year 2 CDFs of income for the franchise and grant treatments are still slightly to the right of the control group distribution, both Kolmogorov-Smirnov tests and distribution regression models fail to reject the hypothesis of no impact of either treatment in Year 2, and also in Year 6 (see the p-values associated with the null hypothesis of no effect reported in the middle and bottom panels of Figure 2). Thus, the treatment effects on income observed in Year 1 appear temporary, and disappear by the second year after treatment – possibly because control group incomes increased substantially over time.[9]

While the effects begin to dissipate in the second year, there is some evidence that income benefits remain but only for part of the population. In Year 2, the distribution regression p-values associated with a test of the null hypothesis of constant treatment effects are 0.036 for the franchise treatment and 0.059 for the grant treatment. We interpret this as suggestive evidence that treatment effects on income persisted into the second year after treatment, but only for a small part of the distribution. However, we find no evidence of treatment effects on income for any part of the distribution by Year 6.

3.2.1 Treatment Effect Heterogeneity

Average treatment effects can mask substantial differences in impacts across treated individuals. Our results suggest that the franchise and grant treatments shift approximately one in ten women into entrepreneurship, suggesting that treatment may have a lasting impact on a small fraction of the population. To the extent that these women can be identified *ex ante* based on observable characteristics, policymakers might be able to better target the interventions, increasing average impacts. At the same time, the observed absence of treatment effects on income might mask important heterogeneity if programs increase incomes for some subpopulations (e.g. those shifted into entrepreneurship) but decrease incomes for

[9]In a population of unemployed youth (both men and women) in northern Uganda, Blattman, Fiala, and Martinez (2020) also find that early impacts of cash grants on income disappear over time.

others.

We test for treatment effect heterogeneity using the causal forest approach proposed in Athey and Imbens (2016) and Wager and Athey (2018). Causal forests are a machine learning technique that involves repeatedly partitioning the data based on covariates into sub-samples that show relatively little variation in treatment effect, generating estimates of individual-level conditional average treatment effects (CATEs). Chernozhukov, Demirer, Duflo, and Fernandez-Val (2018) and Athey and Wager (2019) propose using the best linear predictor of the CATE to construct a summary measure of the presence of detectable treatment effect heterogeneity that is explained by the observed covariates. Using their approach, we do not detect meaningful treatment effect heterogeneity in impacts on self employment, involvement in income-generating activities, or earned income. Online Appendix Figure A1 illustrates this, showing the distribution of estimated CATEs on self-employment. While we do observe some variability in estimated treatment effects, it likely results from noise (given our relatively small sample size). We never observe statistically significant differences between the estimated individual-level CATEs at the twentieth and eightieth percentiles of the distribution.

3.3 Impacts on Wellbeing

Impacts on an index of overall wellbeing, measured in Years 2 and 6, are reported in Table 2.[10] The index aggregates four distinct measures of objective and subjective wellbeing: an (objective) index of living conditions and household assets; a food security index, adapted from USAID's Household Food Insecurity Access Scale;[11] and measures of current and anticipated future (subjective) life satisfaction (Cantril 1965).[12]

In Year 2, neither treatment had a statistically significant impact on wellbeing. The

[10] The Year 1 follow-up did not include wellbeing-related outcomes.

[11] The Household Food Insecurity Access Scale asks about a 30-day recall period, and includes objective questions, such as "Did you go a whole day without eating anything because there was not enough to eat?" as well as more subjective questions, such as "Did you worry that you would not have enough food?"

[12] Additional information on the construction of our outcome variables is included in Online Appendix C. Online Appendix Table A4 reports impacts on the four individual components of the wellbeing index. Online Appendix Table A5 replicates Table A4 but omits the baseline covariates selected by lasso, including only randomization stratum fixed effects as covariates.

coefficient estimate suggests that the franchise treatment increased wellbeing by 0.10 standard deviations, but the impact is not significant (p-value 0.170) and we cannot reject the hypothesis that the franchise and grant treatments had comparable (non)effects on wellbeing in Year 2 (p-value 0.145).[13] Estimates of Year 2 impacts on the individual components of the wellbeing index indicate that the franchise treatment improved anticipated future wellbeing by 0.27 standard deviations, but the effect is only marginally significant (Online Appendix Table A4, q-value 0.056). Other components of the wellbeing index are not impacted in Year 2, and we can never reject the hypothesis that the impacts of the franchise and grant treatments are equal.

In contrast, results from Year 6 indicate that the franchise treatment improved overall wellbeing while the grant treatment did not. In Year 6, random assignment to the franchise treatment improved overall wellbeing by 0.19 standard deviations (p-value 0.010). Assignment to the grant treatment did not improve wellbeing (coefficient estimate -0.05, p-value 0.518), and we can reject the hypothesis that the franchise and grant treatments had the same impact on overall wellbeing (p-value 0.002). The franchise treatment had a statistically significant positive impact on food security, current subjective wellbeing, and anticipated future wellbeing (Online Appendix Table A4, q-values 0.020, 0.020, and 0.004, respectively), though it did not have a positive impact on objective living conditions. Point estimates suggest that the grant treatment had a negative impact on living conditions, food security, and current wellbeing, though results are not statistically significant. We can, however, reject the hypothesis that the franchise and grant treatments had the same impacts on food security and current and future wellbeing (Online Appendix Table A4, q-values 0.016, 0.016, and 0.018, respectively). Thus, in spite of their comparable impacts on labor market outcomes, the franchise treatment made women better off over the medium-to-long run, while the grant treatment alone did not.

[13] We report unadjusted p-values when discussing impacts on our aggregate index of overall wellbeing, but use adjusted q-values when discussing impacts on each of the four components in isolation.

3.3.1 Causal Mediation Analysis

Causal mediation analysis offers a set of statistical tools that can be used to explore potential mechanisms underlying treatment effects – but with the caveat that these methods rely on strong assumptions about the exogeneity of both the treatment and the mediator. Baron and Kenny (1986), early pioneers of mediation analysis, suggest testing the extent to which a variable, M_i, mediates the causal impact of a treatment, T_i, on outcome Y_i by estimating the regressions

$$M_i = \alpha_1 + \beta_1 T_i + \varepsilon_i \tag{2}$$

and

$$Y_i = \alpha_2 + \beta_2 T_i + \gamma_2 M_i + \varepsilon_i. \tag{3}$$

They argue that "perfect mediation holds if the independent variable [in Equation 3] has no effect when the mediator is controlled" (Baron and Kenny 1986, p. 1177).[14] In other words, they argue that the expected value of the regression coefficient β_2 should be zero when M_i explains all of the impact of T_i on Y_i. As many have noted, this claim is typically incorrect unless the treatment and the mediator are both randomly assigned (cf. Angrist and Pischke 2009, Vanderweele 2015); but their proposed approach has nevertheless influenced much of the subsequent literature on causal mediation. For example, Imai, Keele, and Yamamoto (2010) define an average causal mediation effect (ACME) which is equivalent to $\beta_1\gamma_2$ when Equations 2 and 3 characterize the data-generating process; they then show that the ACME is non-parametrically identified under a "sequential ignorability" assumption (in essence, that both the treatment and the mediator are as-good-as-random conditional on observables and, in the case of the mediator, treatment status).

Economists typically question the assumption of sequential ignorability that underlies causal mediation analysis: if it were reasonable to assume that treatment and potential mediators were as-good-as-random conditional on observables, one would not need to run randomized trials to estimate program impacts. Even when treatment is randomly assigned,

[14] See Vanderweele (2015) for discussion.

mediators are not. If, within the treatment or control group, the observed value of the mediator is correlated with potential outcomes in the absence of treatment, causal mediation analysis will be subject to a standard set of concerns about selection bias – invalidating its conclusions.

Those caveats notwithstanding, Table 3 reports the results of a causal mediation analysis of the impacts of the franchise treatment. We estimate the ACME, which is used to calculate the share of the overall treatment effect that could be attributed to a potential mediator (Imai, Keele, and Tingley 2010, Hicks and Tingley 2011). When we examine take-up of the program as a potential mediator (Panel A), we can never reject the hypothesis that observed treatment effects are fully mediated by participation in the microfranchising program, as expected. When we consider self-employment in Year 1 as a potential mediator (Panel B), we cannot reject the hypothesis that self-employment fully mediates the observed treatment effects on income in Year 1. However, results suggest that self-employment in Year 1 does not fully explain subsequent impacts on self-employment, and that observed impacts on wellbeing in Year 6 are not mediated by Year 1 self-employment.

This pattern may help to explain why the franchise treatment impacts wellbeing when the grant treatment does not. If the impacts of the franchise treatment on wellbeing in Year 1 were fully mediated by self-employment (in either Year 1 or Year 6), the franchise program would have to increase wellbeing among each of the one in ten women shifted into self-employment by approximately two standard deviations to generate the observed average treatment effect on wellbeing (of 0.19 standard deviations). This is not impossible – indeed, a comparison of the histograms of Year 6 wellbeing in the franchise and control arms suggests that it is plausible (Online Appendix Figure A2). However, it is also possible that the multifaceted package of life skills and vocational training and ongoing mentoring (that made up the franchise treatment) had positive impacts on some women who did not ultimately remain in self-employment, or on those who would have been self-employed regardless of the intervention.

4 Participant Beliefs about Treatment Effects

Given the tremendous lengths one must go to in order to produce credible estimates of a program's impacts, an important question is whether participants themselves understand the effects of the programs in which they participate. It is not uncommon for labor market programs to survey participants *ex post*; however, Smith, Whalley, and Wilcox (2012) find that such *ex post* assessments of a program's impact are not highly correlated with objective measures of program effects. Understanding participants' beliefs about program impacts is important for two reasons. Most obviously, if — through their participation — participants obtain reasonable estimates of program impacts, this information may be a feasible, low-cost alternative to formal impact evaluation. On the other hand, if program participants do not understand a program's impacts, even after they have participated in the program, it is hard to imagine that they are making optimal decisions about whether or not to participate. Participant beliefs can also provide insights into the psychological impacts of the program on women's attitudes and beliefs.

4.1 Empirical Approach and Practical Considerations

As Smith, Whalley, and Wilcox (2012) point out, one reason participant evaluations of programs may differ from rigorous estimates of program impacts is that participant evaluation questions are often quite open-ended. For example, participants in the National Job Training Partnership Act program were asked "Do you think that the training or other assistance that you got from the program helped you get a job or perform better on the job?" (Smith, Whalley, and Wilcox 2012, p. 9). This question is obviously problematic because it is not at all clear whether better on-the-job performance should be linked to any measurable outcome (e.g. income); moreover, the link between the fraction of participants who believe that the program had a positive impact and the estimated treatment effect of the program is unclear, making it difficult to test whether participants' subjective evaluations are accurate. Smith, Whalley, and Wilcox (2012) suggest replacing such subjective evaluation questions with alternatives that (i) clearly specify the outcomes and time periods of interest, (ii) ask for continuous (as opposed to binary) responses that can be directly compared to ITT estimates, and (iii) make the counterfactual nature of the question transparent.

We follow their recommendations and ask participants in the franchise and grant treatments to estimate the counterfactual probabilities of self-employment and paid work for a reference group of women similar to themselves. Specifically, in our Year 2 follow-up, we ask women in each of the two treatment arms the question: "I would like you to imagine 100 women from [your neighborhood] who applied to the [name of treatment arm] program but who were **not** admitted into it. In other words, please think about 100 women similar to yourself who were **not** selected to the [name of treatment arm] program. Out of 100 women, how many do you think are currently running or operating their own business?" We also ask an analogous question about involvement in paid work for others. Smith, Whalley, and Wilcox (2012) suggest using this question to construct a perceived counterfactual, which can then be compared with the empirically-observed average outcome in the treatment group. We take a different approach, asking each participant to estimate how many of 100 women similar to themselves who "applied for **and were admitted into**" the program were (at the time of the survey) operating their own business (and, in a subsequent question, we ask how many were doing paid work for others). We calculate each participant's implied belief about the treatment effect of the program (on, for example, self-employment) by taking the difference between the perceived frequency of self-employment among women invited to participate in the program and the perceived frequency of self-employment among similar women who were not invited to participate.

We also test a second method proposed by Smith, Whalley, and Wilcox (2012): asking participants about the probability that they would be self-employed (or doing paid work for others) in the absence of the program. These individual-level beliefs about one's own counterfactual can then be combined with data on actual outcomes to construct estimates of perceived treatment effects. However, as Smith, Whalley, and Wilcox (2012) emphasize, there are several drawbacks to this approach. First, program participants may find it inherently difficult to imagine what their lives would have been like in the absence of the program. For example, psychological studies of "hindsight bias" suggest that people have a difficult time remembering the beliefs they held in the past and tend to assume that realized outcomes were always foreseeable (Fischhoff 1975, Madarász 2012). In our context, we might expect that those who have received vocational training and gained self-employment experience might have a difficult time remembering how they had perceived themselves before; thus, hindsight bias might inflate participants' estimates of their own counterfactual, particularly among successful microentrepreneurs. Estimates of one's own counterfactual may also be biased by the tendency to attribute one's own success to individual agency as opposed to external factors (Miller and Ross 1975). This would lead those who have benefited from business or vocational training to overstate the likelihood that they would have started a successful business in the absence of the program. Consistent with this, McKenzie (2016) finds that program participants do a poor job of estimating own counterfactuals.

In the context of our evaluation, an additional problem with questions designed to elicit beliefs about one's own counterfactual probability of self-employment (or paid work) is that they are unlikely to work well when respondents have low levels of numeracy. Though almost 92 percent of the women in our sample completed primary school, a relatively large number are not familiar with the concept of percentages. Roughly one in four cannot (correctly) answer the question: "If there is a 75 percent chance of rain and a 25 percent chance of sun, which type of weather is more likely?" While it is possible to elicit probabilistic expectations from subjects with no prior knowledge of probability, it is costly and time-consuming to do so. Instead, we asked every subject categorical questions about their counterfactual probabilities of self-employment and paid work, and collected more specific data on counterfactual probabilities from those who successfully answered the screening question described above.[15]

[15] We worded the categorical question to make responses directly comparable to probability estimates. Respondents chose one of the following options: (1) *In the absence of the program, I would definitely be self-employed*, (2) *In the absence of the program, I would probably be self-employed but it is not certain*, (3)

4.2 Framework for Interpreting Empirics

To facilitate comparisons between different approaches to belief elicitation, we introduce a simple conceptual framework that formalizes the measurement issues highlighted above. First, consider an outcome, y, and a program whose causal effect on that outcome is to increase its expected value by $\beta > 0$. Let γ denote the expected value of y in the absence of the program: $E[y_j|T_j = 0] = \gamma$.

We wish to know whether program participants hold accurate beliefs about β. Let

$$\tilde{\beta}_i = \tilde{\beta} + \phi_i \tag{4}$$

denote participant i's belief about the impact of the program, and let

$$\tilde{E}[y_j|T_j = 0] = \tilde{\gamma} + \nu_i \tag{5}$$

be participant i's belief about the expected value of the outcome of interest for an untreated individual j who is outwardly similar to her. $\tilde{\beta}$ is the average belief about the impact of the program, and $\tilde{\gamma}$ is the average belief about the outcome of interest in the eligible population in the absence of the program. ϕ_i is the idiosyncratic component of beliefs about the impact of the program; without loss of generality, we assume that the distribution of ϕ_i is mean zero. ν_i can be decomposed into a mean-zero error term and a term which reflects the perceived difference between the population average of y and one's own counterfactual:

$$\nu_i = \tilde{\alpha}_i \cdot \mathbb{1}(j = i) + \epsilon_i. \tag{6}$$

As discussed above, asking participants about their own counterfactuals may be problematic (for example, because of hindsight bias), and the population mean of these $\tilde{\alpha}_i$ values, $\tilde{\alpha} =$

In the absence of the program, the chances of me being self-employed or not self-employed are equal, (4) *In the absence of the program, I would probably not be self-employed but it is not certain*, or (5) *In the absence of the program, I would definitely not be self-employed.*

$E[\tilde{\alpha}_i]$ may not be equal to 0.[16] Combining and generalizing these expressions, respondents report:

$$\tilde{E}[y_j|T_j] = \tilde{\beta} \cdot T_j + \tilde{\gamma} + \tilde{\alpha}_i \cdot \mathbb{1}(j = i) + \phi_i \cdot T_j + \epsilon_i \tag{7}$$

Specifically, when asked to report the rate of self-employment among 100 potential program participants who were not invited to participate in the program, a respondent in our study reports:

$$\tilde{E}[y_j|T_j = 0] = \tilde{\gamma} + \epsilon_i. \tag{8}$$

When asked to report the rate of self-employment among 100 potential program participants who were invited to participate in the program, she reports:

$$\tilde{E}[y_j|T_j = 1] = \tilde{\beta} + \tilde{\gamma} + \phi_i + \epsilon_i. \tag{9}$$

Finally, when asked to report her own counterfactual probability of self-employment, a participant reports:

$$\tilde{E}[y_i|T_i = 0] = \tilde{\gamma} + \tilde{\alpha}_i + \epsilon_i. \tag{10}$$

The framework presented above helps to clarify the distinctions between the different approaches to estimating participant beliefs. First, consider an estimate of participant beliefs constructed by taking the average belief about one's own counterfactual (in our context, the counterfactual probability of self-employment) and subtracting this from the observed outcome in the treatment group. The expected value of this estimator is:

$$\begin{aligned} E[y_j|T_j = 1] - E[\tilde{E}[y_i|T_i = 0]] &= \beta + \gamma - (\tilde{\gamma} + \tilde{\alpha} + E[\epsilon_i]) \\ &= \beta + (\gamma - \tilde{\gamma}) - \tilde{\alpha} \end{aligned} \tag{11}$$

since $E[\epsilon_i] = 0$. Thus, this estimator will be biased if participants hold inaccurate beliefs about the counterfactual probability of self-employment, and it will be biased when psychological factors such as hindsight bias lead participants to overstate their own counterfactual

[16] This may be thought of as a "Lake Wobegon" effect.

probability of self-employment. The second estimator proposed by Smith, Whalley, and Wilcox (2012) is constructed by subtracting the mean rate of self-employment in a reference group of untreated women from the observed rate of self-employment in the treatment group. The expected value of this estimator is given by:

$$\begin{aligned} E[y_j|T_j = 1] - E[\tilde{E}[y_j|T_j = 0]] &= \beta + \gamma - (\tilde{\gamma} + E[\epsilon_i]) \\ &= \beta + (\gamma - \tilde{\gamma}) \end{aligned} \tag{12}$$

This estimator overcomes the behavioral issues inherent in estimating one's own counterfactual. However, when estimates of participant beliefs constructed in this manner diverge from actual program impacts, it is impossible to determine whether participants hold inaccurate beliefs about the impact of the program or inaccurate beliefs about the counterfactual.

The outcomes of interest in impact evaluations are often difficult to measure, and considerable effort goes into the design and pre-testing of questionnaires. Nonetheless, there is no guarantee that outcome measures derived from survey questions (for example, about labor market participation) and participant responses to belief-elicitation questions will line up, particularly in low-income settings where formal, full-time employment is relatively uncommon (and there is continuous variation in the number of hours worked, and labor supply varies substantially from week to week).[17] Impact evaluation questions designed to measure beliefs about the counterfactual may reveal systematic deviations between participants' beliefs about outcome levels and actual outcome levels; however, such measurement error is only problematic if it cannot be separated from the quantity of interest. To address this issue, we propose an estimate of participant beliefs that is calculated by taking the difference between beliefs about the mean outcome of interest in a reference population of

[17] Smith, Whalley, and Wilcox (2012) are aware of this issue and recommend asking extremely specific questions: for example, what fraction of participants meet a well-specified criterion for employment — for example, working more than 35 hours per week — which can then be used to construct the empirical estimate of the programs impact. However, such precisely worded questions are not always feasible. In our context, we worried that any question of the form "Out of 100 women, how many spend at least X hours operating their own business?" would be substantially more difficult to answer than a less specific question because few people work full-time and there is no obvious break in the distribution of hours worked at any point.

treatment versus control individuals:

$$\begin{aligned}E[\tilde{E}[y_j|T_j = 0]] - E[\tilde{E}[y_j|T_j = 0]] \\ &= \tilde{\beta} + \tilde{\gamma} + E[\phi_i] + E[\epsilon_i] - (\tilde{\gamma} + E[\epsilon_i]) \\ &= \tilde{\beta}\end{aligned} \tag{13}$$

Such an estimator allows for a direct test of the hypothesis that participants hold accurate beliefs about program impacts; moreover, collection of the relevant data necessarily also allows researchers to assess the related issue of whether participants can estimate the counterfactual — allowing for a comparison of the different approaches of belief estimation.

4.3 Results

Our results, which are summarized in Figures 3 and 4, suggest that participants hold remarkably accurate beliefs about program impacts. Figure 3 compares ITT estimates of program impacts to estimates of participant beliefs about program impacts calculated by taking the difference in reference group probabilities for the treatment and control groups.[18] For example, the ITT estimates suggest that the franchise treatment increased the likelihood of self-employment in Year 2 by 11.2 percentage points; and the upper panel of Figure 3 shows that those assigned to the program believe that it increased the likelihood of self-employment by 12.3 percentage points. Similarly, those assigned to the cash grant treatment believe that it increased the likelihood of self-employment by 10.6 percentage points (shown in the lower panel of Figure 3); the ITT estimates suggest a 12.3 percentage point increase. Women assigned to both treatments also hold remarkably accurate beliefs about each program's impact on the likelihood of paid employment: women assigned to the franchise arm believe treatment reduced the probability of doing paid work by 3.6 percentage points, when the ITT estimate of the treatment effect suggests a 1.6 percentage point

[18] In other words, beliefs were estimated by asking women assigned to each treatment group to estimate reference group probabilities (frequencies) for both the treatment and comparison groups. Women assigned to the control group were not asked to estimate a reference group probability for those assigned to the treatment groups since they were not familiar with the details of each treatment.

reduction; and women assigned to the grant arm believe treatment reduced the probability of doing paid work by 6.6 percentage points, when the ITT estimate of the treatment effect suggests a 5.2 percentage point reduction. Thus, our results suggest that participants do a remarkably good job of estimating the impact of programs that they have participated in.

Figure 4 compares beliefs about the probability of self-employment and paid work to levels observed in the treatment and control groups, and compares beliefs about one's own counterfactual to beliefs about a reference population of untreated women. Several patterns are apparent. First, though beliefs about the levels of self-employment in the treatment and control groups are accurate, women in the franchise treatment group (but not the grant arm) underestimate the probability of paid work in both the treatment and the control group. Consequently, an estimate of the impact of the franchise program on the probability of paid work that compared counterfactual beliefs to observed levels in the treatment group would perform poorly. Moreover, differences between observed outcome levels and participant beliefs appear to be systematic, with women in either over- or under-estimating the levels in both the treatment and control groups. This suggests that it will typically be better to estimate program beliefs by comparing beliefs about the control group to beliefs about the treatment group (rather than the observed outcome levels in the treatment group).

The figure also demonstrates that estimates of one's own counterfactual are indeed biased: the average of own counterfactual estimates is consistently higher than the estimated outcome for a reference population of untreated women. Interestingly, the pattern is particularly pronounced when women assigned to the franchise treatment are asked to report their own counterfactual probability of self-employment. Women in the franchise arm believe their own counterfactual likelihood of self-employment was 39.9 percent (on average), which is actually higher than the rate of (Year 2) self-employment observed among women assigned to the franchise treatment (36.1 percent), and is substantially higher than either the observed rate of (Year 2) self-employment in the control group (24.3 percent) or women in the franchise treatment's belief about the likelihood of self-employment in a reference population of women who were not invited to participate in the program (25.9 percent).

Thus, though participants hold accurate beliefs about the level of self-employment in both the treatment and control groups, own counterfactual estimates are so inflated that they would suggest a negative impact of the program on self-employment – so, our evidence clearly supports the view that own counterfactual estimates are of little use in estimating treatment effects.[19] However, the evidence for hindsight bias is far less clear among women assigned to the grant treatment, or when women in either arm are asked to estimate their own counterfactual probability of paid work.[20] This finding provides further support for the idea that the franchise treatment had long-lasting impacts on women that are not fully explained by the observed increase in self-employment: though the franchise and grant arms have similar impacts on self-employment, the franchise treatment led to more pronounced changes in women's beliefs about themselves and their identities as entrepreneurs.

4.4 Vignettes

To further explore participants' beliefs about the impacts of entrepreneurship programs, we presented respondents in the Year 6 follow-up with the following vignette:

> *Please think about two women who are like you: from your neighborhood, the same age as you. One gets help from an NGO that allows her to start a small business like a small salon. The other woman gets help from a different NGO that helps her get a job. For both women there are challenges. It can be hard to keep your business open, and it can be hard to work for someone else.*

They were then asked to indicate which woman would be better off overall after one month, and which women would be better off overall after five years. The overwhelming majority

[19]This finding is consistent with recent work by McKenzie (2016); he finds that program participants (business owners) do a very poor job of estimating the counterfactual. Our results support his conclusion, but suggest that an alternative approach to eliciting participants' beliefs performs substantially better.

[20]Women in the grant arm believe that their own counterfactual probability was 30.9 percent. As discussed above, this is higher than the level observed in the control group. It is also higher than beliefs (among women in the grant arm) about the likelihood of self-employment among 100 women like them who did not receive grants (27.2 percent). However, it is substantially lower than the actual likelihood of self-employment among women in the grant treatment arm (37.1 percent), suggesting that hindsight bias is far more muted than in the franchise arm.

of women believed that the women who started her own business would be better off now and in the future. Across all three treatment arms, 81.9 percent of respondents said the self-employed woman would be better off after one month, and 87.6 percent said that she would be better off in five years. However, we also observe a positive impact of the franchise treatment: women in that arm were seven percentage points more likely to say that the self-employed woman was better off in the short-term (Online Appendix Table A7, p-value 0.037), though they were no more likely to say that she would be better off in five years. In contrast, women assigned to the grant arm were no more likely (than women in the control group) to say that the self-employed woman would be better off (in either one month or five years). We interpret this as additional evidence that the franchise treatment changed women's beliefs about self-employment, and potentially strengthened their identities as entrepreneurs.

5 Conclusion

We evaluate two labor market interventions intended to promote self-employment among women in three of Nairobi's poorest neighborhoods. The first is a multifaceted franchise treatment that provided life skills training, vocational education, physical capital, and ongoing mentoring. The second was a cash grant of 20,000 Kenyan shillings (equivalent to 239 US dollars in 2013). Though other multifaceted programs have been evaluated in a variety of contexts (cf. Banerjee, Duflo, Goldberg, Karlan, Osei, Parienté, Shapiro, Thuysbaert, and Udry 2015), and long-term studies are becoming more commonplace (cf. Banerjee, Duflo, and Sharma 2021), the microfranchising model has not previously been evaluated in this way.

Both the franchise and grant treatments increased the likelihood of self-employment: roughly one in ten women in each of the treatment arms was shifted into self-employment relative to the comparison group. Estimated impacts are nearly identical in Years 1, 2, and 6, suggesting that they are likely to persist for the foreseeable future. As in other studies of cash transfers, we see that if anything, cash grants temporarily induced an increase in labor force participation, with no evidence of a decrease in either the short or long term (Banerjee, Hanna, Kreindler, and Olken 2017). However, neither treatment leads to a long-run increase in the likelihood of being involved in any income-generating activity. Both the franchise treatment and the grant treatment had large and statistically significant impacts on income in the year after the program. However, the impacts on income disappeared after the first year, and did not return – though incomes among treated women are no lower than those in the control group.

Though the cash grant and the microfranchising program are similar in having persistent impacts on self-employment and only temporary effects on income, they differ in terms of their impacts on overall wellbeing. More than five years after treatment, women randomly assigned to the IRC's franchise treatment are significantly better off than those in the control group and better off than those offered unrestricted cash grants in lieu of the multifaceted program. This finding can be understood as a shift in perceptions of those in the franchise arm, documented through several empirical patterns: (1) those in the franchise arm believe that in the absence of the program, they would have been microentrepreneurs anyway, to an extent not seen in the grant arm; (2) vignettes reveal that those in the franchise arm believe self-employment to be preferable to paid work, to an extent not seen in the grant arm; (3) the more subjective elements of our wellbeing index are the ones on which there are differential impacts in the franchise arm; and (4) mediation analysis suggests that the shift into self-employment is not the mechanism through which wellbeing effects are brought about: that is, for example, it may be that the wellbeing of those who were going to enter self-employment regardless of the program are those who have experienced an increase in wellbeing.

Taken together, this study shows that though cash grants may, in many contexts, be easier to implement than multifaceted programs, it is not necessarily the case that all their impacts will be lasting (cf. Baird et al. 2019) or persistently similar (cf. McIntosh and Zeitlin 2022), even in a single economic environment. Multifaceted programs, such as the one we study, are designed to address psychological and social barriers via training and mentoring, while also including elements intended to address economic barriers. These kinds of complex, multifaceted programs, designed in coordination with local community organizations, can have subtle impacts – for example, by changing individual aspirations, or creating social connections to mentors and role models – that translate into benefits long after treatment.

References

Adoho, F., S. Chakravarty, D. T. Korkoyah Jr., M. Lundberg, and A. Tasneem (2014): "The Impact of an Adolescent Girls Employment Program: The EPAG Project in Liberia," World Bank Policy Research Working Paper 6832.

Alfonsi, L., O. Bandiera, V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, and A. Vitali (2020): "Tackling youth unemployment: Evidence from a labor market experiment in Uganda," *Econometrica*, 88(6), 2369–2414.

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.

Angrist, J. D., and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Athey, S., and G. Imbens (2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360.

Athey, S., and S. Wager (2019): "Estimating Treatment Effects with Causal Forests: An Application," *Observational Studies*, 5(2), 37–51.

Balboni, C., O. Bandiera, R. Burgess, M. Ghatak, and A. Heil (2022): "Why do people stay poor?," *The Quarterly Journal of Economics*, 137(2), 785–844.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2020): "Women's empowerment in action: evidence from a randomized control trial in Africa," *American Economic Journal: Applied Economics*, 12(1), 210–59.

Bandiera, O., A. Elsayed, A. Smurra, and C. Zipfel (2022): "Young Adults and Labor Markets in Africa," *Journal of Economic Perspectives*, 36(1), 81–100.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry (2015): "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries," *Science*, 348(6236).

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

Banerjee, A., P. Niehaus, and T. Suri (2019): "Universal basic income in the developing world," *Annual Review of Economics*, 11, 959–983.

Banerjee, A. V., R. Hanna, G. E. Kreindler, and B. A. Olken (2017): "Debunking the stereotype of the lazy welfare recipient: Evidence from cash transfer programs," *The World Bank Research Observer*, 32(2), 155–184.

Baron, R. M., and D. A. Kenny (1986): "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, 51(6), 1173–1182.

Benjamini, Y., and Y. Hochberg (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.

Bernhardt, A., E. Field, R. Pande, and N. Rigol (2019): "Household Matters: Revisiting the Returns to Capital among Female Microentrepreneurs," *American Economic Review: Insights*, 1(2), 141–60.

Blattman, C., N. Fiala, and S. Martinez (2014): "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda," *Quarterly Journal of Economics*, 129(2), 697–752.

——— (2020): "The Long-Term Impacts of Grants on Poverty: Nine-Year Evidence from Uganda's Youth Opportunities Program," *American Economic Review: Insights*, 2(3), 287–304.

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

Bossuroy, T., M. Goldstein, B. Karimou, D. Karlan, H. Kazianga, W. Parienté, P. Premand, C. C. Thomas, C. Udry, J. Vaillant, et al. (2022): "Tackling psychosocial and capital constraints to alleviate poverty," *Nature*, 605(7909), 291–297.

Brooks, W., K. Donovan, and T. R. Johnson (2018): "Mentors or Teachers? Microenterprise Training in Kenya," *American Economic Journal: Applied Economics*, 10(4), 196–221.

Cantril, H. (1965): *The Pattern of Human Concerns.* New Brunswick, NJ: Rutgers University Press.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018): "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India," Discussion paper, National Bureau of Economic Research.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81(6), 2205–2268.

——— (2020): "Quantile and distribution regression in Stata: algorithms, pointwise and functional inference.," Discussion paper, working paper.

De Mel, S., D. McKenzie, and C. Woodruff (2008): "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 123(4), 1329–1372.

——— (2009): "Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns," *American Economic Journal: Applied Economics*, 1(3), 1–32.

Dupas, P. (2011): "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya," *American Economic Journal: Applied Economics*, 3(1), 1–34.

Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2011): "Microenterprise Growth and the Flypaper Effect: Evidence from a Randomized Experiment in Ghana," *Journal of Development Economics*, 106, 211–226.

Filmer, D., and L. Fox (2014): "Youth Employment in Sub-Saharan Africa: Overview," Washington, DC: World Bank.

Fischhoff, B. (1975): "Hindsight Is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty.," *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288.

Foresi, S., and F. Peracchi (1995): "The Conditional Distribution of Excess returns: An Empirical Analysis," *Journal of the American Statistical Association*, 90(430), 451–466.

Hamory, J., M. Kremer, I. Mbiti, and E. Miguel (2016): "Start-Up Capital for Youth," AEA RCT Registry.

Haushofer, J., and J. Shapiro (2016): "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya," *Quarterly Journal of Economics*, 131(4), 1973–2042.

Hicks, R., and D. Tingley (2011): "Causal Mediation Analysis," *Stata Journal*, 11(4), 605–619.

Hussam, R., E. M. Kelley, G. Lane, and F. Zahra (2022): "The Psychosocial Value of Employment: Evidence from a Refugee Camp," *American Economic Review*, 112(11), 3694–3724.

Imai, K., L. Keele, and D. Tingley (2010): "A General Approach to Causal Mediation Analysis," *Psychological Methods*, 15, 309–334.

Imai, K., L. Keele, and T. Yamamoto (2010): "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25(1), 51–71.

International Rescue Committee (2016): "Cost Analysis Methodology at the IRC," available online at `https://rescue.box.com/s/co7xgj2vvohgzir3ejnr2e5mwbmqhvp7`, accessed 9 January 2017.

Jayachandran, S. (2021): "Microentrepreneurship in Developing Countries," *Handbook of Labor, Human Resources and Population Economics*, pp. 1–31.

Kraay, A., and D. McKenzie (2014): "Do Poverty Traps Exist? Assessing the Evidence," *Journal of Economic Perspectives*, 28(3), 127–48.

Madarász, K. (2012): "Information Projection: Model and Applications," *The Review of Economic Studies*, 79(3), 961–985.

McIntosh, C., and A. Zeitlin (2021): "Cash versus Kind: Benchmarking a Child Nutrition Program against Unconditional Cash Transfers in Rwanda," *arXiv preprint arXiv:2106.00213*.

McKenzie, D. (2012): "Beyond Baseline and Follow-Up: The Case for More T in Experiments," *Journal of Development Economics*, 99(2), 210–221.

——— (2016): "Can Business Owners Form Accurate Counterfactuals? Eliciting Treatment and Control Beliefs about Their Outcomes in the Alternative Treatment Status," World Bank Policy Research Working Paper 7768.

McKenzie, D., and C. Woodruff (2014): "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?," *World Bank Research Observer*, 29(1), 48–82.

Miller, D. T., and M. Ross (1975): "Self-Serving Biases in the Attribution of Causality: Fact or Fiction?," *Psychological Bulletin*, 82(2), 213.

Riley, E. (2022): "Resisting social pressure in the household using mobile money: Experimental evidence on microenterprise investment in Uganda," CSAE Working Paper 2022-04.

Smith, J., A. Whalley, and N. Wilcox (2012): "Are Participants Good Evaluators?," working paper.

Vanderweele, T. J. (2015): *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, New York, NY.

Wager, S., and S. Athey (2018): "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," *Journal of the American Statistical Association*, 113(523), 1228–1242.

Table 1: Impacts on Income-Generating Activities

	Year 1				Year 2				Year 6			
Outcome	Mean	Franchise	Grant	F=G	Mean	Franchise	Grant	F=G	Mean	Franchise	Grant	F=G
Self-employed	0.24	0.10	0.11		0.24	0.11	0.12		0.38	0.12	0.10	
		(0.04)	(0.04)			(0.04)	(0.04)			(0.03)	(0.04)	
		[0.006]	[0.009]	[0.692]		[0.002]	[0.006]	[0.817]		[0.000]	[0.011]	[0.585]
		{0.030}	{0.022}	{0.865}		{0.009}	{0.031}	{0.817}		{0.001}	{0.054}	{0.707}
Working for others	0.38	-0.07	-0.09		0.50	-0.02	-0.05		0.59	-0.04	0.01	
		(0.04)	(0.04)			(0.04)	(0.05)			(0.03)	(0.04)	
		[0.066]	[0.053]	[0.673]		[0.685]	[0.272]	[0.442]		[0.203]	[0.800]	[0.173]
		{0.110}	{0.067}	{0.865}		{0.685}	{0.453}	{0.817}		{0.508}	{0.911}	{0.431}
Currently working	0.59	0.02	0.02		0.66	0.09	0.06		0.80	0.01	0.03	
		(0.04)	(0.05)			(0.03)	(0.04)			(0.02)	(0.03)	
		[0.621]	[0.709]	[0.972]		[0.010]	[0.181]	[0.452]		[0.602]	[0.251]	[0.502]
		{0.735}	{0.709}	{0.972}		{0.025}	{0.453}	{0.817}		{0.753}	{0.418}	{0.707}
Hours worked	18.13	0.75	6.84		19.14	2.40	1.10		25.75	0.42	4.74	
		(2.21)	(3.09)			(2.16)	(2.67)			(1.85)	(2.41)	
		[0.735]	[0.027]	[0.045]		[0.267]	[0.679]	[0.639]		[0.821]	[0.050]	[0.072]
		{0.735}	{0.046}	{0.224}		{0.445}	{0.849}	{0.817}		{0.821}	{0.124}	{0.360}
Earned income	495.80	168.01	298.36		961.76	-58.53	-14.34		1582.80	96.92	23.16	
		(76.12)	(110.11)			(119.17)	(159.47)			(171.80)	(207.13)	
		[0.028]	[0.007]	[0.245]		[0.623]	[0.928]	[0.753]		[0.573]	[0.911]	[0.707]
		{0.069}	{0.022}	{0.612}		{0.685}	{0.928}	{0.817}		{0.753}	{0.911}	{0.707}

Coefficients in the Franchise and Grant columns are the intent-to-treat effect of the franchise and grant treatment on each outcome, estimated through an OLS regression controlling for randomization stratum fixed effects and the set of baseline controls selected by lasso (see Online Appendix Table A1). Standard errors in parentheses, unadjusted p-values in square brackets, and Benjamini-Hochberg q-values in curly brackets. We winsorize the top and bottom 0.5 percent of the data to limit the influence of a small number of outliers. Self-employed is an indicator for having done any self-employment activity in the past month. Working for others is an indicator for having done any paid work, either for a firm or for an individual, in the past month. Currently working is an indicator for being self-employed or working for others. Hours worked is the total number of labor hours worked over the past week. Earned income is wages, profits, and income (including in-kind income) from all labor activities over the past week. See Online Appendix C for additional information on the construction of outcome variables.

Table 2: Impacts on Wellbeing

	Year 2			Year 6		
	Franchise	Grant	F=G	Franchise	Grant	F=G
Wellbeing index	0.10	-0.02		0.19	-0.05	
	(0.07)	(0.08)		(0.07)	(0.08)	
	[0.170]	[0.794]	[0.145]	[0.010]	[0.518]	[0.002]

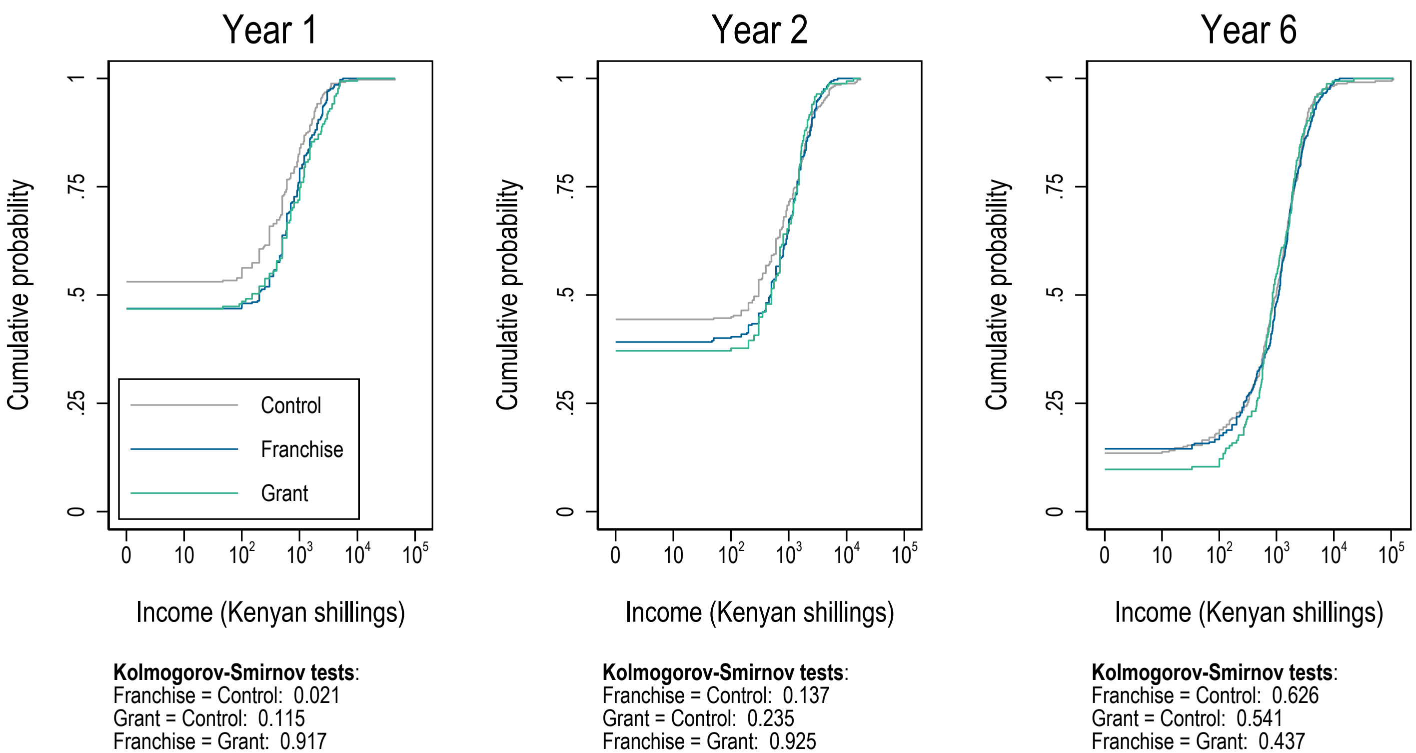
Coefficients in the Franchise and Grant columns are the intent-to-treat effect of the franchise and grant treatment on each outcome, estimated through an OLS regression controlling for randomization stratum fixed effects and the set of baseline controls selected by lasso (see Online Appendix Table A1). Standard errors in parentheses, and unadjusted p-values in square brackets. Wellbeing index is a normalized z-score which averages measures of living conditions, food security, and present and anticipated future life satisfaction. See Online Appendix C for additional information on the construction of outcome variables.

Table 3: Causal Mediation Analysis of Impacts of Franchise Treatment

	Estimated Effects		Share of Total Effect Mediated	
Outcome	ACME	Total	Estimate	Confidence Interval
Panel A. Does Take-Up of the Program Mediate the Effects of the Franchise Treatment?				
Self-employment in Year 1	0.12	0.11	1.14	[0.70,3.10]
Labor income in Year 1	59.9	170.7	0.34	[0.17,1.64]
Self-employment in Year 2	0.10	0.12	0.85	[0.55,1.97]
Self-employment in Year 6	0.13	0.14	0.93	[0.65,1.70]
Wellbeing in Year 6	0.14	0.16	0.89	[0.44,4.94]
Panel B. Does Self-Employment in Year 1 Mediate the Effects of the Franchise Treatment?				
Labor income in Year 1	82.9	170.7	0.47	[0.25,2.51]
Self-employment in Year 2	0.03	0.12	0.28	[0.18,0.64]
Self-employment in Year 6	0.02	0.14	0.17	[0.12,0.33]
Wellbeing in Year 6	0.01	0.16	0.06	[0.03,0.34]
Panel C. Does Self-Employment in Year 6 Mediate the Effects of the Franchise Treatment?				
Wellbeing in Year 6	0.03	0.16	0.16	[0.08,0.89]

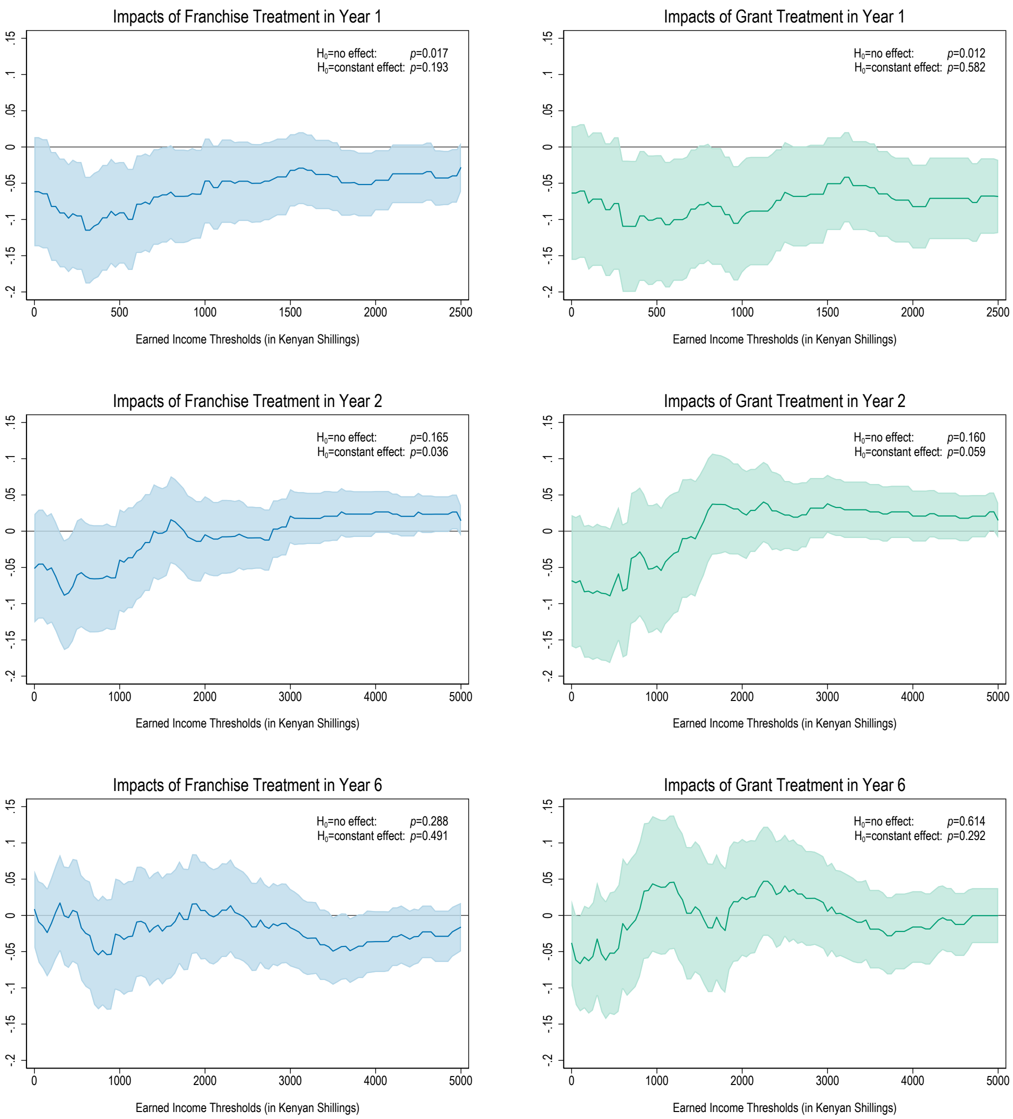
The ACME (average causal mediation effect) is defined in Imai, Keele, and Yamamoto (2010). Both ACME and total effects are estimated following (Hicks and Tingley 2011). We winsorize the top and bottom 0.5 percent of the data to limit the influence of a small number of outliers. Wellbeing index is a normalized z-score which averages measures of living conditions, food security, and present and anticipated future life satisfaction. See Online Appendix C for additional information on the construction of outcome variables.

Figure 1: CDFs of Earned Income by Treatment



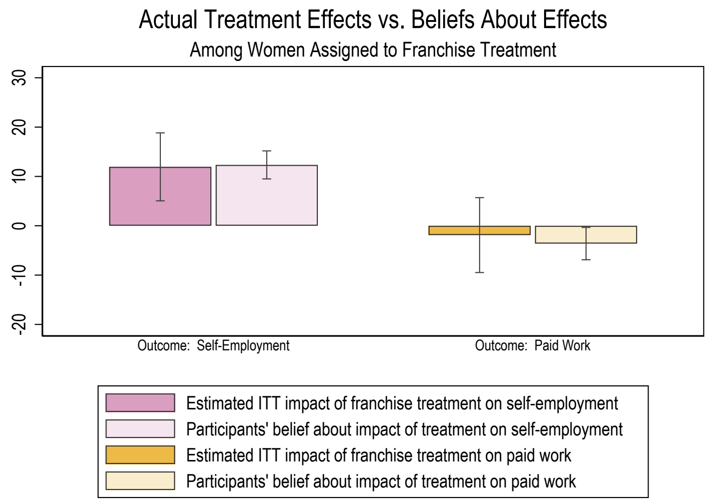
Earned income is wages, profits, and income (including in-kind income) from all labor activities over the past week.

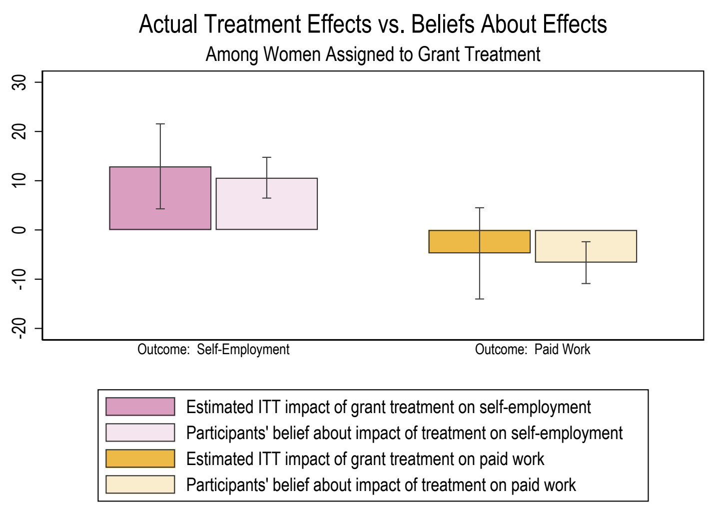
Figure 2: Distribution Regressions of the Impact of Treatment on Earned Income



Earned income is wages, profits, and income (including in-kind income) from all labor activities over the past week. Figures plot the estimated impacts of treatment (the franchise treatment in the left column, the grant treatment in the right column) on the probability of having an income below each of 100 grid points between the minimum and maximum values on the x-axis (shaded areas represent 95 percent confidence intervals). Cramer-von Mises p-values calculated following Chernozhukov, Fernández-Val, and Melly (2020), based on tests of the hypotheses that (1) there is no effect of treatment on the distribution of earned income and (2) that the effect of treatment on earned income is constant across the distribution.

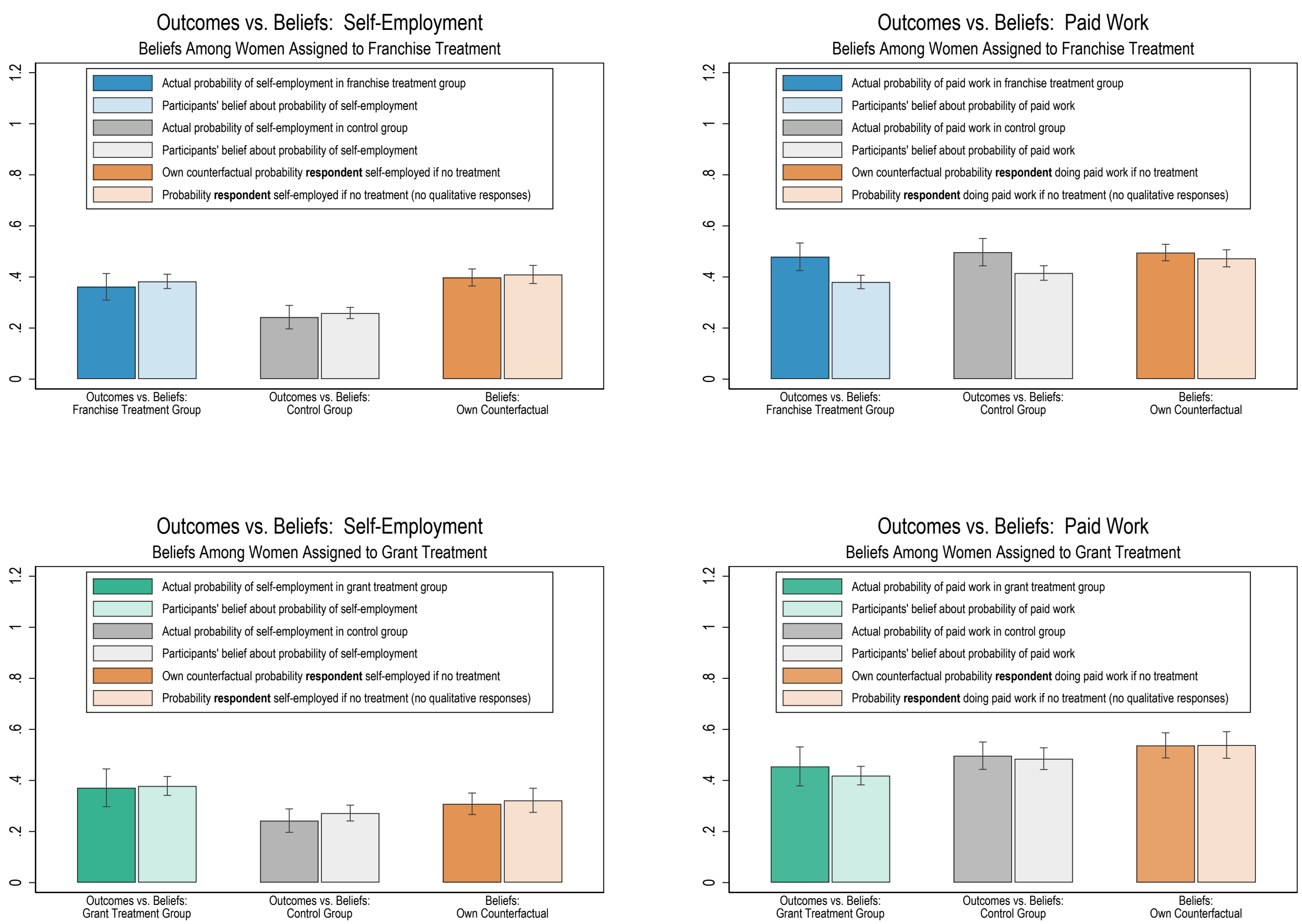
Figure 3: Participant Beliefs About Treatment Effects





ITT estimates of treatment are estimated via OLS, controlling for stratum fixed effects and the set of covariates selected by lasso. Beliefs are estimated using estimates of the frequency of outcomes in a reference class of young women similar to oneself. For example, the estimate of the impact of the franchise treatment on the probability of self-employment is constructed using average responses to two questions: (1) "I would like you to imagine 100 women from [your neighborhood] who applied to the [name of treatment arm] program and were admitted into it, just as you were. In other words, please think about 100 women similar to yourself. Out of 100 women, how many do you think are currently running or operating their own business?" and (2) "Now I would like you to imagine 100 women from [your neighborhood] who applied to the [name of treatment arm] program and but who were not admitted into it. In other words, please think about 100 women similar to yourself who were not selected to the [name of treatment arm] program. Out of 100 women, how many do you think are currently running or operating their own business?" The difference in responses to these two questions (divided by 100) is the individual-level estimate of the average treatment effect of the program on self-employment.

Figure 4: Participant Beliefs About Self-Employment, Paid Work, and Own Counterfactuals



The figure compares observed levels of self-employment and paid work in the treatment groups and the control group to beliefs about levels held by women assigned to the franchise and grant treatment arms. See Figure 3 for a description of the belief elicitation questions. The probability that a respondent would be doing paid work or in self-employment in the absence of treatment is the average response to a question about the counterfactual likelihood of involvement in the labor market.

Online Appendix: not for print publication

A Additional Tables and Figures

Table A1: Baseline Covariates

		By Treatment Arm			
Variable	All	Control	Franchise	Grant	Selected by Lasso?
Age	18.78	18.76	18.80	18.78	No
	(0.03)	(0.04)	(0.04)	(0.06)	
Kikuyu	0.37	0.39	0.36	0.35	No
	(0.02)	(0.03)	(0.03)	(0.04)	
Luo	0.26	0.27	0.25	0.27	No
	(0.01)	(0.02)	(0.02)	(0.03)	
Born outside Nairobi	0.48	0.45	0.52	0.47	Yes
	(0.02)	(0.03)	(0.03)	(0.04)	
Years of education	10.26	10.40	10.30	9.89	Yes
	(0.07)	(0.11)	(0.11)	(0.18)	
Took primary school leaving exam	0.92	0.93	0.91	0.90	No
	(0.01)	(0.01)	(0.02)	(0.02)	
Took secondary school leaving exam	0.41	0.44	0.41	0.34	Yes
	(0.02)	(0.03)	(0.03)	(0.04)	
Numeracy index	0.85	0.86	0.84	0.84	Yes
	(0.01)	(0.01)	(0.01)	(0.02)	
Any vocational training	0.34	0.37	0.32	0.35	No
	(0.02)	(0.03)	(0.02)	(0.04)	
Mother's education (if known)	4.78	4.62	4.84	4.99	No
	(0.16)	(0.26)	(0.27)	(0.36)	
Mother or father alive	0.88	0.89	0.88	0.88	No
	(0.01)	(0.02)	(0.02)	(0.02)	
Married or cohabitating	0.16	0.15	0.19	0.15	No
	(0.01)	(0.02)	(0.02)	(0.03)	
Has given birth	0.41	0.36	0.44	0.44	No
	(0.02)	(0.03)	(0.03)	(0.04)	
Household size	4.89	5.13	4.71	4.76	Yes
	(0.07)	(0.12)	(0.11)	(0.17)	
Lives with a parent	0.50	0.56	0.46	0.46	Yes
	(0.02)	(0.03)	(0.03)	(0.04)	
Lives with own child	0.36	0.32	0.41	0.35	Yes
	(0.02)	(0.02)	(0.03)	(0.04)	
Lives with other relatives (not parents)	0.31	0.28	0.32	0.38	Yes
	(0.02)	(0.02)	(0.02)	(0.04)	
Household has electricty	0.75	0.75	0.76	0.74	No
	(0.01)	(0.02)	(0.02)	(0.03)	
Household has piped water	0.49	0.49	0.49	0.48	No

Continued on next page

Table A1 – *continued from previous page*

Variable	All	By Treatment Arm: Control	Franchise	Grant	Selected by Lasso?
	(0.02)	(0.03)	(0.03)	(0.04)	
Household owns a television	0.57	0.57	0.57	0.55	No
	(0.02)	(0.03)	(0.03)	(0.04)	
Household owns a computer	0.03	0.02	0.03	0.03	No
	(0.01)	(0.01)	(0.01)	(0.01)	
Owns a personal mobile phone	0.73	0.74	0.73	0.73	No
	(0.01)	(0.02)	(0.02)	(0.03)	
Food security	-0.00	0.02	-0.03	0.03	No
	(0.03)	(0.06)	(0.05)	(0.07)	
Has any savings (including jewelry)	0.33	0.34	0.34	0.30	No
	(0.02)	(0.02)	(0.02)	(0.03)	
Value of savings (in USD)	5.47	5.32	5.79	5.14	No
	(0.63)	(1.02)	(1.04)	(1.25)	
Has a personal bank account	0.09	0.09	0.09	0.08	No
	(0.01)	(0.01)	(0.02)	(0.02)	
Any (paid) work experience	0.55	0.54	0.54	0.57	No
	(0.02)	(0.03)	(0.03)	(0.04)	
Currently working	0.15	0.12	0.17	0.15	No
	(0.01)	(0.02)	(0.02)	(0.03)	
Self-employed	0.05	0.04	0.06	0.05	No
	(0.01)	(0.01)	(0.01)	(0.02)	
Working for others	0.10	0.09	0.11	0.10	No
	(0.01)	(0.01)	(0.02)	(0.02)	
Over 20 hours housework last week	0.63	0.60	0.66	0.65	No
	(0.02)	(0.03)	(0.03)	(0.04)	
Observations	905	363	360	182	

Standard errors in parentheses. Selected by Lasso indicates baseline covariates chosen by lasso as predictors of either the grant treatment or the franchise treatment. Lasso is implemented using adaptive cross-validation to choose the penalty parameter.

Table A2: Compliance and Attrition

	Control Mean	Franchise	Grant	F=G
Attended business training	0.00	0.61	0.00	
		(0.03)	(0.01)	
		[0.000]	[0.620]	[0.000]
Launched microfranchise	0.01	0.44	0.00	
		(0.03)	(0.01)	
		[0.000]	[0.878]	[0.000]
Received grant	0.00	-0.00	0.95	
		(0.00)	(0.02)	
		[0.344]	[0.000]	[0.000]
Attritted from Year 1 Survey	0.06	0.01	0.01	
		(0.02)	(0.02)	
		[0.616]	[0.806]	[0.871]
Attritted from Year 2 Survey	0.07	0.01	0.01	
		(0.02)	(0.02)	
		[0.639]	[0.568]	[0.848]
Attritted from Year 6 Survey	0.08	0.02	0.02	
		(0.02)	(0.03)	
		[0.435]	[0.547]	[0.976]

Coefficients in the Franchise and Grant columns are the intent-to-treat effect of the franchise and grant treatment on each outcome, estimated through an OLS regression controlling for randomization stratum fixed effects. Standard errors in parentheses, and unadjusted p-values in square brackets.

Table A3: Impacts on Income-Generating Activities: Regressions Excluding Baseline Covariates

	Year 1				Year 2				Year 6			
	Mean	Franchise	Grant	F=G	Mean	Franchise	Grant	F=G	Mean	Franchise	Grant	F=G
Self-employed	0.24	0.10	0.12		0.24	0.12	0.13		0.38	0.14	0.11	
		(0.03)	(0.04)			(0.04)	(0.04)			(0.03)	(0.04)	
		[0.003]	[0.005]	[0.666]		[0.001]	[0.003]	[0.831]		[0.000]	[0.003]	[0.573]
		{0.014}	{0.015}	{0.862}		{0.004}	{0.017}	{0.881}		{0.000}	{0.017}	{0.594}
Working for others	0.38	-0.07	-0.08		0.50	-0.02	-0.05		0.59	-0.04	0.01	
		(0.04)	(0.04)			(0.04)	(0.05)			(0.03)	(0.04)	
		[0.060]	[0.060]	[0.743]		[0.625]	[0.314]	[0.545]		[0.153]	[0.682]	[0.103]
		{0.099}	{0.075}	{0.862}		{0.625}	{0.523}	{0.881}		{0.383}	{0.781}	{0.258}
Currently working	0.59	0.03	0.03		0.66	0.10	0.07		0.80	0.02	0.04	
		(0.04)	(0.05)			(0.03)	(0.04)			(0.02)	(0.03)	
		[0.497]	[0.465]	[0.862]		[0.006]	[0.111]	[0.528]		[0.436]	[0.130]	[0.415]
		{0.578}	{0.465}	{0.862}		{0.015}	{0.278}	{0.881}		{0.727}	{0.217}	{0.594}
Hours worked	18.13	1.22	7.33		19.14	2.83	1.50		25.75	0.33	4.43	
		(2.19)	(3.01)			(2.14)	(2.58)			(1.84)	(2.36)	
		[0.578]	[0.015]	[0.041]		[0.185]	[0.561]	[0.621]		[0.857]	[0.061]	[0.082]
		{0.578}	{0.025}	{0.205}		{0.309}	{0.701}	{0.881}		{0.857}	{0.152}	{0.258}
Earned income	495.80	166.29	295.02		961.76	-66.64	-46.32		1582.80	45.96	-55.64	
		(74.12)	(107.37)			(119.48)	(154.88)			(168.01)	(199.73)	
		[0.025]	[0.006]	[0.246]		[0.577]	[0.765]	[0.881]		[0.784]	[0.781]	[0.594]
		{0.063}	{0.015}	{0.615}		{0.625}	{0.765}	{0.881}		{0.857}	{0.781}	{0.594}

Coefficients in the Franchise and Grant columns are the intent-to-treat effect of the franchise and grant treatment on each outcome, estimated through an OLS regression controlling for randomization stratum fixed effects. Standard errors in parentheses, unadjusted p-values in square brackets, and Benjamini-Hochberg q-values in curly brackets. We winsorize the top and bottom 0.5 percent of the data to limit the influence of a small number of outliers. Self-employed is an indicator for having done any self-employment activity in the past month. Working for others is an indicator for having done any paid work, either for a firm or for an individual, in the past month. Currently working is an indicator for being self-employed or working for others. Hours worked is the total number of labor hours worked over the past week. Earned income is wages, profits, and income (including in-kind income) from all labor activities over the past week. See Online Appendix C for additional information on the construction of outcome variables.

Table A4: Impacts on Living Conditions and Wellbeing

	Year 2			Year 6		
	Franchise	Grant	F=G	Franchise	Grant	F=G
Wellbeing index	0.10	-0.02		0.19	-0.05	
	(0.07)	(0.08)		(0.07)	(0.08)	
	[0.170]	[0.794]	[0.145]	[0.010]	[0.518]	[0.002]
Living conditions	-0.04	-0.01		-0.06	-0.15	
	(0.07)	(0.09)		(0.07)	(0.09)	
	[0.553]	[0.912]	[0.712]	[0.397]	[0.074]	[0.235]
	{0.553}	{0.912}	{0.712}	{0.397}	{0.295}	{0.235}
Food security	0.06	-0.13		0.21	-0.04	
	(0.08)	(0.09)		(0.08)	(0.09)	
	[0.402]	[0.175]	[0.040]	[0.010]	[0.648]	[0.006]
	{0.537}	{0.670}	{0.161}	{0.020}	{0.672}	{0.016}
Current wellbeing	0.07	-0.01		0.18	-0.04	
	(0.08)	(0.09)		(0.07)	(0.09)	
	[0.389]	[0.897]	[0.392]	[0.015]	[0.672]	[0.008]
	{0.537}	{0.912}	{0.523}	{0.020}	{0.672}	{0.016}
Future wellbeing	0.18	0.09		0.25	0.04	
	(0.07)	(0.09)		(0.07)	(0.09)	
	[0.014]	[0.335]	[0.299]	[0.001]	[0.641]	[0.014]
	{0.056}	{0.670}	{0.523}	{0.004}	{0.672}	{0.018}

Coefficients in the Franchise and Grant columns are the intent-to-treat effect of the franchise and grant treatment on each outcome, estimated through an OLS regression controlling for randomization stratum fixed effects and the set of baseline controls selected by lasso (see Online Appendix Table A1). Standard errors in parentheses, unadjusted p-values in square brackets, and Benjamini-Hochberg q-values in curly brackets. Results in the top panel replicate those reported in Table 2 in the main text. Outcomes in the bottom panel are the four components of the Wellbeing Index (from the top panel). Living conditions is an index that takes the first principle component of indicators for having piped water, having (grid-based) electricity, having a television at home, and having a computer at home, and owning your own mobile phone. Food security is the Household Food Insecurity Access Scale, inverted so that higher numbers indicate greater food security. Current wellbeing is is a ranking of current life satisfaction on an ordinal scale (Cantril's Ladder of Life), and Future wellbeing is anticipated life satisfaction in five years. All outcomes are normalized z-scores.

Table A5: Impacts on Wellbeing: Regressions Excluding Baseline Covariates

	Year 2			Year 6		
	Franchise	Grant	F=G	Franchise	Grant	F=G
Wellbeing index	0.07	-0.07		0.15	-0.10	
	(0.07)	(0.09)		(0.07)	(0.08)	
	[0.370]	[0.401]	[0.105]	[0.040]	[0.211]	[0.001]
Living conditions	-0.06	-0.04		-0.09	-0.20	
	(0.07)	(0.09)		(0.07)	(0.09)	
	[0.440]	[0.660]	[0.850]	[0.180]	[0.022]	[0.194]
	{0.650}	{0.727}	{0.850}	{0.180}	{0.087}	{0.194}
Food security	0.03	-0.17		0.20	-0.07	
	(0.08)	(0.09)		(0.08)	(0.09)	
	[0.650]	[0.063]	[0.025]	[0.013]	[0.474]	[0.003]
	{0.650}	{0.253}	{0.101}	{0.025}	{0.632}	{0.012}
Current wellbeing	0.05	-0.03		0.15	-0.07	
	(0.08)	(0.09)		(0.07)	(0.09)	
	[0.506]	[0.727]	[0.363]	[0.041]	[0.384]	[0.006]
	{0.650}	{0.727}	{0.484}	{0.054}	{0.632}	{0.012}
Future wellbeing	0.15	0.05		0.20	0.00	
	(0.07)	(0.09)		(0.07)	(0.09)	
	[0.042]	[0.556]	[0.288]	[0.006]	[0.961]	[0.016]
	{0.168}	{0.727}	{0.484}	{0.025}	{0.961}	{0.022}

Coefficients in the Franchise and Grant columns are the intent-to-treat effect of the franchise and grant treatment on each outcome, estimated through an OLS regression controlling for randomization stratum fixed effects and the set of baseline controls selected by lasso (see Online Appendix Table A1). Standard errors in parentheses, unadjusted p-values in square brackets, and Benjamini-Hochberg q-values in curly brackets. Outcomes in the bottom panel are the four components of the Wellbeing Index from the top panel. Living conditions is an index that takes the first principle component of indicators for having piped water, having (grid-based) electricity, having a television at home, and having a computer at home, and owning your own mobile phone. Food security is the Household Food Insecurity Access Scale, inverted so that higher numbers indicate greater food security. Current wellbeing is is a ranking of current life satisfaction on an ordinal scale (Cantril's Ladder of Life), and Future wellbeing is anticipated life satisfaction in five years. All outcomes are normalized z-scores.

Table A6: Causal Mediation Analysis of Impacts of Grant Treatment

	Estimated Effects		% of Total Effect Mediated	
Outcome	ACME	Total	Estimate	Confidence Interval
Panel A. Does Self-Employment in Year 1 Mediate the Effects of the Grant Treatment?				
Labor income in Year 1	109.0	301.6	0.36	[0.21,1.14]
Self-employment in Year 2	0.03	0.12	0.25	[0.15,0.82]
Self-employment in Year 6	0.02	0.12	0.21	[0.13,0.64]

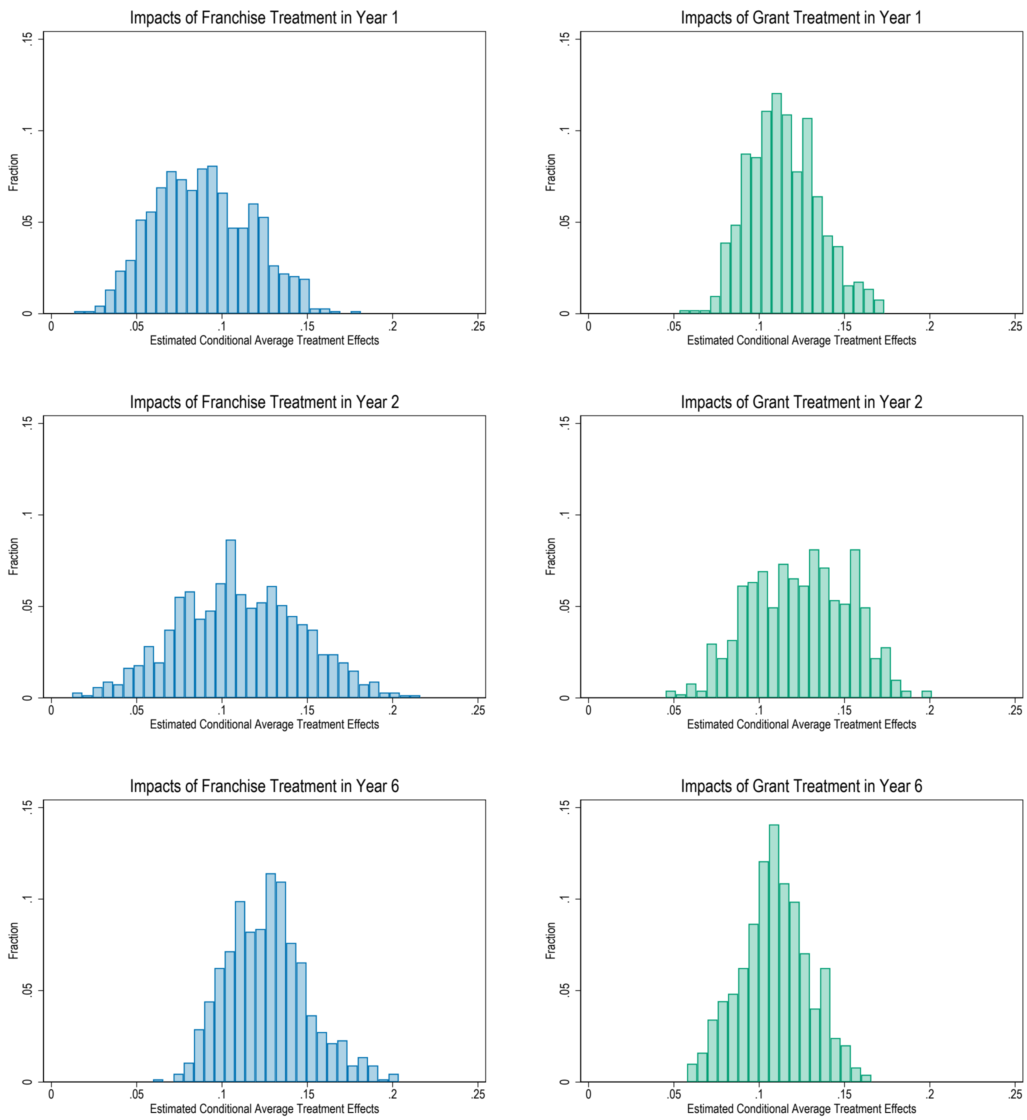
The ACME (average causal mediation effect) is defined in Imai, Keele, and Yamamoto (2010). Both ACME and total effects are estimated following (Hicks and Tingley 2011). We winsorize the top and bottom 0.5 percent of the data to limit the influence of a small number of outliers. See Online Appendix C for additional information on the construction of outcome variables.

Table A7: Impacts of Treatments on Reactions to Vignettes

	Control Mean	Franchise	Grant	F=G
Self-employment program makes woman better off after one month	0.79	0.07	-0.01	
		(0.03)	(0.04)	
		[0.037]	[0.821]	[0.046]
Self-employment program makes woman better off after five years	0.88	0.01	-0.03	
		(0.03)	(0.03)	
		[0.710]	[0.400]	[0.257]

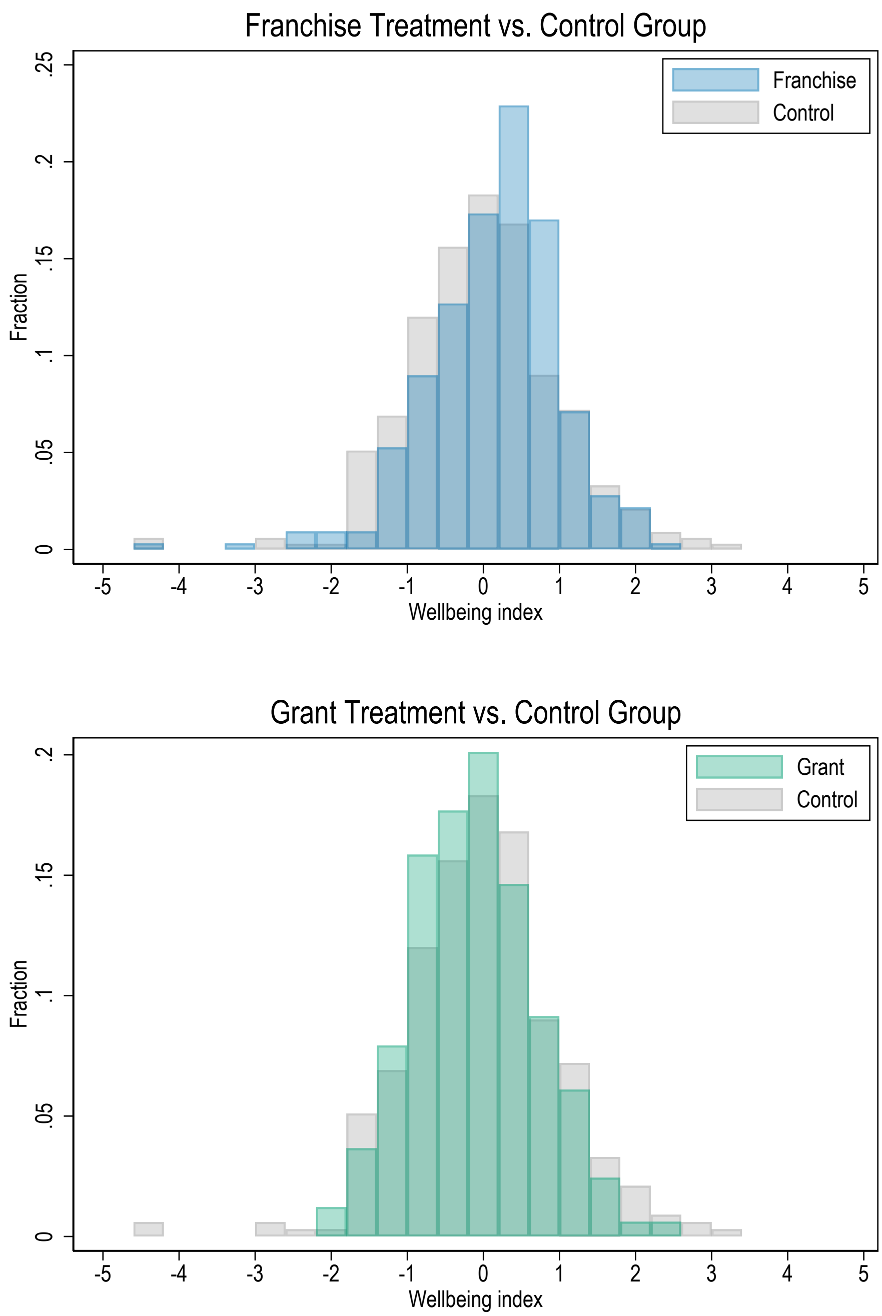
Standard errors in parentheses, and p-values in square brackets. Coefficients in the Franchise and Grant columns are the intent-to-treat effect of the franchise and grant treatment on each outcome, estimated through an OLS regression controlling for randomization stratum fixed effects and the set of baseline controls selected by lasso (see Online Appendix Table A1). Respondents in the Year 6 follow-up were read the vignette: *Please think about two women who are like you: from your neighborhood, the same age as you. One gets help from an NGO that allows her to start a small business like a small salon. The other woman gets help from a different NGO that helps her get a job. For both women there are challenges. It can be hard to keep your business open, and it can be hard to work for someone else.* They were then asked to indicate which woman would be better off overall after one month, and which women would be better off overall after five years.

Figure A1: Histograms of Estimated Conditional Average Treatment Effects



Conditional Average Treatment Effects on SELF-EMPLOYMENT estimated using the Generalized Random Forest package in R.

Figure A2: Histograms of Year 6 Wellbeing Index



Wellbeing index is a normalized z-score which averages measures of living conditions, food security, and present and anticipated future life satisfaction. Histograms represent the distribution of Wellbeing index in Year 6.

B Intervention Details

B.1 Recruitment

The microfranchising program we study was geared toward young women in Nairobi's poorest neighborhoods. Applications for the program came from women between the ages of 16 and 19; in practice, only 14.6 percent of applicants were below 18 years of age when they applied. Only those who had attained the age of legal majority were eligible to receive cash grants, so our analysis focuses on those in the two oldest age cohorts (randomization was stratified by age). The cash grant treatment was not announced in advance; women applied for a business training program and were randomized into one of three treatment arms.

B.2 Implementation of Treatment Arms

The program helped young women launch branded franchise businesses: either salons or mobile food carts. The intervention combined a number of distinct elements: business skills training, franchise-specific vocational training, start-up capital (in the form of the specific physical capital required to start the franchise), and ongoing business mentoring. Several of the intervention's components are common to many entrepreneurship promotion and job skills programs; what distinguishes microfranchise programs from other interventions is the focus on a small number of specific franchise business models that are tailored to the skills and constraints of program participants (i.e. poor young women in urban Nairobi) and to local market conditions. In this case, the implementing organization (the IRC) partnered with two Kenyan businesses looking to expand their presence in slum neighborhoods.

The first component of the franchise program was a two-week training course. In addition to a standard curriculum of business and life skills training topics, it included modules about the two specific franchise models. At the end of the course, participants indicated their preference between the two franchise partners and were then matched with one of them (almost always their first choice).

After the business skills course, program participants received training from the franchise business partner with whom they had been matched. Women assigned to the salon franchise received six weeks of classroom training followed by a two-week internship with a local salon. After the internship, participants received their business start-up kits (which included branded aprons, a hair washing sink, a hair dryer, and a variety of hair styling products). For women assigned to the food cart franchise, the franchise-specific training was a one-day session where franchisees were introduced to the brand, available products, and appropriate preparation methods. Following the franchise training, program participants received business start-up kits that included a mobile cart, an apron or t-shirt displaying the company logo, and an initial stock of smoked chicken sausages. Regardless of which franchise model they were matched with, the program did not require any borrowing by participants.

Each franchise business launched through the program was assigned a mentor who visited the business every few weeks. Mentors helped the young women in the program get their businesses off the ground — for example, by coordinating additional training with the franchise partners, helping the businesses set up bank accounts, or assisting with financial management and record keeping.

B.3 Comparing Implementation Costs

The two treatment arms of our study allow for natural cost comparisons, complementing our overall estimates of each program's impacts. Costs in the cash grant arm are relatively straightforward. The cash grant itself was worth 239 US dollars. Because compliance was slightly below 100 percent, the average disbursement per respondent in the cash grant arm was 228 dollars. Besides simply transferring the money, administrative tasks supporting this arm included having field team members meet participants twice (once to explain the no-strings-attached grant, once for the actual transfer);

confirming, via fingerprint reader, that the individuals our team met with were indeed the intended recipients; and data, accounting, and other indirect costs. These administrative tasks cost a total of roughly 82 dollars per intended recipient. Thus, the total cost of the cash grant arm, per intended recipient, was roughly 310 dollars.

Costs in the microfranchising intervention are more complicated. We begin with all costs that the IRC incurred implementing the program over three fiscal years. This study evaluates only the final calendar year of the program, but other participants were involved in the prior calendar year, and setup costs were required beforehand to make the program possible. Once we arrive at a total cost figure (the numerator), we divide by the total number of participants across all program years (the denominator). We face a number of decisions in both arriving at a total cost figure and in arriving at the number of participants, so we report upper and lower bounds on our cost estimates.[21]

One of the smallest cost items in the IRC budget is international staff support costs. We exclude this for simplicity. A larger cost is internationally hired staff in Kenya, including portions of the country director's time. Our upper bound includes these costs; our lower bound excludes them on the basis that they are needed most intensely for the startup phase of a project. The rest of the costs (national staff time, business support, trainings, office expenses, etc.) are concentrated in the two fiscal years in which the program trained most participants, but there are some costs from the first fiscal year in which the program began and in which the first participants started training. Our upper bound includes these costs; our lower bound includes only half of the first fiscal year's costs, on the basis that continued program operation or operation at larger scale would involve lower startup costs. The upper bound figure for the total cost of the program is roughly 763,000 dollars; the lower bound is 637,000 dollars. Either way, half of the costs come from providing trainings, including the (substantial) costs of providing refreshments for hundreds of participants each day.

These total cost estimates translate into a cost of between 616 dollars and 809 dollars per participant in the microfranchising arm. However, this figure is the cost associated with the treatment on the treated — not the cost for the intention to treat. This distinction matters because while 95 percent of those assigned to the grant treatment received a grant, only 61 percent of those assigned to the microfranchising treatment actually started the training. The intervention costs per individual *assigned* to the relevant treatment are thus roughly 286 dollars for the grant arm, and between 376 dollars and 494 dollars for the microfranchising arm.

[21] In order to determine cost per activity, each project expense was allocated, completely or partially, to either entrepreneurship activities, cash dispersements, or other non-treatment activities, and summed to determine total cost per activity. Total values were then divided by number of clients served to get an average cost per client. See International Rescue Committee (2016) for a detailed discussion of the costing methodology.

C Outcome Variables

- Working for others is an indicator for having done any paid work, either for a firm or for an individual, in the past month.
- Currently working is an indicator for being self-employed or working for others.
- Hours worked is the total number of labor hours worked over the past week.
- Earned income is wages, profits, and income (including in-kind income) from all labor activities over the past week.
- Wellbeing index is a normalized z-score which averages measures of living conditions, food security, and present and anticipated future life satisfaction.
- Living conditions is an index that takes the first principle component of indicators for having piped water, having (grid-based) electricity, having a television at home, and having a computer at home, and owning your own mobile phone.
- Food security is the Household Food Insecurity Access Scale, inverted so that higher numbers indicate greater food security.
- Current wellbeing is is a ranking of current life satisfaction on an ordinal scale (Cantril's Ladder of Life).
- Future wellbeing is anticipated life satisfaction in five years.