Changing aspirations through poverty measurement: The Poverty Stoplight Program

September 15, 2017

Katharina Hammler, PhD[1] / Martin Burt, PhD[2]

Abstract

Fundación Paraguaya's (FP) Poverty Stoplight (PS) is a multidimensional poverty measurement tool and mentoring approach aiming to empower people to lift themselves out of poverty by changing their aspirations. We use FP's administrative database to evaluate whether the PS program is effective in helping program participants overcome poverty. We argue that combining the PS, which is a dashboard metric, with the Alkire-Foster (AF) methodology provides advantages for different types of users. Using the AF methodology, we construct two multidimensional poverty indices based on the 50 indicators of the PS, one of which capturing moderate poverty, the other one extreme poverty. Based on the results of multilevel estimations, we find that participation in the PS program is indeed associated with a decreased likelihood of being moderately poor, and with a decreased number of deprivations. However, we also find that the overall improvement effect cannot be confirmed for clients suffering from extreme poverty. This suggests that the program is most successful in helping those who are closer to the poverty-cut-off, while more work—or possibly simply more time—is necessary to serve those who suffer from extreme poverty.

1 Background

Fundación Paraguaya (FP) is Paraguay's largest non-governmental developmental organization. FP works in the areas of microfinance, entrepreneurship, financial literacy, and self-sufficient vocational education, with the overarching goal of eliminating

[1] Director of Monitoring and Evaluation at Fundación Paraguaya, khammler@fundacionparaguaya.org.py
[2] Founder and CEO of Fundación Paraguaya, burt@fundacionparaguaya.org.py

multidimensional poverty, both within Paraguay, and in many other countries where FP has been active. To support its work, FP developed the Poverty Stoplight (PS, "the Stoplight") in 2011. The PS is both a multidimensional poverty measurement tool and a mentoring approach which claims to make the overwhelming reality of poverty digestible and actionable by allowing families to measure their own multidimensional poverty and to develop and implement a clear plan to overcome it.

The PS survey is designed to be an empowering process carried out as a collaboration between loan officer and client[3]: Through a tablet-based, visual survey which uses a series of graphics, clients self-assess their level of poverty in 50 indicators grouped into 6 dimensions (Income & Employment, Health & Environment, Housing & Infrastructure, Education & Culture, Organization & Participation and Interiority & Motivation). Each indicator has three pre-defined levels: Red (extreme deprivation[4]), Yellow (moderate deprivation) and Green (not deprived). The loan officer guides the client through the survey, presenting the three levels for each indicator and asking the client to choose which level best represents the situation of her family. The results are presented immediately after the survey is completed as a poverty dashboard that summarizes in the stoplight colors red, yellow, and green where the client is deprived. Based on these results, the client and her family then select their own priority areas for improvement, and FP helps them identify practical solutions to their problems in an integrated and empowering mentoring program. Together with the loan officer and mentor,

---

[3] The survey tool can be and is used in many different contexts and by many types of respondents. However, as this evaluation is concerned with the program targeting FP's women microfinance clients, the term "client" refers to this group.

[4] The PS refers to the three levels of each indicator as extreme *poverty*, moderate *poverty* and non-*poverty*. In the language of the AF methodology, these are referred to as extreme or moderate *deprivation*, respectively, as they present the individual indicators and not the overall welfare level. In the interest of enhanced clarity, this paper adopts the language of the AF methodology, speaking of *deprivations* for individual indicators and of *poverty* for the overall welfare assessment of a client.

the family first identifies the most likely source of the problem (for instance, whether it is due to internal or external factors, whether it is a matter of a lack of knowledge and skills, a lack of motivation, a lack of resources, and so on). Then, together they work on appropriate ways of addressing the problems that where identified, drawing on resources from all sectors (private companies, government support, NGOs, family and community…). For this mentoring process, the loan officer has at least one monthly face-to-face interaction and one weekly contact, which can be in person, but also via phone, with the client.

The central idea behind this approach is that the poor are not necessarily poor because they lack resources, but (also) because of aspiration failures (Appadurai 2004; Ray 2006): In this view, preferences and behaviors are determined by the social environment. According to Ray, individuals form aspirations windows containing the states they think they can attain, which is heavily influenced by the experiences and lives they can observe from people who seem similar enough, or relatable, to themselves. The difference between an individual's current state and their aspirations window is referred to as the aspirations gap. The theory predicts that individuals will only work to overcome their aspirations gap if the efforts required to close it appear small compared to the improvements they expect for their lives. This implies that individuals will only be driven to improving their situation if they believe that an improvement is achievable (extended aspirations window), and if the aspirations gap is neither too small (which would mean small benefits from improvement) nor too big (which would mean a lot of effort is required). This theory has been further developed by Dalton et al. (2016), who show formally that poor individuals can be stuck in behavioral poverty traps due to aspiration failures, and that under certain circumstances helping individuals to increase their aspirations can be sufficient for them to be able to overcome poverty. Similarly, psychologist Albert Bandura, and based on his work Grenny et al. (2013), argue that change only happens if individuals can answer two questions affirmatively: First, *is it worth it*? And second, *can I do it*? The PS mentoring program is designed to support the processes that facilitate such positive change, by a) expanding the aspirations window (showing "green" as an attainable goal, demonstrating that the goal is in fact within reach through positive deviants in the community or in the relatable environment of the individual, promoting empowerment); and b) decreasing the (perceived or actual) cost and/or increasing the perceived benefit of achieving a goal (showing the value of being "green", helping to identify resources and strategies to achieve goals). Through this process, the voices, actions, and aspirations of the poor themselves become the essential motors for transformation.

This approach does not imply that the PS shifts all responsibility for overcoming poverty onto the poor's shoulders. Rather, the approach's conceptual framework identifies several levels on which changes are necessary, such as on the level of the individual, the community, or the municipality or even state. For problems whose cause is out of the immediate influence of clients, loan officer and client search for ways of accessing the necessary resources, such as applying for certain benefits or petitioning the government.

There is some evidence that the PS program helps families overcome poverty. However, most of this evidence is anecdotal, and the available quantitative studies are based on administrative data from clients who were purposefully selected to participate in the program (Budzyna and Magnoni 2013; Burt 2014). This study will be the first one to evaluate the PS program using rigorous econometric techniques and a dataset of clients that were randomly selected for participation.

The paper contributes to the literature in several ways. First, it provides empirical evidence that interventions focusing on aspiration failures can play an important role in global poverty elimination efforts. Second, it demonstrates the synergies between two poverty measurement methodologies, the Poverty Stoplight and the Alkire-Foster (AF) method, for the purpose of program evaluation. Expanding on the AF method, the paper includes an added analysis of poverty severity, based on the Poverty Stoplight's "extreme deprivation" (red) and "moderate deprivation" (yellow) levels, illustrating the use of a second "dual cut-off": in additional to the dual cut-off that is characteristic to the AF methodology (one deprivation cut-off for each indicator, and one overall poverty cut-off), we distinguish for each indicator between a moderate and an extreme deprivation cut-off, thus creating extreme poverty and moderate poverty metrics. Hence, in addition to the concepts of poverty incidence (the number of poor) and poverty intensity (number of deprivations of the poor), we consider a third concept, namely, poverty severity (moderate or extreme level of the deprivations)[5]. This approach enables further insights into which client groups benefit most from this program.

2 Methodology

2.1 Combining the Poverty Stoplight with the The Alkire/Foster (AF) Methodology

Fundación Paraguaya's work is focused on eliminating multidimensional poverty. The poverty measurement tool that FP developed for this purpose, the Poverty Stoplight (PS), is a multidimensional dashboard metric: It gathers information on 50 indicators and displays the results in an intuitive format, using stoplight colors to quickly signal deprivations. This format

[5] The AF class of poverty metrics also includes metrics that are sensitive to poverty severity: The Adjusted Poverty Gap Index, M_1, and the Adjusted Squared Poverty Gap Index, M_2, both consider the shortfall or gap in each indicator that individuals would have to overcome in order not to be deprived. However, these metrics have important and often prohibitive shortfalls, especially for practitioners wishing to use them for program evaluation purposes. Most notably, they have stringent data requirements in that data needs to be on a continuous scale, they are much less intuitive than the M_0 metric discussed below, and they imply trade-offs between indicators that may be hard to justify.

makes the results easily understandable and useful for users in the field, such as for the poor themselves or for field workers of NGOs. However, the PS has no built-in way of aggregating the information. Because the Alkire-Foster (AF) methodology provides that possibility (Alkire et al. 2015), it is a natural addition to the Stoplight. The AF class of poverty metrics follow an axiomatic tradition of poverty measurement, meaning that AF metrics are designed to fulfill a predefined number of desirable characteristics. This axiomatic tradition is combined with a practical and intuitive counting approach, which makes the AF measures so useful for program evaluation purposes in general and for a combination with the PS dashboard approach in particular. As has been pointed out by Ferreira and Lugo (2013), the choice between dashboard approaches and scalar indices of multidimensional poverty is sometimes presented as an either-or decision, but presents a false dichotomy. Both approaches have distinct advantages, and combining the Poverty Stoplight with the AF method is one way to reap the benefits of both.

For this study, we only use a subset of AF measures, namely the headcount ratio, H, and the adjusted headcount ratio, M_0 (or rather their constitutive elements, the poverty identification and the censored deprivation counts, see below). An extensive description of these measures can be found, for instance, in Alkire et al. (2015); a short summary follows. AF metrics are based on a dual cut-off approach: First, for each indicator j (out of $j = 1, \dots, d$), a deprivation cut-off z_j is defined, and it is determined whether an individual i (out of $i = 1, \dots, n$) is deprived in indicator j by comparing x_{ij}, the achievement of individual i in indicator j, with the deprivation cut-off z_j. These deprivations are collected in the deprivation matrix g^0 such that $g_{ij}^0 = 1$ whenever $x_{ij} < z_j$ and $g_{ij}^0 = 0$ otherwise. Second, the number of weighted deprivations that an individual suffers is added up to the deprivation score c_i, defined as $c_i =$

$\sum_{i=1}^{d} w_j g_{ij}^0 = \sum_{j=1}^{d} \bar{g}_{ij}^0$, where w_j is the weight assigned to indicator j. Third, the identification function $\rho_k(x_{i.}; z)$ is used to identify individuals as poor if they suffer from at least k (weighted) deprivations: $\rho_k(x_{i.}; z) = 1$ if $c_i \geq k$ and $\rho_k(x_{i.}; z) = 0$ otherwise. The (unadjusted) headcount ratio H is then defined as $H = {}^{q}/_{n}$, where n is the total number of individuals, and q is the number of individuals identified as poor by $\rho_k(x_{i.}; z)$.

In order to obtain the adjusted headcount ratio M_0, one first has to go back to the deprivation matrix g^0 and censor all deprivations of individuals not identified as poor. This is done so as to satisfy the desired property that a poverty measure should change if and only if the achievement of a poor person changes; censoring the deprivations of the non-poor assures that improvements in their situation do not influence the poverty metric. Formally this censored deprivation matrix $g^0(k)$ is obtained by multiplying each element of the deprivation matrix g^0 with the identification function $\rho_k(x_{i.}; z)$: for all i and for all j, $g_{ij}^0(k) = g_{ij}^0 \times \rho_k(x_{i.}; z)$. This matrix now contains only the deprivations of those individuals who have been identified as being poor. A censored deprivation score $c_i(k)$ for each individual i can now be obtained as $c_i(k) = \sum_{i=1}^{d} w_j g_{ij}^0(k)$; it is the weighted sum of all censored deprivations that an individual suffers. Thus, $c_i(k) = c_i$ when $c_i \geq k$ and $c_i(k) = 0$ if $c_i < k$. These censored deprivation scores are collected in the censored deprivation vector $c(k)$.

From the censored deprivation matrix, one can now obtain the adjusted headcount ratio as the mean of the censored deprivation score vector: $M_0 = \mu\big(c(k)\big) = \frac{1}{n} \times \sum_{i=1}^{n} c_i(k)$. This is mathematically equivalent to multiplying the (unadjusted) headcount ratio H with the average intensity of poverty that is suffered by those identified as being poor, which is defined as $A = \frac{1}{q} \sum_{i=1}^{q} c_i(k)$. Note that intensity is the number of deprivations suffered, not the poverty severity (whether a deprivation is moderate or extreme).

Table 1 Structure of the Poverty Stoplight-AF Measure

Dimension/Indicator			**Weight within dimension**	**Weight** w_j
Dimension: Income & Employment				**Sum: 1/6**
(1) Income above the poverty line	(2) Stable Income	(3) Credit Facility	1/6 each	1/36 each
(4) Savings	(5) More than one source of income	(6) ID card		
Dimension: Health & Environment				**Sum: 1/6**
(7) Access to drinking water	(8) Nearby health post	(9) Nutrition (malnutrition and/or obesity)	1/9 each	1/54 each
(10) Personal Hygiene and Sexual Health	(11) Eye and Dental Health	(12) Vaccinations		
(13) Garbage Disposal	(14) Unpolluted Environment	(15) Insurance/ Community Help		
Dimension: Housing & Infrastructure				**Sum: 1/6**
(16) Safe home	(17) Sanitary latrine and cloaca	(18) Electricity	1/12 each	1/72 each
(19) Refrigerator and other household appliances	(20 Separate bedrooms	(21) Elevated cook stove and ventilated kitchen		
(22) Comfort of the home	(23) Regular means of transportation	(24) Roads accessible in all weather		
(25) Fixed line or cellular telephone	(26) Security	(27) Sufficient and appropriate clothing		
Dimension: Education & Culture				**Sum: 1/6**
(28) Literacy	(29) Children with schooling up to 12th grade	(30) Knowledge and skills to generate income	1/11 each	1/66 each
(31) Ability to Plan and Budget	(32) Communication and Social Capital	(33) School Supplies and Books		
(34) Access to information (radio and TV)	(35) Entertainment and Leisure	(36) Value cultural traditions and heritage		
(37) Respect for other Cultures	(38) Human rights for vulnerable/ defenseless people			
Dimension: Organization & Participation				**Sum: 1/6**
(39) Forms part of a self-help group	(40) Ability to influence the public sector	(41) Problem and conflict-solving ability	1/4 each	1/24 each
(42) Registered to vote and vote in elections				
Dimension: Interiority & Motivation				**Sum: 1/6**
(43) Awareness of needs: life map	(44) Self-esteem	(45) Moral Conscience	1/8 each	1/48 each
(46) Emotional affective capacity	(47) Aesthetic self-expression, beauty and art	(48) Violence against women		
(49) Entrepreneurial spirit	(50) Autonomy and Ability to make decisions			

Our AF measure follows the basic structure of the Stoplight: six equally-weighted dimensions, each with a varying number of equally-weighted indicators, adding up to a total of $d = 50$ indicators. This weighting structure implies that the hierarchy of grouping matters for the final weight of an indicator. For instance, as Table 1 above shows, in order to assure equal weight of all six dimensions, indicators assigned to the dimension "Organization and Participation" end up with a final weight of 1/24 each, while indicators in the dimension "Housing and Infrastructure" have each a final weight of 1/72.

In line with the concept of the PS which distinguishes between "extreme deprivation" (red) and "moderate deprivation" (yellow) as well as "no deprivation" (green), there are two measures that capture varying degrees of poverty severity: an "Extreme Poverty" measure that uses the level "red" as the deprivation cut-off ($z_j^1 = red$), and a "Moderate Poverty" measure that uses the level "yellow" as the deprivation cut-off ($z_j^2 = yellow$). The poverty cut-off k is 1/6 (i.e., one dimension): clients who are deprived in at least 1/6 of the weighted indicators are defined to be poor. The structure of this AF measure is described in some more detail in Table 1 above.

2.2 Data

The analysis is based on administrative data from FP. Starting in August of 2015, new program participants for the Stoplight program have been selected randomly each month from all active FP women microfinance clients that are in village banking groups that have not defaulted on their loans. Hence, all PS participants are also microfinance clients (but not all microfinance clients are also PS participants). This selection process implies that participants in the PS program are not necessarily representative of Paraguay´s population (they are all FP microfinance clients), nor of FP´s client base (they are in committees that have not defaulted).

While not ideal from an evaluation perspective, the decision to randomize based on the no-default criterion was taken because of program management requirements, as village banking groups that are defaulting are typically disbanding and clients drop out of the program, thus becoming unavailable for mentoring activities and for data collection. As we will argue in the evaluation methodology section below, this does not pose a threat to the internal validity of our results as the counterfactual (participants who started the program later) is defined by the same group of non-defaulting clients.

Our database consists of the PS results of over 8,900 of FP's women microfinance clients who did their Stoplight baseline survey between August 2015 and June 15, 2017[6]. New participants enter the program every month, and we refer to the program entry survey of each individual as "baseline" and to subsequent surveys as "follow up". Note that the database does not contain data on "true" non-participants, only on earlier and later entrants. Program participants do a follow-up survey after a year, or when their *asesora* (loan officer) thinks that the family has met the program goals, whichever comes first. Such follow-up data is currently available for around 2,400 women. In about 60% of these cases, more than 100 days elapsed between the rounds; in about 25% of cases, more than 200 days elapsed (see Figure 1). In only around 14% of the cases, eleven months or more elapsed between the survey rounds. Also note that despite program policies, in around 10% of cases 500 days or more passed between survey rounds.

[6] The actual number is close to 9,500, yet around 600 of these women clients were purposefully selected for participation by their loan officer instead of being randomly selected. These clients are excluded from this analysis.

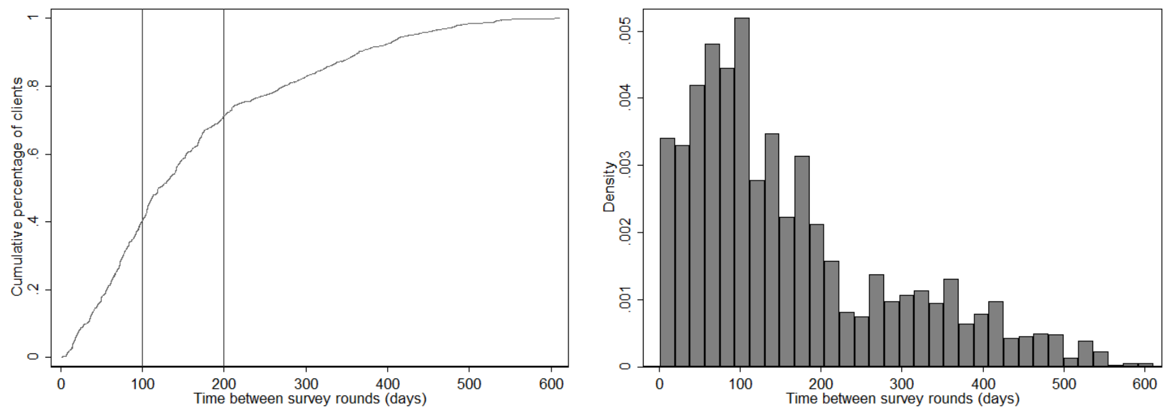

Figure 1 Time difference between baseline and follow-up survey. Left panel: Cumulative distribution; Right panel: histogram

The dataset contains ordinal data on the 50 poverty indicators (coded in the three levels green, yellow, and red indicating non-deprivation, deprivation and extreme deprivation, respectively) together with some background information such as zone of residence, date of survey, and loan officer. The dataset also contains data on program exposure: for each PS indicator, the number of contacts that a program officer had with the client with the goal of overcoming that specific deprivation was recorded, disaggregated by interaction type. The number of *monthly in-person* contacts was censured at 18 (1.5 per month for a year); the number of *weekly check-in* contacts was censored at 78 (1.5 per week for a year). While this data cannot provide any information on the quality of the mentoring activities, we will use it as a proxy to measure program exposure in the robustness section of this paper. Additionally, there is information on family income per capita, which will be used as a control variable.

2.3 Evaluation methodology

Estimation strategy

In order to evaluate whether the Poverty Stoplight program has been successful in decreasing multidimensional poverty, we use multiple regression to compare the poverty level of program participants before and after the intervention with the poverty level of comparable FP clients who are newly entering the program. As beneficiaries are randomly selected for program participation on an ongoing basis (see description in the data section), comparing early and later program entrants gives an approximation of where program participants may be without the program. Using later entrant's respective baseline surveys as a counterfactual effectively creates a control group: including the date (measured in days since the first survey in the sample) in the regression analysis allows us to estimate a secular trend that indicates how participants' poverty levels might have improved without the program.

Note that while clients in defaulting village banking groups are not eligible for program participation, excluding these clients does not affect the internal validity of our results as the counterfactual is also defined by the very same group of non-defaulting clients. However, the effect that the program may have on other types of poor individuals, such as non-microfinance clients or microfinance clients in default, cannot be estimated from this data.

The analysis is done with the help of a multiple regression model that uses program participation (mentoring) as the main explanatory variable, and controls for a general improvement trend over time, together with other control variables. We use multilevel estimation (MLE) for this analysis, modeling the individual observations of the poverty status of our clients as occasions nested within clients, which in turn are nested within loan offices. This approach has some important advantages for the present analysis (Snijders 1996; Snijders and Bosker 2011; Rabe-Hesketh and Skrondal 2012). First, because of the use of client-level random effects, differences in the timing of the observations do not pose a problem. This is important because such differences exist in our database due to the "rolling" character of the program: both the timing of the individual survey rounds and the spacing between the rounds differ between clients. Second, the MLE approach allows to use a varying number of observation rounds for each client. This means we can include all observations from clients with only a baseline or with two, three or four survey rounds. Third, MLE is an effective way of accounting for clustering effects: Our clients are distributed across loan office in the entire country, which may differ from each other in observable and unobservable ways. MLE allows us to test if such clustering effects exist, and to account for the clustering in the estimations to decrease the likelihood of type-I errors (false positives). Additionally, MLE enables us to analyze which effect the clustering might have on poverty and on program outcomes, for instance, whether the characteristics of loan offices are important for program success. This information is valuable from a program management perspective.

There are also some shortcomings to using a multilevel estimation approach. Most importantly, this method is particularly susceptible to omitted variable biases, as the assumption that all predictors are independent from *all* random terms is prone to be violated. A Hausman test is commonly used to test for the presence of such a bias and judge the appropriateness of MLE estimation. Unfortunately, this test is not designed to handle three levels, observations nested within clients nested within loan offices (Kim and Swoboda 2011). We will therefore present the results of alternative specifications (a fixed effects model and a MLE model with clustered standard errors) in the robustness section.

The model

Using MLE, we formulate a random intercept-model with three random slopes, two for time and one for mentoring. Intuitively, this means that we think that a) there are cluster effects at the client- and loan office level (hence, the random intercepts at each level); b) the time trend differs between clients and between loan offices (hence, the random slopes of time at both levels); and c) the effect of mentoring differs between loan offices (hence, the random slope of mentoring at the highest level). Formally, using a three-stage formulation (Raudenbush and Bryk 2002) our multilevel model of individual i's poverty status is constructed as follows. At the first level (observation-level), the poverty level of individual i from loan office s at time t can be described as that individual's average observed poverty level plus a time-specific error term e_{tis}, modified by any time-varying regressors (in our case, $time$, $mentoring$, and $income$). Formally, this level-1 model is defined as:

$$poverty_{tis} = \beta_{0is} + \beta_{1is} time_{tis} + \beta_{2is} mentoring_{tis} + \beta_{3is} income_{tis} + e_{tis}$$

where $poverty$ is the measure of poverty (see explanation below), $time$ is the number of days since the first observation in the entire sample (August 7th, 2015); $mentoring$ identifies the intervention/program exposure (see explanation below); and $income$ is the family per capita income level, in 10,000 current Paraguayan Guaraní (currently approximately USD1.8), centered at the sample mean[7].

Note how the intercept β_{0is} and the slopes β_{1is}, β_{2is}, and β_{3is} are client-specific coefficients. These client-specific coefficients can be explained by level-two models, which describe the level-1 coefficients as the result of level-2 fixed and random effects. In our model, we assume that the level-one intercept β_{0is} (i.e., individual i's average poverty level) can be explained by the average poverty level of all individuals of a given loan office s (γ_{00s}), the effect of the level-2 regressor $rural$ (which is a dummy), plus a client-specific error term (r_{0is}). The level-

[7] Family income per capita is also included in the outcome variable, because the ordinal variable "income above the [extreme/moderate] poverty line" enters with a weight of 1/36 into the index. However, we decided to also include family income per capita in its continuous form on the right-hand side because the concepts are distinct enough as to where the concern becomes relevant whether changes in income alone (rather than program participation) would lead to decreased multidimensional poverty.

1 slopes are modeled to be fixed at level-2, with the exception of the slope of time, which is assumed to vary randomly among clients. Formally, the level-2 models are defined as follows:

$$\beta_{0is} = \gamma_{00s} + \gamma_{01s} rural_{is} + r_{0is}$$

$$\beta_{1is} = \gamma_{10s} + r_{1is}$$

$$\beta_{2is} = \gamma_{20s}$$

$$\beta_{3is} = \gamma_{30s}$$

In these level-2 models, the intercepts ($\gamma_{.0s}$) as well as the slope γ_{01s} are office-specific, as indicated by the subscript s. These office-specific effects can be explained by level-3 models which are formally defined as follows:

$$\gamma_{00s} = \gamma_{000} + u_{0s}$$

$$\gamma_{01s} = \gamma_{010}$$

$$\gamma_{10s} = \gamma_{100} + u_{1s}$$

$$\gamma_{20s} = \gamma_{200} + u_{2s}$$

$$\gamma_{30s} = \gamma_{300}$$

Each level-3 model has one fixed effect; additionally, there are level-3 random effects for the level-1 intercept as well as for the level-1 slopes of time and mentoring. In other words, we assume that the average level of poverty as well as the time trend and the effect of the mentoring program differ across offices. Substituting all level-3 models into the level-2 models and then into the level-1 model and rearranging gives the reduced-form model:

$$\begin{aligned} poverty_{tis} = {} & \gamma_{000} + \gamma_{100} time_{tis} + \gamma_{200} mentoring_{tis} + \gamma_{300} income_{tis} + \gamma_{010} rural_{is} \\ & + u_{1s} time_{tis} + u_{2s} mentoring_{tis} + r_{1is} time_{tis} + u_{0s} + r_{0is} + e_{tis} \end{aligned}$$

The fixed parameters are denoted by the Greek letter γ, the level-3 random effects by the letter u, the level-2 random effects by the letter r, and the observation-level residual is e_{tis}. Our main interest lies in estimating the program effect γ_{200} and in determining if this effect varies between loan offices (as a statistically significant random coefficient u_{2s} would indicate). The other fixed and random effects are mainly included as control variables and to account for the clustering of the data.

Outcome variable

The outcome variable, $poverty$, is operationalized in two different ways. In the first specification, $poverty$ is a dichotomous variable indicating whether or not the individual i has been identified as multidimensionally poor by the identification function $\rho_{1/6}(x_{i.}; z)$ at the measurement occasion t, i.e., whether or not the individual i is deprived in at least 1/6 of all weighted indicators at the measurement occasion t. This multilevel logistic model allows to estimate changes in the probability of being poor that are correlated with participating in the Poverty Stoplight program. In the second specification, $poverty$ is the censored deprivation score $c_i(1/6)$, multiplied by 100 in order to improve readability of the results (i.e., the values can fall theoretically between 0 and 100). This censored deprivation score is the weighted sum of deprivations that the client suffers, censored to be 0 if the number of weighted deprivations falls below the poverty cut-off. The *censored* score is used so as to focus our attention on the multidimensionally poor.

In both specifications, we use two versions of the outcome variable: one for extreme poverty, which is based on the deprivation cut-off $z_j^1 = red$ in each indicator; and for moderate poverty, which is based on the deprivation cut-off $z_j^2 = yellow$ in each indicator.

Main explanatory variable

The effect of the program, $mentoring$, is the time difference (in days) between a client's baseline and her follow-up survey. Rather than being a simple dichotomous identifier of program participation, this time-of-exposure indicator can capture how long a client has participated in the program, thus providing a more nuanced perspective on the program effect. Note, however, that this indicator cannot account for potential differences in program implementation, such as the intensity of mentoring activities. The robustness section of this paper contains the results of identifying program exposure by the total number of mentoring contacts instead of by the number of days in the mentoring program.

3 Results

3.1 Empty model

In a first step, we estimate the empty model, i.e. the model containing only the fixed and random intercepts without any dependent variables, which results in estimations for the variance at the different levels. This is important information from a methodological standpoint (are all levels of the model indeed necessary?) as well as for the practical question of what accounts for most differences in poverty (variation across time, between individuals, or between regions/loan offices). Table 2 shows the results for the empty models with and without random intercepts for the linear model based on the censored deprivation count. Two important conclusions can be drawn from the results. First, there is, indeed, significant variance at all three levels: The lower part of the table contains the results of the likelihood ratio tests that were carried out to test the null hypotheses that the between-client variance, $\psi_{(2)}$, and the between-office variance, $\psi_{(3)}$, are zero, respectively—where $\psi_{(2)}$ denotes the variance of the level-two error term r_{0is} and $\psi_{(3)}$ denotes the variance of the level-three error term u_{0s}. The χ^2statistic for all cases is clearly statistically significant[8]. We therefore conclude that there is variance at all three levels, and that a three-level mixed model is appropriate.

Table 2 Results of null models. Outcome variable: censored deprivation count

	Moderate Poverty			Extreme Poverty		
	(E1) One-level model	(E2) Two-level model	(E3) Three-level model	(E4) One-level model	(E5) Two-level model	(E6) Three-level model
Fixed part						
Intercept	13.66***	14.09***	14.06***	1.884***	1.883***	1.901***
	(0.179)	(0.191)	(0.902)	(0.0864)	(0.0870)	(0.264)
Random Part, standard deviations						
Loan office level						
u_{0s} (Random intercept)			8.950			1.367
			(0.310)			(0.315)
Client level						
r_{ois} (Random intercept)		9.892	4.306		1.469	1.211
		(0.296)	(0.654)		(0.305)	(0.196)
Measurement Occasion level						
e_{tis} (Residual)	16.04	12.64	12.63	7.740	7.597	7.515
	(0.127)	(0.209)	(0.207)	(0.0611)	(0.0835)	(0.0819)
T (observ.)	8028	8028	8028	8028	8028	8028
I (clients)	6355	6355	6355	6355	6355	6355
S (offices)	24	24	24	24	24	24
ll	-33668.9	-33539.3	-33320.8	-27819.7	-27816.7	-27738.4
chi2#		259.22	436.91		5.97	156.51
df#		1	1		1	1
p#		<0.0001	<0.0001		0.0145	<0.0001
ICC (clients)		0.38	0.31		0.04	0.03
ICC (offices)			0.07			0.02

Standard errors in parenthesis. For the fixed part of the model: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
These statistics refer to the likelihood ratio test that the variance of the newly added random intercept is zero.

[8] The reported results are significant even before halving the p-value to account for the fact that the null hypothesis is on the boundary of the parameter space. As ψ cannot be negative, the asymptotic sampling distribution under the null hypothesis is a 50:50 mixture of $\chi^2(0)$ and $\chi^2(1)$, i.e. a distribution with a spike a 0. The p-value obtained from the LR-test is therefore conservative, and the correct p-value is obtained by dividing this value by 2 (Rabe-Hesketh and Skrondal 2012, 88–89).

Second, the table shows how the variation in the data is distributed between the three levels. Intra-class correlation coefficients (ICC) show the percentage of the overall variance that is due to the variance at a specified level. By far the largest share of the variance is due to differences between measurement occasions: in the case of moderate poverty, around 31% of the variance is found between clients, 7% is due to differences between loan offices, and the remainder (over 60%!) is due to differences across time. The case is even stronger for extreme poverty, where around 95% of overall variance is due to differences across measurement occasions, while the higher levels split the remaining variance. Thus, while the variance at the higher levels is statistically significant, this table shows that there are important changes in poverty across time. The goal of the following section is to explain this variance, testing whether participation in the PS program accounts for parts of it.

3.2 The effect of the Poverty Stoplight program

After having determined that most of the variance in the data stems from differences across measurement occasions (survey rounds) rather than from differences across clients or loan offices, the key question becomes whether this time variance can be explained by the Poverty Stoplight mentoring program, or whether it is entirely due to time trends that would also have been present without the PS program.

It is most useful to start with the question whether the PS program is associated with a decrease in the (unadjusted) headcount ratio, or, in other words, whether clients can decrease their probability of being poor by participating in the PS program. To answer this question we estimated a multilevel logit model[9], using the dichotomous poverty identification

[9] As the model would not converge otherwise, this model was estimated without r_{1is}, the client-level random coefficient of time.

as the outcome variable (i.e., if individual i is deprived in at least 1/6 of weighted indicators at measurement occasion t she is assigned a 1, otherwise a 0). The results are reported in the first two columns of Table 3 as the log of the odds ratios. The results show a statistically significant, negative time trend (i.e., the odds of being poor decrease over time, even in the absence of the PS intervention). For moderate poverty, the model suggests that participating in the PS program is associated with a decreased likelihood of being poor. We do not find a statistically significant effect of the PS program for extreme poverty. With regard to our control variables, our models show, as expected, that clients living in a rural area and clients with lower incomes have a higher probability of being multidimensionally poor, both for moderate and for extreme poverty.

Figure 2 helps to interpret the estimation coefficients. The figure depicts the predicted probabilities of being poor by plotting the fitted values from the multilevel logit model against the days of program participation. The left panel, which shows the estimated probabilities of being moderately poor, clearly shows that program participation is correlated with a lower probability of being poor, both for urban and for rural areas. The right panel, which depicts the likelihood of being in extreme poverty, shows no such clear pattern, which is in line with the lack of a statistically significant coefficient in the logit model.

Table 3 Estimation results

	(1)	(2)	(3)	(4)
Outcome variable:	Poverty identification		Censored deprivation count	
Method:	Multilevel logit		Multilevel linear	
Reported as:	Log of odds ratios		Estimation coefficients	
Poverty level:	Moderate	Extreme	Moderate	Extreme
Fixed Part				
Time (days)	-0.00530***	-0.00348***	-0.0211***	-0.00399***
	(0.00067)	(0.00060)	(0.00223)	(0.000932)
Days in program (PS mentoring)	-0.00429***	0.0000184	-0.00650*	0.00298
	(0.0011)	(0.0015)	(0.00262)	(0.00182)
Rural	0.437***	0.465**	0.908*	0.750**
	(0.13)	(0.15)	(0.451)	(0.232)
Income p.c.	-0.0287***	-0.0366***	-0.0735***	-0.0145***
	(0.0015)	(0.0029)	(0.00270)	(0.00144)
Intercept	0.718**	-4.300***	18.32***	2.384***
	(0.28)	(0.35)	(1.144)	(0.415)
Random part: standard deviations				
Loan office level				
u_{1s} (Random effect of time)	0.00252	0.00153	0.00914	0.00338
	(0.00054)	(0.00051)	(0.00180)	(0.000899)
u_{2s} (Random effect of PS mentoring)	0.00366	0.00399	0.00771	0.00666
	(0.0010)	(0.0016)	(0.00264)	(0.00164)
u_{0s} (Random intercept)	1.231	0.840	5.285	1.797
	(0.21)	(0.16)	(0.850)	(0.332)
Client level				
r_{1is} (Random effect of time)	.	.	2.28e-10	0.00612
			(6.89e-11)	(0.000474)
r_{0is} (Random intercept)	2.251	1.585	7.860	8.16e-08
	(0.14)	(0.24)	(0.293)	(2.23e-08)
Measurement Occasion level				
e_{tis} (Residual)	.	.	11.67	7.259
			(0.190)	(0.0666)
T (occasions)	8028	8028	8028	8028
I (clients)	6355	6355	6355	6355
S (offices)	24	24	24	24
ll	-4314.2	-1757.7	-32599.2	-27610.5
chi2	408.0	187.2	915.0	140.8
df	4	4	4	4
p	<0.0001	<0.0001	<0.0001	<0.0001

Standard errors in parentheses. For the fixed part of the model: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

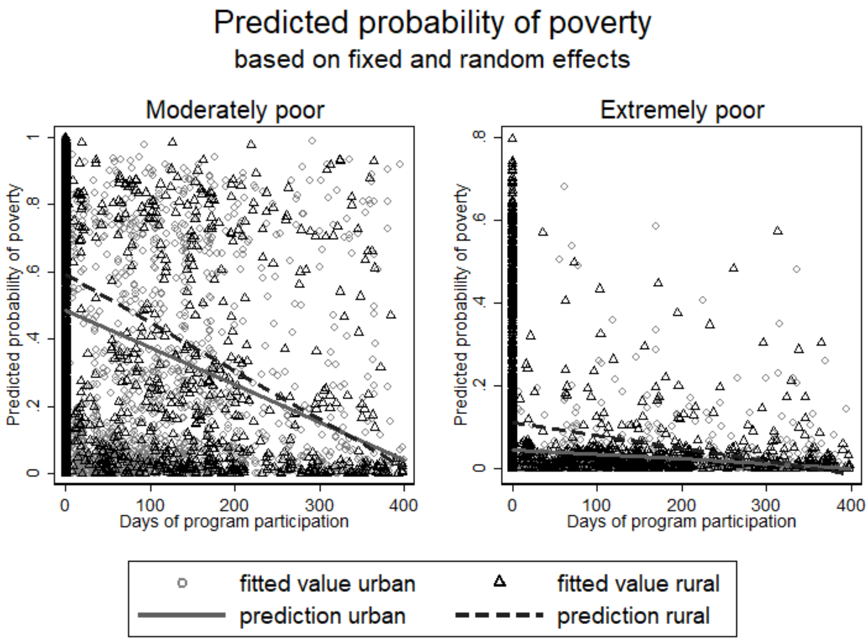

Figure 2 Predicted probability of poverty, based on models 1 and 2

The strength of the AF method is that it can go well beyond simply establishing whether the number of poor people is decreasing, which is a very limited way of measuring poverty. It is, for instance, possible that some of those closest to the poverty line escape poverty and that at the same time poverty intensity among the rest increases; the result would still be a decreased unadjusted poverty headcount ratio, even though overall suffering might increase. The M_0, or adjusted headcount ratio, adjusts the simple headcount ratio by the intensity of poverty that the poor people suffer and thus provides a better picture of poverty. In a second step, we therefore analyze whether the PS program has an effect on M_0 by using the censored deprivation score of each individual as the outcome variable.

The results for these M_0 models are presented in the last two rows of Table 3 and are most easily interpreted with reference to the intercept, which is the estimated adjusted headcount poverty ratio (M_0) at the beginning of the observation period, after zero days in the program, for clients in urban areas with a per capita family income equal to the total sample average. This reference adjusted headcount ratio is estimated at around 18.3% for moderate poverty, and at around 2.4% for extreme poverty. The results show a strong negative generall time trend: the adjusted poverty headcount ratio for moderate poverty decreases for the entire study population by over 0.02 percentage points per day (or almost 4% points over a half-year period), while the adjusted poverty headcount ratio for extreme poverty decreases by around 0.004 percentage points per day (or over 0.7 percentage points over a half-year period). These changes are highly statistically significant. The Poverty Stoplight mentoring program accelerates the path out of poverty for moderate poverty: Each day in the program decreases the censored deprivation count vector by over 0.006 percentage points; after half a year in the program, clients can thus expect to be deprived in one less weighted indicator (in addition to the general improvement trend). This effect is statistically significant, albeit considerably smaller than the estimated decrease in the likelihood of being poor that was established in model 1 (see the steep fitted line in Figure 2). The results for the program effect on extreme poverty are not statistically significant. The control variables show the expected signs: higher income is associated with lower deprivations counts, while living in a rural setting is associated with higher deprivation counts (all these effects are statistically significant in both models).

The random part of the model shows the variance of the results at the measurement occasion-, client-, and loan office-levels. Note that the reported standard errors are not a reliable way for judging the statistical significance of random parameters. Instead, we carried out a series likelihood ratio test for each model in order to test whether each of the random coefficients is in fact statistically significant and improves model fit. All of the parameters reported in Table 3 were found to be individually statistically significant (results not reported).

The results of the Likelihood ratio tests reported in the lower part of the table show that the random coefficients are jointly statistically significant.

These estimates are again best interpreted with reference to the estimated intercept. Using as an example model (3), we learn that the estimated adjusted poverty headcount ratio for this group is around 18.3, but that there is ample variance across measurement occasions: on average, within clients the observed censored deprivation count varies with a standard deviation of around 11.7, which is substantial. At the same time, the average observed censored deprivation count varies within loan-offices, between clients with a standard deviation of around 7.9, which still is considerable: there are stark differences between clients, even within the same loan office. Additionally, the average censored deprivation count varies between loan offices with a standard deviation of 5.3, again indicating substantial differences between them. With regard to random coefficients, time trends differ significantly both between clients and between loan offices, as indicated by the estimates for the random effect of time. Most importantly for our interests: the effect of the PS program differs across loan offices, as demonstrated by the estimated variance in the random coefficient u_{2j}: while the average effect of being in the program for 100 days is a decrease in the censored deprivation score of about 0.6 points, this effect varies between loan offices with a standard deviation of 0.7 points per 100 days, a large variance compared to the point estimate: in around two thirds of the loan offices, being in the PS program for 100 days is associated with a change on the censored deprivation score of anywhere between -1.4 points and +0.1 points. This variance across loan offices may indicate various things. For instance, it may be that some offices have found more effective mentoring strategies, or that the program is not uniformly implemented across offices, or that the program is more appropriate in specific contexts. In any case, a closer looks seems warranted. By calculating the fitted values for the model, we can obtain more precise predictions for the program effect in each individual loan office, and use this information for further research, taking a closer look at the most and least successful loan offices. The spaghetti plots presented in Figure 3 help with that. Each panel represents one loan office, and each line represents one client. Negative slopes indicate a fall in the predicted censored deprivation score, holding everything else constant. Overall, the figure illustrates that more days in the program are associated with a fall in the deprivation score. Additionally, one can see that this improvement is less pronounced in some offices, for instance in number 4, and more pronounced in others, for instance in office number 19.

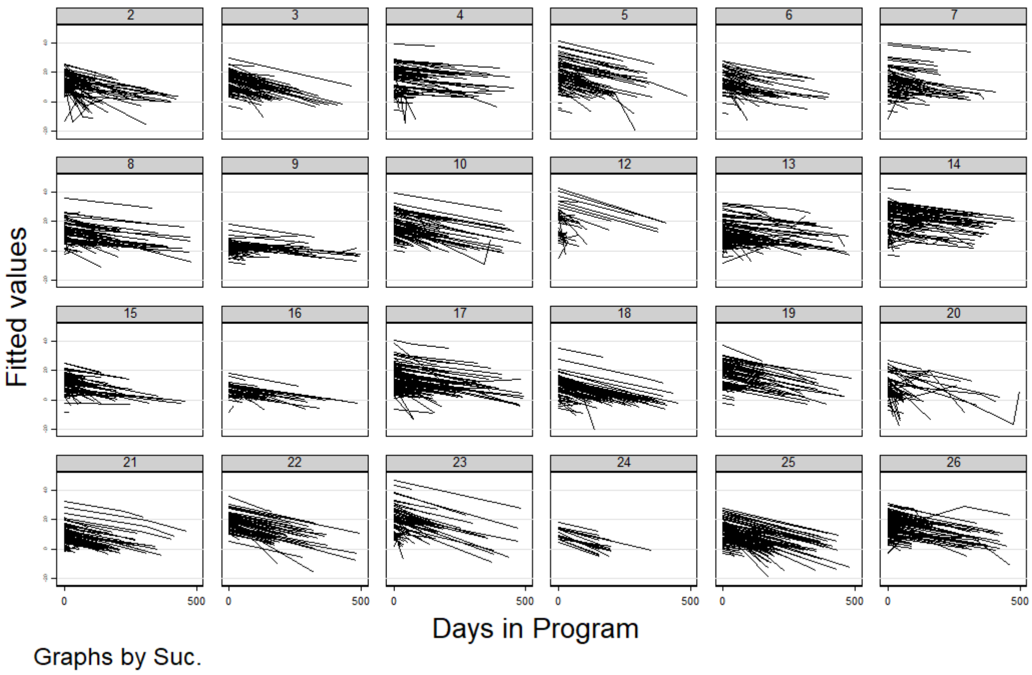

Figure 3 Spaghetti plots for moderate poverty. Each plot represents one loan office, each line one client.

4 Discussion

4.1 Main findings

We found evidence that participating in the Poverty Stoplight program is associated with better odds of overcoming poverty: our data suggests a strong general improvement trend in our study population, and participation in the PS program adds significantly to this positive trend. Note that the entire study population are micro finance clients, which might explain at least part of the large overall decrease; the results suggest that participating in the PS program in addition to the micro finance program increase participants' welfare even further. This overall result is highly encouraging. It supports the theory that changing people's aspirations and providing mentoring can help people overcome poverty.

There are, however, some caveats. Our analysis suggests that the PS program is only successful in reducing *moderate* poverty, while no statistically significant effects were found for *extreme* poverty, neither in the unadjusted nor in the adjusted headcount ratio (that is, at the individual level, neither when looking at the dichotomous poverty identification nor when looking at the censored deprivation counts). This suggests that the PS program is most successful in helping clients overcome poverty whose deprivations are less severe, i.e., who are "yellow" and not "red" in a given indicator. Furthermore, even focusing only on clients in moderate poverty, we find that the decrease in the adjusted headcount ratio is less steep than the decrease in the unadjusted headcount ratio (equivalently, at the level of the program participants, that the censored deprivation score decreases less than the likelihood of being poor). If the moderately poor clients closest to the poverty cut-off line become non-poor without equally strong improvements among those suffering from more deprivations, the poverty headcount ratio decreases, but at the same time the average poverty intensity of those remaining poor increases because of a decomposition effect. This is what is observed in the results, suggesting that the PS tool is most effective in supporting clients who are relatively close to the *poverty* cut-off line, i.e. those whose poverty is less intensive (in that they suffer from relatively fewer deprivations). The self-assessment and mentoring intervention might thus not be sufficient to help those clients lift themselves out of poverty who are suffering from more severe deprivations or more intensive poverty.

It is important to point out here that the average time difference between baseline and follow-up surveys was only around 5.5 months, and only in about 10% of the cases were baseline and follow-up data was available, a year or more passed between the two rounds. This time difference may well be too short to detect a noticeable difference in extreme poverty or in more intense moderate poverty. Additionally, it is worth pointing out that the PS program is part of FP's microfinance programs, which is not targeted at those in extreme poverty. Hence, the strategies used in the mentoring program might not be adequate to address the needs of those suffering from the most severe deprivations.

4.2 Robustness of Results

As briefly discussed in the methods section, multilevel estimation has an important shortcoming: because it includes many error terms, it is particularly prone to endogeneity or omitted variable biases. We therefore want to test the robustness of our results by carrying out a residual analysis and by applying different estimation methods (fixed effects (FE) estimation and MLE with robust standard errors). Furthermore, we also want to check how robust our results are to an alternative measurement of program participation (number of mentoring contacts instead of time-of-exposure). We will start with the latter.

Table 4 shows the results of our MLE models when using the number of mentoring contacts instead of the number of days in the program to identify program participation. The results replicate the findings from the main part of the paper: For moderate poverty, participation in the PS program is associated with a lower probability of being poor (model 5) and a lower censored deprivation score (model 7). For extreme poverty, just as in the main part of the paper, no statistically significant effect can be identified. Hence, we conclude that the positive effect of the PS program is robust to the way in which the intervention is measured.

The second threat to robustness is the use of MLE itself. If the assumption that all error terms are uncorrelated with all explanatory variables and normally distributed does not hold, the results may be biased. We hence carry out some residual analyses (following Kim, Anderson, and Keller 2013; and Rabe-Hesketh and Skrondal 2012). Figure 4 depicts box plots and histograms of all random terms for model 3 (dependent variable: censored deprivation count vector for moderate poverty). The box plot shows some statistical outliners. The histograms suggest that the assumption of normality is violated: Most strikingly, the client-level random intercept r_{0ij} shows a bimodal distribution, and the client-level random coefficient of time, r_{1ij}, has a spike around the mean[10]. This may point towards omitted variables which may in turn bias the estimation results. We therefore estimate our models again using sandwich estimators to obtain robust standard errors which do not rely on the model being correctly specified. The results (presented in Table 5) confirm our conclusions from the main part of the paper: Despite the larger standard errors, the decrease in the censored deprivation score for moderate poverty associated with program participation remains statistically significant. As in the main model, we find no statistically significant effect for extreme poverty.

[10] The outliers and spiked distribution are even more pronounced for extreme poverty (model 4).

Table 4 Robustness: Alternative Identification of program participation, using number of mentoring contacts

	(5)	(6)	(7)	(8)
Outcome var.:	Poverty identification		Censored deprivation count	
Method:	Multilevel logit		Multilevel linear	
Reported as:	Log of odds ratios		Estimation coefficients	
Poverty level:	Moderate	Extreme	Moderate	Extreme
Fixed Part				
Time (days)	-0.00357***	-0.00286***	-0.0212***	-0.00423***
	(0.00045)	(0.00047)	(0.00227)	(0.000939)
PS mentoring contacts	-0.00836***	0.000483	-0.0159*	0.0104
	(0.0016)	(0.0025)	(0.00664)	(0.00550)
Rural	0.286**	0.360**	0.853	0.747**
	(0.088)	(0.12)	(0.450)	(0.232)
Income p.c.	-0.0218***	-0.0330***	-0.0735***	-0.0147***
	(0.00098)	(0.0023)	(0.00271)	(0.00144)
Intercept	0.432*	-3.579***	18.36***	2.414***
	(0.19)	(0.22)	(1.140)	(0.416)
Random part: standard deviations				
Loan Office Level				
u_{1j} (Random effect of time)	0.00180	0.00131	0.00950	0.00349
	(0.00037)	(0.00043)	(0.00183)	(0.000864)
u_{2j} (Random effect of PS)	0.00502	0.00555	0.0227	0.0226
	(0.0016)	(0.0031)	(0.00624)	(0.00444)
u_{0j} (Random intercept)	0.833	0.666*	5.274	1.807
	(0.14)	(0.12)	(0.847)	(0.327)
Client level				
r_{1ij} (Random effect of time)			5.31e-12	0.00591
			(1.44e-12)	(0.000484)
r_{0ij} (Random intercept)	0.971	0.703	7.924	5.15e-09
	(0.089)	(0.17)	(0.291)	(1.48e-09)
Measurement Occasion level				
e_{tij} (Residual)			11.61	7.253
			(0.190)	(0.0664)
T (occasions)	8028	8028	8028	8028
I (clients)	6355	6355	6355	6355
S (offices)	24	24	24	24
ll	-4376.2	-1768.3	-32591.4	-27595.8
chi2	571.9	260.9	897.2	145.4
df	4	4	4	4
p	<0.0001	<0.0001	<0.0001	<0.0001

Standard errors in parentheses. For the fixed part of the model: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

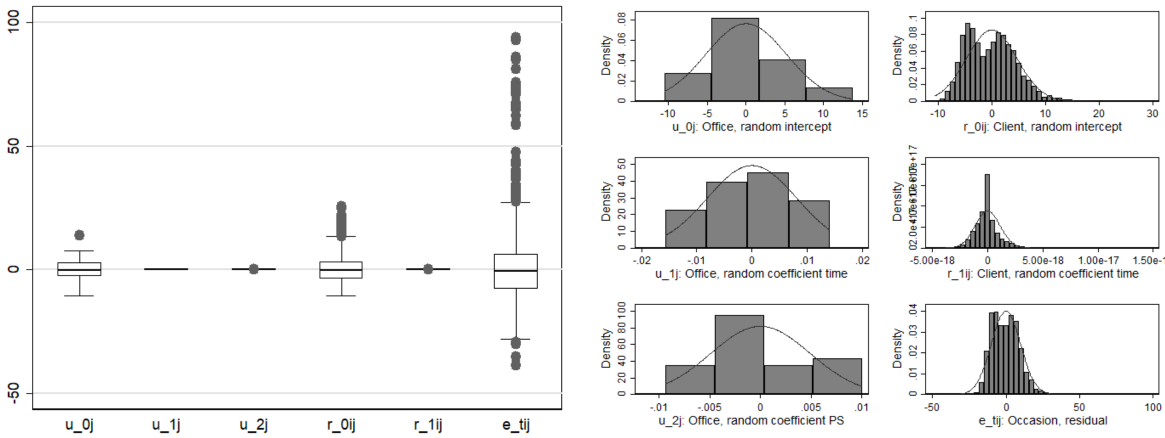

Figure 4 Residual analysis: Box plot and histogram of all random terms of model 3

Table 5 also presents the results of fixed effects models. Fixed effects models eliminate the variation between higher level units, thereby controlling for any cluster-specific effects. By only looking at the within-variation, one can eliminate any potential endogeneity that may arise from omitting unobserved variables that affect both the outcome variable and the explanatory variables at any level. This, however, has the disadvantage that one loses important information because only variables with within-cluster variation and thus only clusters with more than one observation can be used. In our case, we had to drop a) the variable rural as it doesn't vary at the client level, and b) participants for whom only one (baseline) observation is available. Additionally, the time-variable needed to be dropped because of perfect collinearity with days-in-program after the within-transformation. Despite this major reduction in observations, the conclusions from the multilevel model can be confirmed: Participation in the PS program is associated with a decreased censored deprivation score for moderate poverty, while no such effect can be found for extreme poverty.

Table 5 Robustness

	(9)	(10)	(11)	(12)
Outcome variable:	Censored deprivation count			
Method:	MLE with robust SE		Fixed effects	
Poverty level:	Moderate	Extreme	Moderate	Extreme
Fixed Part				
Time (days)	-0.0211***	-0.00399***		
	(0.00271)	(0.00110)		
Days in program (PS mentoring)	-0.00650*	0.00298	-0.0293***	-0.00158
	(0.00286)	(0.00174)	(0.00213)	(0.00171)
Rural	0.908	0.750**		
	(0.949)	(0.276)		
Income p.c.	-0.0735***	-0.0145***	-0.0310***	-0.00458
	(0.00581)	(0.00263)	(0.00615)	(0.00495)
Intercept	18.32***	2.384***		
	(1.315)	(0.490)		
Random part (standard deviations)				
Loan Office Level				
u_{1j} (Random effect of time)	0.00914	0.00338		
	(0.00178)	(0.00136)		
u_{2j} (Random effect of PS mentoring)	0.00771	0.00666		
	(0.00188)	(0.00246)		
u_{0j} (Random intercept)	5.285	1.797		
	(1.173)	(0.595)		
Client level				
r_{1ij} (Random effect of time)	2.28e-10	0.00612		
	(1.38e-08)	(0.00225)		
r_{0ij} (Random intercept)	7.860	8.16e-08		
	(0.930)	(0.00000414)		
Measurement Occasion level				
e_{tij} (Residual)	11.67	7.259		
	(0.890)	(0.493)		
T (occasions)	8028	8028	3259	3259
I (clients)	6355	6355	1583	1583
S (offices)	24	24	24	24
ll	-32599.2	-27610.5		
chi2	347.3	63.91		
df	4	4		
p	<0.0001	<0.0001		
R-squared			0.711	0.455
Within R-squ.			0.161	0.00171

Standard errors in parentheses. For the fixed part of the model: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Interestingly, the effect size is much larger in the FE model compared to the MLE model. In the former, 100 days of program participation are associated with a decrease in almost three weighted deprivations, while in the latter the effect was only estimated to be around 0.65 deprivations over the same period of time. However, given that we cannot control for a general time trend in this specification, the FE model cannot distinguish between an overall improvement trend and the improvement trend associated with the PS program, which likely inflates the coefficient on mentoring.

4.3 Limitations

The main shortcoming of this evaluation study is the lack of two rounds of observations for clients who did not participate in the program, which makes it harder to isolate the program effect. Despite the fact that all participants, including later entrants who provide the counterfactual for earlier ones, are randomly selected from the same pool of clients, we cannot preclude the possibility that earlier and later entrants differ systematically in unaccounted-for ways due to potential shifts in FP´s client base. However, the shortcoming is mitigated by the possibility of measuring program exposure in a nuanced way (showing that longer or more intensive exposure is associated with a stronger decrease in poverty).

An additional shortcoming of the present study is the lack of further control variables that may reasonably be assumed to influence poverty status, and in some cases program participation. For instance, at the client level the database does not allow us to control for hard-to-measure concepts such as motivation or effort; at the loan office level, one would like to control for factors such as the characteristics of the loan officers or the social and economic environment in the respective region. Note, however, that the fixed effects estimation in the robustness section controls for these client-level and office-level effects by only analyzing variation within clients. While being based on a much smaller sample and lacking a control group, the FE results suggest that longer program exposure is associated with decreased poverty.

5 Conclusions

This paper set out to evaluate the Poverty Stoplight program, estimating its effect on multidimensional poverty with the help of the Alkire-Foster poverty measurement method. Our results indicate that participation in the PS program is indeed associated with a decrease in multidimensional poverty, which suggests that the integrated mentoring approach can be a valuable tool to eliminate poverty. The PS's program theory assumes that poor people can overcome their own poverty if they can affirmatively answer the two questions "is it worth it?" and "can I do it?", which can be reframed in the language of the emerging literature on aspiration failures as problems related to the size of the aspirations window and the perceived costs and benefits of closing it. The results of this study support the notion that aspiration failures can be addressed with a targeted mentoring program, enabling people to overcome poverty. Much more research is needed, however, to be able to draw firm conclusions. First, the conclusions of this study should be replicated using a true experimental design in which the poverty levels of both the treatment and the control group are measured at program start and at the follow-up, which allows to identify the program effect more accurately. Second, in order to learn more about the mode of action of the Poverty Stoplight, and about whether it truly can be a tool to overcome aspiration failures, more targeted research is necessary that explicitly measures how and if the PS influences aspirations, and if this truly is the mechanism through with the PS decreases poverty. Third, this study focused on a specific population, i.e., on active women microfinance clients in Paraguay. It is unclear how the PS's effectiveness might differ when applied to another population—especially because microfinance clients might be more receptive to motivational interventions than other poor people. Finally, this research suggested that participation in the PS program is associated with a decrease in moderate poverty, but not in extreme poverty. Given that the average time between baseline and follow-up surveys was only half a year (and only in some instances exceeded one year), it seems likely that a longer-term study is necessary to study the effect of the PS program on extreme poverty. Such a longer-term study would also allow a closer look at the differences in the role that aspiration failures, as opposed to other challenges, play for people in extreme and in moderate poverty, respectively.

6 References

Alkire, Sabina, James Foster, Suman Seth, Maria Emma Santos, and Jose Manuel Roche. 2015. *Multidimensional Poverty Measurement and Analysis*. New York, NY: Oxford University Press.

Appadurai, Arjun. 2004. “The Capacity to Aspire: Culture and the Terms of Recognition.” In *Culture and Public Action*, edited by Vijayendra Rao and Michael Walton, 59–84. Stanford, CA: Stanford Univ. Press.

Budzyna, Laura, and Barbara Magnoni. 2013. “Measuring the Social Impact of Fundación Paraguaya.” EA Consultants.

Burt, Martin. 2014. “The Poverty Stoplight: Does Personalized Coaching in Microfinance Help Clients Overcome Poverty?” Unpubslished internal document. Asunción, Paraguay: Fundación Paraguaya.

Dalton, Patricio S., Sayantan Ghosal, and Anandi Mani. 2016. “Poverty and Aspirations Failure.” *The Economic Journal* 126 (590): 165–88. doi:10.1111/ecoj.12210.

Ferreira, Francisco H. G., and Maria Ana Lugo. 2013. “Multidimensional Poverty Analysis: Looking for a Middle Ground.” *World Bank Research Observer* 28 (2): 220–35.

Grenny, Joseph, Kerry Patterson, David Maxfield, Ron McMillan, and Al Switzler. 2013. *Influencer: The New Science of Leading Change, Second Edition*. 2 edition. New York: McGraw-Hill Education.

Kim, Jee-Seon, Carolyn J. Anderson, and Bryan Keller. 2013. “Multilevel Analysis of Assessment Data.” In *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*, edited by Leslie Rutkowski, Matthias von Davier, and David Rutkowski. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. CRC Press Taylor & Francis Group.

Kim, Jee-Seon, and Chris M. Swoboda. 2011. "Handling Omitted Variable Bias in Multilevel Models: Model Specification Tests and Robust Estimation." In *Handbook of Advanced Multilevel Analysis*, edited by Joop J. Hox and J. Kyle Roberts, 197–218. New York NY: Routledge.

Rabe-Hesketh, Sophia, and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata, Volume I: Continuous Responses, Third Edition*. 3 edition. College Station, Tex: Stata Press.

Raudenbush, Stephen W, and Anthony S Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks: Sage Publications.

Ray, Debraj. 2006. "Aspirations, Poverty, and Economic Change." In *Understanding Poverty*, edited by Abhijit Banerjee, Roland Bénabou, and Dilip Mookherjee, 409–21. New York, NY: Oxford University Press.

Snijders, Tom A. B. 1996. "Analysis of Longitudinal Data Using the Hierarchical Linear Model." *Quality and Quantity* 30 (4): 405–26. doi:10.1007/BF00170145.

Snijders, Tom A. B., and Roel J. Bosker. 2011. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE.