Essays in Development Economics

by Virginia Robano

B.A. in Economics, June 2000, Universidad de la República, Uruguay
M.A. in Economics, May 2004, University of Maryland at College Park

A dissertation submitted to

The Faculty of The Columbian College of Arts and Sciences of The George Washington University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

May 20, 2012

Dissertation directed by

Stephen C. Smith
Professor of Economics and International Affairs

M. Shahe Emran
Coordinator, Tax Force Initiative for Policy Dialogue, Columbia University

The Columbian College of Arts and Sciences of The George Washington University

certifies that Virginia Robano has passed the Final Examination for the degree of Doctor of Philosophy as of March 26, 2012. This is the final and approved form of the dissertation.

Essays in Development Economics

Virginia Robano

Dissertation Research Committee:

Stephen C. Smith, Professor of Economics and International Affairs, Dissertation Co-Director

M. Shahe Emran, Coordinator, Tax Task Force, Iniciative for Policy Dialogue, Columbia University, Dissertation Co-Director

Paul E. Carrillo, Assistant Professor of Economics, Committee Member

James E. Foster, Professor of Economics and International Affairs, Committee Member

Dedication

To Christian, with all my love and gratitude

Acknowledgements

I am deeply grateful to my advisor, Professor Stephen C. Smith for his patient and thoughtful advice during my graduate studies. I acknowledge his support and valuable insights on my research questions and on the dissertation writing process. My hearty thanks also to Professor M. Shahe Emran for his generosity in sharing his knowledge and from whom I learned to be rigorous in my analysis. I have been especially fortunate to have them as my advisors.

I acknowledge financial support from the Institute for International Economic Policy of the Elliot School of International Affairs.

Abstract of Dissertation

Essays in Development Economics

In this dissertation, I write three essays in Development Economics. In essay 1, I provide robust evidence on the effects of BRAC's Targeting the Ultra-Poor Program in Bangladesh (TUP). In addition to BRAC's own classification, I exploit type-1 errors in assignment (leaving eligible participants outside the program) to create an alternative treatment-comparison pair. This allows me to estimate the program effects on the target group not contaminated by mistargeting. The panel structure of the data enables me to eliminate the time-invariant unobserved heterogeneity at household level (i.e. innate entrepreneurial ability and risk preference) by using household fixed effects. Moreover, I carefully address selection on unobservables by implementing the heteroskedasticity based identification technique developed by Klein and Vella (2009) and two additional recently developed matching estimators. The results show that program participation had significant positive effects on food security, clothing, shoes, livestock and cash savings, but there is weak or no evidence of a significant impact on a number of household durables and assets, and on indicators of health and women's empowerment. Furthermore, when one takes into account the differences in initial conditions, the effects of the TUP program on the alternative treatment-comparison group are much larger (as measured by the program effect normalized by the initial mean value of an outcome).

In essay 2, I examine the design of social program targeting mechanisms and corresponding evaluation frameworks when poverty has multiple dimensions. I examine the sensitivity of the group identified as poor to the type and number of screens used, connecting multidimensional targeting to evaluation criteria. I apply this methodology to an assessment of Phase I of BRAC's Targeting the Ultra-Poor Program in Bangladesh, comparing characteristics of selected participants using alternative participation criteria, and compare program outcome measures associated with different targeting mechanisms. I relate this analysis to the multidimensional poverty measurement technology developed by Alkire and Foster (2011), and show that the results represent an approach to conducting multidimensional poverty eval-

uation that parallels their framework. The approach offers an alternative way to examine the heterogeneity of the program impact across poverty levels. Findings of the application confirm that the BRAC TUP program has a significantly larger impact on health-related outcomes for the less poor households. On the other hand, I find that the poorest households ($k = 5$) have a larger impact than the less poor households ($k = 1$) on the net income increase variable (3481.95 Bangladeshi taka vs. 1759.97), on the probability of having a roof of good quality (0.24 vs. 0.13), on food availability (1.02 vs 0.67), on the probability of having meals twice a day (0.58 vs 0.37) and on the probability of owning shoes (0.27 vs. 0.15). Had the program concentrated on the poorest households, the average program impacts would have been larger in magnitude.

In essay 3, I employ unconditional quantile-decomposition methods to conduct a careful accounting exercise analyzing the gender wage gap in the urban sector of twelve Latin American countries (Argentina, Bolivia, Brazil, Colombia, Costa Rica, Chile, Honduras, Mexico, Paraguay, Peru, Uruguay and Venezuela). Unconditional quantiles allow the researcher to compute marginal effects. By contrast, conditional quantiles only calculate wage effects for a subgroup with specific combination of years of education and experience. The data come from harmonized household surveys that contain information on employment status, wages and household's demographic characteristics. I control for the effect of individual characteristics (education and experience) and decompose the gender wage gap into an explained (due to differences in the endowment levels by gender) and an unexplained component (when the same endowment level is paid differently according to which gender the individual belongs) using a recent econometric technique developed by Firpo, Fortin and Lemieux (2009). This technique provides the user with asymptotic results instead of relying on simulations procedures as done by the method developed by Machado and Mata (2005) -based on conditional quantiles and resampling- which results in a loss of efficiency for a small number of calculations or in a computation burden with a large number. More importantly, by reweighting the data, a counterfactual distribution is generated and it is thus possible to calculate unconditional quantiles and therefore account for the contribution of each explanatory variable into the wage gap. I find that the wage gap is larger at the extremes of the distribution, which suggests the presence of *sticky floors* (defined as a gender wage gap favorable to males which is larger at the tenth percentile than at median levels) and *glass ceilings* (defined as a gender wage gap favorable to males which is larger at the ninetieth percentile than at median levels). The former are more frequent. Second, I find that the magnitude of the sticky floors is generally larger than that of the glass ceilings. Third, working women are more educated than working men all along the wage distribution, which partially hides the existent and 'unexplained' pay difference. Thus, my estimates provide a lower-bound for the true gender wage gap. Fourth, the size of the gender wage gap at the bottom is highly correlated with measures of economic development, per capita GDP and income inequality (Gini). With only twelve observations, I am just suggesting some correlations. Nevertheless, I do find suggestive evidence that poorer countries and countries with higher levels of income inequality have higher unexplained gender wage gap differentials at the tenth percentile of the wage distribution. On the other hand, the unexplained portion of the wage gap at the top of the wage distribution (90^{th} percentile) is larger in richer countries and in countries where income distribution is more even, although not statistically significant on the latter.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Much attention has been given in the last decades to evaluating the consequences of program participation in order to assess their effectiveness both in terms of effects, (do they achieve a significant and substantial impact in terms of outcomes?) as well as of targeting (do they reach the desired subgroup of the population?). My dissertation is an original contribution to knowledge in these two aspects.

In chapter 2, I carefully assess a specific anti-poverty program and evaluate its effects in terms of outcomes. To provide credible evidence on the treatment effect of participation in BRAC's TUP program, I address the possible biases due to omitted heterogeneity. I find evidence of errors in assignment and thus construct an alternative treatment-comparison group that represents the target households (the poorest of the poor, henceforth SB1) more faithfully, while also identifying an appropriate comparison group (called SB0). An advantage of panel data is that I can credibly address the time invariant household level heterogeneity (assumed to be additive) by using household fixed effects. This is important for understanding the effects of the TUP or any other microfinance program, as two of the salient omitted variables in this context are innate entrepreneurial ability and risk preferences, both of which are time invariant. The main results of this chapter thus utilize household fixed effects. To estimate the effects of TUP participation on household outcomes, I use a recent econometric technique developed by Klein and Vella (2009). I also report estimates from two recently developed matching estimators. The results from the 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 (2009) and the Difference-in-Difference Matching estimator (DIDM) due to Heckman *et al.* (1998). The treatment effects from the DIDM approach can be biased if the counterfactual trend in the treatment group is different from the actual trend found among the comparison households. The MB-IPW estimator 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 information available for the TUP program. In the absence of credible exclusion restrictions, I 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))[1]. In particular, I implement the heteroskedasticity based instrumental variables approach due to Klein and Vella (2009). I evaluate the causal effects of program participation on a broad set of outcomes. I find significantly positive effects on cash savings, housing quality (homestead land and roof made of tin), food security, clothing, livestock and household durables. There is no evidence of a significant effect on total land ownership, number of duck/hens, productive assets, subjective health outcomes, child labor and female empowerment[2]. I also find that there are some differences in the program effects between the two treatment groups: SUP (selected ultra poor, BRAC's classification) and

[1]For recent applications of heteroskedasticity based identification, see Rigobon (2003), Rigobon and Rodrik (2005), Farre *et al.* (2010), Schroeder (2010), Millimet and Tchernis (2009), Emran and Hou (forthcoming), Emran and Shilpi (forthcoming), and Emran and Sun (2011).

[2]I use a limited set of indicators for this outcome. The conclusion that there is no program impact on women's empowerment thus should be interpreted with appropriate caveat.

SB1. For example, while the evidence that participation increases net income is very strong for the SUP group, it is weaker for the SB1 group. Furthermore, the households in both treatment groups invest in productive assets such as livestock, only the SUP households acquire luxury goods such as radio/TV. In fact, the program effect on radio/TV is negative for the SB1 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 SB1 treatment group compared to the BRAC treatment group SUP.

In chapter 3 I delineate a set of general guidelines to select the criteria for participation into a poverty program. My contribution is connecting the Alkire and Foster (2007) multidimensional poverty methodology to the program evaluation literature. First, I propose to examine the sensitivity of the group identified as the target for participation to the type and number of screens used. I use Alkire and Foster's (2007) Headcount Ratio (H) and Adjusted Headcount Ratio (M_0) to evaluate the performance of the targeting criteria. Then I show how to link the Alkire and Foster (2007, 2011) multidimensional techniques to program evaluation. In the process, I assess the heterogeneity of effects along the different degrees of poverty identified (where poverty would go from the intersection case of all deprivation indicators present -extreme poverty- to the less stringent case of only one criterion present). As an example, I apply the methodology proposed to the BRAC CFPR/TUP dataset. Using the poverty measure M_0 I find that poverty decreased from 28 percent in 2002 to 23 percent in 2005. While for the treated households the decrease was from 36 percent to 25 percent, the M_0 measure for the control group remained at their initial 21 percent level. When analyzing the heterogeneity of the impact, I find a much larger dispersion on the outcomes of the poorest households. For example, the impact on the probability of having cash savings varies from 0.75 to 0.95 (the point estimate is 0.85 percent). For the less poor it varies from 0.81 to 0.86. Nevertheless, the poorest households present a larger program impact than the less poor households on the net income increase variable (3481.95 Bangladeshi Taka vs. 1759.97), on the probability of having a roof of good quality (0.24 vs. 0.13), on food availability (1.02 vs. 0.67), on the probability of having meals twice a day (0.58 vs. 0.37) and on the probability of owning shoes (0.27 vs. 0.15). Had the program concentrated on the poorest households, the average program impacts would have been larger in magnitude.

The last chapter conducts a careful accounting exercise, where I decompose the gender wage gap using the counterfactual female wages[3]. I study the labor markets of twelve Latin American countries and measure male and female hourly wages. Using a recent econometric technique developed by Firpo, Fortin and Lemieux (2009), I decompose the gender wage gap (GWG) and assess its composition effect (the difference in human capital endowments) and the wage structure effect. I find that males earn more than females in all countries and at all points of the wage distribution. Furthermore, the difference is accentuated at the extremes of the distribution, termed *sticky floors* (left tail) and *glass ceilings* (right tail). I find a high correlation between the presence of sticky floors and overall measures of economic development (per capita GDP and income inequality (Gini). Poorer countries and countries with higher levels of income inequality have higher unexplained gender wage gap differentials at the tenth percentile of the wage distribution. On the other hand, the unexplained portion of the wage gap at the top of the wage distribution (90^{th} percentile) is larger in richer countries and in countries where income distribution is more even, although not statistically significant on the latter.

[3]Fortin, Lemieux and Firpo (2010) make the analogy of the unexplained component of the GWG to the average treatment on the treated in the program evaluation literature.

Chapter 2

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

jointly with M. Shahe Emran and Stephen C. Smith

2.1 Introduction

It is widely appreciated, both by practitioners and academics alike, 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), WDR (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). The possible complementarity among the different dimensions can potentially result in multiple

[1]Although there is a growing consensus that 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 s/he 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 s/he 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. While in chapter three we analyze BRAC's identification strategy, in this chapter we do not focus on how to define or identify the ultra-poor, taking 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)

mutually reinforcing poverty traps, thus making ultra-poverty an especially difficult problem to address.

The experience of last few decades suggests that while the poverty programs of NGOs including microcredit programs, have in general been successful in reaching the moderate poor (those households below the poverty line, but relatively close to it), the poorest of the poor are more often inadequately served or completely bypassed by such programs[2]. This appreciation led to the development and implementation of innovative anti-poverty programs that are designed especially for the ultra-poor. These programs are tailored 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 Communities), 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[3]. TUP is a multidimensional program that incorporates both livelihood protection and promotion components and features significant innovations 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 with the goal of helping them graduate to

[2]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 schedule and high time costs involved in regular meetings etc. 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. the moderate poor)

[3]A second phase of the TUP program covering 40 districts was initiated in 2007. 863,000 households are expected to participate in the second phase over five years (2007-2011). This paper provides evidence of the effects of the first phase of the TUP program. For more details on the second phase, see BRAC Annual Report 2007.

standard microcredit programs of BRAC. The program provides training in enterprise activities using the transferred assets, 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[4].

This paper uses a two period panel dataset (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 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 careful 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 panel data is that we can credibly address the time-invariant household level heterogeneity (additive) by using household fixed effect. 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 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)

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

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 (2009), and the difference-in-difference matching estimator due to Heckman *et al.* (1998) (henceforth called DIDM estimator[5]). The treatment effects from the DIDM approach can be biased if the counterfactual trend in the treatment group is different from the actual trend found among the comparison households. The MB-IPW estimator developed by Millimet and Tchernis (2009) 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))[6]. 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 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 mis-targeting (assignment errors) in the TUP program when judged by the inclu-

[5] DIDM approach reduces the bias in estimated treatment effects compared to cross section studies with or without matching (Heckman and Smith (1998), Blundell and Dias (2010). The DIDM approach has been used by Petkova (2009), Berlinski and Galiani (2004) among others.

[6] For recent applications of heteroskedasticity based identification, see Rigobon (2003), Rigobon and Rodrik (2005), Farre *et al.* (2010), Schroeder (2010), Millimet and Tchernis (2009), Emran and Hou (forthcoming), Emran and Shilpi (forthcoming), and Emran and Sun (2011).

sion 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[7]. 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[8]. The assignment errors can be 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 SB1, or 'should be, one', and SB0 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 SB1 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 effect, shows that there is significant positive effect of participation in the TUP program on cash savings, food security, housing (homestead land and roof made of tin), livestock (cow/bull), and shoe/sandal ownership for both the treatment groups: SUP and SB1. The evidence is also consistent with a positive effect of TUP membership on net

[7]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 (1995, 2008), Kanbur (2010)).

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

income of a participant household[9]. There is, however, no evidence of a significant effect of the TUP program on total land ownership, number of duck/hens, productive assets, subjective health outcomes, child labor and women's empowerment[10].

We also find that there are some differences in the program effects between the two treatment groups: SUP and SB1. For example, while the evidence that participation increases net income is very strong for the SUP group, it is weaker for the SB1 group. While the households in both treatment groups seem to invest in productive assets such as livestock, only the SUP households seem to acquire luxury goods such as radio/TV. In fact, the program effect on radio/TV is negative for the SB1 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 SB1 treatment group compared to the BRAC treatment group SUP.

The rest of the paper is structured as follows: Section 2.2 provides a brief discussion of the BRAC TUP program. Section 2.3 describes our data set. Section 2.4 discusses the empirical strategy for identification and estimation of the treatment effects in full detail. Section 2.5 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.

2.2 The BRAC TUP 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

[9] The income measure is self-reported change in income from 2004 to 2005.

[10] 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.

Program serves over one 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 of 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: ownership of less than ten decimals of land (a tenth of an acre), lack of a male earner at home, presence of child labor, 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 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 [11].

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, 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 committee 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 helping them to recover lost assets.

2.3 The Data and Variables description

For the empirical analysis, we use the BRAC TUP 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

[11]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 (column SB1-SNB0). 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 2.3 below.

the initial set of households. Attrition was moderate and was due to migration, death and marriage. The final matched panel contained 5067 households.

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), financial (cash savings) capital. The data includes information about basic needs (food security, clothing and shoes/sandals), stock of household durables and net income change over last year. 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 and imperfect and a partial view of women's status in a household.

The net income change variable is a self-reported measure of change in net household income from 2004 to 2005. This is thus the change in income over one year, and does not provide us a measure of the cumulative impact of the program participation from 2002 to 2005. Unfortunately, we do not have the information in the data set to calculate the level of income at different years. The cash savings is a binary variable that takes on the value of 1 if a household has cash savings in a given year. Food security is measured by three indicators. The first indicator of food security 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], 'food surplus' [4]. The second indicator is 'grain stocks' which is the amount of food grain stocks (kg) in a household in a given year. The third and last indicator of food security 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 during monga season, 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.

The physical assets include livestock (cow/bull, duck/hen goat/sheeps) and other productive assets (such as a fishing net, rickshaw van, and "big trees"). The asset measures do not include any assets transferred from the TUP program[12]. Household durable goods include tubewells 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 ones 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]. 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[13]. 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[14]. 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

[12]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, we do not have information on the transferred assets in 2005 in the data set.

[13]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.

[14]For middle class and richer households in Bangladesh, a particularly revealing indicator of 'feminine goods' is gold jewelry.

variables.

Table 2.1 presents the description of the variables used. Table 2.2 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 of the indicators including large gains in net income, food availability, housing (tin roof), livestock, and most of the assets. There is, however, some evidence that the ownership of homestead land has worsened on average from 2002 to 2005.

2.4 Empirical strategy

Since BRAC did not incorporate any randomized control trials in the phase I of the TUP program, 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: *(i)* we use alternative treatment and comparison groups; and *(ii)* we pay careful attention to potential selection issues, and use a rich set of econometric techniques to tackle them.

2.4.1 Alternative 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 (subset called the "should be, one" (SB1) group henceforth), *(ii)* the eligible households not selected (called the "should be, zero" group (SB0)), *(iii)* households ineligible according to formal criteria but selected in the program (called the "should not be, one" group (SNB1)), and *(iv)* households ineligible and not selected (called the "should not be, zero" group (SNB0)). For details on the construction of these four subsets, please see Appendix 1. As discussed before, our empirical analysis is based on two pairs of treatment-comparison groups: BRAC's own SUP-NSUP and SB1- SB0.

2.4.2 The potential selection issues

There are two levels of selection problems 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 entirely 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 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., SB0) 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 SB0 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 cancel themselves on average or 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 2.3 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 SB1-SB0, the second for SUP-NSUP, the third for SNB1-SNB0 and the last for SB1-SNB0. The first striking feature in Table 2.3 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 than 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 2.3 also shows that the difference in means is, in general, much lower for the treatment-comparison pair SB1-SB0. 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)[15]. Consider for example, total land and number of cow and bull owned by a household, which are among the most important assets in rural Bangladesh. The differences in means for total land owned are: -0.59 (SB1-SB0) and -3.94 (SUP-NSUP), and for the number of cows and bulls are: -0.03 (SB1-SB0) and -0.15 (SUP-NSUP). The same

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

pattern holds for most of the other variables in table 2.3.

It is reassuring that the subsamples SB1 and SB0 closely resemble each other according to the observable characteristics reported in Table 2.3. It thus seems most appropriate to use SB0 as the comparison group to estimate the effects of TUP program on the treatment SB1[16]. 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 2.4 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 clothing, the SB1 group is the poorest among them. An average SUP household owns 35% more land than an average SB1 household, and the SNB1 households own more land than the SB1 households. While the percentage of households who own their homestead land is 40 percent for the SB1 group, the corresponding numbers for SUP and SNB1 are 47 percent and 53 percent respectively. The increase in net income from 2001-2002 was *Tk.* 6350 for an average SB1 household, *Tk.* 8150 for SUP and *Tk.* 9658 for SNB1. Given the above analysis, our focus is on the estimates of treatment effects (ATT) from two alternative pairs of treatment-comparison groups: SB1 (treatment) and SB0 (comparison); and SUP (treatment) and NSUP (comparison). We do not discuss the treatment effect estimates for the treatment group SNB1 as it is composed of the richer households, and thus clearly not the target of the TUP program.

2.4.3 Econometric approach

Matching estimators

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

[16] Note that the treatment group SB1 and the comparison group SNB0 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 2.3, last column).

invariant additive heterogeneity at the individual level by time differencing (i.e, household fixed effect); 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.

An increasingly popular approach to understanding the implications of possible 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 using the DIDM estimator. We, however, chose to omit those results from the paper, because the 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, in our case for a variety of reasons. First, arguably the most important factor that can give rise to a positive selection effect in our context is innate entrepreneurial ability, which is time invariant and thus is taken care of by the household fixed effect (assuming additivity). Second, and more important, in Table 2.3 we find evidence consistent with negative selection in our data set. A negative selection bias is also consistent with the rationale of the TUP program that it is specifically tailored to target the ultra-poor households bypassed by the standard microcredit programs and thus can overcome the constraints on their participation. Third, the existing econometric evidence in fact supports the notion that selection may be negative in case of the standard microfinance programs in Bangladesh. See, for example, Pitt and Khandker (1998), Schroeder (2010)[17]. Fourth, sensitivity analysis using Rosenbaum bounds does not address the possible bias from measurement error. In fact, there is very little known in the literature about the properties of matching estimators in the presence of measurement error.

We use two alternative approaches to address possible biases in the DIDM estimates. First,

[17] 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 Kander (1998) concerns household consumption as the relevant outcome. We are not aware of any formal evidence that shows positive selection into microfinance programs in Bangladesh.

we implement the 'minimum bias inverse probability weighted' (MP-IPW) estimator developed by Millimet and Tchernis (2009). The MB-IPW estimator starts from the inverse probability weighted matching estimator developed by Hirano and Imbens (2001), 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[18]. 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).

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, forthcoming)). 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 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 (2009), Millimet (2011)). To provide intuition behind the approach, we

[18]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)).

consider the following triangular model:

$$Y_{it} = \alpha_0 + \alpha_i + \alpha M_{it} + X'_{it}\gamma_1 + \varepsilon_{it} \tag{2.1}$$

$$M_{it} = \mathbb{1}\{\beta_0 + \beta_i + X'_{it}\beta_1 + u_{it} > 0\} \tag{2.2}$$

where Y_{it} is an outcome of interest (such as income, assets) of household i, M_{it} is a dummy that equals one if the household is a participant of TUP program, and zero otherwise, and α_i and β_i are household fixed effects. The focus is the identification and estimation of the parameter α. The model does not impose any exclusion restrictions on equation (2.1), and identification of the causal effect α is not possible if the error terms are homoskedastic[19]. Assume that the error term in the treatment equation is heteroskedastic of the following form:

$$u_{it} = S_u(\tilde{X}'_{it}\pi)\tilde{u}_{it} \tag{2.3}$$

where $\tilde{u}_{it}$ is a zero mean homoskedastic error, $\tilde{X}'_{it} \subseteq X'_{it}$ 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):

$$Pr(M_i = 1) = P\left(\frac{X_{it}\beta}{S_u(\tilde{X}'_{it}\pi)}\right) \tag{2.4}$$

where $P(\cdot)$ is the distribution function for $\tilde{u}_{it}$. With homoskedastic errors, $S_u(\tilde{X}'_{it}\pi)$ is a constant, and the only source of identification is possible non-linearity of $P(\cdot)$ 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, $S_u(\tilde{X}'_{it}\pi)$ is not a constant function, and identification also

[19] 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 *et al.* (2005).

exploits data from the region where $P(\cdot)$ is approximately linear. The predicted probability of treatment from estimating equation (2.2) 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 (2.1). Note that if the amount of heteroskedasticity is not substantial, then there is little identifying variations in $S_u(\tilde{X}'_{it}\pi)$, and the predicted probability may be a weak instrument. For the specification of the $S_u(\tilde{X}'_{it}\pi)$, we follow the parametric approach developed in Farre, Klein and Vella (2010) which is based on the model of heteroskedastic probit due to Harvey (1976).

$$S_u(\tilde{X}_{it}) = e^{\tilde{X}'_{it}\pi} \tag{2.5}$$

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 more inefficient than the usual IV estimates, because identification here relies on information about the second moment of the data (Lewbel (forthcoming)).

2.5 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 and health-related indicators. As a benchmark, we report estimates from an augmented difference-in-difference specification:

$$Y_{it} = \alpha_0 + \alpha_1 d_{05} + \alpha_{1R}(d_{05} \times d_R) + \alpha_{1K}(d_{05} \times d_K) + \alpha_2 d_T + X'_{02}\Pi + \beta(d_T \times d_{05}) + \varepsilon_{it} \tag{2.6}$$

where Y_{it} is the outcome variable of interest for household i in year t; d_{05} is a dummy that equals one for the year 2005 and d_T is a dummy that equals one when household i belongs to an appropriately defined treatment group (i.e. SB1, SUP) and equals zero when a household belongs to the corresponding comparison group (i.e. SB0, 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 (2.1). 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[20]. 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 2.5 (for SUP) and 2.6 (for SB1). 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 BRAC's 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 and the indicator of women working as daylaborers). As emphasized by Emran, Morshed and Stiglitz (2007), the outside option of a household and thus the net return

[20] The omitted district is Nilphamari.

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 credit, given the wage rate (i.e. separability holds). In contrast, in an imperfect labor market, low labor endowments *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 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 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 SB1 treatment households are reported in Table 2.9 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 assets) have significant positive effects on the probability of selection into the program. The results from probit regressions for the SB1 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 case of the SB1 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 of our samples, SB1-SB0 and SUP-NSUP. In the SB1-SB0 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 from 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 (3) of Tables 2.5 and 2.6. We use a wide radius (0.25) for trimming the sample around the bias minimizing propensity score, as the Monte Carlo results in Millimet and Tchernis (2009) show that MB-IPW with a wide radius effectively corrects for endogeneity bias.

Finally, in column (4) of Tables 2.5 and 2.6, we report estimates from the Klein and Vella (2009) estimator with household fixed effect that corrects for selection on unobservables and measurement error. Following Klein and Vella (2009) and Farre *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 heteroskedasticity probit model is strong enough to identify the effects of the TUP program participation[21]. The Kleibergen-Paap F statistic for SUP and SB1 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 SB1, much higher than the Stock *et al.* rule of thumb value of 10 for one

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

endogenous variable[22].

2.5.1 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 SB1 that excludes the relatively better-off can be particularly illuminating. Table 2.5 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 net income, cash savings, food security, land ownership, quality of housing (with tin roof), livestock (cow and bull, goat and sheep, duck and hen), other productive assets (number of big trees and fishing nets, for example), and household durables (with the exception of radio/TV[23]). 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 members of a household have shoes/sandals. The evidence on health outcomes is conflicting. While there is improvement compared to last year, we do not find evidence of an improvement in over-all health status that can be attributed to the TUP program participation.

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. On the other hand, a comparison between the benchmark DID estimates and the estimates

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

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

from the DIDM and MB-IPW reveals a few notable differences. 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.

The last column in Table 2.5 reports the estimates from the Klein and Vella (2009) approach, implemented with household fixed effect. The first thing to note is that the effect of TUP membership on the net income of a 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 net income in Table 2.5 is that the evidence is not very strong that TUP participation has a positive effect on the net income of a SUP household. An alternative interpretation which in our view is more plausible is that, if anything, the evidence suggests that TUP membership increases net income, because all four estimates in Table 2.5 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 effect on grain stocks is positive, but numerically smaller, and statistically insignificant at the 10 percent level. 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) to 0.22 (K-V), and it is no longer significant at the 10 percent level.

Among the livestock, the effect on the number of cow/bull 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 goat, sheep, duck and hen. The estimated effect for other livestock is numerically smaller compared to 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 (chair, table and bed), 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. Interestingly, the K-V estimates suggest that there is significant program impact on a set of household durables: radio, TV, quilt, blankets and tubewells. While the effects on radio, TV and quilt/blankets are positive, the effect on tubewells 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 tubewells) in 2005 do not include transfers from the TUP program. Thus even though the TUP program might have installed many new tubewells 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 tubewells (presumably not working properly) it would be measured as a negative program effect.

2.5.2 Effects of the TUP Program on the Poorest of the Poor: SB1 Treatment Group

We begin with the benchmark DID estimates reported in column (1) of Table 2.6. 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 net 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 2.5, the DID results in Table 2.6 also suggest that there is no significant effect of TUP membership on the SB1 households regarding women's empowerment and child labor. However, there are two important differences between the results for SUP and SB1. First, the effect on rickshaw van, bicycles and fishing net are not significant at the 10 percent level for SB1, although they were found to be significant for the SUP. Second, the program effect on health status is significant at the 5 percent level for SB1, but it is numerically smaller and not significant at the 10 percent level for SUP. Interestingly, the magnitudes of the effects for most of the outcomes are comparable across the two treatment groups.

The DIDM and MB-IPW estimates for SB1 are reported in columns (2) and (3) respectively in Table 2.6. 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 2.6 reports the estimated program effects from the heteroskedasticity based instrumental variables approach of Klein and Vella (2009) with household fixed effects for the SB1 households. A comparison with the estimates from the matching estimators in columns (2) and (3) shows some interesting differences. The numerical magnitude of the effect of TUP program participation on net income of a SB1 household is significantly smaller (about 50 percent of the estimate from MB-IPW). Moreover, the standard error is large and the estimate is not significant at the 10 percent level. The K-V estimate thus raises doubts about a strong positive effect on income of the SB1 households[24].

The K-V estimates for food availability and grain stocks are consistent with the conclusion

[24] We, however, emphasize that our evidence does not point to any negative effect. All four estimates in Table 2.6 are positive.

that participation in the TUP program enhances food security of an average SB1 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 effect on grain stocks is more than three times larger (compared to the matching estimates) according to the K-V estimate. However, the point estimate for grain stocks has a large standard error, and the effect is not significant at conventional levels.[25] If one takes into consideration the inefficiency of the K-V estimator, a plausible interpretation of the evidence from all the different estimators is that TUP participation has a positive effect on grain stocks of a SB1 household. However, a more conservative interpretation is that the effect on grain stocks lacks robustness.

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 (cow, bull, goat 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, duck and hen, and some of the durables (chair, table, quilt/blanket, tubewells). 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[26]. 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[27]. The results on female clothing and radio/TV thus indicate, that, in some cases, the estimates

[25] The numerical magnitude for SB1 (6.37) is much larger than that for SUP (0.99).

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

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

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 variable bias) would clearly be counterproductive in such cases.

2.5.3 Comparison Between the Treatment Groups, and Broader Implications of the Empirical Results

The evidence presented in Tables 2.5 and 2.6 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 treatment groups; *(ii)* the TUP program effect may be different for different treatment groups (for example, on income and expenditure on luxury goods); *(iii)* the estimates from the sophisticated difference-in-difference or matching estimators may be biased in some cases.

The evidence in Tables 2.5 and 2.6 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 (2009)).

The evidence strongly suggests that TUP participation confers significant benefits to a participant household, irrespective of the treatment group considered. A comparison of Tables 2.5 and 2.6 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 SB1 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 SB1 households. For example, the mean stock of cow/bull was 0.01 for SB1 and 0.04 for SUP in 2002. The estimated program effect is 1.57 for SB1 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 cow/bull become 167.33 times of the initial stock for SB1 and 49.15 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 2.7 (for SUP treatment group) and 2.8 (for SB1 treatment group).

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 SB1 treatment group) do not spend their money on luxury goods such as radio/TV. The relatively better-off participants (SUP) in contrast seem to use microcredit as a substitute for the missing markets for consumer credit; their expenditure on radio/TV goes up significantly.

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. The absence of an effect on child labor is consistent with a situation where the income effect approximately cancels out the labor demand effect of microcredit through an increase in the marginal productivity of child labor.

2.6 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-comparisons groups, we utilize the type 1 errors in assignment (mistargeting) in BRAC's selection 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 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 poor households i.e., SB1.

An interesting finding from our analysis 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. 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.

2.7 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 three out of five 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 a 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 a 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 of 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 the 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 exclusion criterion 3 we used the variable 'disab1', which identified those women who presented a disability. This selected 48 observations, of which 22 previously had been selected as SUP members. Overall, according to the three 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 ten decimals of land, including 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 uncultivated. This criterion selects as eligible 4624 out of the 5067 households 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 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 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 for 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 subgroups: SB1 (selected as SUP and fulfilling both inclusion and exclusion criteria), composed of 1086 observations; SNB1 (selected as SUP, not fulfilling the criteria), composed of 1289 observations; SNB0 (correctly not selected as SUP, criteria not met), composed of 2051 observations; and SB0 (not selected as SUP but fulfilling criteria) with 641 observations.

Table 2.1: Variable Description

Increase in net income	Summary variable to the answer of 'Last year employment and income related information - Increased net income in tk' for the TUP member
Cash savings (dummy)	Binary variable equal to one if the answer to the question 'Do you have any cash savings?' is yes.
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]
Grain stocks (kg)	Stock of grain in kilograms owned by the household
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?
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).
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].
Number of cow/bulls	Number of assets owned, not including program transfers
Number of goat/sheeps	Number of assets owned, not including program transfers
Number of duck/hens	Number of assets owned, not including program transfers
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 2.1 : Variable description, cont.

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 tubewells	Number of assets owned
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].
Ratio of saree to lungi	Ratio of the female clothing to male clothing.
Presence of girls working (dummy)	Presence of female child labor.
Ability of girls to read and write a letter (dummy)	
Years of schooling of girls	
Presence of child labor (dummy)	Binary variable equal to one if the household declares that there are children under 15 working
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 2.2: Summary Statistics

	Year	Mean	Std. Dev.	Obs
Income and savings				
Net income increase	2002	9170.85	7899.87	5035
	2005	16279.71	10845.05	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
Grain stocks (kg)	2002	0.00	0.00	5067
	2005	1.66	17.82	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 cow/bulls	2002	0.11	0.51	5067
	2005	0.94	1.21	5067
Number of goat/sheeps	2002	0.11	0.49	5067
	2005	0.34	0.97	5067
Number of duck/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 2.2: 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 tubewells	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 girls working (dummy)	2002	0.07	0.26	5067
	2005	0.11	0.32	5067
Ability of girls to read and write (dummy)	2002	0.08	0.27	5067
	2005	0.07	0.26	5067
Years of schooling of girls	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 2.3: Test of differences in mean characteristics between treatment and control groups in 2002

	SB1-SB0	SUP-NSUP	SNB1-SNB0	SB1-SNB0
Income and savings				
Net income increase0201	-308.49	-1924.30***	-1466.97***	-4775.16***
	(266.23)	(221.44)	(305.78)	(287.69)
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)
Grain stocks (kg)	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
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 cow/bulls	-0.03***	-0.15***	-0.17***	-0.22***
	(0.01)	(0.01)	(0.02)	(0.02)
Number of goat/sheeps	-0.01	-0.04**	-0.03	-0.09***
	(0.02)	(0.01)	(0.02)	(0.02)
Number of duck/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 2.3: 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 tubewells	-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 girls working (dummy)	-0.02	0.02*	0.01	0.09***
	(0.02)	(0.01)	(0.01)	(0.01)
Ability of girls 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 girls	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 2.4: Mean values of outcome variables in 2002 for different treatment groups

	SB1	SUP	SNB1
Income and savings			
Net income increase0201	6350.14	8150.42	9658.34
	(160.90)	(152.50)	(237.80)
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)
Grain stocks (kg)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
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 cow/bulls	0.01	0.04	0.06
	(0.00)	(0.01)	(0.01)
Number of goat/sheeps	0.06	0.09	0.12
	(0.01)	(0.01)	(0.02)
Number of duck/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 2.4: 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 tubewells	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 girls working (dummy)	0.13	0.08	0.04
	(0.01)	(0.01)	(0.01)
Ability of girls to read and write (dummy)	0.07	0.07	0.07
	(0.01)	(0.01)	(0.01)
Years of schooling of girls	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 2.5: 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				
Net income increase	3013.82	2864.33	3083.26	5140.73
	(326.22)***	(332.44)***	[2665.00, 3474.21]	(6960.00)
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)***
Grain stocks (kg)	1.38	1.73	1.74	0.69
	(0.45)***	(0.54)***	[1.00, 2.57]	(4.06)
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]	
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% 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 2.5: Effects of the TUP Program on Treatment Group 'Selected Ultra Poor' (SUP) (cont.)

	DID	**DIDM**	**MB-IPW**	**K-V**
	(1)	(2)	(3)	(4)
Livestock				
Number of cow/bulls	1.74	1.71	1.71	1.31
	(0.05)***	(0.03)***	[1.67, 1.75]	(0.41)***
Number of goat/sheeps	0.46	0.45	0.45	0.36
	(0.08)***	(0.03)***	[0.39, 0.50]	(0.29)
Number of duck/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)
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 tubewells	0.09	0.14	0.15	-0.53
	(0.02)***	(0.02)***	[0.12, 0.17]	(0.13)***
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% 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 2.5: Effects of the TUP Program on Treatment Group 'Selected Ultra Poor' (SUP) (cont.)

	DID (1)	**DIDM** (2)	**MB-IPW** (3)	**K-V** (4)
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 girls working (dummy)	0.00	0.01	0.01	
	(0.01)	(0.01)	[-0.01, 0.02]	
Ability of girls to read and write (dummy)	0.00	0.01	0.01	
	(0.01)	(0.01)	[-0.01, 0.03]	
Years of schooling of girls	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% 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 (2009); 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 2.6: 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				
Net income increase	3409.32	3512.38	3523.65	1757.16
	(512.43)***	(471.79)***	[2704.98, 4263.91]	(3747.97)
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)**
Grain stocks (kg)	1.98	1.89	1.89	6.37
	(0.72)**	(0.79)***	[0.25, 3.46]	(5.30)
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]	
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% 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 2.6: 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)
Livestock				
Number of cow/bulls	1.69	1.68	1.69	1.57
	(0.07)***	(0.04)***	[1.62, 1.75]	(0.31)***
Number of goat/sheeps	0.45	0.46	0.45	0.59
	(0.08)***	(0.04)***	[0.39, 0.52]	(0.34)*
Number of duck/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)
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 tubewells	0.15	0.16	0.16	-0.08
	(0.03)***	(0.02)***	[0.12, 0.29]	(0.23)
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% 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 2.6: 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)
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 girls working (dummy)	0.02	0.01	0.01	
	(0.02)	(0.02)	[-0.03, 0.03]	
Ability of girls to read and write (dummy)	-0.01	-0.00	0.00	
	(0.01)	(0.02)	[-0.03, 0.03]	
Years of schooling of girls	-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% 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 (2009); 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 2.7: Normalized Program Effects on Treatment Group 'Selected Ultra Poor' (SUP)

	DID	DIDM	MB-IPW	K-V
	(1)	(2)	(3)	(4)
Income and savings				
Net income increase	0.37	0.35	0.38	0.63
Cash savings (dummy)	10.21	9.96	9.93	
Food security				
Food availability	0.47	0.47	0.48	0.71
Grain stocks (kg)	n/a	n/a	n/a	n/a
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 cow/bulls	49.15	48.31	48.36	37.01
Number of goat/sheeps	4.94	4.83	4.78	3.87
Number of duck/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 2.7: 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 tubewells	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 girls working (dummy)	0.00	0.12	0.10	
Ability of girls to read and write (dummy)	0.00	0.14	0.14	
Years of schooling of girls	0.06	0.06	0.04	-0.37
Child labor				
Presence of child labor (dummy)	0.00	0.19	0.17	

Table 2.8: Normalized Program effects on the poorest of the poor SB1 treatment group

	DID (1)	**DIDM** (2)	**MB-IPW** (3)	**K-V** (4)
Income and savings				
Net income increase	0.54	0.55	0.55	0.28
Cash savings (dummy)	15.58	14.49	14.49	
Food security				
Food availability	0.50	0.50	0.52	0.62
Grain stocks (kg)	n/a	n/a	n/a	n/a
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 cow/bulls	167.33	166.34	167.33	155.45
Number of goat/sheeps	7.64	7.81	7.64	10.02
Number of duck/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 2.8: 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 tubewells	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 girls working (dummy)	0.16	0.08	0.08	
Ability of girls to read and write (dummy)	-0.15	0.00	0.00	
Years of schooling of girls	-0.07	-0.07	-0.03	0.43
Child labor				
Presence of child labor (dummy)	0.08	0.04	-0.04	

Table 2.9: Probit Regressions for Selection into the TUP Program

	SUP/NSUP	SB1/SBO
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

Chapter 3

Designing Targeting Criteria and Impact Assessment with Multiple Poverty Dimensions: Framework and Application to a BRAC Program in Bangladesh

with Stephen C. Smith

3.1 Introduction

This paper contributes to the analysis of the design of targeting mechanisms and evaluation frameworks for program intended to reduce poverty, and in which multiple initial screens for eligibility, program activities, and outcome objectives may all be relevant. Starting from a multidimensional conceptualization of poverty, and paralleling the methodology developed by Alkire and Foster (2011), we examine first the sensitivity of the group identified as poor to the type and number of screens used. Second, we offer then a framework for evaluating multiple program outcomes both in terms of the implied (screening) measure of poverty, and in terms of other outcomes deemed objectives of the program.

We apply our methodology to an assessment of Phase I of BRAC's Targeting the Ultra-poor Program (TUP), comparing characteristics of selected participants using alternative participation criteria, and compare program outcome measures associated with different targeting mechanisms. The approach also offers an alternative way to examine the heterogeneity of the program impact across poverty levels. Findings of the application confirm that the BRAC TUP program has a significantly larger impact on health outcomes and female clothing for the less extremely poor among its selected participants. On the other hand, we find that the poorest households ($k = 5$) have a larger impact than the less poor households ($k = 1$) on the net income increase variable (3481.95 Bangladeshi taka vs. 1759.97), on the probability of having a roof of good quality (0.24 vs. 0.13), on food availability (1.02 vs 0.67), on the probability of having meals twice a day (0.58 vs 0.37) and on the probability of owning shoes (0.27 vs. 0.15).

In general, we will refer to individuals as deprived if they meet at least one of these criteria; individuals are identified as poor if they are multidimensionally deprived in a designated number of $k > 0$ dimensions, and identified as ultra-poor if deprived in a sufficiently large number $k + j$ dimensions, $j \geq 0$. Related to the work of Atkinson (2003) we vary the poverty threshold k to consider a recipient as fulfilling the criteria from the case $k = D$ where the recipient is deprived in all dimensions, to the case $k = 1$ where it suffices to be deprived in any one of the designated dimensions to be considered poor (eligible). An analogous structure applies to identification of the ultra-poor.

Moreover, we connect multidimensional targeting screens to evaluation criteria. We show how our analysis can be grounded in the multidimensional poverty measurement technology developed by Alkire and Foster (2011), and facilitates and clarifies analysis of program impact and assessment of the presence of multiple relevant outcome indicators, and heterogeneity in the program effects. For a given extent of deprivations, the program may be evaluated as successful or unsuccessful depending on the number of significant impacts and the size of those impacts. In a simple application, the program is deemed successful to the extent that poverty (as implicitly defined multidimensionally by the participant identification method) fell (calculated relative to that of the control group where applicable).

We then apply our methodology to an assessment of CFPR/TUP. We simulate different program thresholds by varying the numbers of deprivations defining the multidimensional poverty line. Despite the possible targeting errors, households selected for the program have higher levels of poverty in general than the group selected as control, and their initial measure of poverty decreases with the number of criteria used in defining the poverty line (from $k = 5$, the poorest, to $k = 1$, the less poor among the sample).

In addition, we measure the change in the poverty rates before and after the program using Alkire and Foster (2007, 2011) methodology and varying the poverty threshold $k > 0$. The proportion of households identified as ultra-poor in 2002 by the poverty cutoff ($k = 3$) is 28 percent. The 2005 proportion is 23 percent. Interestingly, at the same poverty cutoff, the decrease in the proportion of treated participants goes from 36 percent in 2002 to 25 percent in 2005, while the proportion of households identified as ultra-poor in the control group does not vary from its 2002 level of 21%. When we analyze the poverty change of the less poor households ($k = 1$), poverty diminishes from 46 to 40 percent, and the same pattern holds: the decrease for treated households is from 53 to 40 percent while the control group's poverty rate is 40 percent in 2002 and 40 percent in 2005. The change in poverty for the poorest households ($k = 5$) is from 3 to 1 percent. While in 2005 the treated participants are all out of such extreme poverty condition, the poverty rate of the control group remains at the 2002 level of 2%. We perform a robustness check of the targeting criteria by evaluating poverty by initial levels of asset ownership. We confirm that the households that are deprived according to all five criteria are also the poorest in terms of asset ownership. For example, in 2002, they did not own productive assets such as cow, bulls, goat and sheep, duck and hens; the probability of having two meals a day was 0.46; the probability of owning shoes was 0.48. In contrast, the households identified with only one (any) deprivation criterion owned some productive assets, the probability of having two meals a day was 0.59 and the probability of owning shoes was 0.62 (these numbers are statistically different at 95% confidence interval).

The example shows how one can apply the Alkire and Foster (2007, 2001) framework to evaluate programs in a multidimensional way. We examine the program impact and its heterogeneity across poverty levels using Difference-in-Difference. When we analyze the heterogeneity in the program impact in relation to outcomes, we find that the poorest households ($k = 5$) have a larger impact than the less poor households ($k = 1$) on the net income increase variable (3481.95 Bangladeshi taka vs. 1759.97), on the probability of having a roof of good quality (0.24 vs. 0.13), on food availability (1.02 vs 0.67), on the probability of having meals twice a day (0.58 vs 0.37) and on the probability of owning shoes (0.27 vs. 0.15). Had the program concentrated on the poorest households, the average impacts would have been larger in magnitude.

3.2 Targeting Mechanisms and Evaluation Framework

A multidimensional poverty analysis is called for when individuals or families face multiple deprivations simultaneously, while individuals are understood to be poorer as the number of deprivations increases. In this case, income cannot even be thought of as a sufficient statistic for poverty status, a conclusion that is amplified when these deprivations cannot be purchased directly on markets at prices close to feasible for the poor, or acquired in a reasonable period. Beyond this, in developing countries, income measurement is notoriously imprecise; it might be more efficient to rely on complementary indicators in targeting poverty programs.

A means approach is commonly used, which uses other indicators often with the purpose of predicting income. In practice, prediction of income can be highly error-prone, though it can be more effective when there is a sufficient, representative sample of households whose incomes are measured with relatively high precision, and other variables that are more readily visible are collected for those households to predict their income in a careful regression framework; using the resulting regression coefficients can be reasonably effective at predicting out of sample potential participant incomes. But again, even when error rates are low, usually program eligibility depends on income, missing poverty multidimensionality. Our approach is related to proxy means tests, but rather than seeking a prediction of a single measure, we maintain a multidimensional perspective. In these cases, the use of a single indicator may be inherently much more controversial. A problem is to think about the right number as well as the type of indicators to be used for such screening purposes. We include consideration of whether a very small number of indicators may do a reasonable job of predicting when a broader class of indicators is present.

An alternative approach is geographical targeting, in which poor regions are identified and programs are implemented for all households in that region. This approach, an example of "narrow targeting" (Besley and Kanbur (1990), van de Walle (1998)), has both advantages and disadvantages. It excludes people living in poverty outside of these regions; but it is less costly than universal coverage (that would seek to include poor people outside these regions). Another point in its favor is that people living in poverty are more likely in general to remain chronically poor, and with greater poverty severity, when living in areas in which a large fraction of neighboring people are also poor; so the approach may help reach the ultra-poor who may otherwise be overlooked.

Our more general hypothesis is that requiring more poverty indicators to be present leads to selecting a set of people who are poorer than would be the case if we relied on fewer (or on a single) indicators. This is not for a proxy of income but for a more reliable estimate of what would be broadly understood as deprivation. Our working assumption is that the greater the number of indicators used to assess poverty, the smaller the likelihood of identifying an individual as ultra-poor when she is not (in our application, we find support for this assumption by examining a variety of supplemental poverty indicators). The relevant set of poverty indicators will differ according to the context and the different sets of poverty problems and potentially poverty traps operating locally. In this example, we primarily follow indicators selected by a local NGO that has a deep understanding of the nature of poverty in the region we study.

3.2.1 Targeting

Targeting is an instrument for concentrating program benefits on the poor. To better target recipients, there has to be an agreement of what are the standards of living in the region, of what constitutes poverty, and of what will be the poverty line(s) used to qualify individuals as poor (Besley and Kanbur 1990).

Sen (1976) stressed the importance of both identification of the poor and aggregation of their characteristics into an overall indicator. Alkire and Foster (2007) emphasize that the identification step has been overlooked in the literature, and stress its importance in their multi-dimensional methodology.

During the targeting process, recipients[1] must meet several conditions in order to qualify. The identification step in our methodology parallels the basic ideas of Alkire and Foster (2007, 2011). Identification is done in a sequential process that also may be viewed as analogous to means tests:

(i) In a preliminary stage, utilized in only some instances, is a pre-cutoff. In this aspect, poverty is, in the first place, an absence of some characteristics; if such characteristics are present then an individual cannot be considered eligible as a poor person for these purposes. These have been termed exclusion criteria: if the individual does not have these features, she can be considered as even potentially poor or potentially satisfying the means test.

(ii) Then a set of D deprivation indicators ($d = 1, \ldots, D$) are selected such that if the individual is deprived of one of these indicators (or in a continuous case falls below a threshold) then she is identified as poor in that dimension.

(iii) Let $n = 1, \ldots, N$ represent the number of individuals in a population; in that case, the deprivation data generates a deprivation matrix $g = [g_{ij}]$ of dimensions $(N \times D)$ where if the

[1] In this paper we sometimes talk about individuals and sometimes about households. This is because in general microfinance programs are addressed towards the main female of the household (household head or his spouse), thus we would be able to calculate poverty rates at individual level. But in practice, programs do not cover more than one female per household. Moreover, sometimes data is available at household and not individual level, therefore we are forced to calculate poverty rates at household level, with the caveat that household size might be endogenous.

individual is deprived in one dimension, she is assigned a value of one; if the individual is not deprived, she is assigned a value of zero. Thus, for each dimension, a dimension-specific threshold or poverty line (l) must be specified. For identification, a threshold number of deprivations (zero-valued indicators) must be present. Formally, the identification of individual as deprived is a function $\rho(\cdot)$ of the individual deprivation vector and the cutoff vector: $\rho(d_i;z) = 1$ if a person is deprived in dimension d_i and zero if not.

(iv) Then, if the individual is deprived in the designated number k (or more) dimensions then she is deemed multidimensionally poor (and if deprived in $k+j$ dimensions then "ultra-poor"). The poverty line here would be given by the cutoff $k+j$ number of dimensions to be considered ultra-poor. Thus, identification of the multidimensionally poor individuals is given by the function $\rho_k = (z,w,k)$ [2]. If the person is identified as multidimensionally poor, the identification function takes a value of one; otherwise, it takes a value of zero. By multiplying each row of matrix g by the identification function ρ_k, a censored-deprivation matrix ($g^0(k) = [g^0_{ij}]$) can be generated, where now, if the person is not identified as multidimensionally poor, s/he is assigned a value of zero, even if in the previous matrix g had a positive value in one specific dimension. Thus, the matrix $g^0(k)$ only displays information for the multidimensionally poor.

Note that Category *(iv)* reflects what is called the dual cutoff in the Alkire and Foster approach (2007, 2011, AF henceforth). Sometimes data is available that captures the exact concept of the deprivation (such as a blood test indicating whether a person has a specific health deficiency such as anemia); when such data is not available, a proxy must be used.

Note further that Category *(i)* represents what we may call a pre-cutoff, adding an additional layer to identification. This will not be found in all targeting schemes, nor is it always appropriate for poverty identification; but when viewed as appropriate in the design of a program it needs to be accounted for in the broader identification, targeting and evaluation system. The usual case has what we call 'excluded-up' elements. This feature would indicate that

[2] We work with the benchmark case of equal weights w, such that their sum equals the number of dimensions considered

if a person has some characteristic, such as guaranteed access to other government assistance, she could not be deemed poor for a certain ('ultra-poor') exercise. If the measure is for use in determining eligibility for a particular program for the poor, we may also encounter a more unusual case -what we may call 'excluded-down' elements, typically a person not capable of participating in the program activity. The latter makes sense for certain kinds of targeting: For example, if a person truly cannot work there is no point involving that person in a program designed to improve work outcomes. For a second example, if a family structurally cannot meet conditional cash transfer requirements, (except perhaps with extreme hardship), then unconditional transfers would be more appropriate.

The aggregation step in our methodology also follows AF, which in turns builds on the standard Foster, Greer and Thorbecke ((1984), FGT henceforth) class of decomposable poverty measures P_α. If all deprivation dimensions are ordinal, it is not possible to work with values of α greater than one. In this paper, we calculate the headcount ratio (H) and the adjusted headcount ratio (M_0) poverty indicators, which correspond to the P_0 and P_1 FGT poverty measures respectively. The headcount ratio is the mean of the deprivation matrix g, and it indicates the proportion of the population who are poor; but this measure does not indicate the depth of poverty for those individuals in any given dimension; nor does it conform to dimensional monotonicity in that the measure does not change if an already identified poor person becomes deprived in additional dimension(s). In contrast, the adjusted headcount ratio combines two measures, H and the average share of deprivations A (the number of deprivations that each household has divided by the total number of deprivations considered). Computationally, it is the mean of the censored-deprivation matrix g^0 (where those not identified as multidimensionally poor have their deprivation data replaced with a zero). The resulting adjusted headcount ratio measure also can be written as HA (the product of the headcount ratio and the average intensity of poverty). In contrast to H, the adjusted headcount ratio satisfies dimensional monotonicity (if the average share of deprivations increases, so does M_0) Moreover, as M_0 is defined as the frequency of poverty times the average share of deprivations, it is both easy to compute and to interpret. Last, as AF stress, M_0 can be viewed as a measure of "unfreedom" in the sense of Amartya Sen (as M_0 counts the deprivations, by making the analogy that deprivation in a dimension suggests capability deprivation).

Although we are not proposing a particular set of indicators that are best in all contexts, we propose to use a set of deprivation criteria that relate to basic needs: health, education, housing and work. The deprivation criteria reflect the capabilities (or absence of) the household. We examine a way of considering the type, number, and specific sets of combinations of these indicators to be satisfied, given the choice. The resulting approach equally may be useful when communities are offered the opportunities to enumerate and explain the set of important aspects of what it means to be poor within the context of their community. But even when a community itself selects the poverty indicators -the features that determine if a person is deemed poor- an iterative process similar to the one described here may be needed to result in what would be viewed as a reasonable selection.

3.2.2 Sensitivity Analysis

We examine the sensitivity of the group identified as ultra-poor to the type and number of screens used. For each dimension, we proceed to calculate the proportion of households to be identified, as one dimension that would select almost the entire sample, despite being able to identify an important deprivation, would not be able to select the poorest households. For example, land ownership in the Bangladeshi context is an indicator of minimal security, but ultra-poor and poor households often lack such minimal security. When trying to screen the ultra-poor from the poor households, an indicator of lack of land might not be able to identify the most deprived. For a second example, shoe ownership, which is a capabilities indicator (as it is related to health, mobility for school and for income generating activities, and social status within the community) might do a reasonable screening of the poorest among the poor.

3.2.3 Robustness Check: Initial Level of Asset Ownership

As a robustness check, we analyze the poverty status of households according to an alternative measure: the initial level of asset ownership. "Assets" include any tangible or intangible good that enables the poor to draw a stream of income or consumption. A focus on assets of the poor helps us to know whether a person is temporarily poor, or stuck in a poverty trap, and it clarifies what a family would need to permanently escape from poverty. Thus, if the deprivation dimension used is effective in sorting households according to their poverty status, then those households will present lower levels of asset ownership. We perform targeting tests by evaluating the headcount ratio and assessing the effectiveness of the criteria in finding the ultra-poor by calculating their initial level of asset ownership for the different subgroup combinations that can be constructed with the eligibility criteria.

3.2.4 Analysis of the Program Impact in Relation to Targeting

We relate multidimensional targeting screens to evaluation criteria. We show how our analysis connects to the multidimensional poverty technology developed by AF. Consider a program treatment known to help the ultra-poor on average. However, even in this case, we may still not know if the treatment will help the poorest of the ultra-poor as much as the average (or even at all); the same uncertainty prevails concerning the least poor of the ultra-poor. We use the AF methodology to evaluate the heterogeneity of the program impact, by varying the cutoff $k+j$ number of deprivation dimensions that must be present in a household to be identified as ultra-poor[3].

To evaluate the program impact, we start from the program's classification of the sample intro treatment (T) and control (C) subgroups. For an initial cutoff value of $k+j$ deprivations, we construct a second sub classification: we select those households that satisfy the cutoff $k+j$ number of deprivations and were part of the initial treatment group, and compare the

[3]With available data on continuous measures of deprivation levels, we can also perform analogous sensitivity analysis on varying cutoffs in this dimension.

program outcomes with the subgroup of households that satisfy the cutoff $k+j$ number of deprivations and were part of the initial control group.

To assess the presence of heterogeneity in the program impact, we evaluate the magnitude of the outcomes when $k+j$ varies.

We assess both the effectiveness of the targeting criteria as measured by the headcount ratio; and the heterogeneity of the program impact on them, as their poverty status ranges from fewest to most deprivations[4]. We evaluate the both the program impact and the presence of heterogeneous effects using DID techniques, as well as by calculating the difference between the pre and post-intervention poverty measure.

Furthermore, benefitting from the AF approach that enables us to track down the poverty assessment towards what deprivations are in fact present, we might explain the program impact (or the absence of it) by the contribution of each specific deprivation.

3.3 An empirical application: analysis of the CFPR/TUP I household panel data

3.3.1 Background

BRAC is one of, if not the, largest NGOs in the world (its acronym currently stands for Building Resources Across Communities, but was formerly known as the Bangladesh Rural Advancement Committee). BRAC has very extensive experience designing and implementing programs to alleviate the deprivations of the poorest households. It has a special focus on microfinance, although, as in the program that we analyze here, after concluding that their standard microcredit programs do not engage most of the poorest, has been the pioneer in developing transitional programs to improve the readiness of the ultra-poor to participate

[4]That is, where the identification of poverty ranges from the intersection case - using the terminology of Atkinson (2003) - where households deprived in all dimensions are qualified to participate, to the union case, where households would be classified as ultra-poor even if they are deprived in only one dimension.

in microcredit. In Bangladesh two types of programs were historically designed to alleviate poverty (Matin 2004). When a situation of structural poverty was found, with households permanently living below the poverty line, an anti-poverty program would provide them with enough income to escape from poverty. The other situation identified was that of households facing a negative transitory income shock; the proposed solution was to help them with one-time grants in order to return households above the poverty line[5].

Over time, the analysis of such programs showed that once the programs were finished, households would return to their former poverty situation. Thus, BRAC designed the CFPR/TUP program with a comprehensive approach in mind, providing households with asset transfer, enterprise development training, social development and health care in order to sustainably graduate them out of poverty and to empower them to fully benefit from potential adverse future shocks with standard microfinance programs.

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. The TUP 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. A goal of mainstreaming these clients into microfinance is to enable them to maintain and expand their businesses over time. 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

[5] Matin (2004, WP2, page 7) describes that the proposed solution to a temporary shock would be one-time grants (usually food, but sometimes cash) to help return household income/consumption above poverty levels.

program participants (SUPs) 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.

To select participants, first, all members of treatment and comparison groups are nominated by villagers as among the poorest local families. Second, a subset is selected by BRAC according to exclusion and inclusion criteria of the general type described in the previous section. 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 find such 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 villagers agree on a wealth ranking among the households. For example, those who can afford tin plate walls or roofs were viewed less poor than those with straw walls or thatched roofs. 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 schoolteachers, were actively involved in all stages of the process.

Our panel data set is comprised of women nominated by villages through this process -some of whom were ultimately selected to participate in the program- and some not.

3.3.2 Targeting

We start with the working hypothesis that BRAC, knowing Bangladesh conditions as they do, has selected a set of deprivation criteria that are most relevant for the ultra-poor, or that correlate well with those problems. One concern may arise if the deprivation criteria have a high degree of correlation among them, which might be a consequence of including redundant deprivation characteristics.

The conceptualization of the CFPR/TUP phase I program was on narrow targeting, and as such had sequential layers of identification of the desired recipients. According to Matin (2004) the steps in identifying the poorest households were: (i) Rapport Building; (ii) Participatory Rapid Appraisal meetings; (iii) Survey and preliminary selection; and (iv) Final selection. They did not randomize who would get the treatment; actually, they tried to select the poorest of the poor to participate as 'treated' households in the CFPR/ TUP phase I program. Thus, the control group had (despite having about the same number of deprivation dimensions and therefore qualifying as ultra-poor) a lower depth of poverty, as measured by the means test of both characteristics and outcomes at the start of the program[6]. Nevertheless, the initial poverty status of the households selected to participate have different numbers of deprivation, which suggests imperfect targeting. Given the different degrees of initial poverty of the recipients, we are able to examine the sensitivity of the group identified as poor to the

[6]Clearly, however, that fact does not prevent the use of Difference-in-Difference to evaluate the program impact since all that is needed for the use of DID is that the selected treatment and control groups would follow the same trend over time, not that their starting points would be identical. Additional corrections can be used also for additional potential selection concerns (see Emran, Robano and Smith 2011).

type and number of screens used.

3.3.3 Sensitivity of the Identification Strategy to the Type and Number of Screens Used

As mentioned, after the household survey was conducted, two sets of explicit criteria for program participation was utilized: three criteria for excluding households if any of these conditions were met, and five criteria for including households as poor if a sufficient number (3) of these criteria are met.

Pre-cutoff

There were three conditions, which if met, would automatically exclude the household under consideration, irrespective of whether the household presented any deprivation indicator. The exclusion criteria were: (EC1) participating in another NGO; (EC2) were recipient of a VGD (Vulnerable Group Development) food card; (EC3) there was no female able to work. Of the 5067 households in the 2002 dataset, 444 were participating in another NGO; 127 were recipients of the VGD card, and 48 had no women able to exert labor. Table 3.1 shows the headcount ratio of each criterion, which is respectively 9, 3 and 1%.

The first two criteria aim at excluding women because they have access to other programs. The rationale was to focus on women who were too poor to have sufficient influence to receive the ration cards, or too marginalized for (other) NGOs to find and work with them. The exclusion criterion three aims at excluding women who were disabled, or that for some reason could not use an asset productively. These women might need relief rather than development. Table 3.4 presents the number of deprivation criteria met by the selected households despite disqualification by the exclusion criteria. Among them, almost all households selected by EC3 that participated as part of the treatment group present a sufficient number of deprivation criteria[7]. Unfortunately, while the poverty status of the women who were not able to exert labor is of more poverty according to the number of criteria to be included in the program, the specific design of the CFPR/TUP I program could not alleviate their condition. Overall, 12% of the participants met at least one of the conditions for exclusion from the program, although only 2 percent were actually selected for participation as part of the 'treated' group.

Eligibility Criteria

The second layer for eligibility was a set of five criteria of inclusion into the program. They were designed to identify deprivation conditions along each category. They were: (IC1) ownership of less than 10 decimals of land, (IC2) no male income earner at home, (IC3) children of school age having to work, (IC4) household dependent upon female domestic work outside the household, and (IC5) households having no productive assets. Table 3.2 presents the number of households selected in 2002 by each poverty indicator. Although in theory, there might be good reasons to select those characteristics as ultra-poor identifiers, many of them did not result optimal at distinguishing the ultra-poor households from the poor ones. While lack of land, even homestead land is an indicator of lack of minimal security, it selected more than 90% of the 5067 households in the dataset. In practice, there is a high correlation between the criteria IC2 and IC4, as both indicate that the household female is likely to be the household head and she might need to work outside home to sustain her family. The correlation level is 0.89. The criterion aiming to identify households where there is child labor selects about 15% of the households present in the 2002 dataset, which is consistent with the International Labor Organization estimates for Bangladesh. IC5 selects about half of the full sample as extremely poor.

[7] In the robustness assessment performed in next section, Table 3.6 shows that households in EC1 and EC2 classification have a larger level of asset ownership. In particular, households in EC1 present the largest values of asset ownership.

Ultra-poor status

The final layer for identification of ultra-poor households set the poverty line at $k = 3$, that is, to be chosen, households must meet at least three of the five eligibility criteria, while not being disqualified by any of the three criteria of exclusion. In Table 3.3, we present a disaggregation of the conditions met by the finally selected households. BRAC selected 2375 households as participants, hereafter called SUP (selected ultra poor) and 2692 as part of the control group (hereafter called NSUP, non-selected ultra-poor). They are shown in the first row of the table.

However, the analysis of the conditions established by BRAC for household selection shows that of the 5067 households in the dataset, only 1936 met the conditions for qualifying as ultra-poor. Of them, 1152 were selected for program participation and 784 as part of the control group. They are shown on the second row of Table 3.3. The remaining blocks of Table 3.3 present the numbers of households either selected for treatment or control that meet exactly the criteria established. The sum of households in blocks (2) to (7) equals the number in block (1). Block (2) shows that there are 143 households that met all five-inclusion criteria. Of them, 96 were selected for program participation[8]. Thus, if as stated by BRAC, the poorest households were selected for program participation, we would have expected all 143 poorest households to belong to the treatment group. Block (3) shows the different combinations of four criteria met by the households. The larger group is the one formed by criteria 1, 2, 4 and 5, where 834 households meet the conditions, and of them 547 were selected for program participation. Similarly, block (4) presents the combination of three criteria of inclusion. Again, the larger subgroups are those satisfying criteria 1, 2 and 4 and 1, 3 and 5. If BRAC had relied on that classification, the sum of blocks (2), (3) and (4) would give the SUP/NSUP

[8] We show in next section, Table 3.5 that they are indeed the poorest households according to their initial asset ownership levels.

group in block (1), as those are the households that satisfy the ultra-poverty cutoff of $k = 3$[9].

Thus, by looking at the number of households selected by the combination of eligibility criteria it is apparent that BRAC has made trade-offs implicitly or explicitly regarding relative weights to be placed on the inclusion criteria. It might have been the case that there were more resources available, as the poverty line was implicitly set to a lower number, selecting for participation 782 households that met only two of the five inclusion criteria and 415 households that only fulfilled one criterion. Nevertheless, if the objective of the targeting was to select the poorest recipients, all households with $k = 3$ should belong to the "treated" group, and this is not strictly what we observe. Block (5) presents the combination of households that meet only two of the five criteria of inclusion. The first row in block (5) shows that of the 79 households that satisfy only criteria 1 and 2, 25 were selected for program participation. The combination that selects the largest subgroup is the one formed by criteria 1 and 5, which are the households that have less than ten decimals of homestead land and have no productive assets. There were 1343 households meeting that criterion; and 668 were selected for program participation. Among them, the first row in block (6) shows that 1127 households owned less than ten decimals of land, and 384 of them were selected for participating in the program. IC2 identifies 24 households, and 6 of them were selected for treatment; IC3 identifies 27, and 4 were selected for treatment. Criterion IC4 in itself is not a sufficient indicator of deprivation, as it notes whether the family is dependent upon female work outside the household; there is no household selected either for treatment or for control that meets only this criterion. IC5 identifies 75 households and 21 were selected for treatment. Block (7) presents the number of households in 2002 that do not meet any criteria; however, 26 were selected for participation. The largest proportion of households identified through the

[9] We calculate the proportion of type-1 errors of assignment as the difference between the selected households and the ones who belong to the control group, but according to the program objectives, should have been chosen as recipients. The proportion is 30% (784 eligible households in the control group/2375 selected recipients). Analogously, the proportion of type-2 errors of assignment is 50% (1223 ineligible households in the treatment group/2375). This apparently high proportion of type-1 errors might in part be explained by the fact that some eligible households were intentionally left in the control group in order to increase the comparability between the subsamples. With respect to the type-2 errors of assignment (leakage), they might be explained by the fact that availability of funds made it possible to incorporate more households into the program, lowering the cutoff value $k + j$ to 2 deprivations present instead of the original three. Another plausible explanation is a different view on the ground on the part of staff about what ultra-poverty is.

deprivation indicators was found by the combination of IC1 and IC5: 27 percent of households are in that classification in 2002. Finally, 22 percent of households are identified only by IC1, and 16 percent of households by the combination of IC1, IC2, IC4 and IC5. Thus, lack of land and lack of assets -which *per se* signal deprivations, are the criteria that select the largest proportion of households as participants.

Table 3.4 presents the disaggregation of the number of households selected according to how many criteria they met despite the fact that they met as well the exclusion criteria[10]. First, a very low proportion (2 percent) of households that should have been disqualified received assistance from the CFPR/TUP I program and ten percent participated as part of the control group. There were 49 households selected for program participation that also were participating in another NGO program (the remaining 395 households were part of the control group); 35 households that were recipient of a government food card; and 22 households where there was no female able to work. Second, among the 106 treated participants, 69 satisfy a sufficient number of criteria for participation, thus, they might have been included because of that reason[11]. The households that meet EC3 'no female healthy enough to participate in the program' indeed have a high incidence of inclusion criteria. However, only half of them were selected for participation in the treatment group[12].

Using BRAC targeting criteria we find that *(i)* The ranking of criteria importance among the five deprivations used (counting the number of households selected by each criterion or by a combination of them) is: IC1 'ownership of less than ten decimals of land', followed by IC5 'lack of productive assets'; then IC2 'no presence of a male income earner in the household'; IC4 'household dependent upon female working outside' and IC3 'presence of child labor'. *(ii)* The previous ranking of criteria helps find the poorest households in terms of initial levels

[10]There is only one extra observation that satisfies exclusion criterion 1 -belonging to another NGO- but does not satisfy any of the inclusion criteria and was nevertheless selected for program participation.

[11]Matin (2004) explains that in some cases, working females were forced by their employers to take NGO loans that were destined to the employers, not to the applicants. This reason, and the fact that most of them satisfied a sufficient number of deprivation criteria, might provide a partial explanation of the inclusion of households despite apparent program disqualification. However, the robustness check performed in next section, where we analyze the level of asset ownership shows that the households selected by EC1 present the highest levels of asset ownership.

[12]Again, Table 3.6 shows that households selected by EC3 have the lowest level of asset ownership.

of asset ownership (see next section). *(iii)* But 12% of households were selected into the program despite the fact that they should have been disqualified because they met at least one exclusion criterion (they were either less poor than the objective recipients or unable to exert work). *(iv)* The households selected despite disqualification by the exclusion criteria were not among the poorest according to their initial level of asset ownership.

Moreover, the CFPR/TUP I program was set up in a partial equilibrium analysis, thus it is not possible to compute gains from spillovers of program implementation to other members of the community not included in the sample. Beneficiaries were selected after a sequential process of narrow targeting. Although the poorest households were looked for, a mixture of households with different degrees of poverty (defined by the number of deprivations present) was finally selected to participate as recipients. This fact poses a challenge when trying to evaluate the success of the program, as the different initial conditions might intervene in the absorption of benefits thus generating heterogeneous results. The decision making process was done centralized at BRAC, outside from group beneficiaries. As the money came from elsewhere, there was no trade-off inside the group beneficiaries of where to allocate the funds. The poverty line (identifying as ultra-poor those individuals with at least three criteria of eligibility and not being disqualified by the restrictions imposed), was imprecisely measured, in the sense that not all households with at least three deprivations (or two) were selected for treatment; some households presented disqualification criteria but nevertheless participated. However, the presence of heterogeneity in the initial conditions allows us to assess the role they play in the final program impact.

3.4 Robustness check: finding the ultra-poor by their initial asset level

So far, we have used BRAC's definition of ultra-poor households as those that had at least three deprivation characteristics of the five previously identified (IC1 to IC5). Following Smith (2005), we test whether such definition identifies as well those households who have the lowest initial level of assets (in ownership of several assets and in the magnitude of each one of them). Even though the program has a specific deprivation component (IC5 "lack of assets") to identify households, the present analysis is important in part because several of the key program outcomes also represent productive assets. While it remains the case that poverty itself is multidimensional, assets are important not only because they represent a permanent consumption stream, but also because they can facilitate efforts to pursue remedies for other deprivations over time. Note that we focus on the change in household assets other than those (that remain) transferred from the program itself; this gives a better sense of sustainability of the program.

Table 3.5 presents mean initial asset values for treated households with their 95% confidence intervals. In the results presented, households are sorted in two subgroups: those that have all five deprivations and those that have (only) any one deprivation (no matter which one). The households that are deprived according to all five criteria are also the poorest according to other indicators; and in addition, most of the outcome indicators were significantly lower at the outset of the program (and none were significantly higher). Analogously, Table 3.6 presents the mean initial asset values for the households that met the disqualification criteria. The households in columns EC1 (participant in another NGO program) and EC2 (recipient of a VGD card) present significantly higher mean levels of most assets (note in particular that the food availability outcome[13] is larger (1.78 in column EC1 vs. 1.43 in column EC2 and 1.31 in column EC3)).

From Tables 3.5 and 3.6 we confirm that *(i)* requiring more poverty indicators to be present selects the poorest group of households as measured by the initial mean levels of asset ownership (at least for the set of assets we examined in this paper). *(ii)* Moreover, the households selected despite meeting the exclusion criteria are not asset-poorer than the other subgroups[14].

[13]Food availability is measured as the answer to the following question: '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]. Additional variables description is presented before the start of Table 3.5.

[14]Of households meeting the exclusion criteria, those that meet EC3 ('the household has no healthy female able to work') are the poorest *by the asset criteria*, but due to this deprivation,they also have far less opportunity to benefit from program participation.

(iii) For the sake of brevity, we do not report all mean tests for the seventeen different subgroups of households in the text but we find that the most useful criteria in identifying the poorest households are those that combine IC1, ownership of less than ten decimals of land and IC5 lack of productive assets (results available from the authors).

3.5 Analysis of Program Impact in Relation to Targeting

We proceed to assess which subgroups benefited most from the program by calculating the difference between the pre- and post-intervention poverty measure. Additionally, we present evidence of a heterogeneous program impact in terms of the number of positive outcomes and on their magnitude by degree of poverty.

3.5.1 Poverty measure

In Table 3.7, we present calculations of both AF poverty measures (H and M_0) at each poverty level as defined by k. The upper part of Table 3.7 shows calculations for the full sample (5067 observations). Both H and M_0 coincide in showing that poverty decreased during the period. Using BRAC's poverty definition ($k = 3$), the headcount ratio varied from 38% in 2002 to 34% in 2005. Setting up the poverty requirement at greater levels of deprivation (larger values of k) initially identified a lower number of households as poor, that decreased in proportion in 2005, thus one might say that the program effectively lifted the ultra-poor out of their extreme poverty condition. The intersection case of $k = 5$ identified 3% of households as ultra-poor. That ratio decreased in 2005 to 1%, which suggests that the program was effective in moving the individuals from their extreme poverty condition to a less deprived condition. The last column of Table 3.7 presents the variation in the measures, and the largest variation is for the households identified through the intersection case ($k = 5$), where M_0 decreases 67 percent. While the headcount ratio at $k = 1$ remained at 96% of households, the adjusted-headcount ratio fell from 46% in 2002 to 40% in 2005. This reinforces the idea that the program was successful in alleviating the conditions of the ultra-poor households, because it shows that the intensity of poverty (the average number of deprivation that each household suffers) lessened.

The lower part of Table 3.7 (below the dashed-line) presents the disaggregation of the full sample into BRAC's treatment group (SUP) and the control group (NSUP). There is a reduction in poverty for the SUP group, at each poverty level k, such that at the end of the program there are no households still suffering from all five initial deprivations (in 2002 they were 4%). For the NSUP members, the percentage remained at 2%. At BRAC's poverty line of ($k = 3$), there was a reduction in poverty from 36% to 25% while for the NSUP the proportion of households identified as poor did not vary from the 2002 level of 21%[15].

It is not theoretically required that both H and M_0 were consistent across all poverty lines (Duclos *et al.* 2007), because even though the proportion of households below the poverty line might, say, increase, if the average share of deprivations that they face diminishes (because of an improvement in living conditions due to program participation), the M_0 measure will either increase or diminish, depending on the actual size of the variations. In the case of the TUP group, we find that both H and M_0 measures are consistent, except on the last line ($k = 1$) of NSUP members, where poverty increased by 1% according to the headcount ratio, but as the average share of deprivations decreased slightly, the adjusted headcount ratio decreased in 2%. As a follow-up, we present in Table 3.8 the disaggregation of both H and M_0 according to which were the most effective criteria[16] in *(i)* finding the ultra-poor and *(ii)* moving them out of poverty. As is often the case , the headcount ratio varies widely with k: from 3% of the population when $k = 5$ to 96% when $k = 1$. The first line shows $k = 5$, which identified 143 households in 2002 and 50 in 2005. H and M_0 are similar, as the intensity of poverty for them is the highest possible (equal to 1, in which all deprivations are present). Nevertheless, two thirds of the household were successfully lifted out of such extreme poverty conditions. Similarly, taking the poverty line at $k = 4$ shows that the program lifted out of poverty 553

[15] All variations are statistically different from each other at conventional 95% confidence intervals.

[16] Recall that the criteria were: (IC1) ownership of less than 10 decimals of land, (IC2) no male income earner at home, (IC3) children of school age having to work, (IC4) household dependent upon female domestic work outside the household, and (IC5) households having no productive assets.

households.

Taking BRAC's poverty line of $k = 3$, the combination of criteria (1, 3 and 4) moved 97 households out of poverty. The poverty line at $k = 2$ shows that by criteria 1 and 5, 889 households were moved out of poverty status. Finally, for $k = 1$ the number in poverty increased from 1127 households in 2002 to 1822 in 2005 (note that, as the subgroups are mutually exclusive, this mostly reflects households moving from more than one deprivation to only one remaining deprivation. Furthermore, the number of households without any deprivation criteria increased slightly from 202 at the start of the program to 218 in 2005. Of the five deprivation criteria for selection into the program, over the 2002 to 2005 period three of them increased in incidence while the remaining two showed a decrease. The two that decreased were IC4 "female having to work outside the home" (1 percent) and IC5 "lack of productive assets" (35 percent). Recall that the TUP program was supposed to provide assets to the households as part of their anti-poverty measures. Thus, a concern might arise on the external validity of the analysis because the deprivation criterion that BRAC considered for inclusion into the program was also part of the outcomes. (Note however, that this dataset does not incorporate the assets transferred by the program). As a robustness check, we have performed the same analysis presented in table 3.6 but without considering IC5 as part of the inclusion criteria; thus the poverty levels considered range from $k = 1$ to $k = 4$. In this case, the M_0 indicator for $k = 3$ now shows an overall reduction of 2 percent, composed of a reduction in 7% for the SUP poverty levels (from 26% in 2002 to 24% in 2005) and an increase in 7 percent on the poverty level of the control group NSUP, from 14 percent in 2002 to 15 percent in 2005. Table 3.11 in the appendix presents the results. For the intersection case of $k = 4$, poverty decreased by 2 percent for the whole sample, which again comprises a reduction in poverty for the SUP members of 5 percent, and an increase in poverty for the NSUP of 2 percent. Even though the reduction in poverty is not as big as before, the same pattern holds.

We now complement the poverty assessment of the change in the poverty measures with the Difference-in-Difference analysis of the TUP dataset, by varying the poverty cutoff and assessing the program impact on a broad set of outcomes.

3.5.2 Heterogeneity of the program impact

In order to analyze program heterogeneity, we vary the poverty line from $k = 5$ to $k = 1$ and assess: *(i)* whether the program impact is larger on a broader set of outcome variables; *(ii)* whether the program impact varies in magnitude for each outcome variable.

The initial 2002 CFPR/TUP Idata matrix is of the form (5067×5), that is, it contains 5067 households and five deprivation measures (d). They are all ordinal (binary characteristics) that have a value of one when the household is deprived in that dimension and zero when the household is not deprived. All deprivation measures have equal weights (w). BRAC did not set up an explicit weight criterion, although from Table 3.3 it is apparent that not all eligibility criteria have equal screening potential; in particular, IC1 selects 22% of households, while IC3 and IC5 select 1% each.

We present in figures 3.1 to 3.4 the program impact estimated by Difference in Differences[17] (DID). By using panel data we can solve the problem of unobserved additive heterogeneity at the household level, which may be present in different innate entrepreneurial abilities and risk preference. Because they are time invariant, we solve this problem by taking time differences. The equation estimated is

$$Y_{it} = \alpha_0 + \alpha_1 d_{2005} + \alpha_2 d_T + \beta(d_{2005} \times d_T) + \varepsilon_{it} \tag{3.1}$$

where Y_{it} is each outcome; d_{2005} is a dummy variable equal to one in 2005; d_T is a dummy variable indicating that household belonged to treatment group while the coefficient β of $d_{2005} \times d_T$ indicates the program impact. For the binary outcomes, we present marginal results from probit regressions with robust s.e. and clustering at household level, and for the count and continuous outcomes we present results from OLS regression, also correcting standard errors for heteroskedasticity and clustering at household level. Taken as a whole, the set of figures show that there seems to be a heterogeneous program impact. In particular, in all figures, the poorest households ($k = 5$) present a larger dispersion in the program impact

[17] Table 3.10 in the appendix shows the point estimates and their standard errors.

than the less poor.

Figure 3.1 shows that from the poverty line at $k = 4$ towards lower levels of k, the program impact is statistically larger for the poorest households than for the less poor on the set of outcomes related to income, land and housing. As these assets are very liquid, one might conjecture that it was easy for the program to successfully increase both the probability of having cash savings[18] and the increase in net income for their targeted group (the poorest of the poor). In contrast, while still significant, the increase in the proportion of households that own homestead land or have their roof made of tin is not different across poverty levels; they are respectively 0.08 and 0.13.

Figure 3.2 shows the program impact for the livestock and durable assets, where, with the exception of the poorest households ($k = 5$), who do not present a significant impact on a number of assets (number of duck and hens, fishing nets, rickshaw vans, chair, tables or beds), we find a statistically significant impact on most of them, which does not vary across poverty levels.

Figure 3.3 shows that the poorest households benefited the most in terms of the food security outcomes and shoe ownership. With respect to clothing and health-related measure, female clothing (number of sarees) increased as a result of the program for the households at poverty levels lower than $k = 5$, while the health-related measures show a significant impact for the households at poverty levels lower than $k = 3$.

Finally, Figure 3.4 shows the program impact on the empowerment and children-related outcomes, which show no improvement after the program across poverty status.

Connecting impact evaluation with AF multidimensional poverty measurement, we can evaluate the presence of heterogeneous impacts across the different degrees of deprivations (from $k = 5$ to $k = 1$). For all poverty levels, the Adjusted Headcount Ratio (M_0) indicator decreased over the period, and the decrease in poverty is larger for the treated households (the SUP group). The assessment of the program impact shows a larger dispersion for the poorest

[18] Cash savings are mandatory in many microfinance programs (Shams *et al.* 2010).

households (those who present all five deprivation characteristics), although the program impact on them is larger than on the less poor households for some outcomes (the probability of having cash savings (0.85 vs. 0.83), the net income increase (3481.95 Tk vs. 1759.97 Tk), food security outcomes (food availability 1.02 vs. 0.67), and meals twice a day (0.58 vs. 0.36). On the other hand, there is evidence that the health-related outcomes improve for the less poor households.

3.6 Conclusion

3.6.1 Summary and General Discussion of Results

This paper has assessed the design of targeting mechanisms in anti-poverty programs that are conceptualized as multidimensional, where many different deprivation dimensions are simultaneously addressed with the objective of pushing the households sustainably out of poverty. We propose a way of considering the type, number and specific sets of combinations to be satisfied, given the choice. By considering sequential tests of poverty, the targeting mechanism ensures that the individuals found are those looked for. Before conducting the poverty tests, we propose to perform a careful assessment of the relevancy of each eligibility criterion based on the proportion of the sample that is selected, and on the degree of correlation among the different deprivation criteria.

We have connected the multidimensional methodology to the impact evaluation of programs. Using Alkire and Foster (2007, 2011) poverty measures H (Headcount Ratio) and M_0 (Adjusted Headcount Ratio) we show how to evaluate the performance of an anti-poverty program by calculating the change in time of the poverty measures. In addition, the Alkire and Foster methodology (2007, 2011) allows us to decompose the poverty measure into its deprivation components to determine which one is responsible for the variation in poverty levels. Moreover, by varying the poverty line to different degrees of poverty, we assess the presence of heterogeneous impacts.

We show an example of our proposed measures with the analysis of the CFPR/TUP dataset. According to the poverty measure M_0, poverty decreased from 28 percent in 2002 to 23 percent in 2005. The decrease in poverty is larger for the TUP participants, on whom it varies from 36 percent to 25 percent. The poverty rate for the control group remains at their 2002 level of 21 percent.

By connecting the Alkire and Foster (2007, 2011) multidimensional methodology with program evaluation, we find a heterogeneous effect across households. Those with more deprivations experience a greater program impact on the net income increase variable (3481.95 Bangladeshi taka vs. 1759.97), on the probability of having a roof of good quality (0.24 vs. 0.13), on food availability (1.02 vs 0.67), on the probability of having meals twice a day (0.58 vs 0.37) and on the probability of owning shoes (0.27 vs. 0.15). In contrast, theTUP program has a significantly larger impact on health outcomes and female clothing for the less extremely poor among its selected participants.

Heterogeneity in outcomes is also evident in the dispersion of the impacts. The poorest households present a large variation in terms of impacts, (i.e. ownership of homestead land, quality of housing, livestock and some durables). Although the magnitudes are similar for all households, the dispersion of the impact is larger for the poorest.

3.6.2 Extensions and Concluding Remarks

Our more general conclusions are: The deprivation criteria to be used should relate to human capabilities and not be redundant among each other. The headcount ratio of the criteria helps in deciding the relevancy of each indicator, as one criterion that selected as participants all households would not be suitable at identifying the poorest among them. Sequential tests of poverty help identifying the ultra-poor or multidimensionally poor based on the initial deprivation dimensions met. Depending on available resources for the program, the number of deprivation requirements to qualify for a program can be set to a higher or a lower number and the targeted households will be poorer or less poor.

Table 3.1: Targeting, step 1, Identification of households according to the *disqualification* criteria established by BRAC, 2002

	Description	N	Headcount Ratio
EC1	Participant in another NGO	444	0.09
EC2	Recipient of a VGD card	127	0.03
EC3	No healthy female at home able to work for the program	48	0.01
	Full sample	**5067**	1.00

Notes: (1) BRAC established that if the household met any of the three previous conditions, they should not be selected for participating into the CFPR/TUP I program, because the specific purpose of the latter was to focus on household who were overlooked by previous programs (because of their extreme poverty condition or for some other characteristic that would prevent them from fully benefiting). The first two exclusion criteria aim at 'excluding-up' households: if they meet any of the two conditions they should not be qualified as ultra-poor. The last criterion focus on 'excluding-down' households, that is, if there is no healthy female able to work from the program, there is no point in including them for participating. Another solution has to be found for them.
(2) The Headcount Ratio indicates the prevalence of the deprivation across the full sample of households.
(3) The households identified by the exclusion criteria conform part of the type-2 errors in assignment (selecting ineligible participants).

Table 3.2: Targeting, step 2, Identification of households according to the *eligibility* criteria established by BRAC, 2002

	Description	N	Headcount Ratio
IC1	Less than ten decimals of land	4624	0.91
IC2	No male income earner at home	1893	0.37
IC3	Presence of child labor	740	0.15
IC4	Female having to work outside household	1627	0.32
IC5	No productive assets	2791	0.55
	Full sample	**5067**	1.00

To be classified as ultra-poor, the households had to meet at least three of the five eligibility criteria.

Table 3.3: Assessment of CFPR/TUP I targeting criteria, 2002

	Block	Criteria	SUP (T)	NSUP (C)	Headcount N	Ratio
BRAC's classification	(1)	SUP/NSUP	2375	2692	5067	
Satisfy sufficient number of criteria according to official BRAC rules:			**1152**	**784**	**1936**	
Satisfy all five inclusion criteria	(2)	IC1∪ IC2 ∪ IC3 ∪ IC4 ∪ IC5	96	47	143	0.03
Satisfy four criteria	(3)	IC1 ∪ IC2 ∪ IC3 ∪ IC4	63	49	112	0.02
		IC1 ∪ IC2 ∪ IC3 ∪ IC5	7	15	22	0.00
		IC1 ∪ IC2 ∪ IC4 ∪ IC5	547	287	834	0.16
		IC1 ∪ IC3 ∪ IC4 ∪ IC5	1	0	1	0.00
		IC2 ∪ IC3 ∪ IC4 ∪ IC5	5	2	7	0.00
Satisfy three criteria	(4)	IC1 ∪ IC2 ∪ IC3	7	10	17	0.00
		IC1 ∪ IC2 ∪ IC4	252	195	447	0.09
		IC1 ∪ IC2 ∪ IC5	46	63	109	0.02
		IC1 ∪ IC3 ∪ IC4	0	0	0	0.00
		IC1 ∪ IC3 ∪ IC5	114	92	206	0.04
		IC1 ∪ IC4 ∪ IC5	0	1	1	0.00
		IC2 ∪ IC3 ∪ IC4	3	8	11	0.00
		IC2 ∪ IC3 ∪ IC5	0	1	1	0.00
		IC2 ∪ IC4 ∪ IC5	11	14	25	0.00
		IC3 ∪ IC4 ∪ IC5	0	0	0	0.00
Satisfy insufficient number of criteria according to official BRAC rules:			**1223**	**1908**	**3131**	
Satisfy two criteria	(5)	IC1 ∪ IC2	25	54	79	0.02
		IC1 ∪ IC3	69	114	183	0.04
		IC1 ∪ IC4	0	0	0	0.00
		IC1 ∪ IC5	668	675	1343	0.27
		IC2 ∪ IC3	0	1	1	0.00
		IC2 ∪ IC4	16	30	46	0.01
		IC2 ∪ IC5	1	14	15	0.00
		IC3 ∪ IC4	0	0	0	0.00
		IC3 ∪ IC5	3	6	9	0.00
		IC4 ∪ IC5	0	0	0	0.00
Satisfy one criterion	(6)	IC1	384	743	1127	0.22
		IC2	6	18	24	0.00
		IC3	4	23	27	0.01
		IC4	0	0	0	0.00
		IC5	21	54	75	0.01
Do not meet any criterion	(7)	no criterion	26	176	202	0.04

Notes: (a) Block (1) presents BRAC's criteria of selection into SUP/NSUP groups.
(b) IC stands for 'Inclusion Criterion'.
Blocks (2) to (6) present the disaggregation of BRAC's criteria according to which and how many of the inclusion criteria are met. In this table they are also included the 106 households that met the initial exclusion restriction. If those 106 households were excluded, the sum of blocks (2), (3) and (4) would give the SB1/SB0 classification used in chapter 2.
The eligibility criteria are: (IC1) ownership of less than 10 decimals of land, (IC2) no male income earner at home, (IC3) children of school age having to work, (IC4) household dependent upon female domestic work outside the household, and (IC5) households having no productive assets.
Block (7) presents the number of households in the SUP/NSUP group that are selected despite not meeting any of the inclusion criteria.

Table 3.4: Selected households despite disqualification by exclusion criteria, 2002

	EC1	EC2	EC3	Total
Satisfy sufficient number of criteria for participation:	**28**	**21**	**20**	**69**
Satisfy all 5 criteria of inclusion	3	0	2	5
Satisfy 4 (any) criteria of inclusion	11	12	12	35
Satisfy 3 (any) criteria of inclusion	14	9	6	29
Satisfy insufficient number of criteria for participation:	**21**	**14**	**2**	**37**
Satisfy 2 (any) criteria of inclusion	14	9	1	24
Satisfy 1 (any) criteria of inclusion	6	5	1	12
Satisfy no criterion	1	0	0	1
Total treated (t)	49	35	22	106
Total (N)	444	127	48	619
proportion (t/N)	0.11	0.28	0.46	0.17

This table indicates for the year 2002 how many of the households selected for program participation met the exclusion criteria (and therefore should have been rejected as participants). The rows indicate how many of the inclusion criteria the households met. Additionally, one household met the EC1 and did not meet any of the inclusion criteria but was selected for program participation. The second-to-last row indicates how many of the households met each exclusion criterion. Recall that (EC1) participating in another NGO; (EC2) were recipient of a VGD food card; (EC3) there was no female able to work.

Description of Assets used in Robustness Check

Increase in net income	Summary variable to the answer of 'Last year employment and income related information - Increased net income in tk' for the TUP member
Cash savings (dummy)	Binary variable equal to one if the answer to the question 'Do you have any cash savings?' is yes.
Own homestead land (dummy)	Binary variable that equals one if the household owns homestead land
Total land owned	Total amount of land owned by the household (in tenth of acres)
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).
Number of cow/bulls	Number of assets owned, not including program transfers
Number of goat/sheeps	Number of assets owned, not including program transfers
Number of duck/hens	Number of assets owned, not including program transfers
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
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 tubewells	Number of assets owned
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]
Grain stocks (kg)	Stock of grain in kilograms owned by the household
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?

Description of Assets used in Robustness Check, cont.

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].
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.
Ratio of saree to lungi	Ratio of the female clothing to male clothing.
Shoes (dummy)	Answer to the question 'Do all household members have shoes/sandals?' yes[1] no[0].
Presence of girls working (dummy)	
Ability of girls to read and write a letter (dummy)	
Years of schooling of girls	
Presence of child labor (dummy)	Binary variable equal to one if the household declares that there are children under 15 working

Table 3.5: Initial values of assets for treated households, comparing across poverty thresholds, intersection approach ($k = 5$) vs. union approach ($k = 1$), 2002

	∩ case: $k = 5$				∪ case: $k = 1$			
			95% CI				95% CI	
	mean	sd	lower bound	upper bound	mean	sd	lower bound	upper bound
Income, Land and Housing								
Net income increase	14436.73	9022.67	12957.88	15915.57	15,901.07	9,893.10	15623.07	16179.07
Cash savings (dummy)	0.08	0.27	0.03	0.12	0.14	0.34	0.13	0.15
Own homestead land (dummy)	0.40	0.49	0.32	0.48	0.53	0.50	0.51	0.54
Total amount of land owned	1.33	2.14	0.98	1.68	2.94	7.76	2.72	3.16
Roof made of tin (dummy)	0.51	0.50	0.43	0.59	0.49	0.50	0.48	0.51
Livestock and Durables								
Number of cow/bulls	0.00	0.00	0.00	0.00	0.09	0.40	0.08	0.10
Number of goat/sheeps	0.00	0.00	0.00	0.00	0.10	0.46	0.09	0.11
Number of duck/hens	0.00	0.00	0.00	0.00	1.02	2.19	0.96	1.08
Number of fishing nets	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00
Number of big trees	0.54	2.53	0.12	0.95	0.67	3.40	0.58	0.77
Number of rickshaw/vans	0.00	0.00	0.00	0.00	0.03	0.25	0.02	0.04
Number of bicycles	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.01
Number of chair/tables	0.11	0.40	0.05	0.18	0.33	0.73	0.31	0.35
Number of beds	0.71	0.60	0.61	0.81	0.85	0.69	0.83	0.87
Number of radio/TVs	0.00	0.00	0.00	0.00	0.01	0.10	0.01	0.01
Number of quilt/blankets	0.00	0.00	0.00	0.00	0.03	0.18	0.02	0.03
Number of tubewells	0.02	0.14	0.00	0.04	0.03	0.16	0.02	0.03
Number of observations	143				4,865			

Notes: (1) ∩ case ($k = 5$): to qualify, all five deprivations must be present; ∪ case ($k = 1$): to qualify, at least one (any) deprivation must be present.
(2) From the full sample of 5067 observations, the ∪ case considers 4865, discarding the 202 households identified at the last block of table 3.3 that do not meet any criterion of inclusion.

Table 3.5: Initial values of assets for treated households, comparing across poverty thresholds, intersection approach ($k = 5$) vs. union approach ($k = 1$), 2002, cont.

	$\cap$ case: $k = 5$				$\cup$ case: $k = 1$			
			95% CI				95% CI	
	mean	sd	lower bound	upper bound	mean	sd	lower bound	upper bound
Food security, Health and Clothing								
Food availability	1.41	0.57	1.32	1.51	1.53	0.61	1.51	1.54
Grain stocks (kg)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Meals twice a day (dummy)	0.46	0.50	0.38	0.54	0.59	0.49	0.58	0.61
Self-reported health status	2.28	1.02	2.11	2.45	2.32	0.96	2.29	2.34
Health improvement	2.46	1.12	2.28	2.64	2.60	1.10	2.57	2.63
Number of sarees	1.65	0.51	1.57	1.73	1.79	0.57	1.78	1.81
Number of lungis	1.61	0.54	1.44	1.78	1.73	0.53	1.72	1.75
Ratio of saree to lungi	1.09	0.47	0.95	1.24	1.11	0.42	1.09	1.12
Shoes (dummy)	0.48	0.50	0.39	0.56	0.62	0.49	0.60	0.63
Empowerment and Children								
Presence of girls working	0.47	0.50	0.39	0.55	0.08	0.26	0.07	0.08
Ability of girls to read and write a letter	0.08	0.28	0.04	0.13	0.08	0.27	0.07	0.09
Years of schooling of girls	0.12	0.32	0.07	0.17	0.35	0.48	0.33	0.36
Presence of child labor	1.00	0.00	1.00	1.00	0.15	0.36	0.14	0.16
Number of observations	143				4,865			

Notes: (1) $\cap$ case ($k = 5$): to qualify, all five deprivations must be present; $\cup$ case ($k = 1$): to qualify, at least one (any) deprivation must be present.
(2) From the full sample of 5067 observations, the $\cup$ case considers 4865, discarding the 202 households identified at the last block of table 3.3 that do not meet any criterion of inclusion.

Table 3.6: Initial values of assets for households who met the disqualification criteria, 2002

	EC1		EC2		EC3	
	mean	sd	mean	sd	mean	sd
Income, Land and Housing						
Net income increase	17711.43	10335.59	14750.17	8925.30	8496.83	5200.30
Cash savings (dummy)	0.85	0.36	0.19	0.39	0.13	0.33
Own homestead land (dummy)	0.66	0.48	0.55	0.50	0.48	0.50
Total amount of land owned	6.59	13.94	4.60	9.23	2.93	7.42
Roof made of tin (dummy)	0.65	0.48	0.60	0.49	0.48	0.50
Livestock and Durables						
Number of cow/bulls	0.25	0.84	0.13	0.47	0.19	0.70
Number of goat/sheeps	0.14	0.47	0.09	0.37	0.17	0.48
Number of duck/hens	2.32	3.59	1.56	3.06	0.33	0.69
Number of fishing nets	0.00	0.05	0.00	0.00	0.00	0.00
Number of big trees	1.69	5.12	1.46	3.75	1.40	7.28
Number of rickshaw/vans	0.08	0.41	0.02	0.12	0.00	0.00
Number of bicycles	0.02	0.15	0.01	0.09	0.00	0.00
Number of chair/tables	0.64	1.05	0.20	0.60	0.08	0.35
Number of beds	1.20	0.80	0.94	0.67	0.65	0.64
Number of radio/TVs	0.03	0.18	0.01	0.09	0.00	0.00
Number of quilt/blankets	0.07	0.30	0.03	0.18	0.04	0.20
Number of tubewells	0.08	0.27	0.06	0.23	0.02	0.14
Food, Health and Clothing						
Food availability	1.78	0.67	1.43	0.60	1.31	0.55
Grain stocks (kg)	0.00	0.00	0.00	0.00	0.00	0.00
Meals twice a day (dummy)	0.75	0.43	0.67	0.47	0.42	0.50
Self-reported health status	2.30	0.96	2.03	0.93	1.60	0.76
Health improvement	2.55	1.15	2.41	1.09	1.83	1.00
Number of sarees	2.01	0.66	1.87	0.69	1.67	0.60
Number of lungis	1.88	0.50	1.74	0.52	1.67	0.52
Ratio of saree to lungi	1.13	0.44	1.11	0.48	1.17	0.41
Shoes (dummy)	0.65	0.48	0.57	0.50	0.67	0.48
Empowerment and Children						
Presence of girls working	0.07	0.25	0.08	0.27	0.06	0.24
Ability of girls to read and write a letter	0.13	0.33	0.10	0.30	0.00	0.00
Years of schooling of girls	0.40	0.49	0.43	0.50	0.21	0.43
Presence of child labor	0.14	0.35	0.14	0.35	0.13	0.33
Number of observations	444		127		48	

Note: EC stands for Exclusion Criterion; EC1 = participant in another NGO program; EC2 = recipient of a VGD card; EC3 = no female in the household able to exert work.

Table 3.7: Descriptive statistic: fraction of households out of poverty when varying the poverty line k

	2002			2005					
Poverty line	H	A	M_0	H	A	M_0	ΔH	ΔA	ΔM_0
Full sample (5067 observations)									
Intersection case:k=5	0.03	1.00	**0.03**	0.01	1.00	**0.01**	-0.67	0.00	-0.67
k=4	0.22	0.83	**0.18**	0.11	0.82	**0.09**	-0.50	-0.01	-0.50
k=3	0.38	0.73	**0.28**	0.34	0.67	**0.23**	-0.11	-0.08	-0.18
k=2	0.71	0.58	**0.41**	0.58	0.56	**0.32**	-0.18	-0.03	-0.22
Union case: k=1	0.96	0.48	**0.46**	0.96	0.42	**0.40**	0.00	-0.13	-0.13
Non-deprived	0.04	0.00	**0.00**	0.04	0.00	**0.00**			
SUP members (2375 observations)									
Intersection case:k=5	0.04	1.00	**0.04**	0.00	1.00	**0.00**	-0.94	0.00	-0.94
k=4	0.30	0.83	**0.25**	0.07	0.81	**0.06**	-0.76	-0.02	-0.76
k=3	0.49	0.74	**0.36**	0.39	0.64	**0.25**	-0.20	-0.14	-0.31
k=2	0.81	0.60	**0.49**	0.58	0.56	**0.32**	-0.29	-0.07	-0.34
Union case: k=1	0.99	0.53	**0.53**	0.97	0.41	**0.40**	-0.02	-0.22	-0.23
Non-deprived	0.01	0.00	**0.00**	0.03	0.00	**0.00**			
NSUP members (2692 observations)									
Intersection case:k=5	0.02	1.00	**0.02**	0.02	1.00	**0.02**	-0.06	0.00	-0.06
k=4	0.15	0.82	**0.12**	0.14	0.82	**0.11**	-0.08	0.00	-0.08
k=3	0.29	0.71	**0.21**	0.30	0.70	**0.21**	0.04	-0.02	0.02
k=2	0.62	0.55	**0.34**	0.58	0.56	**0.32**	-0.07	0.02	-0.05
Union case: k=1	0.93	0.43	**0.40**	0.94	0.42	**0.40**	0.01	-0.03	-0.02
Non-deprived	0.07	0.00	**0.00**	0.06	0.00	**0.00**			

Notes: (1) H is the headcount ratio, measured at each poverty level k.
(2) 'A' measures the number of deprivations that each household has divided by the total number of deprivations considered at each poverty line (the 'average deprivation share' in Alkire and Foster's terminology)
(3) M_0 is the adjusted headcount ratio defined by Alkire and Foster (2007) which combines the headcount ratio with the average deprivation share. It is analogous to the P_1 poverty gap in FGT class of poverty measures.

Table 3.8: Fraction of households out of poverty according to CFPR/TUP I disaggregation of deprivation criteria

Poverty line		Which criteria	Number of compliers		H		M_0	
			2002	2005	2002	2005	2002	2005
Satisfy sufficient number of criteria according to official BRAC rules:								
k=5	Satisfy all five inclusion criteria	IC1∪ IC2 ∪ IC3 ∪ IC4 ∪ IC5	**143**	**50**	**0.03**	**0.01**	**0.03**	**0.01**
k=4	Satisfy four criteria		**1119**	**542**	**0.22**	**0.11**	**0.18**	**0.09**
		IC1 ∪ IC2 ∪ IC3 ∪ IC4	112	199	0.02	0.04		
		IC1 ∪ IC2 ∪ IC3 ∪ IC5	22	9	0.00	0.00		
		IC1 ∪ IC2 ∪ IC4 ∪ IC5	834	281	0.16	0.06		
		IC1 ∪ IC3 ∪ IC4 ∪ IC5	1	0	0.00	0.00		
		IC2 ∪ IC3 ∪ IC4 ∪ IC5	7	3	0.00	0.00		
k=3	Satisfy three criteria		**1936**	**1737**	**0.38**	**0.34**	**0.28**	**0.23**
		IC1 ∪ IC2 ∪ IC3	17	39	0.00	0.01		
		IC1 ∪ IC2 ∪ IC4	447	960	0.09	0.19		
		IC1 ∪ IC2 ∪ IC5	109	51	0.02	0.01		
		IC1 ∪ IC3 ∪ IC4	0	0	0.00	0.00		
		IC1 ∪ IC3 ∪ IC5	206	109	0.04	0.02		
		IC1 ∪ IC4 ∪ IC5	1	0	0.00	0.00		
		IC2 ∪ IC3 ∪ IC4	11	25	0.00	0.00		
		IC2 ∪ IC3 ∪ IC5	1	2	0.00	0.00		
		IC2 ∪ IC4 ∪ IC5	25	9	0.00	0.00		
		IC3 ∪ IC4 ∪ IC5	0	0	0.00	0.00		

Notes: (1) H is the Heacount Ratio at each poverty level k. M_0 is the Adjusted Headcount Ratio that combines the average share of deprivations at each poverty level with the H. (2) The bold numbers indicate households that meet *at least* the amount of deprivation criteria defined by each poverty line. Thus: the 1119 households identified in 2002 by the poverty line $k = 4$ comprise 976 households who met exactly 4 deprivation criteria and 143 households who met 5 deprivation criteria.

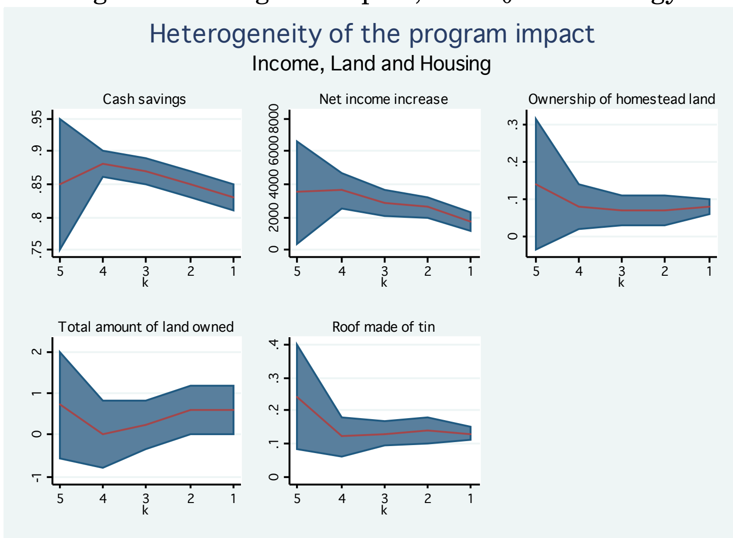
Fraction of households out of poverty according to CFPR/TUP I disaggregation of deprivation criteria, cont

Satisfy insufficient number of criteria according to official BRAC rules:

k=2	Satisfy two criteria	**3612**	**2931**	**0.71**	**0.58**	**0.41**	**0.32**
	IC1 ∪ IC2	79	183	0.02	0.04		
	IC1 ∪ IC3	183	479	0.04	0.09		
	IC1 ∪ IC4	0	0	0.00	0.00		
	IC1 ∪ IC5	1343	454	0.27	0.09		
	IC2 ∪ IC3	1	5	0.00	0.00		
	IC2 ∪ IC4	46	61	0.01	0.01		
	IC2 ∪ IC5	15	5	0.00	0.00		
	IC3 ∪ IC4	0	0	0.00	0.00		
	IC3 ∪ IC5	9	7	0.00	0.00		
	IC4 ∪ IC5	0	0	0.00	0.00		
k=1	Satisfy one criterion	**4865**	**4849**	**0.96**	**0.96**	**0.46**	**0.40**
	IC1	1127	1822	0.22	0.36		
	IC2	24	18	0.00	0.00		
	IC3	27	51	0.01	0.01		
	IC4	0	0	0.00	0.00		
	IC5	75	27	0.01	0.01		
k = 0	No deprivation criteria	**202**	**218**	**0.04**	**0.04**		

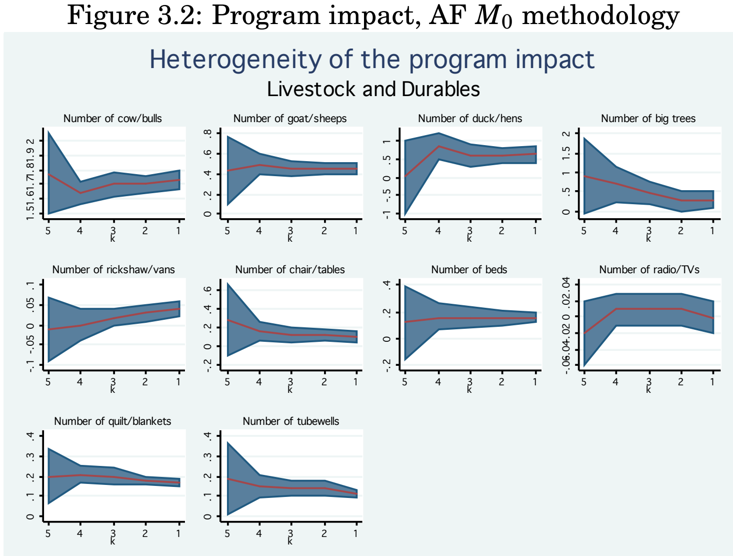
Notes: (1) H is the Heacount Ratio at each poverty level k. M_0 is the Adjusted Headcount Ratio that combines the average share of deprivations at each poverty level with the H. (2) The bold numbers indicate households that meet *at least* the amount of deprivation criteria defined by each poverty line. Thus: the 1119 households identified in 2002 by the poverty line $k = 4$ comprise 976 households who met exactly 4 deprivation criteria and 143 households who met 5 deprivation criteria.

Figure 3.1: Program impact, AF M_0 methodology



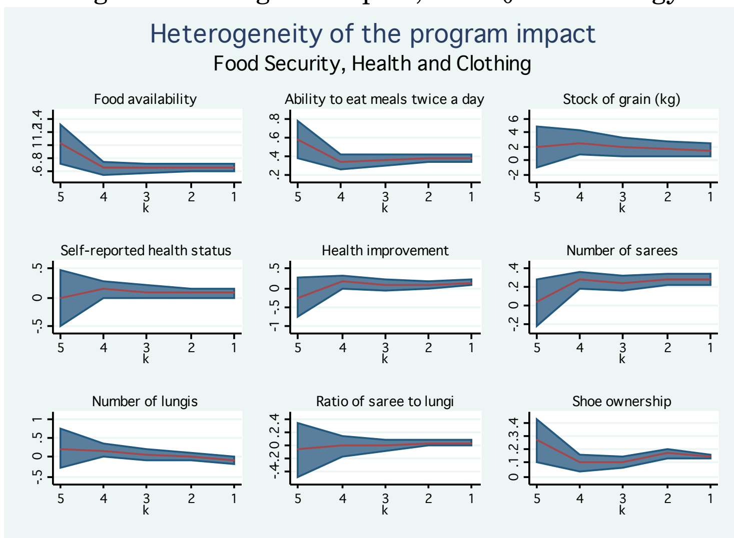
Notes: (1) k is the poverty line. It varies from $k = 5$ (intersection) to $k = 1$ (union). (2) Number of observations in 2002: for $k = 5$ are 143; $k = 4$ are 1119; $k = 3$ are 1936; $k = 2$ are 3612 and $k = 1$ are 4865. (3) M_0 is the adjusted Headcount ratio of Alkire and Foster (2007, 2011) that combines the average share of deprivations (the number of deprivations that each household has divided by the total number of deprivations considered) with the headcount ratio.

Figure 3.2: Program impact, AF M_0 methodology



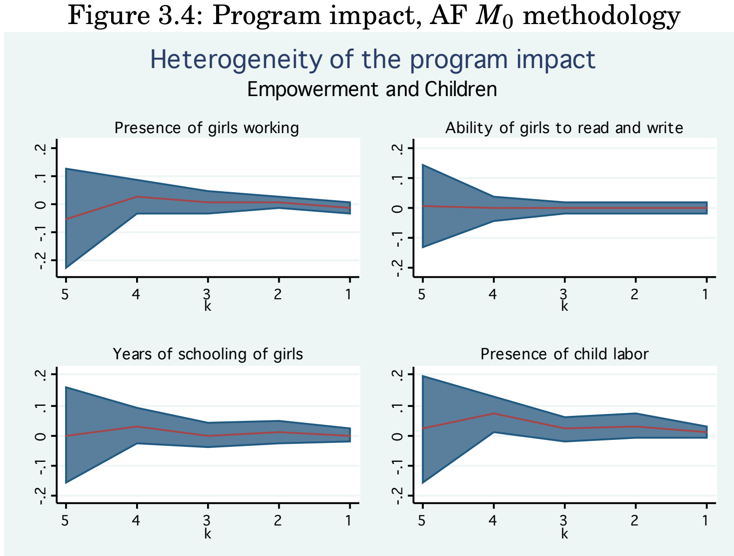
Notes: (1) k is the poverty line. It varies from $k = 5$ (intersection) to $k = 1$ (union). (2) Number of observations in 2002: for $k = 5$ are 143; $k = 4$ are 1119; $k = 3$ are 1936; $k = 2$ are 3612 and $k = 1$ are 4865. (3) M_0 is the adjusted Headcount ratio of Alkire and Foster (2007, 2011) that combines the average share of deprivations (the number of deprivations that each household has divided by the total number of deprivations considered) with the headcount ratio.

Figure 3.3: Program impact, AF M_0 methodology



Notes: (1) k is the poverty line. It varies from $k = 5$ (intersection) to $k = 1$ (union). (2) Number of observations in 2002: for $k = 5$ are 143; $k = 4$ are 1119; $k = 3$ are 1936; $k = 2$ are 3612 and $k = 1$ are 4865. (3) M_0 is the adjusted Headcount ratio of Alkire and Foster (2007, 2011) that combines the average share of deprivations (the number of deprivations that each household has divided by the total number of deprivations considered) with the headcount ratio.

Figure 3.4: Program impact, AF M_0 methodology



Notes: (1) k is the poverty line. It varies from $k = 5$ (intersection) to $k = 1$ (union). (2) Number of observations in 2002: for $k = 5$ are 143; $k = 4$ are 1119; $k = 3$ are 1936; $k = 2$ are 3612 and $k = 1$ are 4865. (3) M_0 is the adjusted Headcount ratio of Alkire and Foster (2007, 2011) that combines the average share of deprivations (the number of deprivations that each household has divided by the total number of deprivations considered) with the headcount ratio.

Table 3.9: Program impact, DID Estimation Results

	∩ k=5	k=4	k=3	k=2	∪ k=1	SUP/NSUP
Income, Land and Housing						
Net income increase	3481.95	3617.54	2847.20	2616.59	1759.97	1172.79
	(1578.73)**	(532.40)***	(424.20)***	(320.80)***	(282.59)***	(301.95)***
Cash savings (dummy)	0.85	0.88	0.87	0.85	0.83	0.82
	(0.05)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Own homestead land (dummy)	0.14	0.08	0.07	0.07	0.08	0.09
	(0.09)	(0.03)**	(0.02)***	(0.02)***	(0.01)***	(0.01)***
Total amount of land owned	0.72	0.02	0.23	0.58	0.61	1.47
	(0.65)	(0.41)	(0.30)	(0.30)*	(0.30)**	(0.36)***
Roof made of tin (dummy)	0.24	0.12	0.13	0.14	0.13	0.14
	(0.08)***	(0.03)***	(0.02)***	(0.02)***	(0.01)***	(0.01)***
Livestock and Durables						
Number of cow/bulls	1.77	1.64	1.70	1.70	1.73	1.74
	(0.14)***	(0.04)***	(0.04)***	(0.03)***	(0.03)***	(0.03)***
Number of goat/sheeps	0.43	0.50	0.45	0.45	0.46	0.46
	(0.17)**	(0.05)***	(0.04)***	(0.03)***	(0.03)***	(0.03)***
Number of duck/hens	0.02	0.83	0.58	0.58	0.61	0.66
	(0.50)	(0.17)***	(0.15)***	(0.11)***	(0.11)***	(0.12)***
Number of fishing nets	0.00	0.01	0.02	0.05	0.03	0.03
	(0.07)	(0.01)	(0.02)	(0.02)***	(0.02)*	(0.02)
Number of big trees	0.90	0.69	0.45	0.26	0.30	0.55
	(0.50)*	(0.23)***	(0.15)***	(0.13)**	(0.10)***	(0.16)***
Number of rickshaw/vans	-0.01	0.00	0.02	0.03	0.04	0.05
	(0.04)	(0.02)	(0.01)	(0.01)***	(0.01)***	(0.01)***
Number of bicycles	-0.01	0.00	0.00	0.00	0.00	0.01
	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of chair/tables	0.28	0.16	0.13	0.13	0.10	0.11
	(0.19)	(0.05)***	(0.04)***	(0.03)***	(0.03)***	(0.03)***
Number of beds	0.12	0.16	0.16	0.15	0.16	0.17
	(0.14)	(0.05)***	(0.04)***	(0.03)***	(0.02)***	(0.02)***
Number of radio/TVs	-0.02	0.01	0.01	0.01	0.00	0.01
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Number of quilt/blankets	0.20	0.21	0.20	0.18	0.17	0.16
	(0.07)***	(0.02)***	(0.02)***	(0.01)***	(0.01)***	(0.01)***
Number of tubewells	0.19	0.15	0.14	0.14	0.11	0.09
	(0.09)**	(0.03)***	(0.02)***	(0.02)***	(0.01)***	(0.01)***
Number of observations	286	2238	3872	7224	9730	10134

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

(2) Estimates come from Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions

Table 3.9: Program Impact, DID Estimation Results, cont.

	∩ k=5	k=4	k=3	k=2	∪ k=1	SUP/NSUP
Food security, Health and Clothing						
Food availability	1.02	0.65	0.65	0.66	0.67	0.67
	(0.15)***	(0.05)***	(0.04)***	(0.03)***	(0.03)***	(0.03)***
Grain stocks (kg)	1.88	2.50	1.91	1.70	1.51	1.28
	(1.52)	(0.90)***	(0.70)***	(0.57)***	(0.51)***	(0.51)**
Meals twice a day (dummy)	0.58	0.33	0.36	0.38	0.37	0.36
	(0.10)***	(0.04)***	(0.03)***	(0.02)***	(0.02)***	(0.02)***
Self-reported health status	-0.01	0.14	0.09	0.08	0.07	0.08
	(0.25)	(0.08)*	(0.06)	(0.04)*	(0.04)*	(0.04)**
Health improvement	-0.25	0.17	0.08	0.10	0.14	0.14
	(0.27)	(0.09)*	(0.07)	(0.05)*	(0.04)***	(0.04)***
Number of sarees	0.04	0.28	0.24	0.28	0.29	0.29
	(0.13)	(0.05)***	(0.04)***	(0.03)***	(0.03)***	(0.02)***
Number of lungis	0.23	0.19	0.08	0.01	-0.08	-0.10
	(0.27)	(0.09)**	(0.07)	(0.05)	(0.04)**	(0.04)***
Ratio of saree to lungi	-0.07	-0.01	0.00	0.04	0.04	0.03
	(0.21)	(0.08)	(0.05)	(0.02)	(0.02)**	(0.02)*
Shoes (dummy)	0.27	0.10	0.11	0.17	0.15	0.15
	(0.08)***	(0.03)***	(0.02)***	(0.02)***	(0.01)***	(0.01)***
Empowerment and Children						
Presence of girls working	-0.05	0.03	0.01	0.01	-0.01	-0.01
	(0.09)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Ability of girls to read and write a letter	0.01	0.00	0.00	0.00	0.00	0.00
	(0.07)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Years of schooling of girls	0.00	0.03	0.00	0.01	0.00	0.01
	(0.08)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Presence of child labor	0.02	0.07	0.02	0.03	0.01	0.00
	(0.09)	(0.03)**	(0.02)	(0.02)*	(0.01)	(0.01)
Number of observations	286	2238	3872	7224	9730	10134

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

(2) Estimates come from Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions

Table 3.10: Program impact, DID Results for BRAC's combination of criteria

	IC1∪ IC2∪ IC3∪ IC4∪ IC5	IC1∪ IC2∪ IC3∪ IC4	IC1∪ IC2∪ IC3∪ IC5	IC1∪ IC2∪ IC4∪ IC5	IC1∪ IC2∪ IC3	IC1∪ IC2∪ IC4
Income, Land and Housing						
Net income increase	3481.95	4638.73	-563.81	4264.01	4928.70	2102.82
	(1578.73)**	(2115.07)**	(5139.17)	(502.88)***	(4091.19)	(806.08)***
Cash savings (dummy)	0.85	0.81		0.90		0.84
	(0.05)***	(0.05)***		(0.02)***		(0.02)***
Own homestead land (dummy)	0.14	-0.03	0.00	0.08	0.14	0.02
	(0.09)	(0.09)	(0.27)	(0.04)**	(0.19)	(0.05)
Total amount of land owned	0.72	0.27	0.26	-0.03	-1.23	0.55
	(0.65)	(0.59)	(1.83)	(0.54)	(3.28)	(0.45)
Roof made of tin (dummy)	0.24	0.17	0.10	0.09		0.16
	(0.08)***	(0.08)**	(0.26)	(0.04)**		(0.04)***
Livestock and Durables						
Number of cow/bulls	1.77	1.67	1.58	1.61	2.34	1.78
	(0.14)***	(0.16)***	(0.47)***	(0.05)***	(0.43)***	(0.09)***
Number of goat/sheeps	0.43	0.41	0.86	0.52	-0.24	0.35
	(0.17)**	(0.19)**	(0.84)	(0.05)***	(0.54)	(0.09)***
Number of duck/hens	0.02	0.65	0.11	0.99	0.24	0.12
	(0.50)	(0.87)	(1.08)	(0.18)***	(1.96)	(0.43)
Number of fishing nets	0.00	0.01	-0.07	0.01	0.29	0.00
	(0.07)	(0.05)	(0.07)	(0.01)	(0.18)	(0.06)
Number of big trees	0.90	1.93	-0.60	0.49	-2.29	-0.02
	(0.50)*	(1.30)	(0.89)	(0.22)**	(3.09)	(0.27)
Number of rickshaw/vans	-0.01	-0.18	0.14	0.02	0.00	0.02
	(0.04)	(0.18)	(0.14)	(0.01)**	(.)	(0.02)
Number of bicycles	-0.01	0.00	0.08	0.00	0.00	-0.01
	(0.02)	(0.02)	(0.16)	(.)	(.)	(0.01)
Number of chair/tables	0.28	0.28	-0.17	0.14	0.13	0.10
	(0.19)	(0.24)	(0.45)	(0.05)***	(0.45)	(0.10)
Number of beds	0.12	0.41	-0.37	0.17	0.37	0.14
	(0.14)	(0.15)***	(0.47)	(0.05)***	(0.47)	(0.08)*
Number of radio/TVs	-0.02	0.03	-0.07	0.01	0.00	0.01
	(0.02)	(0.04)	(0.07)	(0.01)**	(.)	(0.01)
Number of quilt/blankets	0.20	0.28	0.09	0.21	0.19	0.21
	(0.07)***	(0.09)***	(0.23)	(0.03)***	(0.21)	(0.04)***
Number of tubewells	0.19	0.07	0.39	0.18	-0.11	0.16
	(0.09)**	(0.09)	(0.20)*	(0.03)***	(0.25)	(0.05)***
Number of observations	286	224	44	1668	34	894

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

(2) Estimates come from Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions.

(3) The blank spaces occur when there is perfect prediction for such criteria combination.

(4) The intersection of BRAC's treatment/control group (SUP/NSUP) with the criteria combination established in the first row gives us our treatment/control subgroups. Estimations were performed for those subgroups with at least 22 observations.

Table 3.10 Program Impact, DID results, BRAC's combination of criteria, cont.

	IC1∪ IC2∪ IC3∪ IC4∪ IC5	IC1∪ IC2∪ IC3∪ IC4	IC1∪ IC2∪ IC3∪ IC5	IC1∪ IC2∪ IC4∪ IC5	IC1∪ IC2∪ IC3	IC1∪ IC2∪ IC4
Food security, Health and Clothing						
Food availability	1.02	0.61	0.32	0.62	0.07	0.72
	(0.15)***	(0.19)***	(0.44)	(0.06)***	(0.59)	(0.09)***
Grain stocks (kg)	1.88	2.54	0.00	2.66	0.00	1.85
	(1.52)	(2.55)	(.)	(1.12)**	(.)	(1.82)
Meals twice a day (dummy)	0.58	0.32	0.51	0.29	0.41	0.35
	(0.10)***	(0.12)**	(0.24)	(0.05)***	(0.27)	(0.06)***
Self-reported health status	-0.01	0.07	0.73	0.18	0.30	0.00
	(0.25)	(0.25)	(0.49)	(0.09)*	(0.69)	(0.14)
Health improvement	-0.25	0.55	0.25	0.21	1.04	-0.17
	(0.27)	(0.30)*	(0.49)	(0.11)*	(0.66)	(0.14)
Number of sarees	0.04	0.24	0.52	0.32	0.03	0.18
	(0.13)	(0.16)	(0.39)	(0.06)***	(0.34)	(0.08)**
Number of lungis	0.23	0.32	0.91	0.19	1.03	0.04
	(0.27)	(0.29)	(0.81)	(0.10)*	(1.04)	(0.14)
Ratio of saree to lungi	-0.07	-0.19	-0.41	0.11	-0.34	0.11
	(0.21)	(0.17)	(0.31)	(0.11)	(0.49)	(0.11)
Shoes (dummy)	0.27	0.13	-0.45	0.06		0.08
	(0.08)***	(0.10)	(0.25)	(0.04)		(0.05)
Empowerment and Children						
Presence of girls working	-0.05	0.05	0.21	0.01	-0.19	0.01
	(0.09)	(0.10)	(0.28)	(0.02)	(0.23)	(0.03)
Ability of girls to read and write a letter	0.01	-0.05	0.99	0.01		0.00
	(0.07)	(0.08)	(0.01)***	(0.02)		(0.03)
Years of schooling of girls	0.00	-0.02	0.14	0.06	0.00	-0.02
	(0.08)	(0.06)	(0.18)	(0.04)*	(0.00)	(0.05)
Presence of child labor	0.02	0.10	-0.25	0.04	-0.07	-0.04
	(0.09)	(0.09)	(0.22)	(0.02)	(0.25)	(0.03)
Number of observations	286	224	44	1668	34	894

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

(2) Estimates come from Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions.

(3) The blank spaces occur when there is perfect prediction for such criteria combination.

(4) The intersection of BRAC's treatment/control group (SUP/NSUP) with the criteria combination established in the first row gives us our treatment/control subgroups. Estimations were performed for those subgroups with at least 22 observations.

Table 3.10b Program Impact, DID results, BRAC's combination of criteria

	IC1∪ IC2∪ IC5	IC1∪ IC3∪ IC5	IC2∪ IC4∪ IC5	IC1∪ IC2	IC1∪ IC3	IC1∪ IC5
Income, Land and Housing						
Net income increase	5332.48	2105.01	-460.55	5154.39	4301.78	3680.25
	(1477.12)***	(1423.74)	(3250.30)	(2950.32)*	(2327.82)*	(470.85)***
Cash savings (dummy)	0.90					0.83
	(0.05)***					(0.02)***
Own homestead land (dummy)	0.23	0.04	0.17	-0.06	0.03	0.08
	(0.11)**	(0.08)	(0.08)	(0.13)	(0.06)	(0.03)***
Total amount of land owned	-0.24	-0.17	4.53	1.16	-1.56	0.60
	(1.31)	(0.78)	(5.56)	(7.55)	(1.12)	(0.37)
Roof made of tin (dummy)	0.08	0.22		0.15	0.17	0.14
	(0.11)	(0.06)***		(0.10)	(0.07)**	(0.03)***
Livestock and Durables						
Number of cow/bulls	1.70	1.74	2.09	2.03	1.85	1.65
	(0.16)***	(0.13)***	(0.25)***	(0.20)***	(0.15)***	(0.05)***
Number of goat/sheeps	0.27	0.39	0.18	0.16	0.25	0.49
	(0.11)**	(0.12)***	(0.12)	(0.31)	(0.18)	(0.05)***
Number of duck/hens	0.39	0.63	0.57	0.23	0.42	0.68
	(0.42)	(0.36)*	(0.90)	(0.88)	(0.71)	(0.18)***
Number of fishing nets	0.14	0.13	-0.07	0.13	0.00	0.13
	(0.11)	(0.08)	(0.07)	(0.12)	(0.06)	(0.04)***
Number of big trees	0.72	0.06	2.86	-3.81	0.88	0.05
	(0.39)*	(0.41)	(1.95)	(4.04)	(1.08)	(0.12)
Number of rickshaw/vans	0.09	0.08	0.09	0.02	0.07	0.05
	(0.04)**	(0.04)**	(0.09)	(0.08)	(0.06)	(0.02)***
Number of bicycles	0.01	0.00	0.09	0.04	0.01	0.01
	(0.03)	(0.02)	(0.09)	(0.04)	(0.02)	(0.01)
Number of chair/tables	0.10	0.16	1.14	-0.14	0.33	0.12
	(0.16)	(0.15)	(0.37)***	(0.25)	(0.17)**	(0.05)**
Number of beds	0.30	0.16	0.31	0.06	0.22	0.17
	(0.13)**	(0.12)	(0.18)	(0.17)	(0.13)*	(0.04)***
Number of radio/TVs	0.00	-0.01	-0.09	-0.02	0.00	0.02
	(.)	(0.03)	(0.09)	(0.03)	(0.03)	(0.01)
Number of quilt/blankets	0.52	0.06	0.00	0.33	0.21	0.13
	(0.11)***	(0.06)	(0.11)	(0.16)**	(0.06)***	(0.02)***
Number of tubewells	0.26	0.08	0.28	0.04	0.03	0.18
	(0.09)***	(0.07)	(0.20)	(0.11)	(0.08)	(0.03)***
Number of observations	218	412		158	366	2686

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

(2) Estimates come from Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions.

(3) The blank spaces occur when there is perfect prediction for such criteria combination.

(4) The intersection of BRAC's treatment/control group (SUP/NSUP) with the criteria combination established in the first row gives us our treatment/control subgroups. Estimations were performed for those subgroups with at least 22 observations.

Table 3.10b Program Impact, DID results, BRAC's combination of criteria, cont.

	IC1∪ IC2∪ IC5	IC1∪ IC3∪ IC5	IC2∪ IC4∪ IC5	IC1∪ IC2	IC1∪ IC3	IC1∪ IC5
Food security, Health and Clothing						
Food availability	0.82	0.44	0.97	0.95	0.60	0.66
	(0.17)***	(0.12)***	(0.44)**	(0.23)***	(0.13)***	(0.05)***
Grain stocks (kg)	0.00	0.00	-0.71	-7.41	0.00	2.20
	(.)	(.)	(0.73)	(7.46)	(.)	(0.92)**
Meals twice a day (dummy)	0.57	0.42	0.45	0.38	0.39	0.39
	(0.09)***	(0.08)***	(0.19)*	(0.14)**	(0.07)***	(0.03)***
Self-reported health status	0.33	-0.13	-0.43	-0.13	0.05	0.06
	(0.26)	(0.20)	(0.53)	(0.27)	(0.19)	(0.07)
Health improvement	0.37	-0.14	-0.08	-0.20	0.01	0.16
	(0.29)	(0.20)	(0.60)	(0.32)	(0.24)	(0.08)**
Number of sarees	0.49	0.05	0.20	0.10	0.25	0.35
	(0.16)***	(0.12)	(0.29)	(0.18)	(0.12)**	(0.05)***
Number of lungis	0.01	0.27	0.00	0.00	0.05	0.30
	(0.30)	(0.12)**	(0.50)	(0.39)	(0.12)	(0.05)***
Ratio of saree to lungi	0.14	-0.12	0.43	-0.09	0.16	0.04
	(0.14)	(0.07)*	(0.26)	(0.30)	(0.08)*	(0.03)
Shoes (dummy)	0.21	0.15			0.24	0.23
	(0.06)**	(0.09)			(0.06)***	(0.02)***
Empowerment and Children						
Presence of girls working	-0.05	-0.02		0.09	-0.14	0.03
	(0.05)	(0.09)		(0.08)	(0.08)	(0.02)
Ability of girls to read and write a letter	0.79	-0.02		-0.04	0.06	0.00
	(0.10)***	(0.06)		(0.04)	(0.08)	(0.02)
Years of schooling of girls	0.12	-0.06		-0.03	0.08	0.01
	(0.09)	(0.07)		(0.12)	(0.07)	(0.02)
Presence of child labor	-0.04	-0.01		0.11	0.03	0.04
	(0.06)	(0.07)		(0.10)	(0.08)	(0.02)**
Number of observations	218	412		158	366	2686

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

(2) Estimates come from Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions.

(3) The blank spaces occur when there is perfect prediction for such criteria combination.

(4) The intersection of BRAC's treatment/control group (SUP/NSUP) with the criteria combination established in the first row gives us our treatment/control subgroups. Estimations were performed for those subgroups with at least 22 observations.

Table 3.10c Program Impact, DID results, BRAC's combination of criteria

	IC1	IC2	IC3	IC5	SUP/NSUP
Income, Land and Housing					
Net income increase	2059.57	8457.78	2161.59	394.91	1172.79
	(541.45)***	(5138.70)	(7367.27)	(3424.23)	(301.95)***
Cash savings (dummy)	0.75			0.68	0.82
	(0.01)***			(0.09)***	(0.01)***
Own homestead land (dummy)	0.03	0.06	0.05	-0.02	0.09
	(0.03)	(0.18)	(0.12)	(0.16)	(0.01)***
Total amount of land owned	-0.21	10.50	-33.53	0.75	1.47
	(0.42)	(9.11)	(25.96)	(6.62)	(0.36)***
Roof made of tin (dummy)	0.10		0.07	0.05	0.14
	(0.03)***		(0.19)	(0.14)	(0.01)***
Livestock and Durables					
Number of cow/bulls	1.80	2.72	2.52	1.73	1.74
	(0.07)***	(0.56)***	(0.69)***	(0.33)***	(0.03)***
Number of goat/sheeps	0.43	0.83	-0.24	0.50	0.46
	(0.08)***	(0.42)*	(0.55)	(0.30)	(0.03)***
Number of duck/hens	0.51	-0.67	0.85	-0.06	0.66
	(0.30)*	(1.20)	(1.41)	(1.11)	(0.12)***
Number of fishing nets	0.01	0.11	0.08	0.18	0.03
	(0.04)	(0.17)	(0.24)	(0.24)	(0.02)
Number of big trees	0.30	4.78	-9.84	-0.17	0.55
	(0.16)*	(2.55)*	(6.25)	(0.99)	(0.16)***
Number of rickshaw/vans	0.09	-0.11	-0.04	-0.01	0.05
	(0.03)***	(0.08)	(0.04)	(0.06)	(0.01)***
Number of bicycles	0.02	0.00	0.04	0.01	0.01
	(0.01)	(.)	(0.08)	(0.05)	(0.00)
Number of chair/tables	0.11	0.06	-0.15	-0.47	0.11
	(0.07)	(0.77)	(0.52)	(0.22)**	(0.03)***
Number of beds	0.16	0.00	0.00	0.46	0.17
	(0.05)***	(0.60)	(0.46)	(0.24)*	(0.02)***
Number of radio/TVs	0.00	-0.06	0.09	-0.02	0.01
	(0.01)	(0.06)	(0.09)	(0.03)	(0.01)
Number of quilt/blankets	0.09	0.94	-0.17	0.17	0.16
	(0.03)***	(0.49)*	(0.11)	(0.10)*	(0.01)***
Number of tubewells	0.10	0.28	0.14	0.49	0.09
	(0.03)***	(0.22)	(0.26)	(0.12)***	(0.01)***
Number of observations	2254	48	54	150	10134

(1) Standard errors in parentheses. Significance levels are denoted as
$***p < 0.01; **p < 0.05; *p < 0.1$
(2) Estimates come from Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions.
(3) The blank spaces occur when there is perfect prediction for such criteria combination.
(4) The intersection of BRAC's treatment/control group (SUP/NSUP) with the criteria combination established in the first row gives us our treatment/control subgroups. Estimations were performed for those subgroups with at least 22 observations.

Table 3.10c Program Impact, DID results, BRAC's combination of criteria, cont.

	IC1	IC2	IC3	IC5	SUP/NSUP
Food Security, Health and Clothing					
Food availability	0.70	0.39	1.17	0.43	0.67
	(0.06)***	(0.58)	(0.56)**	(0.26)	(0.03)***
Grain stocks (kg)	1.43	-2.22	0.00	0.00	1.28
	(1.19)	(2.28)	(.)	(.)	(0.51)**
Meals twice a day (dummy)	0.37	0.36	0.24	0.50	0.36
	(0.03)***	(0.21)	(0.16)	(0.08)***	(0.02)***
Self-reported health status	0.04	0.50	-0.85	0.06	0.08
	(0.09)	(0.64)	(0.83)	(0.35)	(0.04)**
Health improvement	0.23	0.22	0.37	-0.12	0.14
	(0.09)**	(0.56)	(0.86)	(0.48)	(0.04)***
Number of sarees	0.30	0.56	0.16	0.55	0.29
	(0.06)***	(0.50)	(0.30)	(0.18)***	(0.02)***
Number of lungis	0.19	0.37	-0.14	0.11	-0.10
	(0.06)***	(0.95)	(0.33)	(0.21)	(0.04)***
Ratio of saree to lungi	0.03	0.07	0.00	0.28	0.03
	(0.03)	(0.14)	(0.30)	(0.15)*	(0.02)*
Shoes (dummy)	0.08	-0.99	-0.16	0.16	0.15
	(0.03)**	(0.01)***	(0.13)	(0.13)	(0.01)***
Empowerment and Children					
Presence of girls working	0.02	0.11		0.01	-0.01
	(0.02)	(0.16)		(0.09)	(0.01)
Ability of girls to read and write a letter	0.03	0.06	0.97	-0.06	0.00
	(0.03)	(0.12)	(0.02)***	(0.06)	(0.01)
Years of schooling of girls	0.01	0.16	-0.13	0.08	0.01
	(0.03)	(0.14)	(0.25)	(0.15)	(0.01)
Presence of child labor	0.05	0.22		-0.04	0.00
	(0.02)**	(0.21)		(0.09)	(0.01)
Number of observations	2254	48	54	150	10134

(1) Standard errors in parentheses. Significance levels are denoted as
$***p < 0.01; **p < 0.05; *p < 0.1$
(2) Estimates come from Difference-in-Difference specification and for the binary outcomes they correspond to marginal effects from probit regressions.
(3) The blank spaces occur when there is perfect prediction for such criteria combination.
(4) The intersection of BRAC's treatment/control group (SUP/NSUP) with the criteria combination established in the first row gives us our treatment/control subgroups. Estimations were performed for those subgroups with at least 22 observations.

Table 3.11: Descriptive statistic: fraction of households out of poverty when varying the poverty line k, without considering IC5

	2002			2005					
Poverty line	H	A	M_0	H	A	M_0	ΔHR	ΔA	ΔM_0
Full sample (5067 observations)									
k=4	0.050	0.800	**0.040**	0.049	0.80	**0.039**	-0.02	0.00	-0.02
k=3	0.315	0.632	**0.199**	0.309	0.63	**0.195**	-0.02	0.00	-0.02
k=2	0.443	0.565	**0.250**	0.486	0.55	**0.266**	0.10	-0.03	0.06
Union case: k=1	0.945	0.371	**0.351**	0.952	0.38	**0.359**	0.01	0.02	0.02
Non-deprived	0.055	0.000	**0.00**	0.048	0.00	**0.00**			
SUP members (2375 observations)									
I=4	0.07	0.800	**0.05**	0.06	0.80	**0.05**	-0.05	0.00	-0.05
k=3	0.41	0.632	**0.26**	0.38	0.63	**0.24**	-0.07	0.00	-0.07
k=2	0.53	0.581	**0.31**	0.57	0.56	**0.32**	0.07	-0.04	0.03
Union case: k=1	0.98	0.406	**0.40**	0.97	0.41	**0.40**	-0.01	0.01	0.00
Non-deprived	0.02	0.000	**0.00**	0.03	0.00				
NSUP members (2692 observations)									
k=4	0.04	0.80	**0.03**	0.04	0.80	**0.03**	0.02	0.00	0.02
k=3	0.23	0.63	**0.14**	0.24	0.63	**0.15**	0.07	0.00	0.07
k=2	0.37	0.54	**0.20**	0.41	0.54	**0.22**	0.13	-0.02	0.12
Union case: k=1	0.91	0.34	**0.31**	0.93	0.35	**0.33**	0.02	0.03	0.05

Notes: (1) This table varies the poverty levels from (k = 1) to (k = 4), excluding into the consideration the inclusion criterion 5 (IC5) 'lack of productive assets'.
(2) H is the headcount ratio, measured at each poverty level k.
(3) 'A' measures the number of deprivations that each household has divided by the total number of deprivations considered at each poverty line (the 'average deprivation share' in Alkire and Foster's terminology)
(4) M_0 is the adjusted headcount ratio defined by Alkire and Foster (2007) which combines the headcount ratio with the average deprivation share. It is analogous to the P_1 poverty gap in FGT class of poverty measures.

Chapter 4

Sticky Floors and Glass Ceilings: An Assessment of the Gender Wage Gap in Latin America

4.1 Introduction

Measurement of gender wage gaps (GWG) has long been an important concern among policy makers and in the academic literature. Policy makers generally agree that reducing gender inequality is a top priority. For instance, promoting gender equality and empowering women is one of the UN Millenium Development Goals. In most developing countries gender inequality persists, as the 2010 report of the United Nations declare: *"Despite progress made, men continue to outnumber women in paid employment, and women are often relegated to vulnerable forms of employment [...] Even when women are employed, they are typically paid less and have less financial and social security than men"*[1].

Since the seminal contributions of Oaxaca (1973) and Blinder (1973) the differences in wages are seen as a mixture of differences in endowments and differences in the returns to these endowments. Many papers have looked at these differences in developed (Albrecht *et al.* (2003), Albrecht *et al.* (2006), Arulampalam *et al.* (2007), Nicodemo (2009)) and developing regions (Atal (2009), Badel and Peña (2008) Borráz and Robano (2010), Bucheli and San

[1] `http://www.un.org/millenniumgoals/pdf/MDG_FS_3_EN.pdf`

Román (2004), Wei Chi and Li Bo (2007), Chong and Ñopo (2007), Hoyos *et al.* (2010), Panizza and Qiang (2005), Perticará and Astudillo (2009), Robano (2005), Tenjo *et al.* (2005)). Many of these studies assess what happens to the GWG at different points of the wage distribution, although some of them just analyze the mean difference[2]. Albrecht *et al.* (2003) find that the wage gap at the top of the distribution is notably larger than at the median in Sweden (*glass ceiling*). Arulampalam *et al.* (2005) show evidence that for some european countries the GWG can be larger at the bottom of the wage distribution (*sticky floors*). Measuring the GWG at different points of the wage distribution may have important policy implications, as policies targeted to the average population may fail to affect those at the bottom or top of the distribution.

In this paper we measure the GWG in twelve Latin American labor markets and we explore the source of such differences. With that purpose, we construct a wage gap (the difference between the hourly wage of males and the hourly wage of females) and employ unconditional quantile decomposition methods to assess what part of the wage gap is explained by the different levels of human capital (and of some demographic characteristics) and what part is left unexplained.

We find that in Latin America the GWG is heterogeneous across the wage distribution. After controlling for workers' characteristics, there is a significant wage gap at the bottom of the distribution that is statistically significant in most countries. Interestingly, the size of the GWG at the bottom is highly correlated with measures of economic development (per capita GDP) and income inequality (Gini). Poorer countries and countries with higher levels of income inequality have higher gender wage differentials at the tenth percentile of the wage distribution. On the other hand, the wage gap at the top of the wage distribution (90^{th} percentile) is larger in richer countries and in countries where income distribution is more even (although not statistically significant on the latter).

Badel and Peña (2008) analyze the colombian gender wage gap using an implementation of Machado and Mata (2005) decomposition technique and find the existence of both sticky floors

[2]Along the same lines, Nguyen *et al.* (2007) study the wage distribution in the urban-rural labor market.

and glass ceilings. Hoyos *et al.* (2010) reach the same conclusions by applying the matching methodology developed by Ñopo (2008). Perticará and Astudillo (2009) assess the chilean gender wage gap using a dataset that includes information on the educational attainment of the parents of the respondent -their chosen instrument for education- as well as effective experience. They analyze separately the wage gap for full-time private employers. With the use of effective experience and instrumenting the educational attainment, when they evaluate the situation of private employers, they find that the wage gap disappears. Our results are similar to their estimations. They find the existence of sticky floors -although they do not highlight it. Borráz and Robano (2010) analyze the uruguayan gender wage gap and find the presence of both sticky floors and glass ceilings.

Our contribution is to assess in a comparable form the GWG in Latin America along the wage distribution and relate it with measures of economic development. With that purpose, we work with recent household surveys and use a set of characteristics that are equally defined along all countries. Using unconditional quantile regressions we calculate two estimations of the distribution of the (log of hourly) female wages and compare them with male wages. The first distribution that we calculate is that of the female wages conditional on their own human capital characteristics. The second distribution is the counterfactual distribution of the female wages that would exist had women been paid at the same rate than men were. For the construction of the counterfactual distribution we use the decomposition technique proposed by Firpo, Fortin and Lemieux (2009, FFL henceforth). This technique provides us with asymptotic results instead of relying on simulation procedures as done by the method developed by Machado and Mata (2005) (which results in a loss of efficiency for a small number of simulations or in a computation burden with a large number). More importantly, as opposed to Machado and Mata (2005), the FFL technique permits to calculate the contribution of each explanatory variable into the wage gap, thus it allows us the calculation of unconditional quantiles[3]. The FFL technique is similar to that developed by Di Nardo, Fortin and Lemieux (1996), which uses a reweighting procedure to generate a counterfactual distribu-

[3]Conditional quantiles allow the user to make inference of the effect on wages of a subgroup with specific years of education and experience. By contrast, the unconditional quantiles allow the user to compute marginal effects.

tion. A second step -not implemented in this paper- in the FFL is to further decompose the productivity and structure effects into the contribution of each covariate.

We calculate the resulting GWG as the difference between the male and female wages. Using the counterfactual distribution of females wages, we observe what part of the GWG is explained by the differences in the characteristics of the individuals (subtracting to the male wage distribution the counterfactual female wages, called 'composition' effect) and what portion is left unexplained (the difference between the counterfactual female wages and the actual female wages, called 'structure effect').

There can be several reasons explaining why male wages are higher than female's. If men were more educated than women, then we would expect men to earn higher wages and we would observe a positive GWG. In a similar analysis done for China, Wei Chi and Bo Li (2007) find that there is a sticky floor effect that can be explained by a lower female educational attainment. According to our results, this explanation does not fit, as working women have at least the same educational level, and are often more educated than working men. Moreover, the higher female educational attainment contributes to hid the underlying GWG. For instance, the gap observed at the tenth percentile in the raw data compares women of higher human capital characteristics to men of lower human capital characteristics[4]. With respect to experience levels, while working females have equal or more experience than males at the tenth percentile, at the ninetieth percentile males have equal or more experience than females.

Another explanation for the existence of a GWG favorable to males is the presence of discrimination in the labor market. The term 'discrimination' refers to the fact that workers with similar levels of human capital are paid differently because of the demographic group to which they belong (women in our case). We would expect then a positive GWG, as women of similar human capital characteristics than men would earn lower wages. According to Becker (1971)'s theory, there would be no reason to discriminate against females, as if such

[4] Bolivia is the sole country in the analysis on which we find an explanation analogous to that of Chi Wei and Bo Li (2007).

was the case, one might hire women at, say, $w_f + \epsilon$ and make extra-profits. But then this rent opportunity would not be sustained in equilibrium, as ϵ would increase until $w_f + \epsilon = w_m$. However, this theory relies on a perfectly competitive labor market, an assumption not likely to be validated in Latin America.

A third explanation is that social traditions might segregate women of similar human capital characteristics than men into lower-paying occupations that accommodate female's larger household responsibilities. In particular, as mentioned by Booth (2009), this explanation could explain the existence of sticky floors. Panizza and Qiang (2005) study the occupational segregation into public and private sectors in Latin America and find that the GWG is larger on the private sector. The second and third explanations imply that the returns to human capital investment for women are lower than for men. In this paper we focus on the analysis of the first explanation for the existence of a positive GWG: the evaluation of the human capital characteristics of the working population and the analysis of the retribution to those characteristics.

We find that working women have higher levels of human capital than working men; that difference increases as we move towards the extremes of the wage distribution and contributes to hid the portion of the wage gap that is left unexplained.

4.2 Methodology

We follow Fortin, Lemieux and Firpo (2010) in explaining the methodology for the decomposition of the differences in the distribution of wages, the GWG. For a population of individuals $i = 1, \ldots, N$, let the hourly wages (Y_{ti}) be determined by some observed (X_i) and unobserved (ε_i) components for each gender $t : t = \{M, F\}$ where $g_t(\cdot)$ is a wage setting function for each gender. In order to compute the decomposition, we need to assume that the wage setting function is linearly additive:

$$Y_{ti} = g_t(X_i, \varepsilon_i) = X_i'\beta_t + \varepsilon_{it} \tag{4.1}$$

and that the error term has a zero conditional mean: $\mathbb{E}[\varepsilon_t|X] = 0$. From the observed data it is possible to non-parametrically identify the empirical distribution of Y_t: let it be denoted F_{Y_t}, and its corresponding quantiles τ[5]:

$$\tau = F_{Y_t}(Q_{t,\tau}) = \mathbb{E}[F_{Y|X}(Q_{t,\tau}|X)] = \int F_{Y_t|X_t}(Q_{t,\tau}|X)dF_{X_t}(X) \tag{4.2}$$

where dF_X is the density function of X. At a particular quantile τ, the GWG is the difference between the two groups:

$$\Delta_O^\tau = F_{Y_M}(Q_{M,\tau}) - F_{Y_F}(Q_{F,\tau}) \tag{4.3}$$

To decompose the GWG into its explained (due to differences in human capital characteristics) and unexplained components (due to differences in the wage structure), we need to compute the counterfactual distribution of female wages ($F^c_{Y_M}(Q_{F,\tau})$) had they been paid at the same rate than men were[6]. This first step in computing the decomposition is analogous to DFL, and it consists of first estimating a reweighting factor $\Psi(X)$ to account for the proportion of males in the whole sample. In practice, this reweighting procedure is achieved by first pooling the data and estimating the probability of females in the sample ($Pr(F = 1|X)$) using a probit model. Then the reweighting factor is estimated using the predicted probabilities for males and females and the sample proportions of both groups ($Pr(F = 1);Pr(M = 1)$):

$$\hat{\Psi}(X) = \frac{\hat{Pr}(F = 1|X)/\hat{Pr}(F = 1)}{\hat{Pr}(F = 0|X)/\hat{Pr}(F = 0)} \tag{4.4}$$

and finally the counterfactual quantiles of interest can be computed using $\hat{\Psi}(X)$, which is analogous to a propensity score.

[5]Up to this point, the sole assumption needed is that the groups need to be mutually exclusive.

[6]Additional assumption: simple counterfactual treatment -we are not considering general equilibrium effects.

$$F^c_{Y_M}(Q_F) = \int F_{Y_M,X_M}(Y|X)\hat{\Psi}(X) \cdot dF_{X_F}(X) \tag{4.5}$$

.

The difference between male and female wages can be written as:

$$\Delta^{\tau}_{O} = \underbrace{[F_{Y_M}(Q_{M,\tau}) - F^c_{Y_M}(Q_{F,\tau})]}_{\text{composition effect}} + \underbrace{[F^c_{Y_M}(Q_{F,\tau}) - F_{Y_F}(Q_{F,\tau})]}_{\text{unexplained component}} \tag{4.6}$$

The second step in FFL[7], consists of substituting the outcome variable for a linear approximation based on the influence function (IF, Hampel, 1968, 1974). The IF is a measure of robustness of a distributional statistic to outlier data when F is replaced by its empirical distribution: By definition, $\int_{-\infty}^{\infty} IF(y,\tau)dF(y) = 0$. In the case of the quantiles, the IF can be written as:

$$IF(y,Q_\tau) = \frac{\tau - \mathbb{1}\{y \leq Q_\tau\}}{f_y(Q_\tau)} \tag{4.7}$$

The recentered influence function (RIF) adds up to the IF the distributional statistic of interest, and unlike conditional quantile regressions, the RIF functions have the property that they are linear in expectations: $\mathbb{E}[RIF(y,Q_\tau)] = Q_\tau$. Thus, the expected value of the RIF is just the quantile of interest.

$$RIF(y,Q_\tau) = Q_\tau + \frac{\tau - \mathbb{1}\{y \leq Q_\tau\}}{f_y(Q_\tau)} \tag{4.8}$$

[7] Assumptions:

1. Unconfoundedness (Selection on Observables): Beyond the observed characteristics, there is no other characteristic associated with both the outcome and the treatment.
2. Overlapping: For the observed and unobserved characteristics, both male and female wages are observed.
3. Ignorability (Conditional Independence): In general there exists interaction between X and ε, thus, it is not possible to further decompose the differences in the returns due to X from the difference in the returns due to ε, which are referred to as 'wage structure'. Thus, the obtained decomposition is: $\hat{q_\tau}(Y_M) - \hat{q_\tau}(Y_F) = [(\bar{X_M} - \bar{X_F})\hat{\beta_M} + \hat{R}^{\text{comp}}) + \bar{X_F}(\hat{\beta_M} - hat\beta_F) + \hat{R}^s]$ where $\hat{R}^s$ and $\hat{R}^{\text{comp}}$ are the error terms resulting from the linear approximation.

As explained by Chi and Li, the RIF regressions can be compared to OLS, and thus β can be interpreted as the marginal effect of X on q_τ. To perform the decomposition with this technique, instead of integrating over the whole conditional distribution function[8], FFL show that it is possible to invert locally the proportions and obtain the desired quantiles. Once the sample quantiles are estimated along with their corresponding density function $\hat{f}(Q_\tau)$, it is possible to obtain the RIF estimates by plugging in the results into equation (4.8).

4.3 Data description and the observed gender wage gap

The data come from country population surveys that follow a harmonized methodology and were provided by the Inter-American Development Bank, MECOVI program[9]. The sample comprises the observations for the population aged between 14 an 65 years old that live in urban areas[10] and who reported to be working at least one hour in the week preceding the survey. To get rid of extreme outliers we have trimmed the data by 1% of the hourly wage distribution at both tails for each of the distributions (male and female) and we have discarded the observations that reported more than 98 hours worked in a week. Additionally, we are aware that the rate of female participation differs among the countries and it could diminish the comparability of the results. However, as of today there is no accepted and robust selection estimation method available to account for self-selection in the quantile regression framework (see Melly (2008)). In the case of females, the human capital characteristics of those working are likely to be higher than those not participating (because of the higher opportunity costs for the more educated). Thus, if we were able to account for selection, the

[8] In a previous version of this paper we used Melly's technique (2006), which yields the same results but is computationally more time-intensive, as it needs to compute the whole distribution to obtain the unconditional quantiles of interest.

[9] MECOVI is the 'Programa para el Mejoramiento de las Encuestas y la medición de Condiciones de Vida' the spanish translation for the Program for Improvement of the Surveys of Living Conditions. MECOVI is a regional program of technical assistance for capacity building to improve the household surveys to measure living conditions and poverty in Latin America and the Caribbean that was jointly launched in 1996 by IDB, World Bank and UN-ECLAC.

[10] We have restricted our analysis to the urban sector s of each country because of comparability reasons, as some surveys present only urban data (Argentina and Uruguay). Moreover, we are interested in comparing the situations in the whole Latin America region, believing that, if at all, the urban areas are more comparable among countries.

wage distribution of females would move to the right (as we would allow the entrance of those with a lower reservation wage). This fact would increase the observed GWG. Thus, our estimations could be considered as a lower-bound for the true GWG. We have included some additional variables likely to capture part of the selection effect (such as marital status and number of dependents in the household). They are not our preferred estimation and the corresponding results -available upon request- do not significantly differ, except in some specific cases (i.e. Mexico). To avoid any endogeneity or sample selection problem, we do not include the categories of occupation when explaining the wage regressions, as the category of occupation is likely to be a choice variable that depends on the educational attainment. Additional information on the sources, the number of observations and the variables used is provided in the appendix.

4.3.1 Hourly wages

In Table 4.1 we present the hourly wages measured at constant US prices of 2000 and the unconditional GWG with its 95% statistical significance. Almost all countries present a statistical significant wage gap in favor of males at some point of the wage distribution (Peru is the sole exception). In particular, there is a significant GWG at the tenth percentile in Argentina, Brazil, Colombia, Chile, Honduras and Mexico, although there are no sticky floors in all of them, as Argentina, Brazil and Chile present a higher GWG at median levels. At the ninetieth percentile almost all countries (except Honduras and Peru) present a wage gap favorable to males, which is also larger in magnitude than the observed GWG at median levels. Therefore, just by looking at the unconditional distribution of wages we might conclude that there is a glass ceiling problem in Latin America, while the sticky floors problem is only present in Colombia, Honduras and Peru. The problem with this analysis is that it compares working individuals with different human capital characteristics; in particular, the working females are those with the highest reservation wage. To enrich our analysis we consider the distribution of human capital characteristics for the population under study, conditional on the quantiles of the wage distribution for each gender.

4.3.2 Education

Our first measure of human capital is the variable 'years of schooling'[11]. We report summary statistics for this variable in Table 4.2 *by quantiles of each gender wage distribution*. In general, except Bolivia, working females have the same educational attainment than males. At the tenth percentile, females are more educated in Argentina, Brazil, Colombia, Chile, Honduras and Uruguay. In the other countries studied there are not statistical differences. Thus the observed sticky floors cannot be explained by a higher level of male educational attainment.

At median levels females are significantly more educated in Brazil, Colombia, Costa Rica, Chile, Honduras, Mexico and Uruguay. We report that in Argentina, Bolivia, Brazil, Chile, Mexico and Paraguay there is a GWG in the observed distribution. Therefore, the observed GWG in Chile and Mexico cannot be explained by a higher educational attainment of males.

At the ninetieth percentile of the wage distribution, half of the countries (Argentina, Brazil, Colombia, Chile, Paraguay and Uruguay) present statistically higher levels of female educational attainment. Thus the observed glass ceiling in Table 4.1 cannot be explained by a lower female educational attainment.

Peru and Venezuela stand out as they do not present statistically different educational attainments at any of the percentiles studied. Bolivia presents higher levels of male educational attainment at the tenth percentile of the wage distribution.

In summary, the differences in the educational attainment cannot explain the sticky floors, nor the GWG at median levels and neither the glass ceilings, as working women have at least (if not higher) the same educational attainment than males.

[11] In the estimations we use both years of schooling and its square.

4.3.3 Experience

Our second measure of human capital is experience. Most of the countries do not report a direct measure of experience, therefore we use age as a proxy, with the caveat that age might be an imperfect proxy for female experience, as they take time-off the labor market while childbearing. In a set of estimations -available upon request- we have included the number of dependents present in the household as well as some other demographic characteristics as part of the covariates. They do not significantly differ from the estimations presented here.

We present the differences in experience by gender *conditional on the quantiles of each gender wage distribution* in Table 4.3. In Bolivia, Chile, Paraguay and Uruguay, working females have more experience at the tenth percentile. At the median wage distribution, only the uruguayan females have more experience.

So far we have found that when there are differences in human capital characteristics at some point of the distribution, they are favorable to women. In contrast, at the ninetieth percentile of the experience distribution, Brazil, Chile, Mexico, Peru and Venezuela, working males have significantly more experience. This may in part be due to the lower retirement age for females in most Latin American countries.

The differences in experience can at least partially explain the observed glass ceiling in the wage distribution of Brazil, Chile, Mexico and Venezuela, but they cannot explain the observed sticky floors, nor the GWG at median levels.

4.4 Estimations

So far we have presented the observed GWG and the differences in education and experience by quantiles of the wage distribution. Another factor explaining the GWG could be that males and females were paid differently to the same human capital characteristics. To assess that hypothesis, we run a mincerian equation that includes experience, its square, education, its square and a binary variable for females (F) at the tenth, median and ninetieth percentile of the wage distribution:

$$Y_i(\tau) = X_i'\beta(\tau) + \alpha F_i + \varepsilon_i(\tau) \tag{4.9}$$

Table 4.4 presents the coefficient of the female binary variable from equation (4.9). It shows not only that there is a penalty on being female but also that such penalty is different at different points of the wage distribution. While Argentina, Brazil, Mexico, Paraguay, Uruguay and Venezuela present the largest level their corresponding penalty at the ninetieth percentile, the rest of the countries present the largest penalty level at the tenth percentile.

The unconditional GWG observed in Table 4.1 does not incorporate the differences in education and experience by gender shown in Tables 4.2 and 4.3. Moreover, it assumes that male and female workers are rewarded at the same rate, which, as shown in Table 4.4 such assumption does not hold. To evaluate both the different human capital levels and the different retribution to the same characteristics we use FFL decomposition technique, whose results are described in next section.

4.4.1 Decomposition results

We present in Table 4.5 the results from equation (4.6), that is, the observed GWG after its decomposition into its explained (human capital) and unexplained ('wage structure') components. Panel A presents the observed GWG after accounting for human capital characteristics. The sum of panels B and C accounts for the observed GWG presented in panel A. Panel B accounts for the first term of equation (4.6), that is, the results from subtracting to the male wages the counterfactual female wages that would exist if women were paid at the same rate than men were . In panel C we show the second component of equation (4.6), the unexplained component of the GWG resulting from subtracting to the counterfactual female wages the actual female wages

In contrast to what we observed in Table 4.1, after accounting for human capital characteristics, there are sticky floors in nine countries (they are not observed in Brazil, Paraguay and Uruguay). The largest sticky floors are observed in Bolivia and Peru, where the male wage is about 0.5 log points (65% in levels) larger than the female wage. On the other hand, only in Uruguay there is no significant GWG at the tenth quantile. At median levels all countries present a significant GWG. The largest is observed in Bolivia, of about 0.32 log points (38% in levels) and the smallest are in Uruguay and Venezuela, where men earn 0.03 log points (3% in levels) more than women. At the ninetieth percentile we observe glass ceilings in: Brazil, Colombia, Mexico, Uruguay and Venezuela. Bolivia presents the largest male-female difference, of 0.26 log points (30%), while in Brazil and Uruguay males earn 0.21 log points (23%) more than females. Moreover, the magnitude of the glass ceiling is larger than that of the sticky floors in Brazil, Uruguay and Venezuela. In the rest of the countries, the observed GWG at the bottom of the distribution is larger than that at the top. After taking into account the human capital characteristics we appreciate that what is now more prevalent is the sticky floors problem and not the glass ceiling effect. As a visual aid, we include a set of figures with the decomposition of the GWG analogous to what is presented in panels A, B and C in Table 4.5. Notice the pattern of U-shaped distributions in many countries.

4.5 Relationship with measures of economic development

We use our estimations for the SF and GC components and relate them to economic development indicators. Data come from WDI. In particular, we assess the relationship between the GWG and the GDP PC. We find that sticky floors are more prevalent in countries with lower levels of economic growth (see Figure 4.1, top left panel). That fact would imply that uneducated women face tough labor market conditions, thus policies there should include specific considerations for them. We find that sticky floors are associated with abundance of low-skilled female labor supply. On the contrary, in poorer countries we do not observe glass ceilings because it its more difficult to find a substitute for a highly educated female.

One could extrapolate that once countries achieve economic growth, they will be facing glass ceilings as well (Figure 4.1, low left panel). Suggested policies: protection for females, such as minimum wage enforcement, maternity leave paid by the government and not by the employer in order not to promote the hiring of men.

We also compare our estimations with an inequality measure. We use the Gini Index, and again, we find a significant relationship between the sticky floors and country inequality (top right panel in Figure 4.1). One might think that as countries become more equal in terms of redistribution, sticky floors will go down. On the other hand, it might be that inequality and sticky floors are simultaneously reinforcing each other. But that would imply that there are no glass ceilings when inequality is low. In fact, the unexplained component of glass ceilings decreases along the Gini Index (low right panel, Figure 4.1).

4.5.1 Relationship with country-specific gender indicators

We compare our estimations for the GWG at the tenth, median and ninetieth quantile with some selected indicators on the socio-economic status of women, that are available through the OECD[12]. We analyze whether there is any correlation between some of those indicators and the measures of wage gap by country that we estimate; in particular we review the POLITICAL PARTICIPATION INDICATORS (which provide information on the percentage of women that are members of the parliament and the percentage of females that are either legislators, senior officials or managers) and the SOCIAL INSTITUTIONS STATISTICS (which give information related to Ownership Rights, Civil Liberties, Physical Integrity) as well as FAMILY CODE (traditions and informal practice related to parental authority, inheritance, early marriage, polygamy acceptance). We do not find any substantial difference that could explain the observed GWG nor its decomposed components. For the other indicators provided in the dataset, there is not enough variation on the data. However, as our analysis covers only twelve countries thus any conclusion from this has to be taken with caution.

[12]Gender, Institutions and Development database 2009

4.6 Conclusions

We have decomposed the observed GWG in twelve Latin American countries using a novel quantile regression decomposition technique. We can conclude that *(i)* we observe a significant wage gap favorable to men in all countries. *(ii)* Working women have equal or more endowments than working men and their endowments are less dispersed along the wage distribution. *(iii)* There is a statistically significant and 'unexplained' wage gap favorable to men, in particular at the extremes of the wage distribution, but also present at the middle in all countries, that varies in size. *(iv)* If we were able to account for selection into the labor market, a proportion of low-skilled women would enter the female wage distribution at the left extreme, increasing the observed wage gap, thus our estimations account for a lower-bound of the true wage distribution. *(v)* We find that the size of the observed and unexplained sticky floors is highly correlated with measures of economic development (per capita GDP and income inequality as measured by the Gini Index). Poorer countries and countries with higher levels of income inequality have higher gender wage gap differentials at the tenth percent of the wage distribution. On the other hand, richer countries and countries with lower levels of inequality present higher gender wage gap at the ninetieth percent of the wage distribution.

Table 4.1: Summary statistics, hourly wages and the GWG

	q10	mean	q50	q90	sd	min	max	N
Argentina	**0.12***	**0.16***	**0.11***	**0.19***				
male	0.79	2.43	1.91	4.68	1.80	0.23	11.98	10,739
female	0.67	2.27	1.80	4.49	1.70	0.19	10.78	7,258
Bolivia	**0.11**	**0.33***	**0.22***	**0.65***				
male	0.23	1.20	0.73	2.57	1.41	0.06	13.39	748
female	0.12	0.87	0.51	1.92	1.09	0.01	8.75	681
Brazil	**0.05***	**0.28***	**0.13***	**0.68***				
male	0.39	1.60	0.93	3.62	1.92	0.11	14.47	73,268
female	0.34	1.32	0.80	2.94	1.49	0.09	11.31	53,084
Colombia	**0.07***	**0.16***	**0.04**	**0.23***				
male	0.27	1.20	0.70	2.66	1.52	0.08	13.28	12,117
female	0.20	1.04	0.66	2.43	1.22	0.06	9.44	10,612
Costa Rica	**0.12**	**0.08**	**0.00**	**0.06***				
male	0.87	2.53	1.76	5.33	2.15	0.33	14.04	3,756
female	0.75	2.45	1.76	5.27	1.95	0.23	10.53	2,495
Chile	**0.21***	**0.62***	**0.24***	**1.33***				
male	1.25	3.84	2.36	8.25	3.99	0.44	27.68	36,601
female	1.04	3.22	2.12	6.92	3.15	0.30	22.24	22,222
Honduras	**0.09***	**0.09***	**0.06**	**0.00**				
male	0.46	1.77	1.18	3.68	1.79	0.08	14.04	8,783
female	0.37	1.68	1.12	3.68	1.61	0.10	11.05	6,794
Mexico	**0.18***	**0.39***	**0.14***	**1.16***				
male	0.93	3.15	2.00	6.96	3.35	0.35	26.58	10,940
female	0.75	2.76	1.86	5.80	2.68	0.17	18.74	7,178
Paraguay	**0.03**	**0.14***	**0.07***	**0.32***				
male	0.22	0.96	0.65	1.98	1.03	0.04	8.62	4,391
female	0.19	0.82	0.58	1.66	0.80	0.04	5.95	3,188
Peru	**0.15**	**0.25**	**0.22**	**-0.08**				
male	0.39	1.38	0.84	2.65	1.70	0.21	13.64	529
female	0.24	1.13	0.62	2.73	1.39	0.08	7.87	412
Uruguay	**0.01**	**0.58***	**0.04**	**1.44***				
male	0.82	3.39	2.07	7.23	3.95	0.23	35.95	18,934
female	0.81	2.81	2.03	5.79	2.50	0.18	17.36	15,753
Venezuela	**0.08**	**0.13***	**0.05**	**0.26***				
male	0.63	1.46	1.16	2.64	1.12	0.16	7.93	3,248
female	0.55	1.33	1.11	2.38	0.98	0.12	7.93	2,617

(1) Hourly wages in 2000 constant US prices and the observed gender wage gap (in bold numbers) by quantiles of each wage distribution.
(2) Star denotes significance at $p < 0.05$.

Table 4.2: Average years of schooling, by quantiles of each gender wage distribution

	Quantiles	schooling	male schooling			female schooling		
	of wage	gap	mean	sd	N	mean	sd	N
Argentina	p10	-0.61 *	8.18	3.22	1102	8.79	3.35	763
	p50	-0.18	9.93	3.22	1059	10.11	3.66	694
	p90	-1.45 *	13.15	3.18	1059	14.60	2.36	637
Bolivia	p10	2.96 *	7.79	4.42	75	4.83	4.50	69
	p50	1.32	8.61	4.38	75	7.29	4.94	68
	p90	0.26	11.22	4.00	74	10.96	4.70	68
Brazil	p10	-0.62 *	3.43	3.70	7468	4.05	4.07	5344
	p50	-0.87 *	6.35	4.40	7330	7.22	4.81	5250
	p90	-0.75 *	11.67	5.31	7149	12.42	5.56	5308
Colombia	p10	-0.65 *	6.03	3.69	1212	6.68	4.09	1072
	p50	-0.49 *	8.55	3.97	1401	9.04	3.94	1061
	p90	-0.60 *	13.83	3.65	1166	14.43	3.09	1053
Costa Rica	p10	-0.55	6.85	3.39	376	7.40	3.33	252
	p50	-1.13 *	8.79	3.48	347	9.92	3.71	251
	p90	0.06	13.94	3.35	375	13.88	3.50	245
Chile	p10	-0.44 *	9.08	3.63	3667	9.52	3.47	2225
	p50	-0.33 *	10.74	2.97	3327	11.07	3.04	2258
	p90	-0.28 *	13.75	3.05	3657	14.03	2.92	2209
Honduras	p10	-0.45 *	5.70	2.83	965	6.15	2.84	681
	p50	-1.16 *	7.31	3.21	743	8.47	3.39	676
	p90	-0.47	12.30	4.20	859	12.77	3.76	679
Mexico	p10	0.47	7.82	3.54	1150	7.35	4.14	788
	p50	-0.73 *	9.28	3.66	1089	10.01	4.15	623
	p90	0.07	13.59	3.42	1045	13.52	3.29	717
Paraguay	p10	-0.17	6.50	3.35	440	6.67	3.38	320
	p50	0.03	8.28	3.40	441	8.25	3.48	319
	p90	-0.70 *	13.13	3.39	439	13.83	3.17	317
Peru	p10	-0.50	8.11	4.49	56	8.61	4.93	42
	p50	-0.21	10.70	3.60	52	10.91	4.04	38
	p90	0.82	15.38	1.67	52	14.56	2.54	41
Uruguay	p10	-0.84 *	6.59	3.24	1896	7.43	3.59	1601
	p50	-0.61 *	7.96	3.95	1885	8.57	4.63	1581
	p90	-0.48 *	12.30	4.51	1880	12.78	4.79	1572
Venezuela	p10	-0.56	7.77	3.78	331	8.33	3.48	262
	p50	-0.53	9.18	3.58	418	9.71	3.86	263
	p90	-0.61	13.54	3.78	324	14.15	3.24	261

Star denotes significance at $p < 0.05$.

Table 4.3: Average years of experience (proxied by age), by quantiles of each gender wage distribution

	Quantiles	experience	male experience			female experience		
	of wage	gap	mean	sd	N	mean	sd	N
Argentina	p10	-1.70	34.90	14.16	1102	36.60	14.23	763
	p50	-0.03	38.07	13.48	1059	38.10	12.41	694
	p90	-0.01	43.24	11.19	1059	43.25	10.08	637
Bolivia	p10	-6.10 *	39.57	13.66	75	45.67	12.28	69
	p50	-1.88	38.93	9.86	75	40.81	10.37	68
	p90	-1.32	39.22	10.43	74	40.54	11.63	68
Brazil	p10	-0.66	32.53	14.06	7468	33.19	13.51	5344
	p50	0.35	34.66	11.64	7330	34.31	11.26	5250
	p90	2.10 *	41.63	10.12	7149	39.53	9.96	5308
Colombia	p10	-1.49	35.74	13.90	1212	37.23	12.74	1072
	p50	1.52	37.36	11.04	1401	35.84	10.32	1061
	p90	1.77	42.11	10.60	1166	40.34	9.90	1053
Costa Rica	p10	-0.67	34.44	14.05	376	35.11	12.77	252
	p50	0.77	35.09	12.03	347	34.32	11.05	251
	p90	-0.96	40.27	9.66	375	41.23	10.10	245
Chile	p10	-1.98 *	37.74	13.53	3667	39.72	12.83	2225
	p50	0.77	37.83	12.12	3327	37.06	11.59	2258
	p90	1.21*	42.72	10.91	3657	41.51	11.09	2209
Honduras	p10	0.19	32.18	14.38	965	31.99	13.41	681
	p50	1.49	33.44	11.29	743	31.95	10.84	676
	p90	1.36	41.19	11.34	859	39.83	10.39	679
Mexico	p10	-1.05	34.06	14.05	1150	35.11	13.54	788
	p50	0.97	35.46	11.26	1089	34.49	10.56	623
	p90	2.56 *	41.93	10.34	1045	39.37	9.83	717
Paraguay	p10	-3.73 *	35.18	16.02	440	38.91	13.40	320
	p50	-0.18	33.56	10.94	441	33.74	11.61	319
	p90	1.87	40.98	10.83	439	39.11	10.59	317
Peru	p10	4.49	34.55	15.65	56	30.06	12.69	42
	p50	-0.05	38.07	12.34	52	38.12	11.42	38
	p90	9.95 *	45.23	11.01	52	35.28	8.36	41
Uruguay	p10	-3.37 *	35.94	14.08	1896	39.31	13.93	1601
	p50	-1.71 *	38.17	12.31	1885	39.88	12.28	1581
	p90	0.92	45.49	10.60	1880	44.57	10.00	1572
Venezuela	p10	0.24	35.87	12.43	331	35.63	11.86	262
	p50	-0.96	35.43	11.35	418	36.39	11.61	263
	p90	2.47 *	41.75	11.05	324	39.28	9.82	261

Star denotes significance at $p < 0.05$.

Table 4.4: Female penalty, mincerian equation

	q10	q50	q90
Argentina	-0.18	-0.17	-0.19
	(0.02)***	(0.01)***	(0.02)***
Bolivia	-0.38	-0.25	-0.15
	(0.11)***	(0.07)***	(0.10)
Brazil	-0.20	-0.23	-0.39
	(0.01)***	(0.01)***	(0.01)***
Colombia	-0.26	-0.12	-0.23
	(0.02)***	(0.01)***	(0.02)***
Costa Rica	-0.25	-0.15	-0.14
	(0.03)***	(0.02)***	(0.04)***
Chile	-0.26	-0.22	-0.22
	(0.01)***	(0.01)***	(0.01)***
Honduras	-0.30	-0.14	-0.09
	(0.03)***	(0.01)***	(0.03)***
Mexico	-0.21	-0.11	-0.22
	(0.02)***	(0.01)***	(0.03)***
Paraguay	-0.16	-0.20	-0.21
	(0.04)***	(0.02)***	(0.03)***
Peru	-0.42	-0.30	-0.13
	(0.08)***	(0.07)***	(0.13)
Uruguay	-0.09	-0.14	-0.38
	(0.01)***	(0.01)***	(0.02)***
Venezuela	-0.18	-0.09	-0.25
	(0.04)***	(0.02)***	(0.04)***

(1): Coefficient of the binary variable ‘female’ of quantile regressions of log of hourly wage on age, schooling and their squares

(2) Significance levels are denoted as $***p < 0.01$.

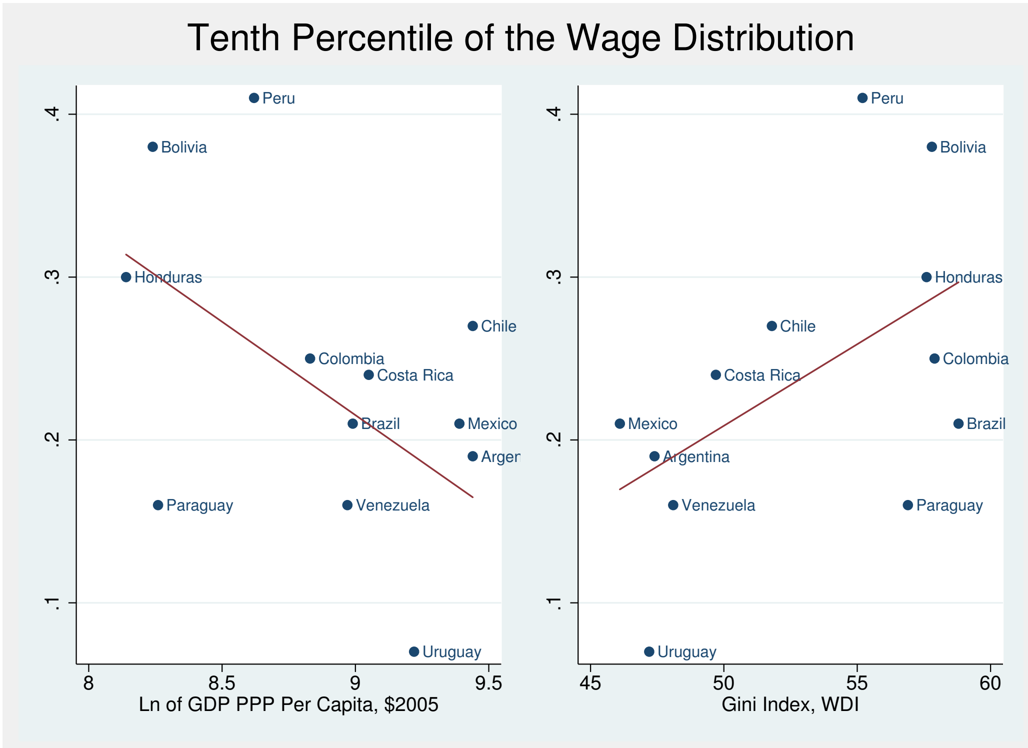
Table 4.5: Decomposition results

	Panel A: Observed GWG			Panel B: Explained GWG			Panel C: Unexplained GWG		
	q10	q50	q90	q10	q50	q90	q10	q50	q90
Argentina	0.14	0.09	0.04	-0.06	-0.09	-0.15	0.19	0.19	0.19
	(0.02)***	(0.01)***	(0.02)***	(0.00)***	(0.01)***	(0.01)***	(0.02)***	(0.01)***	(0.02)***
Bolivia	0.53	0.32	0.26	0.14	0.07	0.09	0.38	0.24	0.17
	(0.10)***	(0.07)***	(0.10)***	(0.06)**	(0.03)***	(0.04)***	(0.12)***	(0.07)***	(0.09)*
Brazil	0.17	0.19	0.21	-0.04	-0.08	-0.25	0.21	0.27	0.45
	(0.01)***	(0.01)***	(0.01)***	(0.00)***	(0.00)***	(0.01)***	(0.01)***	(0.00)***	(0.01)***
Colombia	0.21	0.08	0.17	-0.04	-0.04	-0.08	0.25	0.12	0.25
	(0.02)***	(0.01)***	(0.03)***	(0.01)***	(0.01)***	(0.01)***	(0.02)***	(0.01)***	(0.02)***
Costa Rica	0.18	0.05	-0.02	-0.05	-0.10	-0.20	0.24	0.15	0.18
	(0.03)***	(0.02)***	(0.04)***	(0.01)***	(0.01)***	(0.02)***	(0.03)***	(0.02)***	(0.04)***
Chile	0.23	0.19	0.11	-0.03	-0.06	-0.13	0.27	0.25	0.24
	(0.01)***	(0.01)***	(0.01)***	(0.00)***	(0.00)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Honduras	0.23	0.06	0.04	-0.07	-0.07	-0.12	0.30	0.14	0.17
	(0.03)***	(0.02)***	(0.03)***	(0.01)***	(0.01)***	(0.02)***	(0.02)***	(0.01)***	(0.03)***
Mexico	0.19	0.06	0.17	-0.02	-0.05	-0.10	0.21	0.11	0.27
	(0.02)***	(0.01)***	(0.03)***	(0.00)***	(0.01)***	(0.01)***	(0.02)***	(0.01)***	(0.03)***
Paraguay	0.13	0.16	0.14	-0.03	-0.05	-0.10	0.16	0.21	0.24
	(0.04)***	(0.03)***	(0.04)***	(0.01)**	(0.01)***	(0.02)***	(0.04)***	(0.02)***	(0.04)***
Peru	0.50	0.30	0.33	0.10	0.06	0.18	0.41	0.24	0.15
	(0.07)***	(0.08)***	(0.15)	(0.03)***	(0.03)*	(0.08)**	(0.07)***	(0.07)***	(0.14)
Uruguay	0.00	0.03	0.21	-0.07	-0.11	-0.23	0.07	0.15	0.43
	(0.01)	(0.01)***	(0.02)***	(0.00)***	(0.00)***	(0.01)***	(0.01)***	(0.01)***	(0.02)***
Venezuela	0.12	0.03	0.14	-0.04	-0.06	-0.13	0.16	0.08	0.27
	(0.04)***	(0.02)***	(0.04)***	(0.01)***	(0.01)***	(0.02)***	(0.04)***	(0.02)***	(0.04)***

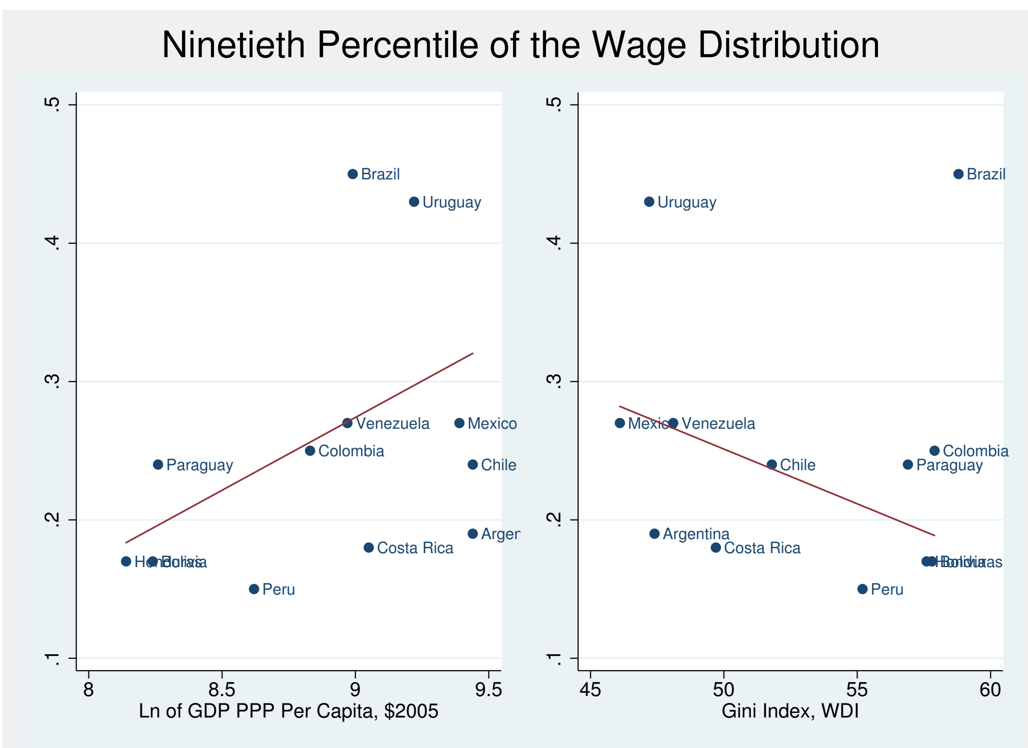
(1) Significance levels are denoted as $***p < 0.01; **p < 0.05; *p < 0.1$

(2) Estimations use Firpo, Fortin and Lemieux (2009) decomposition technique.

Figure 4.1: The Unexplained Component of The Gender Wage Gap



(a) sticky floors



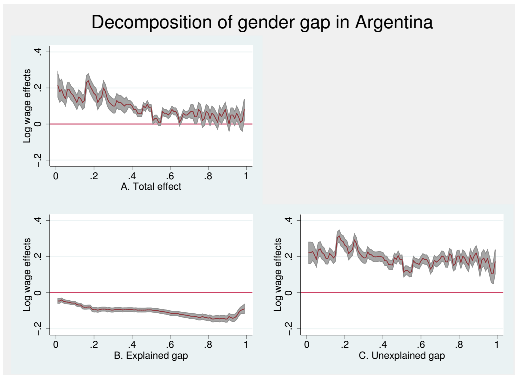
(b) glass ceiling

Source: WDI for the GDP and Gini Index, own calculations for the sticky floors.

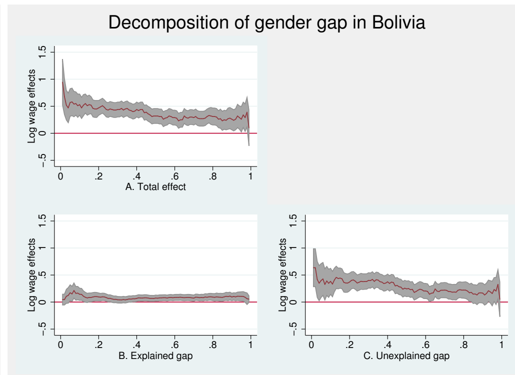
Table 4.6: Appendix: Data sources

Country	Year of survey	Number of observations	CPI	US exchange rate	Comments
Argentina	2007	17997	1.17	3.1	INDEC, 1st trimester
Bolivia	2005	1429	1.13	8.07	MECOVI, Encuesta de Hogares
Brazil	2003	126352	1.17	3.08	PNAD
Colombia	2003	22879	1.07	2877.65	Encuesta de Calidad de Vida
Costa Rica	2003	6251	1.07	398.66	Encuesta de Hogares de propósitos múltiples
Chile	2006	114851	1.17	530.29	CASEN
Honduras	2006	15577	1.17	18.90	Encuesta de Hogares de propósitos múltiples
Mexico	2004	18118	1.10	11.29	Encuesta Nacional de Gastos e Ingresos de los Hogares
Paraguay	2003	7579	1.07	6424.34	Encuesta de Hogares
Peru	2000	941	1.00	3.49	ENAHO (Only metropolitan Lima is considered)
Uruguay	2006	34687	1.17	24.07	INE (Only Montevideo is considered)
Venezuela	2003	5865	1.07	1606.96	INE (Only metropolitan Caracas is considered)

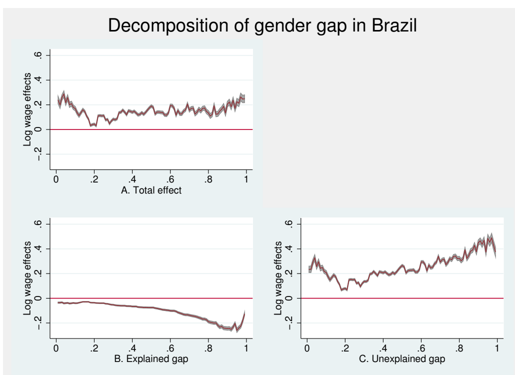
Notes: (1) Source of CPI (Consumer Price Index) and US exchange rate is World Development Indicators (2008).
(2) Official exchange rate, LCU per US, period average.



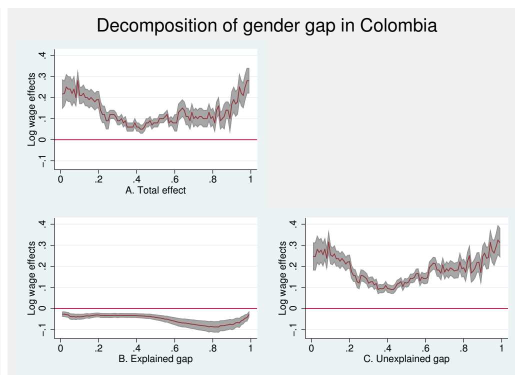
(c) Argentina



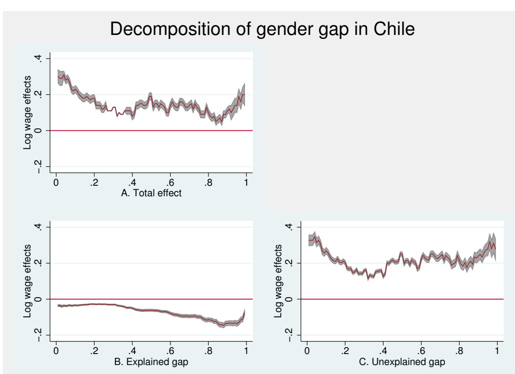
(d) Bolivia



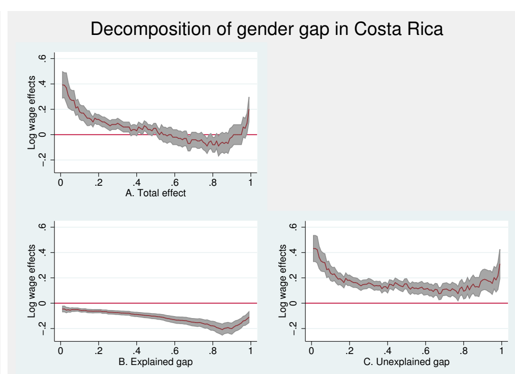
(e) Brazil



(f) Colombia

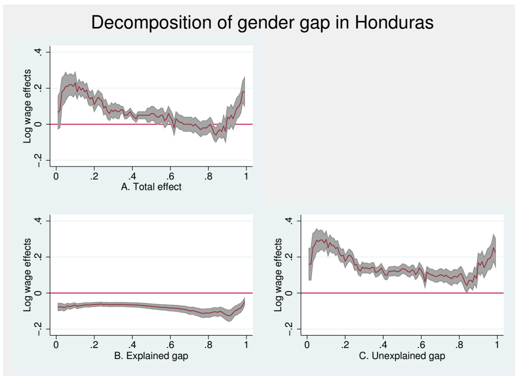


(g) Chile

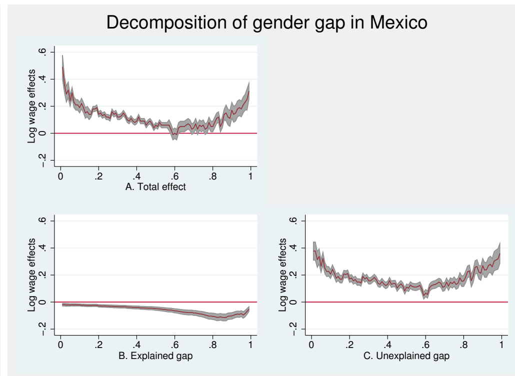


(h) Costa Rica

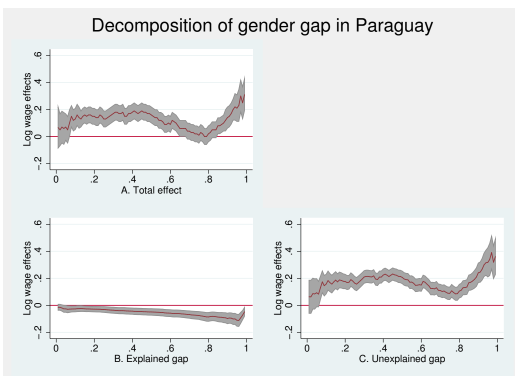
Figure 4.2: Results from wage decompositions



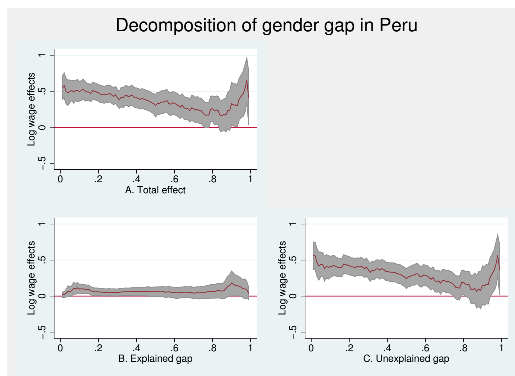
(a) Honduras



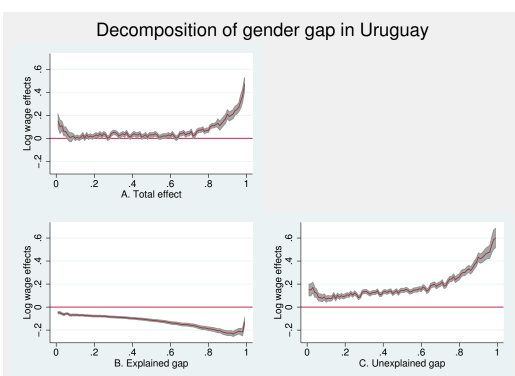
(b) Mexico



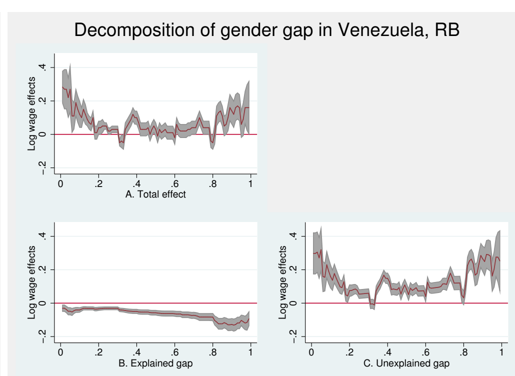
(c) Paraguay



(d) Peru



(e) Uruguay



(f) Venezuela

Figure 4.3: Results from wage decompositions

References

Aakvik, A. (2001). Bounding a matching estimator: The case of a norwegian training program. *Oxford Bulletin of Economics and Statistics*, 63(1):115–143.

Ahmed, A., Hill, R., Smith, L., Frankenberger, T., Gulati, K., Wiesmann, D., Quabili, W., and Yohannes, Y. (2007). *The world's most deprived: Characteristics and causes of extreme poverty and hunger*, volume 43. Intl Food Policy Res Inst.

Albrecht, J., Björklund, A., and Vroman, S. (2003). Is there a glass ceiling in sweden? *Journal of Labor Economics*, 1(21):145–177.

Albrecht, J., Van Vuuren, A., and Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the netherlands. *Labour Economics*, 16(4):383–396.

Alkire, S. and Foster, J. (2010). Counting and multidimensional poverty measurement. *Journal of Public Economics*.

Alkire, S. and Foster, J. (2011). Understandings and misunderstandings of multidimensional poverty measurement. *Journal of Economic Inequality*, pages 1–26.

Alkire, S., Foster, J., and Santos, M. E. (2011). Where did identification go? *Journal of Economic Inequality*, pages 1–5.

Alkire, S. and Santos, M. (2010). Acute multidimensional poverty: A new index for developing countries. *Oxford Poverty & Human Development Initiative (OPHI) Working Paper, No. 38*.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 13(1):151–184.

Arulampalam, W., Booth, A., and Bryan, M. (2007). Is there a glass ceiling over europe? exploring the gender pay gap across the wage distribution. *Industrial and Labor Relations Review*, pages 163–186.

Atal, J., Ñopo, H., Winder, N., and Dept, I.-A. D. B. R. (2009). *New century, old disparities: gender and ethnic wage gaps in Latin America*. Inter-American Development Bank, Research Dept.

Atkinson, A. (2003). Multidimensional deprivation: contrasting social welfare and counting approaches. *Journal of Economic Inequality*, 1(1):51–65.

Badel, A. and Peña, X. (2010). Decomposing the gender wage gap with sample selection adjustment: Evidence from colombia. *Revista de análisis económico*, 25(2):169–191.

Banerjee, A., Duflo, E., Chattopadhya, R., Mukhopadhya, J., and Shapiro, J. (2008). Targeting efficiency: How well can we identify the poor? *presented at CMFR, India*.

Berg, C. and Emran, M. (2011). Does microfinance help the ultrapoor cope with seasonal shocks? evidence from seasonal famine (monga) in bangladesh. Technical report, Working Paper. Available at SSRN: http://ssrn. com/abstract= 1802073.

Berlinski, S., Galiani, S., and Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93(1-2):219 – 234.

Besley, T. and Kanbur, S. (1990). *The principles of targeting*, volume 385. Office of the Vice President, Development Economics, World Bank.

Black, D. A. and Smith, J. A. (2004). How robust is the evidence on the effects of college quality? evidence from matching. *Journal of Econometrics*, 121(1-2):99 – 124. <ce:title>Higher education (Annals issue)</ce:title>.

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455.

Blundell, R. and Dias, M. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44(3):565–640.

Booth, A. (2009). Gender and competition. *Labour Economics*, 16(6):599–606.

Borraz, F. and Robano, C. (2010). Brecha salarial en uruguay. *Revista de Análisis Económico*, 25(1):49–77.

BRAC (2007). Annual report 2007. *Dhaka, Bangladesh*.

Bucheli, M. and Sanroman, G. (2005). Salarios femeninos en el uruguay:¿ existe un techo de cristal? *Revista de economía*, 12(2):63–88.

Center, C. P. R. (2008). Chronic poverty report 2008-09. *Manchester*.

Chi, W. and Li, B. (2008). Glass ceiling or sticky floor? examining the gender earnings differential across the earnings distribution in urban china, 1987-2004. *Journal of Comparative Economics*, 36(2):243–263.

Chong, A. and Ñopo, H. (2007). Beliefs vs. actions. the puzzle of discrimination in latin america. *Research Department, Inter-American Development Bank, mimeo*.

Deaton, A. (1989). Looking for boy-girl discrimination in household expenditure data. *The World Bank Economic Review*, 3(1):1–15.

Duclos, J., Sahn, D., and Younger, S. (2006). Robust multidimensional poverty comparisons. *The Economic Journal*, 116(514):943–968.

Ebbes, P., Wedel, M., and Böckenholt, U. (2009). Frugal iv alternatives to identify the parameter for an endogenous regressor. *Journal of Applied Econometrics*, 24(3):446–468.

Emran, M. and Hou, Z. (forthcoming). Access to markets and rural poverty: Evidence from household consumption in china. *Review of Economics and Statistics*.

Emran, M., Morshed, A., and Stiglitz, J. (2007). Microfinance and missing markets. *The George Washington University, Working Papers*.

Emran, M., Robano, V., and Smith, S. C. (2009a). Assessing the frontiers of ultra-poverty reduction: Evidence from cfpr/tup, an innovative program in bangladesh. *revise and resubmit Economic Development and Cultural Change*.

Emran, M. and Shilpi, F. (forthcoming). *The extent of the market and stages of agricultural specialization*. Canadian Journal of Economics.

Emran, M., Shilpi, F., and Stiglitz, J. E. (2009b). *De Soto vs. Yunus: Microfinance, Property*

Rights and Poor. Paper presented at at the ASSA 2009 Conference at San Francisco.

Emran, M. and Sun, Y. (2011). Magical transition? intergenerational educational and occupational mobility in rural china: 1988-2002. *The George Washington University Working Papers - available on SSRN*.

Fairlie, R. (2005). An extension of the blinder-oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement*, 30(4):305–316.

Farré, L., Klein, R., and Vella, F. (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*, pages 1–23.

Firpo, S., Fortin, N., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. *Handbook of Labor Economics*, 4:1–102.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica: Journal of the Econometric Society*, pages 761–766.

Gandelman, N. (2009). Female headed households and homeownership in latin america. *Housing Studies*, 24(4):525–549.

Goals, M. D. (2010). *The Millenium Development Goals Report 2010*. UN.

Harvey, A. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica: Journal of the Econometric Society*, pages 461–465.

Heckman, J., Ichimura, H., and Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):605.

Heckman, J. and Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics*, 86(1):30–57.

Hirano, K. and Imbens, G. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Out-*

comes Research Methodology, 2(3):259–278.

Hoyos, A., Ñopo, H., and Peña, X. (2010). The persistent gender earnings gap in colombia, 1994-2006. *Inter-American Development Bank*.

Huber, M. and Melly, B. (2011). *Quantile regression in the presence of sample selection*. School of Economics and Political Science, Department of Economics, University of St. Gallen.

Hulme, D. and Mosley, P. (1996). *Finance against poverty*, volume 2. Psychology Press.

Kanbur, R. (2010). Macro crises and targeting transfers to the poor. *Journal of Globalization and Development*, 1(1):9.

Klein, R. and Vella, F. (2009). A semiparametric model for binary response and continuous outcomes under index heteroscedasticity. *Journal of Applied Econometrics*, 24(5):735–762.

Klein, R. and Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, 154(2):154–164.

Koenker, R. (2005). *Quantile regression*, volume 38. Cambridge Univ Pr.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Hallock, K. (2001). Quantile regression. *The Journal of Economic Perspectives*, 15(4):143–156.

Lewbel, A. (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*, pages 1201–1213.

Lewbel, A. (forthcoming). Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economics Statistics*.

Lipton, M. (1986). Seasonality and ultrapoverty. *IDS Bulletin*, 17(3):4–8.

Machado, J. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465.

Matin, I. (1998). Mis-targeting by the grameen bank: A possible explanation. *IDS Bulletin*, 29(4):51–58.

Matin, I. and Halder, S. (2007). Combining methodologies for better targeting of the extreme poor: Lessons from brac's cfpr/tup programme. *Working Papers*.

Matin, I. and Hulme, D. (2003). Programs for the poorest: Learning from the igvgd program in bangladesh. *World Development*, 31(3):647 – 665. Chronic Poverty and Development Policy.

Matin, I., Sulaiman, M., and Rabbani, M. (2008). Crafting a graduation pathway for the ultra-poor: Lessons and evidence from a brac programme. *RED Working Paper, BRAC*.

McMillen, D. (2008). Changes in the distribution of house prices over time: Structural characteristics, neighborhood, or coefficients? *Journal of urban economics*, 64(3):573–589.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12(4):577–590.

Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. *Review of Labor Economics*, 68:543–572.

Millimet, D. and Tchernis, R. (2009). Estimation of treatment effects without an exclusion restriction: with an application to the analysis of the school breakfast program. Technical report, National Bureau of Economic Research.

Mitra, C. et al. (2011). Multidimensional poverty: Measurement, estimation, and inference. *OPHI Working Papers*.

Mundial, B. (2006). *World development report 2006*. Oxford University Press.

Mustafa, S., Ara, I., Banu, D., Hossain, A., Kabir, A., Mohsin, M., Yusuf, A., and Jahan, S. (1996). Beacon of hope. *Dhaka, BRAC*.

Nguyen, B., Albrecht, J., Vroman, S., and Westbrook, M. (2007). A quantile regression decomposition of urban-rural inequality in vietnam. *Journal of Development Economics*, 83(2):466–490.

Nicodemo, C. (2009). *Gender pay gap and quantile regression in European families*, volume Working Paper 3978. IZA.

Noor, M., Shihab, A., Parveen, R., Rahman, H., Ali, T., Zaman, S., and Matin, I. (2004). Stories of targeting. *BRAC TUP Working Paper No. 1*.

Ñopo, H. (2008). Matching as a tool to decompose wage gaps. *The review of economics and statistics*, 90(2):290–299.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 14(3):693–709.

Panizza, U. and Qiang, C. (2005). Public-private wage differential and gender gap in latin america: Spoiled bureaucrats and exploited women? *Journal of Socio-Economics*, 34(6):810–833.

Perticará, M. and Astudillo, A. (2008). ¿ qué tan alta puede ser la brecha de salarios en chile? investigando diferencias salariales entre hombres y mujeres a partir de regresiones de cuantiles. *ILADES-Georgetown University Working Papers*, 211.

Petkova, N. (2009). Does foreign share holding improve firm operating performance? Technical report, University of Oregon, Department of Finance, Working Paper.

Pitt, M. and Khandker, S. (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, M., Prakash, V., and Sulaiman, M. (2008). Impact assessment of cfpr/tup: a descriptive analysis based on 2002-2005 panel data.

Ravallion, M. (2007a). Evaluating anti-poverty programs. *Handbook of development economics*, 4:3787–3846.

Ravallion, M. (2007b). *How relevant is targeting to the success of an antipoverty program?* The World Bank.

Ravallion, M. (2011). On multidimensional indices of poverty. *Journal of Economic Inequality*, pages 1–14.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, 85(4):777–792.

Rigobon, R. and Rodrik, D. (2005). Rule of law, democracy, openness, and income. *Economics of Transition*, 13(3):533–564.

Robano, V. (2005). The female labor supply in uruguay. Technical report, The George Washington University.

Schroeder, E. (2010). The impact of microcredit borrowing on household consumption in bangladesh. Technical report, Mimeo, Georgetown University Economics Department.

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, 44(2):219–231.

Sen, A. (1980). Equality of what? *Tanner lectures on human values*, 1:195–220.

Shams, R., Mahmud, S., and Das, N. (2010). Integration of the ultra poor into mainstream development: How effective is cfpr? *CFPR Working Paper*, (21).

Smith, S. C. (2005). *Ending global poverty: A guide to what works*. Palgrave Macmillan New York.

Song, J. and Manchester, J. (2007). New evidence on earnings and benefit claims following changes in the retirement earnings test in 2000. *Journal of Public Economics*, 91(3-4):669–700.

Sulaiman, M. and Matin, I. (2006). Targeting effectiveness of cfpr/tup in scaled up environment. *CFPR/TUP Working Paper Series*, 8.

Tenjo, J., Ribero, R., and D., L. F. B. (2005). *Evolución de las diferencias salariales por sexo en seis países de América Latina: Un intento de interpretación*. CEDE, Centro de Estudios sobre Desarrollo Económico, Facultad de Economía, Universidad de los Andes.

Todd, P. (2006). Matching estimators. *Palgrave Dictionary of Economics, Palgrave Macmillan, New*.

Van de Walle, D. (1998). Targeting revisited. *The World Bank Research Observer*, 13(2):231–248.

Wikigender, O. (2010). Gender, institutions and development database. Technical report, OECD Development Centre.

Yun, M. (2007). An extension of the oaxaca decomposition using generalized residuals. *Journal of Economic and Social Measurement*, 32(1):15–22.

Zaman, H. (1998). Can mis-targeting be justified?: Insights from brac's micro-credit programme. *IDS Bulletin*, 29(4):59–65.