Institute for International Economic Policy Working Paper Series
Elliott School of International Affairs
The George Washington University

Assessing the Frontiers of Ultra-Poverty Reduction: Evidence from Targeting the Ultra-Poor (CFPR/TUP) program in Bangladesh

IIEP-WP-2009-06
Stephen C. Smith, George Washington University
M. Shahe Emran, IPD, Columbia University
Virginia Robano, George Washington University

Revised, June 2012

Institute for International Economic Policy
1957 E St. NW, Suite 502
Voice: (202) 994-5320
Fax: (202) 994-5477
Email: iiep@gwu.edu
Web: www.gwu.edu/~iiep

IIEP Working Paper

Assessing the Frontiers of Ultra-Poverty Reduction: Evidence from Targeting the Ultra-Poor (CFPR/TUP) program in Bangladesh

M. Shahe Emran, IPD, Columbia University
Virginia Robano, George Washington University
Stephen C. Smith George Washington University

Abstract
This paper uses household panel data to provide robust evidence on the effects of BRAC's Targeting the Ultra-poor Program in Bangladesh. We use alternative treatment-comparison pairs; in addition to BRAC's own classification, we exploit Type 1 errors in assignment in BRAC's selection process to create a second treatment-comparison pair. This allows us to estimate the program effects on the target group, not contaminated by mistargeting. To address selection on unobservables, we implement heteroskedasticity-based identification and two recent estimators based on matching and propensity score reweighting. The results show that participation had significant positive effects on per capita income, food security, clothing, shoes, livestock, and cash savings, but there is weak or no evidence of an impact on a number of household durables and assets, and indicators of health and women's empowerment. The effects on the poorest of the poor (the target group) may be different from the effects on an average participant in TUP program. Taking into account differences in initial conditions, the effects of the TUP program on the poorest of the poor are much larger (as measured by the program effect normalized by the initial mean value of an outcome), suggesting that the TUP program would have had a larger impact with better targeting.

Keywords: Ultra-Poor, CFPR/ TUP, BRAC, Microfinance, Bangladesh, Assignment Error, Difference-in-Difference, Matching, Heteroskedasticity Based Identification
JEL Clodes: O1, I3

Corresponding Author: Stephen C. Smith, Department of Economics, George Washington University, 2115 G ST NW, Washington DC 20052 . Email for correspondence: ssmith@gwu.edu (Stephen C Smith). We are grateful to two anonymous referees and an associate editor for helpful comments on an earlier version, and to Frank Vella and Art Lewbel for helpful discussions on the heteroskedasticity based identification. Special thanks to Munshi Sulaiman for help with data and discussions on the TUP program. We also thank Atonu Rabbani, Rajeev Dehejia, Esther Duflo, Ahmed Mushfiq Mobarak, Arif Al Mamun, Susan McCoskey, Forhad Shilpi, Chao Zhou, and participants at the Northeast Universities Development Conference at Boston University, and seminars at BRAC University/RED, International Food Policy Research Institute, George Washington University, and James Madison University for helpful comments and/or discussions. We would like to thank the Research and Evaluation Division (RED) of BRAC for providing us with the panel data used in this paper and for arranging three field visits of authors to the project. Support from U.S. Department of State (S-ECAAS-04-GR216(MA)), the GW Institute of Public Policy (GWIPP), and the GWU Center for International Business Research (CIBER) and IIEP is gratefully acknowledged. The standard disclaimers apply.

Assessing the Frontiers of Ultra-Poverty Reduction: Evidence from the Targeting the Ultra-Poor (CFPR/TUP) Program in Bangladesh

Abstract

This paper uses household panel data to provide robust evidence on the effects of BRAC's Targeting the Ultra-poor Program in Bangladesh. We use alternative treatment-comparison pairs; in addition to BRAC's own classification, we exploit Type 1 errors in assignment in BRAC's selection process to create a second treatment-comparison pair. This allows us to estimate program effects on the target group, uncontaminated by mistargeting. To address selection on unobservables, we implement heteroskedasticity-based identification (Klein and Vella 2009) and two recent matching estimators. The results show participation had significant positive effects on per capita income, food security, clothing, shoes, livestock, and cash savings, but there is weak or no evidence of an impact on a number of household durables and assets, and indicators of health and women's empowerment. The effects on the poorest (the target group) may be different from the effects on an average participant in TUP program. Taking into account differences in initial conditions, the effects of the TUP program on the poorest of the poor are much larger (as measured by the program effect normalized by the initial mean value of an outcome), suggesting that the TUP program would have had a larger impact with better targeting.

Keywords: Ultra-Poor, CFPR/ TUP, BRAC, Microfinance, Bangladesh, Assignment Error, Difference-in-Difference, Matching, Heteroskedasticity-Based Identification

JEL Codes: O1, I3

(1) Introduction

It is widely appreciated, both by practitioners and academics, that extreme poverty (or ultra-poverty) may be different from other forms of poverty and deprivation (see, for example, IFPRI 2007; Matin et al. 2008; World Bank 2006; Lipton 1983).[1] Ultra-poverty differs from conventional poverty in terms of depth (higher degree of deprivation), length (longer duration of time) and breadth (larger number of dimensions such as illiteracy, malnutrition etc.).[2] The possible complementarity among the different dimensions can potentially result in multiple mutually reinforcing poverty traps, thus making ultra-poverty an especially difficult problem to address.

The experience of the last few decades suggests that while the poverty programs of NGOs including microcredit programs have, in general, been successful in reaching the moderate poor (i.e., households below poverty line, but relatively close to it), the poorest of the poor are more often inadequately served or completely bypassed by such programs.[3] This appreciation led to the development and implementation of innovative anti-poverty programs that are designed especially for the ultra-poor. These programs are supposed to address simultaneously the multitude of interrelated factors that create extreme poverty and make it a trap difficult to escape from.

BRAC (acronym for Building Resources Across Countries), formerly known as the Bangladesh Rural Advancement Committee, is one of the first NGOs to design and implement a program specifically designed to address extreme poverty. In 2002, BRAC implemented an ultra-poverty program called "Challenging the Frontiers of Poverty Reduction: Targeting the Ultra-poor, Targeting Social Constraints" (henceforth TUP). The first phase of the TUP program was implemented over 2002-2006. It covered 100,000 ultra-poor households from 15 of the poorest districts of Bangladesh over a period of five years.[4] TUP is a multidimensional program that incorporates both livelihood protection and promotion components and features significant innovations

[1] Although there is a growing consensus that extreme or ultra-poverty is an important and difficult problem requiring novel intervention strategies, the concept of "ultra-poverty" remains unsettled. There are different definitions in the literature: Lipton (1983) defines ultra-poverty in terms of a calorie intake threshold (a person is ultra-poor if he/she gets 80 percent or less calorie of an appropriate poverty line calorie benchmark); a recent IFPRI report (2007) identifies an individual as ultra-poor if he/she lives on less than 54 cents per day. Emran, Shilpi, and Stiglitz (2009) define ultra-poverty in terms of endowments and access to markets; physical and human capital endowment of the ultra-poor are so low that it results in exclusion from both formal labor and credit markets. In this paper, we take the BRAC identification scheme as given for the empirical analysis. The BRAC definition refers to "not being able to meet even the barest of the basic needs". For recent analysis of issues related to identification and proper targeting of the ultra-poor, see Banerjee et al. (2008) and Sulaiman and Matin (2006).

[2] For discussions, see World Bank (2006), and Chronic Poverty Research Center (2008).

[3] The lack of participation by ultra-poor in standard microfinance may be a result of both self-selection and screening by the NGOs. The ultra-poor may find it difficult to participate because of rigid repayment schedules and high time costs involved, for example, in regular meetings. The NGOs, on the other hand, may try to screen out the poorest because of a lack of complementary inputs such as human capital (little education and ill health). It is easier for an NGO to show effectiveness of their program and thus attract donor funds by concentrating on the households marginally below the poverty line (i.e., moderate poor).

[4] A second phase of the TUP program covering 40 districts was initiated in 2007. 863000 households are expected to participate in the second phase over five years (2007-2011). This paper provides evidence of the effects of the

in targeting (through participatory wealth ranking by the villagers) and harnessing social capital (through village support networks and sponsorship of community leaders). It focuses on developing human capital (health, education, and training) and physical capital (asset transfers) for poor women. The program provides training in enterprise activities using the transferred asset, and also health services. A more complete discussion of the program is provided in Section 2. TUP as a strategy to tackle ultra-poverty has attracted much attention over last few years among NGO communities and academic researchers. Similar programs are being replicated in several other countries including India, China, Uganda, Tanzania, and Peru.[5]

This paper uses a two period panel data set (2002, 2005) to analyze the effects of the TUP program participation on a set of household outcomes including income, food security, clothing, health, child labor, schooling, and asset accumulation. The first phase of the TUP program was not a randomized intervention. Thus one has to carefully address the selection issues to identify and estimate the treatment effects of the program. Given the importance of the program, a carefully done analysis of the TUP program (phase I) with household panel data is of significant value.

To provide credible evidence on the treatment effect of participation in the TUP program, we carefully address the possible biases due to omitted heterogeneity. We utilize BRAC's selection criteria and the assignment errors to construct an alternative treatment-comparison pair that represents the target households more faithfully, and also identifies an appropriate comparison group. An advantage of the panel data is that we can credibly address the time-invariant household level heterogeneity (additive) by using household fixed effects. This is important for understanding the effects of TUP (or any other microfinance program), as two of the salient omitted variables in this context are innate entrepreneurial ability and risk preference, both of which are time invariant. The main results in this paper thus utilize household level fixed effects. To estimate the effects of TUP membership on household outcomes we take advantage of the recent heteroskedasticity-based identification approach due to Klein and Vella (2009). We also report estimates from two different matching estimators. The results from matching estimators are especially useful for the binary outcomes, because the Klein and Vella (2009) approach is not designed for the binary outcomes.

The matching estimators used in this paper are the minimum bias inverse probability weighted estimator (MB-IPW) due to Millimet and Tchernis (2012), and the difference-in-difference match-

first phase of the TUP program. For more details on the second phase of the TUP program, see BRAC Annual Report 2007.

[5]Other examples of programs for ultra poverty include the Grameen beggars program and the Bandhan "Chartering into Unventured Frontiers - Targeting the Hardcore Poor (CUF-THP)" program.

ing estimator due to Heckman et al. (1998) (henceforth called the DIDM estimator).[6] The treatment effects from the DIDM approach can be biased if the counterfactual trend in the treatment group is different from the trend found among the comparison households after matching on observables and household fixed effects are used to control for potential heterogeneity. The MB-IPW estimator developed by Millimet and Tchernis (2012) minimizes the bias that arises from the violation of the conditional independence assumption by appropriately trimming the sample around the bias-minimizing propensity score.

A standard approach to tackling selection on unobservables and measurement error is to develop an instrumental variables strategy that exploits some features of the program design or implementation to isolate exogenous variations in program participation. Unfortunately, it was not possible to develop a credible and strong enough instrumental variables strategy based on the features of the TUP program. In the absence of credible exclusion restrictions, we employ the heteroskedasticity-based identification approach that does not rely on the standard exclusion restrictions (Rigobon 2003; Klein and Vella 2009, 2010; and Lewbel, forthcoming).[7] In particular, we implement the heteroskedasticity-based instrumental variables approach due to Klein and Vella (2009).

An innovative feature of our study is that it uses alternative treatment and comparison groups using the type 1 assignment errors in the TUP selection process. BRAC's own treatment group is called the 'selected ultra-poor' (SUP) that includes all the actual participants in the program irrespective of whether they meet the eligibility criteria set out by BRAC itself. The corresponding comparison group is called 'not selected ultra-poor' ($NSUP$). However, there is significant mistargeting (assignment errors) in the TUP program when judged by the inclusion and exclusion criteria set out by BRAC for the TUP program participants. Such targeting errors usually result in a treatment group that is on average richer than the intended target group of a program.[8] We use the assignment errors in the selection of participants in the TUP program to partition the sample to generate alternative treatment and comparison groups.[9] The assignment errors can be

[6]The DIDM approach reduces the bias in estimated treatment effects compared to cross section studies with or without matching (Blundell and Dias 2009). The DIDM approach has been used by Petkova (2009) and Berlinski and Galiani (2004), among others.

[7]For recent applications of heteroskedasticity-based identification, see Rigobon (2003), Rigobon and Rodrik (2005), Farré et al. (2010), Schroeder (2010), Millimet and Tchernis (2012), Emran and Hou (forthcoming), Emran and Shilpi (2012), Mallick (2012), Berg et al. (2012), Chowdhury et al. (2012) and Emran and Sun (2011).

[8]Mistargeting is common in microcredit programs including most well-known programs such as Grameen. For discussions on mistargeting in Grameen microfinance programs see Matin (1998), and for evidence on BRAC programs see Montogomery et al. (1996), and Zaman (1998). For a recent discussion on issues related to targeting in microcredit programs see Banerjee et al. (2008). Although mistargeting has become a concern in microfinance programs, it is well-known that optimal targeting does not imply zero assignment errors (Ravallion and Lipton 1995; Ravallion 2008; Kanbur 2010).

[9]A descriptive analysis of the TUP program with some preliminary impact evaluation was done by BRAC's in-house research and evaluation division (RED) using the same panel data set (see Rabbani et al. 2006). They use

used to create two treatment-comparison pairs based on type 1 and type 2 errors in assignment. The treatment and comparison groups based on type 1 errors are called SB_1, or 'should be, one', and SB_0 or 'should be, zero'. The treatment group consists of the households who satisfy the BRAC inclusion and exclusion criteria and thus are correctly selected into the program, while the comparison group consists of the households who are incorrectly excluded from the program according to the stated criteria. As we show later, this comparison group is very similar to the treatment group in terms of initial characteristics and thus the possibility of selection bias is less compared to the $SUP - NSUP$ groups. Also, the treatment group SB_1 consists of the poorest of the households in our sample (i.e., the 'true' ultra-poor). This treatment-comparison pair thus allows us to estimate the treatment effect of program participation on the intended beneficiaries of the TUP program. Given that the focus of the study is on the effects of the TUP program on the poorest of the poor, we omit the results from the treatment-comparison pair based on type-2 errors, because the treatment group consists of the richest households in the sample.

The evidence from the matching estimators and Klein and Vella (2009) approach, both implemented with household fixed effects, shows that there is significant positive effect of participation in the TUP program on food security, cash savings, livestock (cows/bulls), housing (homestead land and tin roof house), and shoes for both the treatment groups: SUP and SB_1. The evidence is also consistent with a positive effect of TUP membership on per capita income of a participant household, especially for the SB_1 group.[10] There is, however, no evidence of a significant effect of the TUP program on subjective health outcomes, productive assets (such as fishing nets, big trees and rickshaw vans), child labor, and women's empowerment.[11]

We also find that there are some important differences in the program effects between the two treatment groups: SUP and SB_1. For example, while the evidence that participation increases net income is very strong for the SB_1 group, it is weaker, both in terms of magnitude and statistical significance, for the SUP group. The households in both treatment groups seem to invest in productive assets such as livestock, the effect is stronger for the SB_1 group. Only the relatively richer SUP households seem to acquire luxury goods such as radio/TV. In fact, the program effect on radio/TV is negative for the SB_1 households according to the Klein and Vella (2009) estimate, which is consistent with the idea that the poorest of the poor households cut back their other expenditure in order to make productive investments. When the differences in the initial conditions are taken into account, the program effects are significantly larger for the SB_1

the selected ultra poor (SUP) as the treatment group and the not selected ultra-poor ($NSUP$) as the comparison group.

[10]The definition of income is a broad one that takes into account the change in the value of household assets due to activities such as livestock fattening commonly undertaken by TUP participants. It is thus a comprehensive measure of the household welfare for the TUP participants.

[11]We, however, have limited indicators of women's empowerment. The conclusion that there is no program impact on women's empowerment thus should be interpreted with appropriate caveat.

treatment group compared to the BRAC treatment group SUP. Had the program concentrated on the poorest of the poor, the program effects would have been larger than those found.

The rest of the paper is structured as follows. Section 2 provides a brief discussion of the BRAC TUP program. Section 3 describes our data set. Section 4 discusses the empirical strategy for identification and estimation of the treatment effects in full detail. Section 5 reports the results of our empirical analysis on the treatment effect of program participation in a sequential manner starting from a simple difference-in-difference approach. The paper concludes with a brief summary of the findings.

(2) The BRAC Ultra-poverty Program

One of the most comprehensive approaches to redressing ultra-poverty has been developed and implemented by BRAC in Bangladesh. BRAC is the world's largest NGO when measured in terms of membership, scope, and budget. Founded in 1972, it started microfinance in 1974, which by 2010 includes approximately eight million women members. The BRAC Education Program serves over 1 million (approximately 10%) Bangladeshi primary students in some 35,000 informal schools. Over 110 million individuals receive BRAC health and other services in Bangladesh.

TUP (phase I) was launched in 2002 in three of the poorest districts in Northwest Bangladesh (Rangpur, Kurigram, and Nilphamari) identified on the basis of poverty mapping and selected from a larger group of potential participants, who together form the basis for our panel data set. All members of treatment and comparison groups were selected by villagers as among the poorest local families. A subset was selected by BRAC according to exclusion and inclusion criteria. The exclusion criteria required that participating women must be capable of doing work outside the home, must not belong to another NGO program and must not receive a food benefits card. In the inclusion criteria, participating women have to meet three of the following: child labor is present; ownership of less than 10 decimals of land (a tenth of an acre), lack of a male earner at home, adult women selling labor outside of the household, and lack of any productive assets (Noor et al. 2004, p. ix, BRAC Annual Report 2007, p. 24).

To identify the ultra-poor women, several strategies were used. One is "Participatory Wealth Ranking" that utilizes local information available to the villagers. A meeting is held in which a village map is drawn on the ground with each household labeled. The villagers agree on a wealth ranking among the households, to identify those who are the poorest of the poor. Those who can afford tin plate walls or roofs are less poor than those with straw walls or thatched roofs. Those who are known to have a steady, formal job are categorized as among the well off. To keep the process manageable, only about 150 households were included in each wealth ranking exercise.

There are incentives for people to try to rank themselves as poor to receive assistance; but the multiple checks done on family status means their ability to get away with this is limited. To

supplement community meetings, BRAC staff members walk through the village, looking for any hut that gives the appearance of extreme poverty. They then try to bring potentially overlooked ultra-poor people to the attention of the community meetings. Village leaders, generally people who are relatively well educated such as the school teachers, were actively involved in all stages of the process. Although the BRAC selection mechanism was imperfect, it is important to appreciate that the resulting mistargeting may actually have made BRAC's own treatment and comparison groups ($SUP - NSUP$) more comparable than it otherwise would be.[12]

The TUP program aims to improve the physical, human, and social capital of the poorest of the poor. A core activity of the program is to provide participants with a grant of specific physical assets. The TUP program then provides assistance for using the transferred assets effectively as a microenterprise. In particular, BRAC staff members offer ongoing training in specific enterprise activities notably livestock and poultry rearing, operation of tree nurseries, and village vending such as circulating around the village with a pushcart. Each training program is targeted to the specific asset transferred; periodic refresher training is offered. After enterprises are established, microfinance and related services are eventually provided through the equivalent of BRAC's primary Village Organizations.

The TUP program works to develop human capital through the microenterprise training, as well as general education including functional literacy, and improved health. BRAC provides the program participants (SUP) with health services. BRAC staff including BRAC's village health volunteers known as Shastho Shebikas provide training, basic care, and referrals. Financial assistance for illness is also provided. Direct services include child health, immunization, diarrheal disease control, vitamin A supplements for children under 5, TB control, and family planning services and pregnancy care. Yet another activity is to install tube wells and sanitary latrines which are expected to provide health benefits.

The program also seeks to build social capital through village support networks and sponsorship of community leaders for extremely poor women. The village support committees engage elites, often individuals who are known for public-spirited or religiously motivated charitable works. The committees are expected to assist the TUP participants when they are subjected to various types of shocks, such as by helping them to recover lost assets.

[12]This point can be seen clearly if we compare the subgroups that would result from perfect targeting in the sample of households in our data set. The differences between the treatment and comparison groups are most pronounced in this case when compared to both the treatment-comparison pairs used for the analysis in this paper. Please see Table 2a below.

(3) The Data and Variables Description

For the empirical analysis, we use the BRAC TUP panel data set. This is a two-year panel of about 5000 households. The baseline survey of 5626 households was done in 2002. In 2005, 5288 households were resurveyed, along with 278 newly formed households that had split from the initial set of households. Attrition was moderate and was due to migration, death, and marriage. The final matched panel contained 5067 households and was provided to us by Research and Evaluations Department (RED) of BRAC.[13]

The BRAC TUP panel data set provides information on a wide range of household characteristics and outcomes. The survey contains a rich body of information regarding the asset base of the household that includes physical (land, rickshaw van, fishing nets), human (schooling, child labor, health), and financial (cash savings) capital. The data include information about basic needs (food security, clothing, and shoes/sandals), stock of household durables and income. The data set also includes some information that can be used as indicators of women's empowerment, but they are limited in scope and likely to provide only an imperfect and a partial view of women's status in a household.

The income variable we use includes both the *level* of income in last year and *change* in the value of assets over last year. It is very important to include a measure of the change in the value of assets in this context, because many households fatten livestock and almost all their returns are realized only when they actually sell the livestock. The income variable is thus a comprehensive measure of household welfare, much more informative than the standard income variable that fails to account for changes due to activities such as livestock fattening. We use 'per capita' income as the relevant income measure of household welfare.

The cash savings is a binary variable that takes on the value of 1 if a household has cash savings in a given year.[14] Food security is measured by two indicators; the first is 'food availability' which ranks food availability in a household among four possible outcomes: 'always deficit' [1], 'deficit some times' [2], 'neither deficit nor surplus' [3], or 'food surplus' [4]. The second indicator is a binary measure called 'two meals a day' that takes on the value of one when the household members can have at least two meals a day, and zero otherwise. The clothing variables refer to the main type of male and female clothing in Bangladesh: saree (female clothing), and lungi (male clothing). The 'shoe/sandal' is a binary variable with a value of one when all the household members own shoes or sandals and zero otherwise.

[13] Because only the balanced panel was coded, we cannot analyze the issue of panel attrition.

[14] It is important to note that most of the savings is not voluntary, it is part of the program. One might thus wonder if it useful to look at cash savings as an outcome. Our view is that it depends on what one is interested in. If the interest is to understand the changes in economic lives of the poor due to participation in TUP program, the cash saving is clearly an outcome of the program participation. But if one is interested in behavioral responses in savings due to income effect of the program, the "forced" savings is clearly not informative.

The physical assets include livestock (cows/bulls, ducks, hens, etc.) and other productive assets (such as a fishing nets, rickshaw vans, and "big trees"). The asset measures do not include any assets transferred from the TUP program.[15] Household durable goods include tube wells as well as chairs, beds, radios, TVs and quilts. Some of the durable goods such as beds, chairs and quilts may be considered as basic needs of a household, and others such as TV and radio are 'luxury' goods, given the low level of income and assets of the households. There are two health indicators on subjective health conditions reported by the respondents. The 'health status' variable asks the respondent to rank his/her perceived current health status given five options: Excellent [5], Very good [4], Good [3], Fair [2], Poor/Bad [1]. The second health indicator is 'health improvement' that ranks one's health compared to last year among five possible cases: Much better than one year ago [5]; somewhat better now [4]; about the same [3], somewhat worse [2]; much worse [1]. Note that the subjective health indicators may have systematic measurement errors because the TUP program raises health awareness of the participants. Thus a negative response might reflect the fact that an individual is better aware of the preexisting conditions, rather than a worsening health status.

As indicators of women's welfare and empowerment we use the ratio of saree (female clothing) to lungi (male clothing), child labor among girls, and schooling of girls.[16] In Bangladeshi society, saree is prized by women, and for the poor households, the ratio of saree to lungi is a reasonable indicator of relative expenditure on *feminine goods* in the household.[17] There is a large literature that uses relative expenditure on feminine goods as an indicator of female empowerment (see, for example, Deaton 1989). These indicators of women's welfare and empowerment, although useful, are admittedly limited, and thus the conclusions about women's empowerment in this paper should be treated with appropriate caution.

Our analysis covers both flow and stock variables, and one might plausibly argue that three years may not be enough to capture long term effects of the program, and thus the evidence on the stock variables should be interpreted with appropriate caveats. It is possible that our analysis underestimates of the long-run effects of program participation on the stock variables.

Table 1 presents the summary statistics of the relevant variables used in this paper. One can see some interesting changes from 2002 to 2005 for the sample of households in the panel. There are significant improvements for an average household from 2002 to 2005 in terms of most

[15] As pointed out by an anonymous referee, the assets transferred by the TUP program are part of the endowment of a household in 2005 and thus one should include them as part of the asset measures. However, information on the transferred assets in 2005 were not included in the data set.

[16] We thank an anonymous referee for suggesting child labor and schooling among girls as indicators of women's empowerment. Note, however, that the presence of child labor is one of the five inclusion criteria for TUP eligibility. This means that the treatment households are expected to have more incidence of child labor at the baseline by design.

[17] For middle class and richer households in Bangladesh, a particularly revealing indicator of "feminine good" is gold jewelry.

of the indicators including large gains in per capita income, food availability, housing (tin roof), livestock, and most of the other assets. There is, however, some evidence that the ownership of homestead land has worsened on average from 2002 to 2005.

(4) Empirical Strategy

Since BRAC did not incorporate any randomized control trials in TUP phase I, we have to rely on the non-experimental data to estimate program effects on household outcomes. To address potential bias due to the non-experimental nature of the data, we use a two-pronged strategy: (1) we pay careful attention to potential selection issues, and use a rich set of econometrics techniques to tackle them, and (2) we use alternative treatment and comparison groups.

(4.1) Alternative Treatment and Comparison Groups

Like many other microfinance programs, there is significant mistargeting (or assignment errors) in the TUP program. Based on the formal selection criteria of BRAC, one can partition the sample of households in the panel data set into four subsets. They are: (i) households that are eligible according to the stated criteria and are included in the program (the subset called the "should be, one" (SB_1) group henceforth), (ii) the eligible households not selected (called the "should be, zero" group (SB_0)), (iii) households ineligible according to formal criteria but selected in the program (called the "should not be, one" group (SNB_1)), and (iv) households ineligible and not selected (called the "should not be, zero" group (SNB_0)). For details on the construction of these four subsets, please see Appendix 1. As discussed above, our empirical analysis is based on two pairs of treatment-comparison groups: BRAC's own $SUP - NSUP$ and $SB_1 - SB_0$.

(4.2) The Potential Selection Issues

There are two levels of selection issues for any given treatment-comparison pair: (i) BRAC's selection process, and (ii) the participation decision by a household. As discussed earlier, BRAC's selection process was based on a set of explicit inclusion and exclusion criteria. To understand the nature of potential selection bias arising from BRAC's selection process we need to have an implicit model of the actual decision making by the BRAC employees.

A simple but not implausible model is to assume that BRAC employees were following the set of inclusion and exclusion criteria, and thus the *assignment errors* discovered in the data are largely due to either randomness arising from human fallibility and other factors, or due to the fact that some eligible households declined to participate in the program. If self-selection out of the program by eligible households is important then households in the treatment group may systematically differ from the comparison group. This potential bias arising from self-selection out of the TUP program, however, is not likely to be important, as non-participation in phase I of this program by a selected ultra-poor was reported by BRAC to have been uncommon.

An alternative model is to assume that BRAC employees were using both the formal criteria and private information available to them. In this case, the objective function of the BRAC employees becomes relevant. If the objective was to identify the true ultra-poor, then the group of households who should have been in the program according to the set of formal criteria but were not selected (i.e., SB_0) must be relatively well off (more advantaged) in terms of initial economic conditions and characteristics in 2002. Under the alternative assumption that the objective was to identify and exclude potentially high risk households so as to help ensure the "success" of the program, then the SB_0 group is likely to be systematically more disadvantaged in 2002. In the presence of heterogeneity among the BRAC employees, both positive and negative selections are likely to characterize our data set. The relevant issue is whether such heterogeneity cancels out on average, or if there is either positive or negative selection in net terms. The evidence presented below is consistent with the interpretation that in net terms the negative selection dominates for both the treatment-comparison pairs; the treatment groups had systematically disadvantaged initial conditions in 2002.

Table 2a reports the difference in means and the associated standard errors for a set of observable characteristics in the baseline (2002) across different pairs of treatment-comparison groups. The first column gives the initial difference in means for the $SB_1 - SB_0$, the second for $SUP - NSUP$, the third for $SNB_1 - SNB_0$, and the last for $SB_1 - SNB_0$. The first striking feature in Table 2a is that most of the entries are negative implying that a treatment group in general exhibits adverse initial conditions compared to the respective comparison group. The fact that the treatment group SUP was systematically disadvantaged in 2002 indicates that even in the presence of mistargeting the TUP program participants are on average poorer households among the poor. As noted before this also points to the possibility of negative selection biases under the plausible assumption that selection on unobservables is similar to the selection on observables (Altonji et al. 2005).

The evidence in Table 2a also shows that the difference in the means is, in general, much lower for the treatment-comparison pair $SB_1 - SB_0$. In contrast, there are some significant and relatively large differences in the initial conditions in 2002 between the treatment and comparison groups as defined by BRAC (i.e., the subsets SUP and $NSUP$) and used by BRAC's Research and Evaluation Division (RED) in its "descriptive analysis" of the TUP program (Rabbani, Prakash, and Sulaiman, 2006).[18] Consider for example, total land and number of cows and bulls owned by a household, which are among the most important household assets in rural Bangladesh. The differences in means for total land owned are: -0.59 ($SB_1 - SB_0$), and -3.94 ($SUP - NSUP$),

[18]The standard errors reported in this paper are clustered at the village level. There are 27 villages in the data set.

and for the number of cows and bulls are -0.03 ($SB_1 - SB_0$), and -0.15 ($SUP - NSUP$). The same pattern holds for most of the other variables in Table 2a.

It is reassuring that the subsamples SB_1 and SB_0 closely resemble each other according to the observable characteristics reported in Table 2a. It thus seems most appropriate to use SB_0 as the comparison group to estimate the effects of TUP program on the treatment group SB_1.[19] To tackle any remaining selection biases, we use a battery of recent econometric approaches (see below).

The evidence indicates that there are important differences in the initial conditions across the different treatment groups. Table 2b reports the group averages of a set of variables in 2002 across the treatment groups. Although the groups are similarly situated according to some observables like food availability, and quality of houses as indicated by the roof made of tin, the SB_1 group is the poorest among them. An average SUP household owns 35 percent more land than an average SB_1 household, and the SNB_1 households own more land than the SB_1 households. While the percentage of households who own their homestead land is 40 percent for the SB_1 group, the corresponding numbers for SUP and SNB_1 are 47 percent and 53 percent respectively. The per capita income measure was $Tk.2452.63$ for an average SB_1 household, $Tk.2488.62$ for SUP, and $Tk.2518.78$ for SNB_1. Given the above analysis, our focus is on the estimates of treatment effects (ATT) from two alternative pairs of treatment-comparison groups: SB_1 (treatment) and SB_0 and SUP (treatment) and $NSUP$ (comparison). Note that the SUP consists of both SB_1 and SNB_1, although we do not discuss separately the treatment effect estimates for the treatment group SNB_1 as it is composed of the richer households, and thus clearly not the target group of the TUP program.

(4.3) Econometric Approaches

Matching Estimators and Beyond

Following Heckman et al. (1998), we combine the difference-in-difference approach with matching (the DIDM estimator). As mentioned earlier, the DIDM approach purges any time invariant additive heterogeneity at the individual level by time differencing (i.e., household fixed effects); and then matching on the pre-intervention characteristics takes care of selection on observables in a flexible way without imposing any particular functional form. This, however, may not adequately address the possibility that the estimated treatment effect may be contaminated by selection on unobservables that vary over time. The assumption that most unobservables are time invariant or change very slowly over time may not be appropriate in our context. The households are given training to use the transferred assets and also various forms of information (such

[19] Note that the treatment group SB_1 and the comparison group SNB_0 satisfy the BRAC inclusion and exclusion criteria perfectly. The difference in means in 2002 between these two groups is much more pronounced than the differences across SUP and $NSUP$ (see Table 2a).

as health) are provided on an ongoing basis by the TUP program. The sphere of their social interactions also broadens through the "village elders". The learning by doing effect may also be substantial because many borrowers manage productive assets for the first time in their life.[20] Also, the BRAC field officers went through a learning process themselves, especially because it was the first phase of a new program. It is plausible that there is substantial heterogeneity in the ability of BRAC field officers, and this influenced the learning by the borrowers differentially over time.

We use two alternative approaches to address possible biases in the DIDM estimates. First, we implement the 'minimum bias inverse probability weighted' (MB-IPW) estimator developed by Millimet and Tchernis (2012). The MB-IPW estimator starts from the inverse probability weighted matching estimator developed by Hirano and Imbens (2001), which weights the observations on the treatment group by the probability of being treated (the inverse of the propensity score) and weights the observations in the control group by the probability of not being treated (i.e. one minus the inverse of the propensity score) and minimizes any possible bias in the estimates arising from the failure of the conditional independence assumption by using an appropriately trimmed sample around the bias minimizing propensity score.[21] In our case, we used $\theta = 0.25$, which means that at least 25 percent of both the treatment and control groups have a propensity score in the interval used (described below). Second, and more importantly, we provide estimates of the effects of the TUP program that address possible selection on unobservables (omitted variables bias) by using the heteroskedasticity-based identification approach due to Klein and Vella (2009).[22]

[20] The household fixed effects take care of the heterogeneity in the innate entrepreneurial ability, but not the dynamic learning by doing effect. Also, it is possible that the entrepreneurs themselves do not know their ability and learn about it over time when they participate in the TUP program. For a theoretical model of credit market failure in entrepreneurial development in such cases of discovery of ability, see Emran and Stiglitz (2009).

[21] Millimet and Tchernis call it the Minimum Bias (MB) estimator, but we find it informative to use the term Minimum Bias Inverse Probability Weighted (MB-IPW) estimator, because it builds on the IPW estimator due to Hirano and Imbens (2001).

[22] An increasingly popular approach to understanding the implications of violation of the conditional independence assumption in a matching estimator is to use sensitivity analysis following Aakvik (2001). An earlier version of this paper reported results from such sensitivity analysis in the context of the DIDM estimator. We, however, chose to omit those results from the paper, because such a sensitivity analysis is based on an *arbitrary and implausible assumption that selection on unobservables is always positive.* Such a sensitivity analysis is of little value, if not counterproductive, for two reasons. First, in Table 2a we provide evidence consistent with negative selection in our data set. The existing econometric evidence in fact supports the notion that selection may be negative in the case of the microfinance programs in Bangladesh (see, for example, Pitt and Khandker 1998; Schroeder 2010). Also, the sign of the selection bias (or omitted variables bias) is likely to depend on the outcome under focus. The evidence of negative selection provided by both Schroeder (2010) and Pitt and Khandker (1998) concerns household consumption as the relevant outcome. We are not aware of any formal evidence that shows positive selection into microfinance or TUP programs in Bangladesh. Second, sensitivity analysis using Rosenbaum bounds does not address the possible bias from measurement error.

Heteroskedasticity Based Identification: the Klein and Vella (2009) Approach

There is now a substantial econometric literature that shows that in the absence of credible exclusion restrictions required for an instrumental variables strategy, one can use heteroskedasticity for identification (see Rigobon 2003; Klein and Vella 2009, 2010; Lewbel 1997, 2012). As noted by Rigobon (2003), analogous to the standard instrumental variables, heteroskedasticity can be understood as an exogenous 'probabilistic shifter' of the endogenous treatment variable which helps us trace out the causal relation between a dependent variable (household outcomes such as food availability) and the endogenous treatment variable (microfinance membership). In recent papers, Klein and Vella (2009, 2010) show that when the treatment equation in a triangular model exhibits heteroskedasticity, this effectively induces an exclusion restriction even though there is no standard exclusion restriction available. Monte Carlo evidence from a number of recent studies shows that the Klein and Vella approach (henceforth K-V for short) is effective in correcting for the endogeneity bias, and also the bias from measurement error (Ebbes et al. 2009; Klein and Vella 2009, 2010; Millimet and Tchernis 2012).

To provide intuition behind the approach, we consider the following triangular model:

$$\Delta Y_i = \alpha_0 + \alpha_i + \alpha M_i + X_i'\gamma_1 + \xi_i \quad (1)$$

$$M_i = \mathbf{1}\{\beta_0 + \beta_i + X_i'\beta_1 + u_i > 0\} \quad (2)$$

where ΔY_i is the change from 2002 to 2005 for an outcome of interest (such as income, assets, etc) of household i; M_i is a dummy that equals one if the household i is a participant of TUP program and zero otherwise and α_i and β_i are household fixed effects. The focus is on the identification and estimation of the parameter α. The model does not impose any exclusion restrictions on equation (1), and identification of the causal effect α is not possible if the error terms are homoskedastic.[23] Assume that the error term in the treatment equation is heteroskedastic of the following form:

$$u_i = S_u(\tilde{X}_i'\pi)\tilde{u}_i \quad (3)$$

where $\tilde{u}_i$ is a zero mean homoskedastic error, $\tilde{X}_i \subseteq X_i$ are the variables generating heteroskedasticity, and $S_u(\tilde{X}_i)$ is a positive and nonconstant function. In this case, the probability of treatment (probability of TUP membership) can be written as follows (ignoring the fixed effects):

$$\Pr(M_i = 1) = P\left(\frac{X_i'\beta_1}{S_u(\tilde{X}_i'\pi)}\right) \quad (4)$$

[23] As noted before, one can argue that identification in the above model can be achieved without exclusion restrictions, because the treatment equation is nonlinear. But such identification depends critically on the validity of the Normality assumption and the nonlinearity of the Normal CDF. The model is in general poorly identified. For discussions, see Klein and Vella (2009), Altonji, Elder and Taber (2005).

where $P(\cdot)$ is the distribution function for $\tilde{u}_{it}$. With homoskedastic errors, $S_u(\tilde{X}_i'\pi)$ is a constant, and the only source of identification is possible non-linearity of the $P(\cdot)$ function such as a Normal distribution. However, such identification based on the non-linearity in the tails of the distribution may not be very useful because it relies on a small fraction of the data (for discussions on this point, see Altonji et al. 2005 and Klein and Vella 2009). In contrast, when there is heteroskedasticity, $S_u(\tilde{X}_i'\pi)$ is not a constant function, and identification also exploits data from the region where $P(\cdot)$ is approximately linear. The predicted probability of treatment from estimating equation (2) above becomes a valid instrument in the presence of heteroskedasticity, because it is no longer a linear combination of the control variables in the outcome equation (1). Note that if the amount of heteroskedasticity is not substantial, then there is little identifying variation in $S_u(\tilde{X}_i'\pi)$, and the predicted probability may be a weak instrument. For the specification of the $S_u(\tilde{X}_i'\pi)$, we follow the parametric approach developed in Farré, Klein and Vella (2010) which is based on the model of heteroskedastic probit due to Harvey (1976).

$$S_u(\tilde{X}_i) = e^{\tilde{X}_i'\pi} \tag{5}$$

A limitation of the heteroskedasticity-based identification is that it is not appropriate for the binary outcomes. Another disadvantage of the heteroskedasticity-based instrumental variable approach is that the estimates are likely to be less efficient than the usual IV estimates, because identification here relies on information about the second moment of the data (Lewbel 2012).

(5) Estimated Effects of the TUP Program

In this section, we report the estimated treatment effects on a set of household outcomes including income, land ownership, basic needs such as food availability and clothing, assets, child labor, schooling, and health related indicators. As a benchmark, we report estimates from an augmented difference-in-difference specification.

We use the following difference-in-difference specification:

$$Y_{it} = \alpha_0 + \alpha_1 d_{05} + \alpha_{1R}\left(d_{05} * d_R\right) + \alpha_{1K}\left(d_{05} * d_K\right) + \alpha_2 d_T + X_{02}^{'}\Pi + \beta\ \left(d_T * d_{05}\right) + \epsilon_{it} \tag{6}$$

where Y_{it} is the outcome variable of interest for household i in year t, d_{05} is a dummy that equals 1 for the year 2005, and d_T is a dummy that equals 1 when household i belongs to an appropriately defined treatment group (i.e., SB_1, SUP) and equals zero when a household belongs to the corresponding comparison group (i.e., $SB_0, NSUP$). The parameter of interest is β, which isolates the treatment effect on outcome Y under certain assumptions. The crucial difference-in-difference identification assumption is that the treatment and comparison groups would follow the same trend in the absence of the program. If this assumption is not satisfied, the estimate of the

treatment effect $\hat{\beta}$ will be biased when we use OLS to estimate equation (1). Thus we augment the DID specification in two ways to make it more plausible that the counterfactual trend for the treatment group is well represented by the actual trend in the comparison group. First, we allow for differential time trends in the different districts where the households are situated in: d_R and d_K are dummies for Rangpur and Kurigram districts respectively.[24] In addition, we allow for the possibility that the trends might differ across households with different observable characteristics. Thus, we also control for a set of observables that are likely to be important for selection into the treatment (either because of BRAC's criteria, or the household's own outside option).

The DID estimates, although useful as a benchmark, do not fully exploit the panel dimension of the data; and also rely on the assumption that the control variables do not have any non-linear effects on the outcomes. To address these twin issues, we implement the DIDM approach with household fixed effects. The DIDM estimates are reported in column (3) of Tables 3a (for SUP) and 3b (for SB_1). An important step in implementing the DIDM approach is to choose an appropriate set of observable characteristics that are likely to be important in determining the selection into treatment and may also affect the outcome variables (Heckman and Navarro-Lozano 2004; Blundell and Dias 2009). As discussed above, we need to consider two levels of selection: BRAC's selection process and also the participation decisions of the households. We thus use observables that reflect these two levels of selection problems for matching. To account for the BRAC selection process we use the TUP program's own set of inclusion criteria.

Moreover, we include indicators of physical and human capital (for example, land owned, household size, gender dummy, and the indicator of women working as day laborers). As emphasized by Emran, Morshed and Stiglitz (2007), the outside option of a household and thus the net return from participation in the TUP or other NGO programs depends on the nature of labor market interactions and the shadow value of labor, especially of women's labor. In a perfect labor market, the labor and land endowments would not affect the incentives to take loans, given the wage rate (i.e., separability holds). In contrast, in an imperfect labor market, more labor endowment, *ceteris paribus*, implies low shadow value of labor, and such a household would find it more attractive to borrow from microfinance NGOs. In effect, the microcredit allows a household to create demand for labor within the household; this is especially valuable when they cannot find employment outside, for example, because of high unemployment rate. We thus expect that the effect of labor endowment will be positive on the probability of joining into the microfinance programs. Similar arguments imply that the effects of higher land endowment will be negative, as more land, *ceteris paribus*, increases the shadow price of labor.

We include household size as an indicator of labor endowment of the household, and the variable "day labor" as a measure of labor market participation by women. We also include "land

[24]The omitted district is Nilphamari.

owned," as it is a crucial variable for the determination of the shadow price of labor and also whether a woman is excluded from the formal credit market (lack of collateral) and the labor market (efficiency wage effects). All of the matching variables are from the 2002 baseline survey. The estimated selection equations for the SUP and SB_1 treatment households are reported in Table A.1 in the appendix. The results from the probit regressions show that, for SUP, the participation/selection into the TUP program depends on a household's land and labor endowments. Consistent with the theoretical predictions of Emran et al. (2007), land has a negative and labor endowment a positive effect on the probability of participation in the TUP program. Also, among the five inclusion criteria, the fourth criterion (adult women working outside the home) and the fifth (no productive asset) have significant positive effects on the probability of selection into the program. The results from probit regressions for the SB_1 households show the importance of inclusion criteria four and five for their selection into the program. However, land and labor endowment do not have a significant effect in the case of the SB_1 households, although they bear theoretically consistent signs.

Another important issue in the implementation of any matching estimator is the common support. We chose the set of covariates so as to ensure ex-ante a balanced matching quality, which we tested with a visual inspection of the density function of the propensity score in both treatment and control groups and with a test of differences of the mean propensity score for treatment and control. We imposed the common support region for each or our samples, SB_1/SB_0 and $SUP/NSUP$. In the SB_1/SB_0 group, the common support region is [0.22, 0.80] and used 1656 of the 1657 observations available. In the $SUP/NSUP$ case, the common support region is [0.00, 0.78] and uses 4842 of the 4854 observations available. Thus, there is no selection in the information lost by the common support imposition. The results for the DIDM estimator reported in this paper are based on the radius caliper algorithm for matching. However, note that the estimates from alternatives such as kernel matching are very similar (the results are available from the authors).

The estimates from the MB-IPW estimator with household fixed effects are reported in column (4) of Tables 3a and 3b. We use a wide radius ($\theta = 0.25$) for trimming the sample around the bias minimizing propensity score[25], as the Monte Carlo results in Millimet and Tchernis (2012) show that MB-IPW with a wide radius effectively corrects for endogeneity bias.

Finally, in column (5) of tables 3a and 3b, we report estimates from the Klein and Vella (2009) estimator with household fixed effects that correct for selection on unobservables and measurement error. They are our preferred estimates for continuous outcomes.[26] Following Klein and Vella

[25]In addition, the sample is trimmed around the 2 percent tails of the propensity score to minimize the bias arising from the failure of the conditional independence assumption.

[26]Because K-V is not suitable to use on binary outcomes, our preferred set of estimates in that case are the MB-IPW ones.

(2009) and Farré et al. (2010), we include the full set of explanatory variables in the heteroskedastic probit specification, thus allowing for the possibility that potentially all of the variables can give rise to heteroskedasticity in the microfinance participation equation. An important issue here is whether the heteroskedasticity-based instrument derived from the heteroskedastic probit model is strong enough to identify the effects of the TUP program participation.[27] The LR test for the null of homoskedasticity is convincingly rejected in both the samples; the P-value is 0.00 for SUP and 0.03 for SB_1. The Kleibergen-Paap F statistics for SUP and SB_1 show that the heteroskedasticity-based instrument is strong enough to identify the effects of the TUP program participation; the Kleibergen-Paap F statistic is 35.43 for SUP and 33.98 for SB_1, much higher than the Stock et al. rule of thumb value of 10 for one endogenous variable.[28]

(5.1) Effects of the TUP Program on Treatment Group "Selected Ultra Poor" (SUP)

We first consider the effects of the TUP program participation on the BRAC's own treatment group SUP. Despite the fact that it includes relatively better-off households, the effects of the program on the SUP treatment group can be interesting; a comparison with the effects on the treatment group SB_1 that excludes the relatively better-off households can be particularly illuminating. Table 3a reports the results for the SUP treatment group; the standard errors are clustered at the village level (there are 27 villages in the data set).

The benchmark DID (column 1) estimates suggest that participation in the TUP program for an SUP household had positive effects on a number of outcomes including per capita income, cash savings, food security, land ownership, quality of housing (with tin roof), livestock (cows and bulls, goats and sheep, and ducks and hens), other productive assets (number of big trees and fishing nets, for example), and household durables (with the exception of radio/TV).[29] However, there seems to be no significant effect on the indicators of female empowerment and child labor. Interestingly, the results indicate that TUP participation leads to improvements in women's clothing (number of sarees), but there is no significant effect on men's clothing (lungi). Participation in the TUP program also improves the probability that all the members of a households have shoes/sandals. The evidence on health outcomes is conflicting, while there is improvement compared to the health status in the last year, we do not find evidence of an improvement in over-all health status that can be attributed to the TUP program participation.

[27] It is now widely appreciated in the literature that the IV estimates may be more biased than the OLS if the instrument is weak.

[28] Note that we do not report the estimates for the binary outcomes from the Klein and Vella (2009) estimator, as it not appropriate in such cases.

[29] The effect on fishing net, rickshaw van, and bicycles is numerically small.

The estimates from the two matching estimators, DIDM (column 2) and MB-IPW (column 3), are very close to each other, with the point estimates numerically identical in many cases. While for most of the household outcomes columns 2 and 3 show similar results, the estimates from the matching estimators reveals a few notable differences compared to the DID estimates. Probably the most important is the estimated effect on the total land endowment of a household; the effect is negative and insignificant according to both DIDM and MB-IPW, while it is positive, numerically substantial (1.41) and statistically significant at the 5 percent level in the DID specification in column (1). Moreover, the effects on the ownership of homestead land and number of big trees are much smaller according to the matching estimators when compared to the DID estimate. In contrast, the effect on male clothing is more than three times larger and also statistically significant at the 1 percent level according to the matching estimators. But it is numerically small and not statistically significant according to the DID estimate.[30]

The last column in Table 3a reports the estimates from the Klein and Vella (2009) approach, implemented with household fixed effects. The first thing to note is that the effect of TUP membership on per capita income of an SUP household is numerically much larger according to the heteroskedasticity-based instrumental variable estimate. But the standard error is also large, and the effect is not significant at the 10 percent level. A conservative interpretation of the results on income in Table 3a is that the evidence is not very strong that TUP participation has a positive effect on the income of an average SUP household. An alternative interpretation, which in our view is more plausible, is that, if anything, the evidence suggests that TUP membership increases income, because all four estimates in Table 3a are numerically large, and the large standard error of the K-V estimate probably reflects, at least partly, the inefficiency of the heteroskedasticity-based estimator noted earlier. This interpretation is also consistent with the robust evidence from the matching estimators of a positive effect on cash savings, ownership of homestead land and quality of housing (tin roof) for a SUP household. The K-V estimate also strengthens the conclusion that TUP improves significantly the over-all food availability (the effect is about 50 percent larger according to K-V). The TUP program effect is much larger for male clothing according to the K-V estimate; it increases from 0.22 (DIDM) to 0.70 (K-V) and is significant at the 10 percent level. The estimated effect for female clothing (saree) is somewhat smaller; it declines from 0.28 (DIDM, and MB-IPW) to 0.22 (K-V), and it is no longer significant at the 10 percent level.

Considering the impact on livestock ownership, the effect on the number of cows/bulls a household owns seems to be very robust, but the evidence in favor of a substantial positive effect is much weaker according to the K-V estimates for the other livestock such as goats, sheep, ducks, and hens. The estimated effect for other livestock is numerically smaller compared to

[30]We again emphasize here that both the DIDM and MB-IPW estimates are built on household fixed effects to take care of the time invariant heterogeneity such as innate entrepreneurial ability.

the estimates from DIDM and MB-IPW, and statistically not significant at the 10 percent level. The K-V instrumental variable estimates for productive assets (fishing nets, big trees, rickshaw van, and bicycles), a number of household durables (chairs, tables and beds), self-reported health outcomes, child labor, and women's empowerment indicators also fail to show any significant effect of TUP membership. One should however be careful in interpreting the lack of a program effect on some of these outcomes. For example, the TUP program may not increase the number of fishing nets owned by a household, as fishing is not the main economic activity for most of the households in Bangladesh. With respect to the lack of impact on child labor reduction, we speculate that children participate in the newly formed enterprises during their initial start-up. Interestingly, the K-V estimates suggest that there is significant program impact on a set of household durables: radio, TV, quilts, blankets and tube wells. While the effects on radio, TVs and quilt/blankets are positive, the effect on tube wells is negative according to the K-V estimate. The negative effect may appear to be a bit puzzling. It might, however, be useful to recall that the asset measures (including tube wells) in 2005 do not include the transfers from the TUP program. Thus even though the TUP program might have installed many new tube wells as part of the health intervention to ensure safe drinking water, our data do not capture this effect. This implies that if the TUP replaced some existing tube wells (presumably not working properly), it would be measured as a negative program effect.

(5.2) Effects of the TUP Program on the Poorest of the Poor: SB_1 Treatment Group

Turning to the effects of the TUP Program on the poorest of the poor (the SB_1 treatment group), we begin with the benchmark DID estimates reported in column (1) of Table 3b. Similar to the results on the SUP treatment group, the benchmark estimates suggest beneficial effects of TUP membership on a number of important household outcomes including per capita income, ownership of homestead land, better quality of housing (tin roof), food security, livestock, and most of the household durables. Consistent with the effects on SUP treatment group in Table 3a, the DID results in Table 3b also suggest that there is no significant effect of TUP membership on the SB_1 households regarding women's empowerment and child labor. However, there are three important differences between the results for SUP and SB_1. First, unlike SUP treatment group, the effect on per capita income is significant at the 10 percent confidence level for SB_1. Second, the effect on rickshaw vans, bicycles and fishing nets are not significant at the 10 percent level for SB_1, although they were found significant for the SUP. Third, the program effect on health status is significant at the 5 percent level for SB_1, but it is numerically smaller and not significant at the 10 percent level for SUP.

The DIDM and MB-IPW estimates for SB_1 are reported in columns (2) and (3) respectively in Table 3b. The results from DIDM and MB-IPW estimators are in general close to the DID estimates reported in column (1). Among the few exceptions are the number of big trees and two indicators of health. The effect on male clothing is much larger according to the matching estimators. In contrast, the positive effect on the two health outcomes found in DID becomes smaller and statistically not significant according to both DIDM and MB-IPW.

The last column in Table 3b reports the estimated program effects from the heteroskedasticity-based instrumental variables approach of Klein and Vella (2009) with household fixed effects for the SB_1 households. A comparison with the estimates from the matching estimators in columns (2) and (3) shows some interesting differences. The K-V estimates for per capita income are about 33 percent larger than the matching estimates, and significant at the 10 percent level. The K-V estimates for food availability are consistent with the conclusion that participation in the TUP program enhances food security of an average SB_1 household. The magnitude of the program effect on 'food availability' is about 20 percent larger, and it is significant at the 5 percent level.

The K-V estimates when combined with the estimates from the matching estimators provide strong evidence of a substantial and significant positive effect of TUP participation on livestock (cows, bulls, goats and sheep). However, the positive and significant program impact indicated by the matching estimators is not supported by K-V results for a number of household outcomes such as male clothing, ducks and hens, and some of the durables (chairs, tables, quilt/blankets, tube wells). The results thus suggest that the estimates from the matching estimators may be biased upward for these outcomes.

The K-V estimate of the effect on female clothing (saree) is numerically much larger than the estimates from the matching estimators, and it is also significant at the 5 percent level. The overall evidence thus is very robust that the TUP has a substantial positive effect on female clothing. But there is no evidence, according to the K-V estimates, that male clothing (lungi) or the ratio of female to male clothing (saree to lungi) are positively affected by the program participation in any appreciable manner.[31] Interestingly, the K-V instrumental variable estimates suggest that the TUP program participation increases productive investment such as livestock and big trees, but reduces the expenditure on luxury items such as radio/TV.[32]. The results on female clothing and radio/TV thus indicate that, in some cases, the estimates from the matching estimators may under-estimate the program effect. Applying the sensitivity analysis within a matching approach that arbitrarily assumes a positive selection effect (upward omitted variables bias) would clearly be counterproductive in such cases.

[31]The ratio of saree to lungi is an indicator of women's bargaining power in household expenditure choices.

[32]The negative significant effect on radio/TV is in contrast to the DIDM and MB-IPW estimates that do not show any perceptible effect

(5.3) Comparison Between the Treatment Groups, and Broader Implications of the Empirical Results

The evidence presented in Tables 3a and 3b is interesting for three reasons: (*i*) TUP program seems to have robust beneficial effects on food security, cash savings, and household assets such as livestock and housing for both the treatment groups, (*ii*) the TUP program effect may be different for different treatment groups (for example, on income and expenditure and on luxury goods), and (*iii*) the estimates from the sophisticated difference-in-difference and matching estimators may be biased in some cases.

The evidence in Tables 3a and 3b shows that, in many cases, the estimates of the program effects from difference-in-difference and matching estimators are consistent with those from alternatives such as heteroskedasticity-based identification. However, a comparison with the Klein and Vella (2009) estimates also suggest that the estimates from sophisticated difference-in-difference and matching estimators can potentially under or overestimate the causal effects of a microfinance program. This is consistent with other recent evidence on the performance of the matching estimators (see, in particular, the evidence on the effects of school breakfast program in Millimet and Tchernis 2012).

The evidence strongly suggests that TUP participation confers significant benefits to a participant household, irrespective of the treatment group considered. A comparison of Tables 3a and 3b shows that, for many outcomes, the estimated effects are similar between the two treatment groups, but this apparent similarity hides important differences. This is because of the fact that the households in two treatment groups start from very different initial conditions in 2002. Since households in the SB_1 treatment group are poorer with lower initial income and asset positions, the gains from participation when normalized by the initial conditions are, in general, much higher for the SB_1 households. For example, the mean stock of cows/bulls was 0.01 for SB_1 and 0.04 for SUP in 2002. The estimated program effect is 1.57 for SB_1 and 1.31 for SUP. If we express as proportion of the initial stock to get a measure of 'normalized program effects', the program effects on cows/bulls become 157 times of the initial stock for SB_1 and 32.75 times of the initial stock for SUP. This dramatically illustrates the importance of taking into account the fact that the poorest of the poor start from much worse initial asset positions. We report the normalized program effects in Tables 4a (for SUP treatment group) and 4b (for SB_1 treatment group). Had the program concentrated on the poorest of the poor, the program effects would have been larger than those found.

While all the participant households seem to invest in livestock and housing improvements, the poorest of the poor households who are the target group of the TUP program (i.e., the SB_1 treatment group) do not spend their money on luxury goods such as radio/TV.

On the other hand, the evidence also suggests that the TUP participation does not have any significant effect on our indicators of women's empowerment, health, and child labor.

(6.) Conclusions

Using a two-period household level panel data set, this paper provides robust evidence on the effects of BRAC's Targeting Ultra-poor Program (TUP Phase I) on a set of important household outcomes for the ultra-poor. We use a battery of recent econometric approaches and alternative treatment-comparison groups to identify and estimate the effects of TUP program participation. In addition to BRAC's own treatment-comparison groups, we utilize the type 1 errors in assignment (mistargeting) in BRAC's screening to create an alternative treatment-comparison pair. This allows us to identify a treatment group composed of the poorest of the poor (i.e., ultra-poor) among the sample households and also an appropriate comparison group for this treatment group.

To estimate the effects of the TUP program, we use recently developed matching estimators and a heteroskedasticity based identification approach that takes into account selection on both observables and unobservables. The results show that there is significant impact of TUP program participation on food security, cash savings and livestock of the ultra-poor households. The evidence also indicates that the TUP program may not have any significant effects on health related outcomes, women's empowerment and some of the productive assets examined. When the differences in the initial conditions are taken into account, we find that the normalized program effects are significantly larger for the treatment group consisting of the poorest of the poor households, i.e., SB_1. Had the program concentrated on the poorest of the poor, the program effects would have been larger than those found.

An interesting finding from our analysis with potentially broad policy implications is that the effects of an ultra-poverty program on the poorest of the poor may be different from the effects on an average participant in the program when there is mistargeting. Compared to an average program participant, the poorest participants (i.e., SB_1) gain more in per capita income, invest more in productive assets such as cows/bulls and goats/sheep, and do not increase expenditure on the luxury goods such as radio/TV. This implies that the researchers should carefully define the treatment and comparison groups using program criteria, especially when there is significant mistargeting in a program. The substantive conclusions can differ depending on the treatment group under focus in an analysis. The evidence also indicates that the effectiveness of the TUP program would improve substantially with better targeting.

References

Aakvik, Arild. 2001. "Bounding a Matching Estimator: The Case of a Norwegian Training Program." *Oxford Bulletin of Economics and Statistics* 63(1) (February): 115-143.

Altonji, Joseph G, Todd E. Elder and Christopher R. Taber 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 13(1) (February): 151-184.

Banerjee, Abhijit, Esther Duflo, R. Chattopadhya, J. Mukhopadhya and J. Shapiro. 2008. "Targeting Efficiency: How Well Can We Identify the Poor?" Paper presented at CMFR, India, 2008.

Berg, Claudia and M. Shahe Emran. 2011. "Does Microfinance Help the Ultra-Poor Cope with Seasonal Shocks? Evidence from Seasonal Famine (Monga) in Bangladesh." Working Paper, Department of Economics, The George Washington University and Institute of Microfinance (InM), Dhaka Bangladesh, 2010. Available on SSRN.

Berg, Claudia, M. Shahe Emran and Forhad Shilpi. 2012. "Microfinance and Moneylenders: The Effects of MFIs on Informal Credit Market in Bangladesh." Working Paper, TheGeorge Washington University and World Bank, Washington, DC.

Berlinski, Samuel, Sebastian Galiani and Paul Gertler. 2009. "The Effect of Pre-primary Education on Primary School Performance." *Journal of Public Economics* 93(1-2): 219-234.

Black, Dan A. and Jeffrey A Smith. 2004. "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics*, Elsevier 121(1-2): 99 - 124.

Blundell, Richard and Monica Costa Dias. 2009. "Alternative Approaches to Evaluation in Empirical Microeconomics." *Journal of Human Resources* 44(3) (Summer): 565-640.

BRAC. 2007. BRAC Annual Report, Dhaka, Bangladesh.

Chowdhury, Shamal, M. Shahe Emran, Debdulal Mallick and Forhad Shilpi. 2012. "Long-term Effects of Microfinance on Household Outcomes in Bangladesh: Evidence from Multiple Panel Data Sets." World Bank, Washington, DC.

Chronic Poverty Research Center. 2008. Chronic Poverty Report 2008-09. Manchester

Deaton, Angus. 1989. "Looking for Boy-Girl Discrimination in Household Expenditure Data." *World Bank Economic Review*, Oxford University Press 3(1) (January): 1-15.

Ebbes, Peter, Michel Wedel and Ulf Böckenholt. 2009. "Frugal IV Alternatives to Identify the Parameter for an Endogenous Regressor." *Journal of Applied Econometrics* 24(3): 446-468.

Emran, M. Shahe and Zhaoyang Hou. forthcoming. "Access to Markets and Rural Poverty: Evidence from Household Consumption in China." *Review of Economics and Statistics.*

Emran, M. Shahe and Forhad Shilpi. 2012. "The Extent of the Market and Stages of Agricultural Specialization." *Canadian Journal of Economics* (August).

Emran, M. Shahe and Joseph E. Stiglitz, 2009. "Financial Liberalization, Financial Restraint and Entrepreneurial Development." Working Paper, Columbia University. Available at SSRN.

Emran, M. Shahe and Yan Sun. 2011. "Magical Transition? Intergenerational Educational and Occupational Mobility in Rural China: 1988-2002." Working Paper, The George Washington University. Available on SSRN.

Emran, M. Shahe, AKM Morshed and Joseph E. Stiglitz, 2007. "Microfinance and Missing Markets." Working Paper, The George Washington University.

Emran, M. Shahe, Forhad Shilpi and Joseph E. Stiglitz, 2009. "De Soto vs. Yunus: Microfinance, Property Rights and Poor." Paper presented at the ASSA 2009 Conference at San Francisco.

Farré, Lidia, Roger Klein and Francis Vella. 2010. "A Parametric Control Function Approach to Estimating the Returns to Schooling in the Absence of Exclusion Restrictions: an Application to the NLSY." *Empirical Economics* 1-23.

Harvey, A.C. 1976. "Estimating Regression Models with Multiplicative Heteroscedasticity." *Econometrica: Journal of the Econometric Society*: 461-465.

Heckman, James J. and Salvador Navarro-Lozano. 2004. "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models." *Review of Economics and Statistics* 86(1): 30-57.

Heckman, James J., Hidehiko Ichimura and Petra Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic studies* 65(2): 261-294.

Hirano, Keisuke and Guido W. Imbens. 2001. "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization." *Health Services and Outcomes Research Methodology* 2(3): 259-278.

IFPRI. 2007. "The world's most deprived: Characteristics and Causes of Extreme Poverty and Hunger." Vol. 43 (October). Intl Food Policy Res Inst.

Kanbur, Ravi. 2010. "Macro Crises and Targeting Transfers to the Poor." *Journal of Globalization and Development* 1(1) (January).

Klein, Roger and Francis Vella. 2009. "A Semiparametric Model for Binary Response and Continuous Outcomes under Index Heteroscedasticity." *Journal of Applied Econometrics* 24(5): 735-762.

Klein, Roger and Francis Vella. 2010. "Estimating a Class of Triangular Simultaneous Equations Models without Exclusion Restrictions." *Journal of Econometrics* 154(2) (February): 154-164.

Lewbel, Arthur. 1997. "Constructing Instruments for Regressions with Measurement Error when no Additional Data are Available, with an Application to Patents and R&D." *Econometrica: Journal of the Econometric Society* 1201-1213.

Lewbel, Arthur. forthcoming. "Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models." *Journal of Business and Economics Statistics*.

Lipton, Michael. 1983. "Poverty, Undernutrition, and Hunger." World Bank, Washington, DC.

Mallick, Debulal. 2012. "Microfinance and Moneylender Interest Rate: Evidence from Bangladesh." *World Development* 40(6) (June): 1181-1189.

Matin, Imran. 1998. "Mis-Targeting by the Grameen Bank: A Possible Explanation." *IDS Bulletin* 29(4): 51-58.

Matin, Imran and David Hulme. 2003. "Programs for the Poorest: Learning from the IGVGD Program in Bangladesh." *World Development* 31(3) (March): 647 - 665.

Matin, Imran, Munshi Sulaiman and Mehnaz Rabbani. 2008. "Crafting a Graduation Pathway for the Ultra-Poor: Lessons and Evidence from a BRAC Programme." *Chronic Poverty Research Center* BRAC Research and Evaluation Division (March) no. 109.

Millimet, Daniel L. and Rusty Tchernis. 2012. "Estimation of Treatment Effects without an Exclusion Restriction: with an Application to the Analysis of the School Breakfast Program." *Journal of Applied Econometrics*. (forthcoming)

Montgomery, David R, Bhattacharya, D. and Hulme, David. 1996. "Credit for the Poor in Bangladesh: The BRAC Rural Development Programme and the Government Thana Resource Development and Employment Programme." in D. Hulme and P. MOsley (eds.), Finance Against Poverty, Vol. 2, Routledge, London.

Mustafa, Shams, Ara, Ishrat, Banu, Dilruba, Hossain, Altaf, Kabir, Azmal, Mohsin, Mohammad, Yusuf, Abu and Jahan, Sarwar. 1996. "Beacon of Hope: An Impact Assessment Study of BRAC's RDP." Dhaka, BRAC. (February)

Noor, Marufia, Abu Muhammad Shihab, Rezvina Parveen, Hasanur Rahman, Mamun ur Rashid, Tariq Ali, Shahaduz Zaman and Imran Matin. 2004. "Stories of Targeting." BRAC CFPR/TUP Working Paper Series no. 1.

Petkova, Neviana. 2009. "Does Foreign Share holding Improve Firm Operating Performance?" University of Oregon, Department of Finance, Working Paper.

Pitt, Mark M. and Shahidur R. Khandker. 1998. "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: does the Gender of Participants Matter?" *Journal of Political Economy* 106(5): 958-996.

Rabbani, Mehnaz, VA Prakash, and Munshi Sulaiman. 2008. "Impact Assessment of CFPR/ TUP: a Descriptive Analysis based on 2002-2005 Panel Data." BRAC CFPR/TUP Working Paper Series, no. 12.

Ravallion, Martin. 2007. "Evaluating Anti-poverty Programs." Handbook of Development Economics, 4: 3787-3846.

Ravallion, Martin and Lipton, Michael. 1995. "Poverty and Policy." Handbook of Development Economics, 3: 2551-2657.

Rigobon, Roberto. 2003. "Identification Through Heteroskedasticity." *Review of Economics and Statistics* 85(4): 777-792.

Rigobon, Roberto and Dani Rodrik. 2005. "Rule of Law, Democracy, Openness, and Income." *Economics of Transition* 13(3): 533-564.

Schroeder, Elizabeth. 2010. "The Impact of Microcredit Borrowing on Household Consumption in Bangladesh." Mimeo, Oregon State University Economics Department.

Song, Jae G. and Joyce Manchester. 2007. "New Evidence on Earnings and Benefit Claims following Changes in the Retirement Earnings Test in 2000." *Journal of Public Economics* 91(3-4) (April): 669-700.

Sulaiman, Munshi and Imran Matin. 2006. "Targeting Effectiveness of CFPR/TUP in Scaled Up Environment." BRAC CFPR/TUP Working Paper Series, no. 8.

Todd, Petra. 2006. "Matching Estimators." Entry for New Palgrave Dictionary of Economics, Second Edition, Palgrave Macmillan.

World Bank. 2006. "World Development Report 2006." Oxford University Press.

Zaman, Hassan. 1998. "Can Mis-Targeting be Justified?: Insights from BRAC's Micro-Credit Programme." *IDS Bulletin* 29(4): 59-65.

Appendix 1: Creating Variables for the Errors in Assignment Analysis

Initial eligibility for people living in poverty to join the program is based upon selection at a meeting of the village, which designates households in the lower two socioeconomic strata; but among those selected as potentially eligible ultra-poor by the village, the NGO then selects participants according to three exclusion criteria and the presence of at least 3 out of 5 inclusion criteria (Noor et al. 2004, p. ix). The exclusion criteria are EC1 (the household is not a member of another NGO), EC2 (the household is not a recipient of a government welfare food distribution program), and EC3 (there is no female able to work in the household.)

We created our own designation of those eligible using the survey data. To do so for the case of NGO membership we used the responses to: (i) whether the household had NGO savings (variable "ngos" - selected 340 observations); (ii) whether the household had a loan from an NGO (variable "ngoln" - selected 64 observations); (iii) whether the materials for the house wall and roof were provided by an NGO – tins1=3, (selected 32 observations); (iv) whether the source of a loan was from an NGO (variable srln – selected 1 observation), and (v) whether the household was indicated as a member of more than one NGO (selecting 23 observations). This classification selected 444 observations for the year 2002, of which 49 had been selected as SUP members for the program despite apparent ineligibility.

Exclusion criterion 2 was composed of the following variables: (i) whether the household had government benefits (gprben1=2), which selected 28 observations; (ii) whether main source of income was government benefits, in main source of income, for three primary sources (variables msoi1, msoi2, msoi3), which selected 3, 11 and 7 observations respectively. This classification selected 127 observations, of which 35 had been selected as SUP members for the program.

To create EC3 we used the variable disab1, those women who presented a disability. This selected 48 observations, of which 22 previously had been selected as SUP members. Overall, according to the exclusion criteria, we identified 103 participants who were selected despite being ineligible.

With respect to the inclusion criteria, the household had to meet at least three out of five conditions in order to be considered for the TUP program. They were: IC1: owning less than 10 decimals of land (a tenth of an acre), including homestead; IC2: no male income earner at home; IC3 children of school-age working; IC4: adult women of household selling labor outside homestead; and IC5: household having no productive assets.

With respect to the first inclusion criterion (ownership of less than 10 decimals of land, including for their homestead), we created a dummy variable for whether the household owns self cultivated land, owns land that others cultivate, owns homestead land, or owns land that is un-

cultivated. This criterion selects (as eligible) 4624 out of the 5067 for the year 2002, of which 2279 had been selected for SUP.

For the second inclusion criterion, no male income earner present at home, we first created a dummy variable for the presence of no male income earner at home, as the intersection of males of working age (more than 14 years old) that are not working. There are 66 observations that fulfill this criterion, of which 27 already had been selected as SUP. The second auxiliary variable constructed was a dummy for the presence of no male at home (additional to the previous one, no male earner). This variable selects 1893 observations, of which 1085 had been selected for SUP participation.

For the third inclusion variable, that school-age children present in the household are working, we used questionnaire data to that effect, which selected 740 observations, of which 372 had been selected as an SUP.

For the fourth inclusion criterion, that there are adult women selling labor outside the homestead, we selected those observations for which the main source of income (for the first three primary occupations) were: 5 =daylabor (agriculture), 6=daylabor (non-agriculture), 7=small business/trading, 9=begging, 10=servant, 11=professional. This selected 1627 observations, of which 994 had been already selected as SUP.

For the fifth inclusion criterion that the household had no productive assets, we used the dummy variable "prodasst", which selected 2791 observations, of which 1520 were already SUP members.

Finally, to construct the inclusion criteria, we considered those observations that fulfilled at least three out of the five conditions. According to these data, there were 1727 observations that should have been classified as SUP, of which 641 were not.

According to the exclusion and inclusion criteria, we have created the following groups: SB_1 (selected as SUP, and fulfilling both inclusion and exclusion criteria), composed of 1086 observations; SNB_1 (selected as SUP, not fulfilling the criteria), composed of 1289 observations; SNB_0 (correctly not selected as SUP, criteria not met), composed of 2051 observations and SB_0 (not selected as SUP, but fulfilling criteria), with 641 observations.

Table 1a: Variable description

Income and savings	
Per capita income	Summary variable to the answer of 'Last year employment and income related information - Increased net income in tk' for the TUP member divided by household size
Cash savings (dummy)	Binary variable equal to one if the answer to the question 'Do you have any cash savings?' is yes.
Food Security	
Food availability	What would you say the status of your household is in terms of food availability? Always deficit[1], deficit some times [2], neither deficit nor surplus [3], food surplus [4]
Meals twice a day (dummy)	Binary variable equal to 1 when the answer to the following question is yes: Could your household afford two meals per day most of the time during last year?
Land and Housing	
Total land owned	Total amount of land owned by the household (in tenth of acres)
Own homestead land (dummy)	Binary variable that equals one if the household owns homestead land
Roof made of tin (dummy)	Binary variable that equals one if the material of household's main living room is tin (sign of good quality).
Clothing	
Number of sarees	Number of sarees (female clothing) owned by the TUP member.
Number of lungis	Number of lungis (male clothing) owned by the household head.
Shoes (dummy)	Answer to the question 'Do all household members have shoes/sandals?' yes[1] no[0].
Livestock	
Number of cows/bulls	Number of assets owned, not including program transfers
Number of goats/sheep	Number of assets owned, not including program transfers
Number of ducks/hens	Number of assets owned, not including program transfers
Productive Assets	
Number of fishing nets	Number of assets owned, not including program transfers
Number of big trees	Number of assets owned, not including program transfers
Number of rickshaw/vans	Number of assets owned, not including program transfers
Number of bicycles	Number of assets owned, not including program transfers

Table 1a: Variable description, cont.

Household Durables	
Number of chair/tables	Number of assets owned
Number of beds	Number of assets owned
Number of radio/TVs	Number of assets owned
Number of quilt/blankets	Number of assets owned
Number of tube wells	Number of assets owned
Health	
Health status	Answer to the following question: 'How do you perceive your current health status?' Excellent [5], Very good [4], Good [3], Fair [2], Poor/Bad [1]
Health improvement	Answer to 'How do you consider your health compared to last year?' Much better than one year ago [5]; somewhat better now [4]; about the same [3], somewhat worse [2]; much worse [1].
Female Empowerment	
Ratio of saree to lungi	Ratio of the female clothing to male clothing.
Presence of female children working (dummy)	Presence of female child labor.
Ability of female children to read and write a letter (dummy)	
Years of schooling of female children	
Child Labor	
Presence of child labor (dummy)	Binary variable equal to one if the household declares that there are children under 15 working
Observable Characteristics	
Household size	Size of household
Female working as daylabor (dummy)	Binary variable equal to one if the household female works as a daylabor.
Sex of household head (dummy)	Binary variable equal to one if the household head is a female
Total amount of land owned	In tenth of acres

Table 1b: Summary statistics

	Year	Mean	Std. Dev.	Obs
Income and savings				
Per capita income	2002	2660.30	2258.69	5035
	2005	3980.69	2647.00	5067
Cash savings (dummy)	2002	0.15	0.35	5067
	2005	0.61	0.49	5067
Food security				
Food availability	2002	1.55	0.63	5067
	2005	2.06	0.78	5067
Meals twice a day (dummy)	2002	0.60	0.49	5067
	2005	0.40	0.49	5067
Land and housing				
Total amount of land owned	2002	4.30	14.57	5067
	2005	4.36	15.11	5067
Own homestead land (dummy)	2002	0.54	0.50	5067
	2005	0.53	0.50	5067
Roof made of tin (dummy)	2002	0.50	0.50	5067
	2005	0.78	0.41	5067
Clothing				
Number of sarees	2002	1.81	0.59	5067
	2005	2.21	0.82	5067
Number of lungis	2002	1.75	0.54	3644
	2005	1.59	1.25	5067
Shoes (dummy)	2002	0.62	0.48	5067
	2005	0.90	0.30	5067
Livestock				
Number of cows/bulls	2002	0.11	0.51	5067
	2005	0.94	1.21	5067
Number of goats/sheep	2002	0.11	0.49	5067
	2005	0.34	0.97	5067
Number of ducks/hens	2002	1.15	2.83	5067
	2005	2.53	3.69	5067
Productive assets				
Number of fishing nets	2002	0.00	0.05	5067
	2005	0.15	0.60	5067
Number of big trees	2002	0.89	5.97	5067
	2005	0.61	2.76	5067
Number of rickshaw/vans	2002	0.03	0.27	5067
	2005	0.07	0.28	5067
Number of bicycles	2002	0.01	0.08	5067
	2005	0.02	0.15	5067

Table 1b: Summary statistics, cont.

	Year	Mean	Std. Dev.	Obs
Household durables				
Number of chair/tables	2002	0.37	0.80	5067
	2005	0.65	1.05	5067
Number of beds	2002	0.88	0.73	5067
	2005	1.14	0.76	5067
Number of radio/TVs	2002	0.01	0.12	5067
	2005	0.03	0.18	5067
Number of quilt/blankets	2002	0.03	0.21	5067
	2005	0.16	0.44	5067
Number of tube wells	2002	0.03	0.16	5067
	2005	0.45	0.50	5067
Indicators of health				
Health status	2002	2.32	0.97	5055
	2005	2.50	1.07	5013
Health improvement	2002	2.61	1.10	5055
	2005	2.93	1.06	5013
Indicators of female empowerment				
Ratio of saree to lungi	2002	1.11	0.42	3627
	2005	1.03	0.36	3514
Presence of female children working (dummy)	2002	0.07	0.26	5067
	2005	0.11	0.32	5067
Ability of female children to read and write (dummy)	2002	0.08	0.27	5067
	2005	0.07	0.26	5067
Years of schooling of female children	2002	0.35	0.48	5067
	2005	0.23	0.42	5067
Child labor				
Presence of child labor (dummy)	2002	0.15	0.35	5067
	2005	0.19	0.39	5067

Table 2a: Test of differences in mean characteristics between treatment and control groups in 2002

	SB1-SB0	SUP-NSUP	SNB1-SNB0	SB1-SNB0
Income and savings				
Per capita income	-106.75	-323.75***	-371.40***	-437.55***
	(94.30)	(63.62)	(86.12)	(87.96)
Cash savings (dummy)	-0.01	-0.12***	-0.15***	-0.19***
	(0.01)	(0.01)	(0.01)	(0.01)
Food security				
Food availability	-0.12***	-0.29***	-0.31***	-0.42***
	(0.03)	(0.02)	(0.02)	(0.02)
Meals twice a day (dummy)	-0.15***	-0.18***	-0.18***	-0.22***
	(0.02)	(0.01)	(0.02)	(0.02)
Land and housing				
Total amount of land owned	-0.59**	-3.94***	-4.69***	-5.73***
	(0.21)	(0.41)	(0.62)	(0.66)
Own homestead land (dummy)	-0.08***	-0.14***	-0.12***	-0.25***
	(0.02)	(0.01)	(0.02)	(0.02)
Roof made of tin (dummy)	-0.09***	-0.11***	-0.13***	-0.10***
	(0.02)	(0.01)	(0.02)	(0.02)
Clothing				
Number of sarees	-0.05	-0.15***	-0.16***	-0.23***
	(0.03)	(0.02)	(0.02)	(0.02)
Number of lungis	-0.10*	-0.13***	-0.13***	-0.16***
	(0.05)	(0.02)	(0.02)	(0.03)
Shoes (dummy)	-0.06*	-0.11***	-0.12***	-0.12***
	(0.02)	(0.01)	(0.02)	(0.02)
Livestock				
Number of cows/bulls	-0.03***	-0.15***	-0.17***	-0.22***
	(0.01)	(0.01)	(0.02)	(0.02)
Number of goats/sheep	-0.01	-0.04**	-0.03	-0.09***
	(0.02)	(0.01)	(0.02)	(0.02)
Number of ducks/hens	-0.05	-0.57***	-0.65***	-0.96***
	(0.09)	(0.08)	(0.11)	(0.12)
Productive assets				
Number of fishing nets	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Number of big trees	-0.22	-0.74***	-0.85***	-1.04***
	(0.11)	(0.17)	(0.25)	(0.27)
Number of rickshaw/vans	0.01	-0.03***	-0.04***	-0.04***
	(0.01)	(0.01)	(0.01)	(0.01)
Number of bicycles	0.00	-0.01***	-0.01**	-0.01**
	(0.00)	(0.00)	(0.00)	(0.00)
Number of observations	1727	5067	3340	3137

(1) Standard errors in parentheses
(2) Significance levels are denoted as $***p < 0.01; **p < 0.05; *p < 0.1$

Table 2a: Test of differences in mean characteristics between treatment and control groups in 2002, cont.

	SB1-SB0	SUP-NSUP	SNB1-SNB0	SB1-SNB0
Household durables				
Number of chair/tables	-0.12***	-0.31***	-0.33***	-0.45***
	(0.03)	(0.02)	(0.03)	(0.03)
Number of beds	-0.07*	-0.26***	-0.30***	-0.38***
	(0.03)	(0.02)	(0.03)	(0.03)
Number of radio/TVs	0.00	-0.01**	-0.01**	-0.02***
	(0.00)	(0.00)	(0.00)	(0.01)
Number of quilt/blankets	0.00	-0.03***	-0.04***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)
Number of tube wells	-0.01	-0.02***	-0.03***	-0.02***
	(0.01)	(0.00)	(0.01)	(0.01)
Indicators of health				
Health status	0.03	-0.02	0.02	-0.17***
	(0.05)	(0.03)	(0.03)	(0.04)
Health improvement	0.12*	0.02	0.04	-0.14***
	(0.05)	(0.03)	(0.04)	(0.04)
Indicators of female empowerment				
Ratio of saree to lungi	0.03	-0.01	-0.01	-0.02
	(0.04)	(0.01)	(0.02)	(0.02)
Presence of female children working (dummy)	-0.02	0.02*	0.01	0.09***
	(0.02)	(0.01)	(0.01)	(0.01)
Ability of female children to read and write (dummy)	-0.01	-0.02**	-0.03**	-0.03**
	(0.01)	(0.01)	(0.01)	(0.01)
Years of schooling of female children	0.04	0.00	0.02	-0.07***
	(0.03)	(0.02)	(0.02)	(0.02)
Child labor				
Presence of child labor (dummy)	-0.02	0.02*	-0.02*	0.17***
	(0.02)	(0.01)	(0.01)	(0.01)
Number of observations	1727	5067	3340	3137

(1) Standard errors in parentheses
(2) Significance levels are denoted as $***p < 0.01; **p < 0.05; *p < 0.1$

Table 2b: Mean values of outcome variables in 2002 for different treatment groups

	SB1	SUP	SNB1
Income and savings			
Per capita income	2452.63	2488.62	2518.78
	(1896.60)	(2078.63)	(2219.92)
Cash savings (dummy)	0.06	0.08	0.10
	(0.01)	(0.01)	(0.01)
Food security			
Food availability	1.33	1.39	1.44
	(0.02)	(0.01)	(0.02)
Meals twice a day (dummy)	0.48	0.51	0.53
	(0.02)	(0.01)	(0.01)
Land and housing			
Total land owned	1.64	2.21	2.68
	(0.12)	(0.11)	(0.18)
Own homestead land (dummy)	0.40	0.47	0.53
	(0.01)	(0.01)	(0.01)
Roof made of tin (dummy)	0.45	0.44	0.42
	(0.02)	(0.01)	(0.01)
Clothing			
Number of sarees	1.70	1.73	1.76
	(0.02)	(0.01)	(0.01)
Number of lungis	1.65	1.68	1.68
	(0.03)	(0.01)	(0.02)
Shoes (dummy)	0.56	0.57	0.57
	(0.02)	(0.01)	(0.01)
Livestock			
Number of cows/bulls	0.01	0.04	0.06
	(0.00)	(0.01)	(0.01)
Number of goats/sheep	0.06	0.09	0.12
	(0.01)	(0.01)	(0.02)
Number of ducks/hens	0.67	0.84	0.99
	(0.06)	(0.04)	(0.06)
Productive assets			
Number of fishing nets	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Number of big trees	0.40	0.50	0.58
	(0.06)	(0.06)	(0.09)
Number of rickshaw/vans	0.01	0.02	0.02
	(0.01)	(0.01)	(0.00)
Number of bicycles	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Number of observations	1086	2375	1289

Standard errors in parentheses

Table 2b: Mean values of outcome variables in 2002 for different treatment groups, cont.

	SB1	SUP	SNB1
Household durables			
Number of chair/tables	0.14	0.21	0.26
	(0.01)	(0.01)	(0.02)
Number of beds	0.70	0.74	0.78
	(0.02)	(0.01)	(0.02)
Number of radio/TVs	0.00	0.01	0.01
	(0.00)	(0.00)	(0.00)
Number of quilt/blankets	0.01	0.02	0.02
	(0.00)	(0.00)	(0.00)
Number of tube wells	0.02	0.01	0.01
	(0.00)	(0.00)	(0.00)
Indicators of health			
Health status	2.21	2.31	2.40
	(0.03)	(0.02)	(0.03)
Health improvement	2.52	2.62	2.70
	(0.03)	(0.02)	(0.03)
Indicators of female empowerment			
Ratio of saree to lungi	1.10	1.11	1.11
	(0.02)	(0.01)	(0.01)
Presence of female children working (dummy)	0.13	0.08	0.04
	(0.01)	(0.01)	(0.01)
Ability of female children to read and write (dummy)	0.07	0.07	0.07
	(0.01)	(0.01)	(0.01)
Years of schooling of female children	0.30	0.35	0.39
	(0.02)	(0.01)	(0.02)
Child labor			
Presence of child labor (dummy)	0.26	0.16	0.07
	(0.01)	(0.01)	(0.01)
Number of observations	1086	2375	1289

Standard errors in parentheses

Table 3a: Effects of the TUP Program on Treatment Group 'Selected Ultra Poor' (SUP)

	DID	**DIDM**	**MB-IPW**	**K-V**
	(1)	(2)	(3)	(4)
Income and savings				
Per capita income	1126.69	1067.81	932.58	1377.06
	(110.72)***	(93.1)***	[678.86, 1175.82]***	(968.62)
Cash savings (dummy)	0.83	0.81	0.81	
	(0.01)***	(0.01)***	[0.79, 0.83]***	
Food security				
Food availability	0.66	0.65	0.66	0.99
	(0.03)***	(0.03)***	[0.62, 0.71]***	(0.27)***
Meals twice a day (dummy)	0.37	0.37	0.37	
	(0.02)***	(0.02)***	[0.34, 0.40]***	
Land and housing				
Total amount of land owned	1.41	-0.10	-0.07	10.94
	(0.52)**	(0.42)	[-0.47, 0.28]	(13.14)
Own homestead land (dummy)	0.09	0.05	0.05	
	(0.01)***	(0.02)***	[0.02, 0.08]***	
Roof made of tin (dummy)	0.13	0.13	0.13	
	(0.01)***	(0.02)***	[0.11, 0.16]***	
Clothing				
Number of sarees	0.29	0.28	0.28	0.22
	(0.03)***	(0.03)***	[0.23, 0.32]***	(0.16)
Number of lungis	0.06	0.22	0.20	0.70
	(0.04)	(0.04)***	[0.15, 0.25]***	(0.38)*
Shoes (dummy)	0.15	0.14	0.14	
	(0.01)***	(0.02)***	[0.11, 0.17]***	
Livestock				
Number of cows/bulls	1.74	1.71	1.71	1.31
	(0.05)***	(0.03)***	[1.67, 1.75]***	(0.41)***
Number of goats/sheep	0.46	0.45	0.45	0.36
	(0.08)***	(0.03)***	[0.39, 0.50]***	(0.29)
Number of ducks/hens	0.64	0.56	0.56	0.34
	(0.12)***	(0.13)***	[0.39, 0.73]***	(1.30)
Productive assets				
Number of fishing nets	0.03	0.05	0.06	-0.02
	(0.01)*	(0.02)***	[0.03, 0.09]***	(0.24)
Number of big trees	0.55	0.24	0.25	-1.14
	(0.22)**	(0.18)	[0.09, 0.41]***	(5.00)
Number of rickshaw/vans	0.04	0.04	0.04	0.06
	(0.01)***	(0.01)***	[0.02, 0.06]***	(0.08)
Number of bicycles	0.01	0.01	0.01	0.00
	(0.01)	(0.00)	[0.00, 0.01]	(0.04)
Number of observations	9708	4854	4854	4854
Kleibergen-Paap Wald F				35.43
Likelihood Ratio Test of Heteroskedasticity in the probit model for Selection into the TUP Program				30.89
p-value for the LR test				0.00

(1) Robust Standard errors in parentheses are clustered at the village level. There are 24 clusters.
(2) Significance levels are denoted as $***p < 0.01; **p < 0.05; *p < 0.1$
(3) Column (3) presents 95 percent confidence intervals calculated by the bootstrap percentile method, using 250 replications.
(4) The number of observations for the outcome 'Number of lungis' is 8343 in column (1) and is 3489 in columns (2) to (4).
(5) Additional explanations are provided at the end of the table.

Table 3a: Effects of the TUP Program on Treatment Group 'Selected Ultra Poor' (SUP) (cont.)

	DID	DIDM	MB-IPW	K-V
	(1)	(2)	(3)	(4)
Household durables				
Number of chair/tables	0.10	0.11	0.12	0.08
	(0.03)***	(0.03)***	[0.07, 0.17]***	(0.40)
Number of beds	0.17	0.16	0.16	0.21
	(0.02)***	(0.02)***	[0.12, 0.20]***	(0.21)
Number of radio/TVs	0.00	0.01	0.01	0.14
	(0.01)	(0.01)	[0.00, 0.02]	(0.05)***
Number of quilt/blankets	0.16	0.16	0.16	0.44
	(0.02)***	(0.01)***	[0.14, 0.18]***	(0.11)***
Number of tube wells	0.09	0.14	0.15	-0.53
	(0.02)***	(0.02)***	[0.12, 0.17]***	(0.13)***
Indicators of health				
Health status	0.07	0.06	0.06	-0.16
	(0.04)	(0.04)	[-0.02, 0.12]	(0.47)
Health improvement	0.13	0.11	0.10	-0.03
	(0.06)**	(0.05)***	[0.04, 0.17]***	(0.28)
Indicators of female empowerment				
Ratio of saree to lungi	0.03	0.03	0.02	-0.21
	(0.02)	(0.02)	[0.00, 0.04]	(0.13)
Presence of female children working (dummy)	0.00	0.01	0.01	
	(0.01)	(0.01)	[-0.01, 0.02]	
Ability of female children to read and write (dummy)	0.00	0.01	0.01	
	(0.01)	(0.01)	[-0.01, 0.03]	
Years of schooling of female children	0.01	0.02	0.02	-0.17
	(0.01)	(0.01)	[-0.01, 0.03]	(0.14)
Child labor				
Presence of child labor (dummy)	0.00	0.03	0.03	
	(0.01)	(0.01)*	[0.00, 0.05]	
Number of observations	9708	4854	4854	4854
Kleibergen-Paap Wald F				35.43
Likelihood Ratio Test of Heteroskedasticity in the probit model for Selection into the TUP Program				30.89
p-value for the LR test				0.00

(1) Robust Standard errors in parentheses are clustered at the village level. There are 24 clusters.
(2) Significance levels are denoted as $***p < 0.01; **p < 0.05; *p < 0.1$
(3) Column (3) presents 95 percent confidence intervals calculated by the bootstrap percentile method, using 250 replications.
(4) Estimates presented in columns (2) to (4) use household fixed effects.
(5) Estimates in Column (1) come from the augmented Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions; in column (2) from the difference-in-difference matching estimator due to Heckman et al. (1998); in column (3) from the minimum bias inverse probability weighted estimator due to Millimet and Tchernis (2012); in column (4) from the Klein-Vella (2009) identification through heteroskedasticity method.
(6) The number of observations for the Health outcomes is 9647 in column (1) and is 4795 in columns (2) to (4); it is 6845 for the ratio of saree to lungi in column (1) and is 3230 in the other columns.
(7) The variables included in the selection equation are: household size, dummy if daylabor activities, total amount of land owned, sex of household head (female=1) and the five inclusion criteria (IC) established by BRAC (IC1: less than ten decimals of land including homestead; IC2: no male income earner at home; IC3: presence of child labor; IC4: female working outside the household; IC5: lack of productive assets).

Table 3b: Effects of the TUP Program on the poorest of the poor, SB1 treatment group

	DID	**DIDM**	**MB-IPW**	**K-V**
	(1)	(2)	(3)	(4)
Income and savings				
Per capita income	1313.12	1325.01	1328.36	1779.59
	(159.82)***	(168.15)***	[684.40, 1701.99]***	(910.96)*
Cash savings (dummy)	0.86	0.80	0.80	
	(0.02)***	(0.02)***	[0.76, 0.84]***	
Food security				
Food availability	0.67	0.67	0.69	0.83
	(0.04)***	(0.05)***	[0.60, 0.78]***	(0.38)**
Meals twice a day (dummy)	0.36	0.35	0.36	
	(0.04)***	(0.03)***	[0.29, 0.43]***	
Land and housing				
Total amount of land owned	0.04	-0.06	-0.07	2.45
	(0.31)	(0.37)	[-0.73, 0.59]	(4.41)
Own homestead land (dummy)	0.10	0.07	0.07	
	(0.03)***	(0.03)***	[0.02, 0.12]***	
Roof made of tin (dummy)	0.13	0.13	0.13	
	(0.02)***	(0.03)***	[0.08, 0.18]***	
Clothing				
Number of sarees	0.27	0.26	0.27	0.69
	(0.04)***	(0.04)***	[0.18, 0.34]***	(0.35)**
Number of lungis	0.08	0.29	0.16	-1.28
	(0.07)	(0.12)	[0.01, 0.29]***	(0.80)
Shoes (dummy)	0.10	0.10	0.09	
	(0.02)***	(0.03)***	[0.05, 0.14]***	
Livestock				
Number of cows/bulls	1.69	1.68	1.69	1.57
	(0.07)***	(0.04)***	[1.62, 1.75]***	(0.31)***
Number of goats/sheep	0.45	0.46	0.45	0.59
	(0.08)***	(0.04)***	[0.39, 0.52]***	(0.34)*
Number of ducks/hens	0.60	0.67	0.62	0.13
	(0.15)***	(0.17)***	[0.33, 0.91]***	(1.37)
Productive assets				
Number of fishing nets	0.02	0.02	0.03	-0.06
	(0.02)	(0.02)	[-0.01, 0.06]	(0.12)
Number of big trees	0.27	0.25	0.29	1.63
	(0.18)	(0.16)	[0.04, 0.51]***	(1.23)
Number of rickshaw/vans	0.01	0.02	0.01	0.08
	(0.01)	(0.01)	[-0.01, 0.04]	(0.10)
Number of bicycles	-0.00	-0.00	0.00	-0.04
	(0.01)	(0.01)	[-0.01, 0.01]	(0.05)
Number of observations	3314	1657	1657	1657
Kleibergen-Paap Wald F				33.98
Likelihood Ratio Test of Heteroskedasticity in the probit model for Selection into the TUP Program				18.45
p-value for the LR test				0.03

(1) Robust Standard errors in parentheses are clustered at the village level. There are 23 clusters.
(2) Significance levels are denoted as $***p < 0.01; **p < 0.05; *p < 0.1$
(3) Column (3) presents 95 percent confidence intervals calculated by the bootstrap percentile method, using 250 replications.
(4) The outcome 'Number of lungis' has 2217 observations in column (1) and has 560 observations in columns (2) to (4).
(5) Additional explanations are provided at the end of the table.

Table 3b: Effects of the TUP Program on the poorest of the poor, SB1 treatment group (cont.)

	DID	DIDM	MB-IPW	K-V
	(1)	(2)	(3)	(4)
Household durables				
Number of chair/tables	0.13	0.14	0.15	0.01
	(0.04)***	(0.05)***	[0.08, 0.23]***	(0.42)
Number of beds	0.16	0.17	0.17	0.59
	(0.04)***	(0.04)***	[0.10, 0.24]***	(0.34)*
Number of radio/TVs	0.01	0.01	0.01	-0.09
	(0.01)	(0.01)	[0.00, 0.02]	(0.06)*
Number of quilt/blankets	0.21	0.22	0.21	0.11
	(0.02)***	(0.02)***	[0.18, 0.25]***	(0.19)
Number of tube wells	0.15	0.16	0.16	-0.08
	(0.03)***	(0.02)***	[0.12, 0.29]***	(0.23)
Indicators of health				
Health status	0.11	0.08	0.09	0.07
	(0.05)**	(0.07)	[-0.04, 0.21]	(0.60)
Health improvement	0.11	0.09	0.09	0.92
	(0.06)*	(0.08)	[-0.04, 0.20]	(0.77)
Indicators of female empowerment				
Ratio of saree to lungi	-0.01	-0.03	-0.02	0.56
	(0.05)	(0.06)	[-0.26, 0.16]	(0.48)
Presence of female children working (dummy)	0.02	0.01	0.01	
	(0.02)	(0.02)	[-0.03, 0.03]	
Ability of female children to read and write (dummy)	-0.01	-0.00	0.00	
	(0.01)	(0.02)	[-0.03, 0.03]	
Years of schooling of female children	-0.02	-0.02	-0.01	0.13
	(0.02)	(0.03)	[-0.05, 0.03]	(0.19)
Child labor				
Presence of child labor (dummy)	0.02	0.01	-0.01	
	(0.03)	(0.03)	[-0.04, 0.03]	
Number of observations	3314	1657	1657	1657
Kleibergen-Paap Wald F				33.98
Likelihood Ratio Test of Heteroskedasticity in the probit model for Selection into the TUP Program				18.45
p-value for the LR test				0.03

(1) Robust Standard errors in parentheses are clustered at the village level. There are 23 clusters.
(2) Significance levels are denoted as $***p < 0.01; **p < 0.05; *p < 0.1$
(3) Column (3) presents 95 percent confidence intervals calculated by the bootstrap percentile method, using 250 replications.
(4) Estimates presented in columns (2) to (4) use household fixed effects.
(5) Estimates in Column (1) come from the augmented Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions; in column (2) from the difference-in-difference matching estimator due to Heckman et al. (1998); in column (3) from the minimum bias inverse probability weighted estimator due to Millimet and Tchernis (2012); in column (4) from the Klein-Vella (2009) identification through heteroskedasticity method.
(6) The number of observations for the Health outcomes is 3301 in column (1) and is 1644 in columns (2) to (4); for the ratio of saree to lungi is 1002 in column (1) and is 373 in the other columns.
(7) The variables included in the selection equation are: household size, dummy if daylabor activities, total amount of land owned, sex of household head (female=1) and the five inclusion criteria (IC) established by BRAC (IC1: less than ten decimals of land including homestead; IC2: no male income earner at home; IC3: presence of child labor; IC4: female working outside the household; IC5: lack of productive assets).

Table 4a: Normalized Program Effects on Treatment Group 'Selected Ultra Poor' (SUP)

	DID	DIDM	MB-IPW	K-V
	(1)	(2)	(3)	(4)
Income and savings				
Per capita income	0.45	0.43	0.37	0.55
Cash savings (dummy)	10.21	9.96	9.93	
Food security				
Food availability	0.47	0.47	0.48	0.71
Meals twice a day (dummy)	0.73	0.73	0.73	
Land and housing				
Total amount of land owned	0.64	-0.05	-0.03	4.96
Own homestead land (dummy)	0.19	0.11	0.10	
Roof made of tin (dummy)	0.30	0.30	0.30	
Clothing				
Number of sarees	0.17	0.16	0.16	0.13
Number of lungis	0.04	0.13	0.12	0.42
Shoes (dummy)	0.27	0.25	0.24	
Livestock				
Number of cows/bulls	49.15	48.31	48.36	37.01
Number of goats/sheep	4.94	4.83	4.78	3.87
Number of ducks/hens	0.76	0.67	0.67	0.40
Productive assets				
Number of fishing nets	11.86	19.76	21.74	-7.91
Number of big trees	1.11	0.48	0.51	-2.30
Number of rickshaw/vans	2.31	2.31	2.31	3.47
Number of bicycles	4.74	4.74	3.32	0.00

Table 4a: Normalized Program Effects on Treatment Group 'Selected Ultra Poor' (SUP) (cont.)

	DID	DIDM	MB-IPW	K-V
	(1)	(2)	(3)	(4)
Household durables				
Number of chair/tables	0.49	0.54	0.59	0.39
Number of beds	0.23	0.22	0.22	0.28
Number of radio/TVs	0.00	1.25	0.75	17.50
Number of quilt/blankets	10.00	10.00	10.06	27.50
Number of tube wells	6.12	9.52	9.93	-36.05
Indicators of health				
Health status	0.03	0.03	0.03	-0.07
Health improvement	0.05	0.04	0.04	-0.01
Indicators of female empowerment				
Ratio of saree to lungi	0.03	0.03	0.02	-0.19
Presence of female children working (dummy)	0.00	0.12	0.10	
Ability of female children to read and write (dummy)	0.00	0.14	0.14	
Years of schooling of female children	0.06	0.06	0.04	-0.37
Child labor				
Presence of child labor (dummy)	0.00	0.19	0.17	

Table 4b: Normalized Program effects on the poorest of the poor SB1 treatment group

	DID	DIDM	MB-IPW	K-V
	(1)	(2)	(3)	(4)
Income and savings				
Per capita income	0.54	0.54	0.54	0.73
Cash savings (dummy)	15.58	14.49	14.49	
Food security				
Food availability	0.50	0.50	0.52	0.62
Meals twice a day (dummy)	0.75	0.73	0.75	
Land and housing				
Total amount of land owned	0.02	-0.04	-0.04	1.49
Own homestead land (dummy)	0.25	0.18	0.18	
Roof made of tin (dummy)	0.29	0.29	0.29	
Clothing				
Number of sarees	0.16	0.15	0.16	0.41
Number of lungis	0.05	0.18	0.10	-0.77
Shoes (dummy)	0.18	0.18	0.16	
Livestock				
Number of cows/bulls	167.33	166.34	167.33	155.45
Number of goats/sheep	7.64	7.81	7.64	10.02
Number of ducks/hens	0.90	1.00	0.93	0.19
Productive assets				
Number of fishing nets	n/a	n/a	n/a	n/a
Number of big trees	0.68	0.63	0.73	4.10
Number of rickshaw/vans	0.72	1.45	0.72	5.80
Number of bicycles	0.00	0.00	0.00	-21.74

Table 4b: Normalized Program effects on the poorest of the poor SB1 treatment group (cont.)

	DID	DIDM	MB-IPW	K-V
	(1)	(2)	(3)	(4)
Household durables				
Number of chair/tables	0.95	1.02	1.09	0.07
Number of beds	0.23	0.24	0.24	0.84
Number of radio/TVs	2.17	2.17	2.17	-19.57
Number of quilt/blankets	19.09	20.00	19.09	10.00
Number of tube wells	8.15	8.70	8.70	-4.35
Indicators of health				
Health status	0.05	0.04	0.04	0.03
Health improvement	0.04	0.04	0.04	0.37
Indicators of female empowerment				
Ratio of saree to lungi	-0.01	-0.03	-0.02	0.51
Presence of female children working (dummy)	0.16	0.08	0.08	
Ability of female children to read and write (dummy)	-0.15	0.00	0.00	
Years of schooling of female children	-0.07	-0.07	-0.03	0.43
Child labor				
Presence of child labor (dummy)	0.08	0.04	-0.04	

Table A1: Probit Regressions for Selection into the TUP Program

	SUP/NSUP	SB1/SB0
Household size	0.03	0.04
	(0.01)**	(0.02)
Dummy if daylabor activities	0.14	0.04
	(0.04)***	(0.07)
Total amount of land owned	-0.02	-0.01
	(0.00)***	(0.01)
Sex of household head (female=1)	0.04	0.23
	(0.20)	(0.26)
IC1: Less than ten decimals of land including homestead	0.27	0.29
	(0.11)**	(0.30)
IC2: No male income earner at home	-0.07	-0.36
	(0.16)	(0.23)
IC3: Presence of child labor	0.05	0.02
	(0.06)	(0.09)
IC4: Female working outside the household	0.57	0.52
	(0.11)***	(0.16)***
IC5: Lack of productive assets	0.34	0.21
	(0.04)***	(0.07)***
Constant	-0.81	-0.49
	(0.13)***	(0.37)
Number of observations	4854	1657

(1) Standard errors in parentheses. Significance levels are denoted as
$***p < 0.01; **p < 0.05; *p < 0.1$
(2) Covariates are 2002 pre intervention characteristics.
(3) 'IC' stands for 'inclusion criterion'. IC1 to IC5 are the five BRAC criteria for inclusion into the TUP program